

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2012-226106

(P2012-226106A)

(43) 公開日 平成24年11月15日(2012.11.15)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 O L 11/00 (2006.01)	G 1 O L 11/00 4 O 2 G	5 D O 1 5
G 1 O L 15/10 (2006.01)	G 1 O L 15/10 5 O O Z	

審査請求 未請求 請求項の数 12 O L (全 25 頁)

(21) 出願番号	特願2011-93441 (P2011-93441)	(71) 出願人	000002185
(22) 出願日	平成23年4月19日 (2011. 4. 19)		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	東山 恵祐
			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
		Fターム(参考)	5D015 AA06

(54) 【発明の名称】 楽曲区間検出装置および方法、プログラム、記録媒体、並びに楽曲信号検出装置

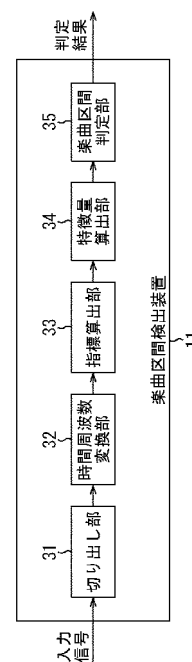
(57) 【要約】

【課題】入力信号から楽曲部分を精度良く検出する。

【解決手段】指標算出部は、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（例えばパワースペクトル）と、信号成分の強さを近似した関数（2次関数）とに基づいて、信号成分のトーンらしさの指標を算出し、楽曲判定部は、トーンらしさの指標に基づいて、入力信号の各領域に楽曲が含まれているか否かを判定する。本技術は、楽曲と雑音とが混在した入力信号から楽曲部分を検出する楽曲区間検出装置に適用することができる。

【選択図】 図1

図1



【特許請求の範囲】**【請求項 1】**

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、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、

前記トーンらしさの指標に基づいて、前記入力信号の各領域に楽曲が含まれているか否かを判定する楽曲判定部と

を備える楽曲区間検出装置。

【請求項 2】

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 2 次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 2 次関数との誤差に基づいて、前記指標を算出する

請求項 1 に記載の楽曲区間検出装置。

【請求項 3】

前記指標算出部は、前記 2 次関数の曲率に応じて前記指標を調整する

請求項 2 に記載の楽曲区間検出装置。

【請求項 4】

前記指標算出部は、前記 2 次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 に記載の楽曲区間検出装置。

【請求項 5】

所定時間分の前記入力信号の各領域の前記トーンらしさの指標に基づいて、所定時間分の前記入力信号の特徴量を算出する特徴量算出部をさらに備え、

前記楽曲判定部は、前記特徴量が所定の閾値より大きい場合、所定時間分の前記入力信号に楽曲が含まれていると判定する

請求項 1 に記載の楽曲区間検出装置。

【請求項 6】

前記特徴量算出部は、所定時間分の前記入力信号の各領域の前記トーンらしさの指標を、周波数毎に時間方向に積算することで前記特徴量を算出する

請求項 5 に記載の楽曲区間検出装置。

【請求項 7】

前記特徴量算出部は、所定時間分の前記入力信号の各領域において、周波数毎に、所定の閾値より大きい前記トーンらしさの指標が時間方向に最も連続する前記領域の前記トーンらしさの指標を積算することで前記特徴量を算出する

請求項 5 に記載の楽曲区間検出装置。

【請求項 8】

前記特徴量を時間方向にフィルタリングするフィルタ処理部をさらに備え、

前記楽曲判定部は、時間方向にフィルタリングされた前記特徴量が所定の閾値より大きい場合、所定時間分の前記入力信号に楽曲が含まれていると判定する

請求項 5 に記載の楽曲区間検出装置。

【請求項 9】

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、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、

前記トーンらしさの指標に基づいて、前記入力信号の各領域に楽曲が含まれているか否かを判定する楽曲判定ステップと

を含む楽曲区間検出方法。

10

20

30

40

50

【請求項 10】

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、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、

前記トーンらしさの指標に基づいて、前記入力信号の各領域に楽曲が含まれているか否かを判定する楽曲判定ステップと

を含む処理をコンピュータに実行させるプログラム。

【請求項 11】

請求項 10 に記載のプログラムが記録されている記録媒体。

【請求項 12】

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、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

を備える楽曲信号検出装置。

【発明の詳細な説明】**【技術分野】****【0001】**

本技術は、楽曲区間検出装置および方法、プログラム、記録媒体、並びに楽曲信号検出装置に関し、特に、入力信号から楽曲部分を検出することができるようにする楽曲区間検出装置および方法、プログラム、記録媒体、並びに楽曲信号検出装置に関する。

【背景技術】**【0002】**

従来、テレビジョン放送やラジオ放送の放送番組では、様々な音楽（楽曲）が用いられている。放送番組には、音楽番組のように、明確に音楽を主体として用いるものもあれば、ドラマ等のように音楽をBGM（Back Ground Music）として用いるものもある。

【0003】

このような放送番組の視聴者にとっては、例えば、音楽番組の楽曲部分だけを再生して視聴したいというニーズがある。

【0004】

また、放送事業者にとっては、どの楽曲をどの放送番組で使用したかを管理することで、著作権料の支払いを容易にしたり、放送番組の編集を行う際の参考にしたいというニーズがある。

【0005】

これは、楽曲のデータベースが用意されている場合には、放送番組の音声信号とデータベースの音声信号とを比較して、放送番組の音声信号に含まれる楽曲を検索する技術を用いることで実現することが可能である。しかしながら、楽曲のデータベースが用意されていない場合や、放送番組の音声信号に含まれる楽曲がデータベースに登録されていない場合には、上述したような楽曲検索の技術を用いることはできない。この場合、人間が放送番組を聴くことで楽曲の有無や一致の確認を行うことになるが、膨大な量の放送番組を聴くには多大な時間と労力がかかってしまう。

【0006】

そこで、放送番組の音声信号から楽曲が含まれる区間を検出する技術が提案されている。

【0007】

例えば、入力信号をスペクトラムに変換したときのピークが時間方向に持続する時間区間に基づいて、楽曲区間を検出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】****【0008】**

【特許文献 1】特開平 10 - 301594 号公報

10

20

30

40

50

【発明の概要】

【発明が解決しようとする課題】

【0009】

上述した特許文献1の手法によれば、音楽番組の音声信号のように、特定の時間に楽曲のみが含まれる入力信号や、楽曲とその楽曲に比べて十分レベルの低い楽曲以外の音（以下、雑音という）とが混在している入力信号に対しては、高い精度で楽曲区間を検出することができる。

【0010】

しかしながら、ドラマのように、BGMとしての楽曲と、その楽曲とほぼ同レベルの音声等の雑音とが混在している入力信号に対しては、スペクトラムのピークを適切に検出することができず、楽曲区間の検出の精度は低下してしまう可能性がある。

10

【0011】

また、放送番組においては、台詞やナレーション等の音声はセンターに定位することが多いという特徴を利用して、入力信号の左チャンネル信号から右チャンネル信号を減算する（または、右チャンネル信号から左チャンネル信号を減算する）ことで、音声（雑音）の影響を除くようにした手法もある。しかしながら、この手法は、モノラル放送には適用することができず、また、楽曲自体がセンターに定位しているような入力信号に対しても適用することができない。さらに、音声圧縮による量子化ノイズは左右のチャンネルで独立して発生するため、この手法においては、減算した信号に本来の入力信号とは相関の小さい量子化ノイズが含まれる恐れがある。

20

【0012】

さらに、スペクトラムにおいて時間方向に持続して形成されるピークは、楽曲によるものに限らず、ノイズ、サイドローブ、干渉、および時間変動トーン等によるものもあるので、ピークに基づいた楽曲区間の検出結果において、楽曲以外の雑音の影響を完全に除くことはできない。

【0013】

このように、楽曲と、その楽曲とほぼ同レベルの雑音とが混在している入力信号から楽曲部分を精度良く検出することはできなかった。

【0014】

本技術は、このような状況に鑑みてなされたものであり、入力信号から楽曲部分を精度良く検出することができるようにするものである。

30

【課題を解決するための手段】

【0015】

本技術の一側面の楽曲区間検出装置は、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、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、前記トーンらしさの指標に基づいて、前記入力信号の各領域に楽曲が含まれているか否かを判定する楽曲判定部とを備える。

【0016】

前記指標算出部には、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2次関数で近似する近似処理部とを設け、前記極大となる点近傍の前記信号成分の強さと前記2次関数との誤差に基づいて、前記指標を算出させることができる。

40

【0017】

前記指標算出部には、前記2次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。

【0018】

前記指標算出部には、前記2次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。

【0019】

前記楽曲区間検出装置には、所定時間分の前記入力信号の各領域の前記トーンらしさの

50

指標に基づいて、所定時間分の前記入力信号の特徴量を算出する特徴量算出部をさらに設け、前記楽曲判定部には、前記特徴量が所定の閾値より大きい場合、所定時間分の前記入力信号に楽曲が含まれていると判定させることができる。

【0020】

前記特徴量算出部には、所定時間分の前記入力信号の各領域の前記トーンらしさの指標を、周波数毎に時間方向に積算することで前記特徴量を算出させることができる。

【0021】

前記特徴量算出部には、所定時間分の前記入力信号の各領域において、周波数毎に、所定の閾値より大きい前記トーンらしさの指標が時間方向に最も連続する前記領域の前記トーンらしさの指標を積算することで前記特徴量を算出させることができる。

10

【0022】

前記楽曲区間検出装置には、前記特徴量を時間方向にフィルタリングするフィルタ処理部をさらに設け、前記楽曲判定部には、時間方向にフィルタリングされた前記特徴量が所定の閾値より大きい場合、所定時間分の前記入力信号に楽曲が含まれていると判定させることができる。

【0023】

本技術の一側面の楽曲区間検出方法は、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、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、前記トーンらしさの指標に基づいて、前記入力信号の各領域に楽曲が含まれているか否かを判定する楽曲判定ステップとを含む。

20

【0024】

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、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、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、前記トーンらしさの指標に基づいて、前記入力信号の各領域に楽曲が含まれているか否かを判定する楽曲判定ステップとを含む処理をコンピュータに実行させる。

【0025】

本技術の一側面の楽曲信号検出装置は、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、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を備える。

30

【0026】

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、トーンらしさの指標に基づいて、入力信号の各領域に楽曲が含まれているか否かが判定される。

【発明の効果】

【0027】

本技術の一側面によれば、入力信号から楽曲部分を精度良く検出することが可能となる。

40

【図面の簡単な説明】

【0028】

【図1】本技術を適用した楽曲区間検出装置の一実施の形態の構成を示すブロック図である。

【図2】指標算出部の機能構成例を示すブロック図である。

【図3】特徴量算出部の機能構成例を示すブロック図である。

【図4】楽曲区間検出処理について説明するフローチャートである。

【図5】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6】ピークの検出について説明する図である。

【図7】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 8】指標の調整関数について説明する図である。

【図 9】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 10】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 11】特徴量の算出について説明する図である。

【図 12】特徴量の算出について説明する図である。

【図 13】特徴量算出部の他の機能構成例を示すブロック図である。

【図 14】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 15】特徴量の算出について説明する図である。

【図 16】従来の手法による判定結果のフィルタリングについて説明する図である。

【図 17】楽曲区間検出装置の他の機能構成例を示すブロック図である。

10

【図 18】楽曲区間検出処理について説明するフローチャートである。

【図 19】特徴量のフィルタリングについて説明する図である。

【図 20】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】

【0029】

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. 楽曲区間検出装置の構成

2. 楽曲区間検出処理

3. その他の構成

20

【0030】

< 1. 楽曲区間検出装置の構成 >

図 1 は、本技術を適用した楽曲区間検出装置の一実施の形態の構成を示している。

【0031】

図 1 の楽曲区間検出装置 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から楽曲部分を検出し、その結果を出力する。

【0032】

楽曲区間検出装置 11 は、切り出し部 31、時間周波数変換部 32、指標算出部 33、特徴量算出部 34、および楽曲区間判定部 35 を備えている。

【0033】

切り出し部 31 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 32 に供給する。

30

【0034】

時間周波数変換部 32 は、切り出し部 31 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 33 に供給する。

【0035】

指標算出部 33 は、時間周波数変換部 32 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、特徴量算出部 34 に供給する。

40

【0036】

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（例えばパワースペクトル）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 33 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。

【0037】

特徴量算出部 34 は、指標算出部 33 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を算出し、楽曲区間

50

判定部 35 に供給する。

【0038】

楽曲区間判定部 35 は、特徴量算出部 34 からの特徴量に基づいて、切り出し部 31 によって切り出された所定時間分の入力信号に楽曲が含まれているか否かを判定し、その判定結果を出力する。

【0039】

[指標算出部の構成]

次に、図 2 を参照して、図 1 の指標算出部 33 の詳細な構成について説明する。

【0040】

図 2 の指標算出部 33 は、時間区間選択部 51、ピーク検出部 52、近似処理部 53、トーン度算出部 54、および出力部 55 を備えている。

10

【0041】

時間区間選択部 51 は、時間周波数変換部 32 から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 52 に供給する。

【0042】

ピーク検出部 52 は、時間区間選択部 51 によって選択された所定の時間区間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。

【0043】

近似処理部 53 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 52 によって検出されたピーク周辺の信号成分の強さ（例えばパワースペクトル）を、所定の関数で近似する。

20

【0044】

トーン度算出部 54 は、近似処理部 53 により近似された所定の関数と、ピーク検出部 52 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。

【0045】

出力部 55 は、トーン度算出部 54 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 55 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 31 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、特徴量算出部 34 に供給する。

30

【0046】

このようにして、切り出し部 31 により切り出された所定時間分の入力信号についての、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。

【0047】

[特徴量算出部の構成]

次に、図 3 を参照して、図 1 の特徴量算出部 34 の詳細な構成について説明する。

【0048】

図 3 の特徴量算出部 34 は、積算部 71、合算部 72、および出力部 73 を備えている。

40

【0049】

積算部 71 は、指標算出部 33 からトーンらしさの指標に対して、単位周波数毎に、所定の条件を満たすトーン度を積算し、合算部 72 に供給する。

【0050】

合算部 72 は、積算部 71 から、単位周波数毎のトーン度の積算値に対して、所定の条件を満たす積算値を合算し、出力部 73 に供給する。

【0051】

出力部 73 は、合算部 72 から合算値に対して所定の演算を施し、その結果を、切り

50

出し部 3 1 により切り出された所定時間分の入力信号の特徴量として、楽曲区間判定部 3 5 に供給する。

【 0 0 5 2 】

< 2 . 楽曲区間検出処理 >

次に、図 4 のフローチャートを参照して、楽曲区間検出装置 1 1 の楽曲区間検出処理について説明する。楽曲区間検出処理は、外部の装置等から入力信号が楽曲区間検出装置 1 1 に入力されると開始される。なお、楽曲区間検出装置 1 1 には、入力信号が、時間的に連続して入力される。

【 0 0 5 3 】

ステップ S 1 1 において、切り出し部 3 1 は、入力信号から所定時間（例えば、2 秒間）分の信号を切り出し、時間周波数変換部 3 2 に供給する。なお、ここで切り出された所定時間分の入力信号を、以下においては、適宜ブロックという。

【 0 0 5 4 】

ステップ S 1 2 において、時間周波数変換部 3 2 は、例えばハニング窓等の窓関数や、例えば離散フーリエ変換等を用いて、切り出し部 3 1 からの所定時間分の入力信号（ブロック）をスペクトログラムに変換し、指標算出部 3 3 に供給する。なお、ここでは、窓関数として、ハニング窓に限らず、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 5 5 】

ステップ S 1 3 において、指標算出部 3 3 は、指標算出処理を実行することで、時間周波数変換部 3 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。

【 0 0 5 6 】

[指標算出処理の詳細]

ここで、図 5 のフローチャートを参照して、図 4 のフローチャートのステップ S 1 3 における指標算出処理の詳細について説明する。

【 0 0 5 7 】

ステップ S 3 1 において、指標算出部 3 3 の時間区間選択部 5 1 は、時間周波数変換部 3 2 からの入力信号のスペクトログラムにおいて、ある 1 フレーム分のスペクトログラムを選択し、ピーク検出部 5 2 に供給する。なお、フレーム長は、例えば 1 6 ミリ秒などとされる。

【 0 0 5 8 】

ステップ S 3 2 において、ピーク検出部 5 2 は、時間区間選択部 5 1 によって選択された 1 フレーム分のスペクトログラムにおいて、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 5 9 】

例えば、図 6 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（1 つの四角形（マス目）は、各フレーム、各周波数のスペクトルを示す）において、太線のマス目で示される、あるフレームのある周波数で、図 6 の下段に示されるようなピーク p（具体的には、ピーク p を示す円で囲まれているスペクトルのうちの最大となるスペクトル）が検出されるようになる。なお、実際には、図 6 の上段に示されるマス目の縦の数は、図 6 の下段に示されるスペクトルの周波数方向（横軸方向）の数（黒丸の数）と等しくなる。

【 0 0 6 0 】

ステップ S 3 3 において、近似処理部 5 3 は、時間区間選択部 5 1 によって選択された 1 フレーム分のスペクトログラムについて、ピーク検出部 5 2 によって検出されたピーク

10

20

30

40

50

周辺のパワースペクトルを、2次関数で近似する。

【0061】

上述したように、図6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、サイドローブ、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。

【0062】

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、あるフレームでのピーク周辺の対数振幅スペクトルの値は、2次関数で近似することが可能であるとされている。

【0063】

そこで、本技術においては、ピーク周辺の対数振幅スペクトルを2次関数で近似する。

【0064】

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。

【0065】

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。

【0066】

したがって、持続性トーンは、図7で示されるような、あるフレームにおける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。

【0067】

【数1】

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

・・・（1）

【0068】

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る。すなわち、以下の式（2）を誤差関数とすることができる。

【0069】

【数2】

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

・・・（2）

【0070】

式（2）において、 $f(k, n)$ は、第nフレーム、第kピンのDFTスペクトルを示しており、 $g(k, n)$ は、上述で示された、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。

【0071】

10

20

30

40

50

【数 3】

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

・・・ (3)

【0072】

なお、式(2)における 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 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。

10

【0073】

図5のフローチャートに戻り、ステップS34において、トーン度算出部54は、近似処理部53により近似された2次関数と、ピーク検出部52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、時間区間選択部51によって選択された1フレーム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。

【0074】

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 は、以下の式(5)のように示される。

【0075】

20

【数 4】

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

・・・ (4)

【0076】

【数 5】

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

・・・ (5)

30

【0077】

式(5)において、aハット(aに^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。

【0078】

このようにして、トーン度 が求められる。

【0079】

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。

40

【0080】

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ブ性のものであったとしても、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 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 ' は、以下の式(6)のように示される。

50

【 0 0 8 1 】

【 数 6 】

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

・・・ (6)

【 0 0 8 2 】

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 8 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、トーン度は 0 となる。なお、関数 $D(x)$ は、図 8 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。

10

【 0 0 8 3 】

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、より適切なトーン度が得られるようになる。

【 0 0 8 4 】

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。

【 0 0 8 5 】

20

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してもよい。具体的には、式 (6) における関数 $D(x)$ において、 (a ハット) - a_{ideal} の項を、 - (b ハット) / 2 (a ハット) - k_p に置き換えるようにしてもよいし、関数 $D\{ - (b ハット) / 2 (a ハット) - k_p \}$ を式 (6) の左辺に乗算した値を、調整したトーン度 η' とするようによい。

30

【 0 0 8 6 】

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。

【 0 0 8 7 】

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。

【 0 0 8 8 】

【 数 7 】

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

40

・・・ (7)

【 0 0 8 9 】

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。

【 0 0 9 0 】

【数 8】

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

・・・ (8)

【0091】

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 は以下の式 (9) で与えられる。

10

【0092】

【数 9】

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

・・・ (9)

【0093】

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 8 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 ハットが 0 または a' ハットが 0 の場合にも、トーン度 は 0 となる。

20

【0094】

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。

【0095】

さて、図 5 のフローチャートに戻り、ステップ S 3 5 において、出力部 5 5 は、トーン度算出部 5 4 によって算出された、1 フレーム分のスペクトログラムについてのトーン度を保持し、1 ブロックにおける全てのフレームについて、上述した処理を行ったか否かを判定する。

30

【0096】

ステップ S 3 5 において、全てのフレーム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 S 3 1 に戻り、次のフレームのスペクトログラムに対して、ステップ S 3 1 乃至 S 3 5 の処理が繰り返される。

【0097】

一方、ステップ S 3 5 において、全てのフレーム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 S 3 6 に進む。

【0098】

ステップ S 3 6 において、出力部 5 5 は、保持しているフレーム毎のトーン度を時系列に配列して、特徴量算出部 3 4 に供給 (出力) し、処理はステップ S 1 3 に戻る。

40

【0099】

図 9 は、指標算出部 3 3 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。

【0100】

図 9 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 (フレーム)、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 9 に示されるように、トーンらしさの指標 S の時間粒度 (フレーム長) は、例えば 16 ミリ秒などとされる。

50

【0101】

このように、入力信号の1ブロックについて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。

【0102】

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8 kHz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。

10

【0103】

図4のフローチャートに戻り、ステップS13の後、ステップS14において、特徴量算出部34は、指標算出部33からのトーンらしさの指標に基づいて、特徴量算出処理を実行し、入力信号の楽曲らしさを表す特徴量を算出する。

【0104】

[特徴量算出処理の詳細]

ここで、図10のフローチャートを参照して、図4のフローチャートのステップS14における特徴量算出処理の詳細について説明する。

【0105】

ステップS51において、積算部71は、指標算出部33からのトーンらしさの指標に対して、周波数毎に、所定の閾値より大きいトーン度を積算し、合算部72に供給する。

20

【0106】

例えば、図11に示されるトーンらしさの指標Sが指標算出部33から供給された場合、積算部71は、まず、トーンらしさの指標Sにおいて最も低い周波数（すなわち、図中最も下の行）のトーン度に注目する。次に、積算部71は、その注目している周波数（以下、注目周波数という）のトーン度において、図中網かけで示されている、所定の閾値より大きいトーン度を、時間方向（図中左から右の方向）に順番に加算していく。なお、所定の閾値は、適宜、適切に設定されるものとし、例えば0であってもよい。そして、積算部71は、注目周波数を1つ上げ、その注目周波数に対して上述の処理を繰り返す。このようにして、注目周波数毎にトーン度の積算値が得られる。このトーン度の積算値は、楽曲の信号成分が含まれる周波数ほど、高い値となる。

30

【0107】

図10のフローチャートに戻り、ステップS52において、積算部71は、周波数毎のトーン度の積算の処理を、全ての周波数について行ったか否かを判定する。

【0108】

ステップS52において、全ての周波数について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、ステップS51、S52の処理が繰り返される。

【0109】

一方、ステップS52において、全ての周波数について処理が行われたと判定された場合、すなわち、図11のトーンらしさの指標Sにおいて、全ての周波数を注目周波数として積算値が算出された場合、積算部71は、周波数毎のトーン度の積算値Sfを合算部72に供給し、処理はステップS53に進む。

40

【0110】

ステップS53において、合算部72は、積算部71からの周波数毎のトーン度の積算値に対して、所定の閾値より大きい積算値を合算し、出力部73に供給する。

【0111】

例えば、図12に示される周波数毎のトーン度の積算値Sfが積算部71から供給された場合、合算部72は、周波数毎のトーン度の積算値Sfにおいて、図中網かけで示されている、所定の閾値より大きい積算値を、周波数方向（図中下から上方向）に順番に加算していく。なお、所定の閾値は、適宜、適切に設定されるものとし、例えば0であってもよ

50

い。そして、合算部 7 2 は、得られた合算値 S_b を出力部 7 3 に供給する。なお、合算部 7 2 は、周波数毎のトーン度の積算値 S_f において、所定の閾値より大きい積算値の数をカウントし、そのカウント値（図 1 1 の例では 5）を、合算値 S_b とともに出力部 7 3 に供給する。

【0 1 1 2】

ステップ S 5 4 において、出力部 7 3 は、合算部 7 2 からの合算値を、合算部 7 2 からのカウント値で除算した値を、切り出し部 3 1 により切り出された 1 ブロック分の入力信号の特徴量として、楽曲区間判定部 3 5 に供給する。すなわち、例えば、図 1 2 における合算値 S_b をカウント値 5 で割った値 S_m が、そのブロックの特徴量として算出される。

【0 1 1 3】

このようにして、入力信号のブロックについての、楽曲らしさを表す特徴量が算出されるようになる。

【0 1 1 4】

図 4 のフローチャートに戻り、ステップ S 1 4 の後、ステップ S 1 5 において、楽曲区間判定部 3 5 は、特徴量算出部 3 4 からの特徴量が、所定の閾値より大きいかなかを判定する。

【0 1 1 5】

ステップ S 1 5 において、特徴量が所定の閾値より大きいと判定された場合、処理はステップ S 1 6 に進み、楽曲区間判定部 3 5 は、切り出し部 3 1 により切り出されたブロックに対応する入力信号の時間区間が、楽曲が含まれる楽曲区間であると判定して、その旨

【0 1 1 6】

一方、ステップ S 1 5 において、特徴量が所定の閾値より大きくないと判定された場合、処理はステップ S 1 7 に進み、楽曲区間判定部 3 5 は、切り出し部 3 1 により切り出されたブロックに対応する入力信号の時間区間が、楽曲が含まれない非楽曲区間であると判定として、その旨の情報を出力する。

【0 1 1 7】

ステップ S 1 8 において、楽曲区間検出装置 1 1 は、全ての入力信号（ブロック）について、上述した処理を行ったか否かを判定する。

【0 1 1 8】

ステップ S 1 8 において、全ての入力信号について処理を行っていないと判定された場合、すなわち、入力信号が時間的に連続して入力され続けている場合、処理はステップ S 1 1 に戻り、これ以降の処理が繰り返される。

【0 1 1 9】

一方、全ての入力信号について処理を行ったと判定された場合、すなわち、入力信号の入力が終了した場合、処理も終了する。

【0 1 2 0】

以上の処理によれば、楽曲と雑音とが混在した入力信号から、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に基づいて、入力信号に楽曲が含まれる区間が検出されるようになる。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ため、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、楽曲と雑音とが混在する入力信号から、楽曲部分を精度良く検出することが可能となる。

【0 1 2 1】

< 3 . その他の構成 >

なお、以上においては、特徴量算出処理において得られる周波数毎のトーン度の積算値は、楽曲の信号成分が含まれる周波数ほど高い値となるものとしたが、ある注目周波数に、不連続的に、値の高いトーン度が含まれる場合であっても、その注目周波数のトーン度の積算値は、高い値となってしまう。トーン度は、フレーム毎の、時間方向に対するトーンの安定性を示すものであるが、複数のフレームについて、連続的にトーン度が高ければ

10

20

30

40

50

、トーンの安定性がより明確に示されるようになる。

【 0 1 2 2 】

そこで、以下においては、複数のフレームについての連続的なトーン度の高さを評価するようにした特徴量算出処理について説明する。

【 0 1 2 3 】

[特徴量算出部の他の構成]

まず、図 1 3 を参照して、複数のフレームについての連続的なトーン度の高さを評価するようにした特徴量算出処理を行う特徴量算出部 3 4 の構成について説明する。

【 0 1 2 4 】

なお、図 1 3 の特徴量算出部 3 4 において、図 3 の特徴量算出部 3 4 に設けられたものと同様の機能を備える構成については、同一名称および同一符号を付するものとし、その説明は、適宜省略するものとする。

10

【 0 1 2 5 】

すなわち、図 1 3 の特徴量算出部 3 4 において、図 3 の特徴量算出部 3 4 と異なるのは、積算部 7 1 に代えて積算部 9 1 を設けた点である。

【 0 1 2 6 】

積算部 9 1 は、指標算出部 3 3 からのトーンらしさの指標に対して、単位周波数毎に、時間的に最も連続して所定の条件を満たすトーン度を積算し、合算部 7 2 に供給する。

【 0 1 2 7 】

[特徴量算出処理の詳細]

次に、図 1 4 のフローチャートを参照して、図 1 3 の特徴量算出部 3 4 による特徴量算出処理の詳細について説明する。

20

【 0 1 2 8 】

なお、図 1 4 のフローチャートのステップ S 9 2 乃至 S 9 4 の処理は、図 1 0 のフローチャートのステップ S 5 2 乃至 S 5 4 の処理と基本的に同様であるので、その説明は省略する。

【 0 1 2 9 】

すなわち、ステップ S 9 1 において、積算部 9 1 は、指標算出部 3 3 からのトーンらしさの指標に対して、周波数毎に、所定の閾値より大きいトーン度が、時間方向に最も連続する時間区間のトーン度を積算し、合算部 7 2 に供給する。

30

【 0 1 3 0 】

例えば、図 1 5 に示されるトーンらしさの指標 S が指標算出部 3 3 から供給された場合、積算部 9 1 は、まず、トーンらしさの指標 S において最も低い周波数（すなわち、図中最も下の行）のトーン度に注目する。次に、積算部 9 1 は、その注目周波数のトーン度において、図中網かけで示されている、所定の閾値より大きいトーン度を、時間方向（図中左から右の方向）に順番に加算していく。このとき、積算部 9 1 は、まず、所定の閾値より大きいトーン度が時間的に連続する時間区間 t1 のトーン度を加算し、その数 2 をカウントする。積算部 9 1 は、同様にして、時間区間 t2、時間区間 t3 についても、トーン度を加算し、その数 3、2 をカウントする。そして、積算部 9 1 は、カウントした数のうち最も大きい数 3 に対応する時間区間 t2 のトーン度を加算した値を、注目周波数毎のトーン度の積算値とする。積算部 9 1 は、全ての周波数について、上述の処理を繰り返す。このようにして、注目周波数毎にトーン度の積算値が得られる。このトーン度の積算値は、楽曲の信号成分が含まれる周波数ほど高い値となり、さらに、トーンの安定性をより明確に示すものとなる。

40

【 0 1 3 1 】

これにより、楽曲らしさを表す特徴量の信頼性を高めることができ、ひいては、楽曲と雑音とが混在する入力信号から、楽曲部分をより精度良く検出することが可能となる。

【 0 1 3 2 】

このように、楽曲区間検出処理によって得られる楽曲区間の判定結果の信頼性は、高いものとなるが、特徴量が閾値近傍の値となる場合、楽曲区間と非楽曲区間とが頻繁に入れ

50

替わった判定結果が得られる可能性がある。そこで、従来、このような楽曲区間と非楽曲区間とが頻繁に入れ替わった判定結果に対して、メディアンフィルタ等を用いてフィルタリングすることにより、安定した判定結果を得るようにしていた。

【0133】

図16は、従来の手法による判定結果のフィルタリングについて説明する図である。

【0134】

図16の上段には、時間方向に対するブロック毎の特徴量が示されており、楽曲区間では高い値をとり、非楽曲区間では低い値をとっている。

【0135】

図16の中段には、図16の上段に示される特徴量を所定の閾値により2値化した、楽曲区間の判定結果が示されている。この判定結果においては、図16の上段に示される、非楽曲区間における特徴量の算出エラーによって、非楽曲区間にもかかわらず楽曲区間であると誤判定されている箇所が見られる。

10

【0136】

図16の下段には、図16の中段に示される判定結果に対してフィルタリングした結果が示されている。図16の下段に示されるように、フィルタリングすることにより、非楽曲区間における特徴量の算出エラーによる影響は除くことができるようになるが、図中右側の楽曲区間の、非楽曲区間との境界付近が、フィルタリングのエラーにより非楽曲区間として扱われてしまっている。

【0137】

20

このように、フィルタリングされた楽曲区間の判定結果の信頼性は、決して高いものとは言えなかった。

【0138】

そこで、以下においては、楽曲区間の判定結果の信頼性をより高めるようにした構成について説明する。

【0139】

[楽曲区間検出装置の他の構成]

図17は、楽曲区間の判定結果の信頼性を高めるようにした楽曲区間検出装置の構成を示している。

【0140】

30

なお、図17の楽曲区間検索装置111において、図1の楽曲区間検索装置11に設けられたものと同様の機能を備える構成については、同一名称および同一符号を付するものとし、その説明は、適宜省略するものとする。

【0141】

すなわち、図17の楽曲区間検索装置111において、図1の楽曲区間検索装置11と異なるのは、特徴量算出部34と楽曲区間判定部35との間に、フィルタ処理部131を新たに設けた点である。

【0142】

フィルタ処理部131は、特徴量算出部34からの特徴量に対してフィルタリングし、そのフィルタリングされた特徴量を、楽曲区間判定部35に供給する。

40

【0143】

なお、図17の楽曲区間検索装置111における特徴量算出部34の構成は、図3を参照して説明した構成であってもよいし、図13を参照して説明した構成であってもよい。

【0144】

[楽曲区間検出処理の詳細]

次に、図18のフローチャートを参照して、図17の楽曲区間検出装置111による楽曲区間検出処理の詳細について説明する。

【0145】

なお、図18のフローチャートのステップS111乃至S114の処理は、図4のフローチャートのステップS11乃至S14の処理と基本的に同様であるので、その説明は省

50

略する。また、図 18 のフローチャートのステップ S 1 1 4 における処理の詳細は、図 10 のフローチャートを参照して説明したものであってもよいし、図 14 のフローチャートを参照して説明したものであってもよい。

【0146】

なお、図 18 のフローチャートのステップ S 1 1 4 においては、特徴量算出部 34 は、算出した特徴量をブロック毎に保持する。

【0147】

ステップ S 1 1 5 において、楽曲区間検出装置 1 1 1 は、全ての入力信号（ブロック）について、ステップ S 1 1 1 乃至 S 1 1 4 の処理を行ったか否かを判定する。

【0148】

ステップ S 1 1 5 において、全ての入力信号について処理を行っていないと判定された場合、すなわち、入力信号が時間的に連続して入力され続けている場合、処理はステップ S 1 1 1 に戻り、ステップ S 1 1 1 乃至 S 1 1 4 の処理が繰り返される。

【0149】

一方、全ての入力信号について処理を行ったと判定された場合、すなわち、入力信号の入力が終了した場合、特徴量算出部 34 は、全てのブロックについての特徴量をフィルタ処理部 1 3 1 に供給し、処理はステップ S 1 1 6 に進む。

【0150】

ステップ S 1 1 6 において、フィルタ処理部 1 3 1 は、特徴量算出部 34 からの特徴量に対して、ローパスフィルタを用いてフィルタリングし、平滑化された特徴量を、楽曲区間判定部 35 に供給する。

【0151】

ステップ S 1 1 7 において、楽曲区間判定部 35 は、特徴量算出部 34 からの特徴量が、所定の閾値より大きいかな否かを、ブロック毎に順番に判定する。

【0152】

ステップ S 1 1 7 において、特徴量が所定の閾値より大きいと判定された場合、処理はステップ S 1 1 8 に進み、楽曲区間判定部 35 は、そのブロックに対応する入力信号の時間区間が、楽曲が含まれる楽曲区間であると判定して、その旨の情報を出力する。

【0153】

一方、ステップ S 1 1 6 において、特徴量が所定の閾値より大きくないと判定された場合、処理はステップ S 1 1 9 に進み、楽曲区間判定部 35 は、そのブロックに対応する入力信号の時間区間が、楽曲が含まれない非楽曲区間であると判定として、その旨の情報を出力する。

【0154】

ステップ S 1 2 0 において、楽曲区間検出装置 1 1 1 は、全ての入力信号（ブロック）の特徴量について、上述した処理を行ったか否かを判定する。

【0155】

ステップ S 1 2 0 において、全ての入力信号の特徴量について処理を行っていないと判定された場合、処理はステップ S 1 1 7 に戻り、次のブロックの特徴量についての処理が繰り返される。

【0156】

一方、全ての入力信号の特徴量について処理を行ったと判定された場合、処理は終了する。

【0157】

図 19 は、上述した楽曲区間検出処理における、特徴量に対するフィルタリングについて説明する図である。

【0158】

図 19 の上段には、図 16 の上段と同様に、時間方向に対するブロック毎の特徴量が示されている。

【0159】

10

20

30

40

50

図 19 の中段には、図 19 の上段に示される特徴量に対してフィルタリングした結果が示されている。図 19 の中段に示されるように、フィルタリングすることにより、図 19 の上段に示される非楽曲区間における特徴量の算出エラーが平滑化されるようになる。

【0160】

図 19 の下段には、図 19 の中段に示される特徴量を所定の閾値により 2 値化した、楽曲区間の判定結果が示されている。この判定結果においては、楽曲区間および非楽曲区間はそれぞれ正しく判定されている。

【0161】

特徴量は、パワースペクトルの時間に対する安定性を定量化したトーンらしさの指標から算出される、楽曲らしさを確実に表現する値であるので、上述したように、その特徴量に対してフィルタリングすることで、より信頼性の高い楽曲区間の判定結果を得ることができるようになる。

【0162】

なお、フィルタリングは、全てのブロックについての特徴量に対して行う必要はなく、目的に応じてフィルタリングするブロックを選択するようにしてもよい。

【0163】

例えば、図 17 の楽曲区間検出装置 111 において、図 4 の楽曲区間検出処理のように、一旦全ての入力信号について楽曲区間であるか否かを判定し、非楽曲区間と判定されたブロックの特徴量のみに対してフィルタリングするようにすることで、楽曲区間の検出漏れを減らして楽曲部分の再現率を高めることが可能となる。

【0164】

また、本技術は、図 1 に示される楽曲区間検出装置 11 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 1 の切り出し部 31 を設けるようにするとともに、サーバに、図 1 の切り出し部 31 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、サーバが楽曲区間検出処理を行うようにしてもよい。この場合、サーバは、判定結果を、インターネットを介して端末装置に送信し、端末装置は、サーバからの判定結果を表示部等に表示するようにもできる。

【0165】

また、以上においては、楽曲区間検出装置 11（楽曲区間検出装置 111）において、ブロック毎のトーンらしさの指標から求められる特徴量に基づいて、そのブロックが楽曲区間であるか否かを判定するようにしたが、切り出し部 31 乃至指標算出部 33 のみを備えるようにして、そのブロックにおける楽曲の信号成分を検出する楽曲信号検出装置として、楽曲区間検出装置 11（楽曲区間検出装置 111）を機能させるようにしてもよい。

【0166】

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。

【0167】

図 20 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。

【0168】

コンピュータにおいて、CPU（Central Processing Unit）901、ROM（Read Only Memory）902、RAM（Random Access Memory）903 は、バス 904 により相互に接続されている。

【0169】

バス 904 には、さらに、入出力インタフェース 905 が接続されている。入出力インタフェース 905 には、キーボード、マウス、マイクロホン等よりなる入力部 906、ディスプレイ、スピーカ等よりなる出力部 907、ハードディスクや不揮発性のメモリ等よりなる記憶部 908、ネットワークインタフェース等よりなる通信部 909、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 911 を駆動するドライブ 910 が接続されている。

【0170】

以上のように構成されるコンピュータでは、CPU 901 が、例えば、記憶部 908 に記憶されているプログラムを、入出力インタフェース 905 およびバス 904 を介して、RAM 903 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。

10

【0171】

コンピュータ (CPU 901) 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 911 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。

【0172】

そして、プログラムは、リムーバブルメディア 911 をドライブ 910 に装着することにより、入出力インタフェース 905 を介して、記憶部 908 にインストールすることができる。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。

20

【0173】

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。

【0174】

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。

30

【0175】

さらに、本技術は以下のような構成をとることができる。

(1) 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、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、

前記トーンらしさの指標に基づいて、前記入力信号の各領域に楽曲が含まれているか否かを判定する楽曲判定部と

を備える楽曲区間検出装置。

(2) 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、

40

前記極大となる点近傍の前記信号成分の強さを 2 次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 2 次関数との誤差に基づいて、前記指標を算出する

(1) に記載の楽曲区間検出装置。

(3) 前記指標算出部は、前記 2 次関数の曲率に応じて前記指標を調整する

(2) に記載の楽曲区間検出装置。

(4) 前記指標算出部は、前記 2 次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

50

(2) または (3) に記載の楽曲区間検出装置。

(5) 所定時間分の前記入力信号の各領域の前記トーンらしさの指標に基づいて、所定時間分の前記入力信号の特徴量を算出する特徴量算出部をさらに備え、

前記楽曲判定部は、前記特徴量が所定の閾値より大きい場合、所定時間分の前記入力信号に楽曲が含まれていると判定する

(1) 乃至 (4) のいずれかに記載の楽曲区間検出装置。

(6) 前記特徴量算出部は、所定時間分の前記入力信号の各領域の前記トーンらしさの指標を、周波数毎に時間方向に積算することで前記特徴量を算出する

(5) に記載の楽曲区間検出装置。

(7) 前記特徴量算出部は、所定時間分の前記入力信号の各領域において、周波数毎に、所定の閾値より大きい前記トーンらしさの指標が時間方向に最も連続する前記領域の前記トーンらしさの指標を積算することで前記特徴量を算出する

(5) に記載の楽曲区間検出装置。

(8) 前記特徴量を時間方向にフィルタリングするフィルタ処理部をさらに備え、

前記楽曲判定部は、時間方向にフィルタリングされた前記特徴量が所定の閾値より大きい場合、所定時間分の前記入力信号に楽曲が含まれていると判定する

(5) 乃至 (7) のいずれかに記載の楽曲区間検出装置。

(9) 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、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、

前記トーンらしさの指標に基づいて、前記入力信号の各領域に楽曲が含まれているか否かを判定する楽曲判定ステップと

を含む楽曲区間検出方法。

(10) 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、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、

前記トーンらしさの指標に基づいて、前記入力信号の各領域に楽曲が含まれているか否かを判定する楽曲判定ステップと

を含む処理をコンピュータに実行させるプログラム。

(11) (10) に記載のプログラムが記録されている記録媒体。

(12) 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、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

を備える楽曲信号検出装置。

【符号の説明】

【 0 1 7 6 】

1 1 楽曲区間検出装置， 3 1 切り出し部， 3 2 時間周波数変換部， 3 3 指標算出部， 3 4 特徴量算出部， 3 5 楽曲区間判定部， 5 2 ピーク検出部， 5 3 近似処理部， 5 4 トーン度算出部

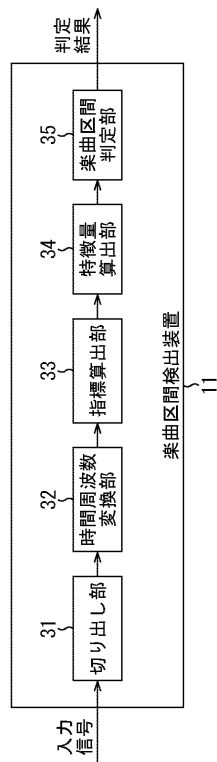
10

20

30

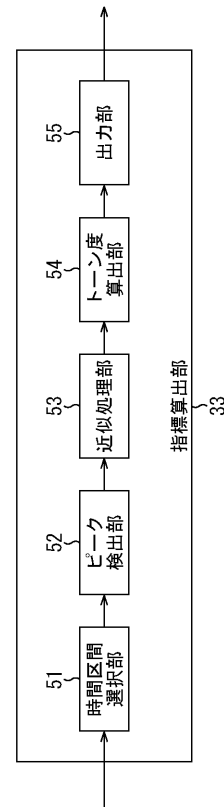
【図 1】

図1



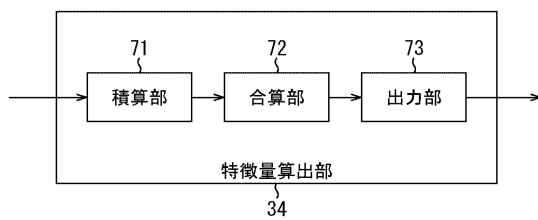
【図 2】

図2



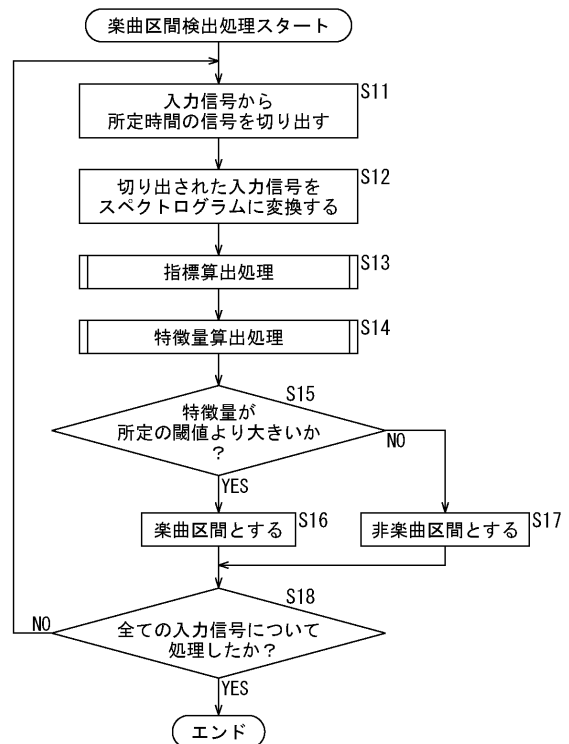
【図 3】

図3



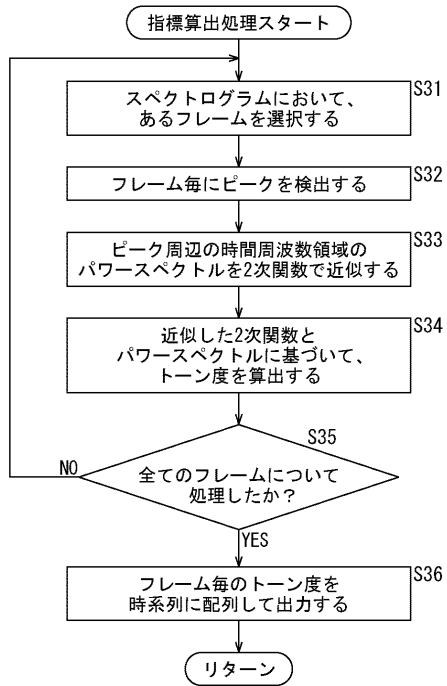
【図 4】

図4



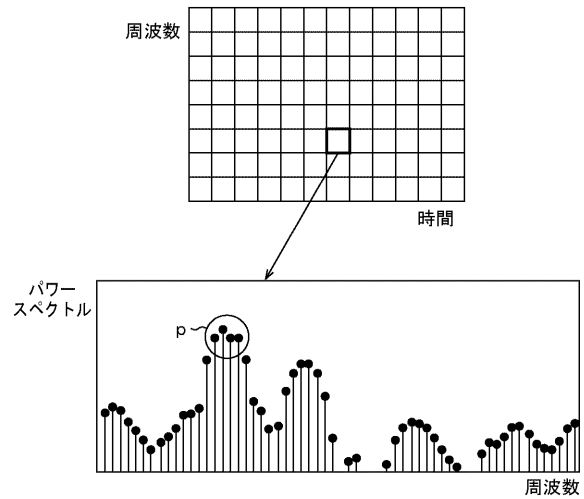
【図5】

図5



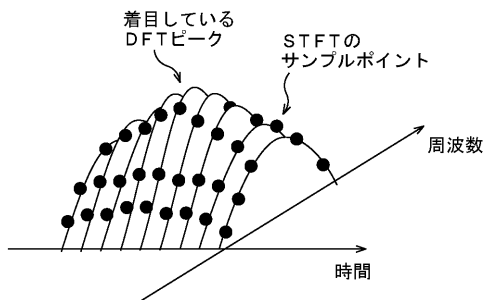
【図6】

図6



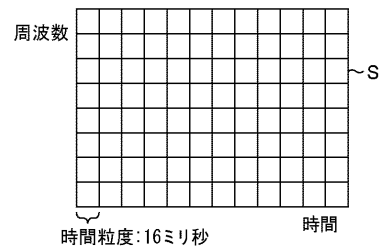
【図7】

図7



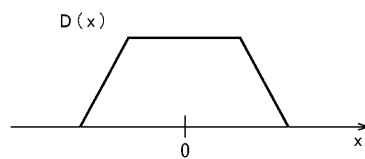
【図9】

図9



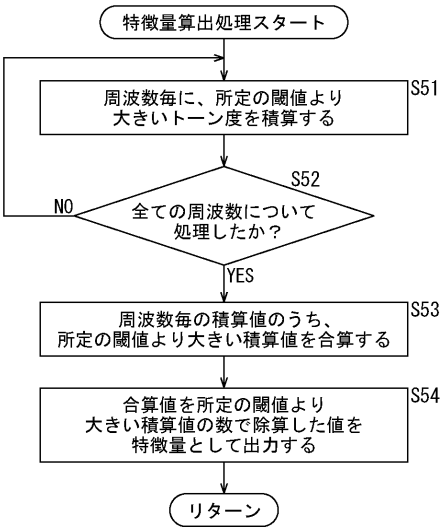
【図8】

図8



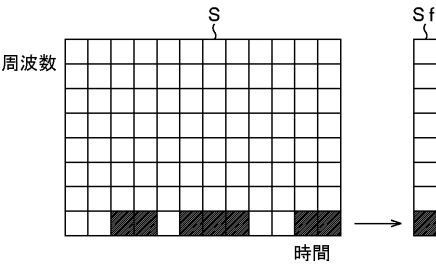
【図 1 0】

図10



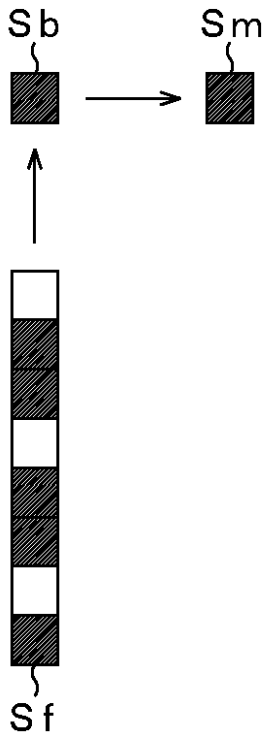
【図 1 1】

図11



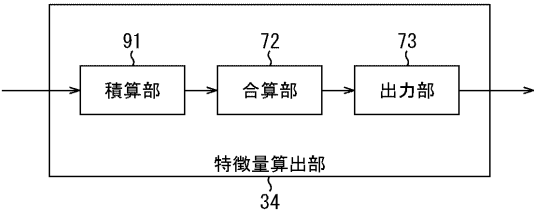
【図 1 2】

図12



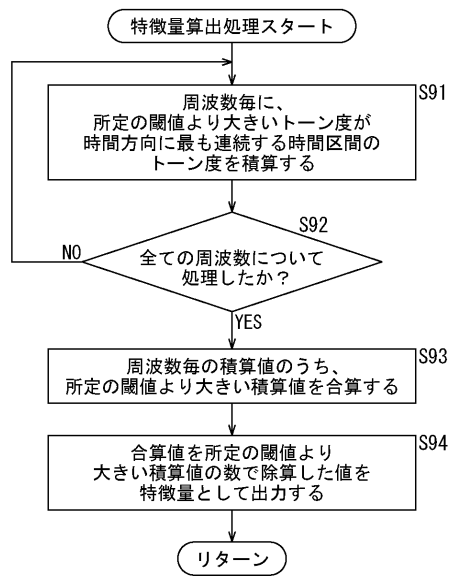
【図 1 3】

図13



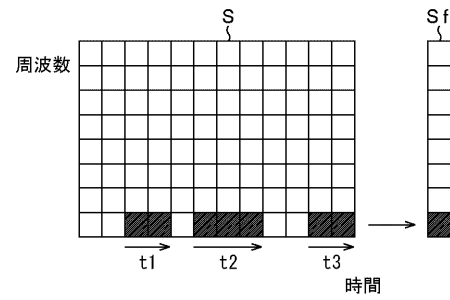
【図 14】

図14



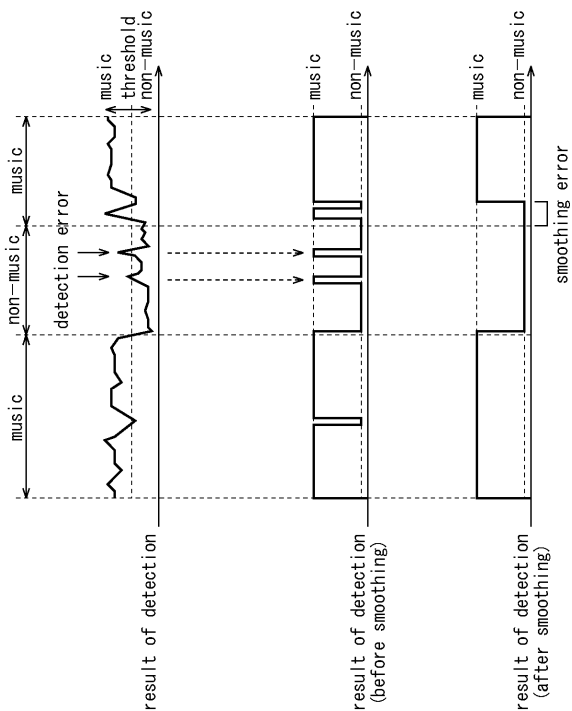
【図 15】

図15



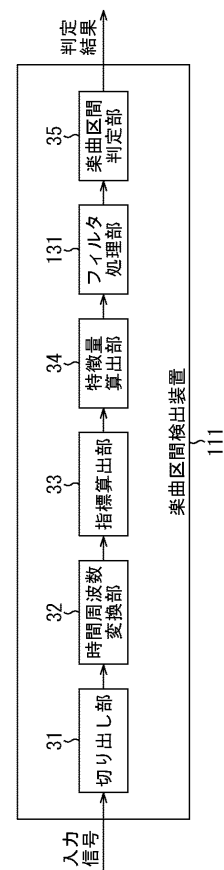
【図 16】

図16

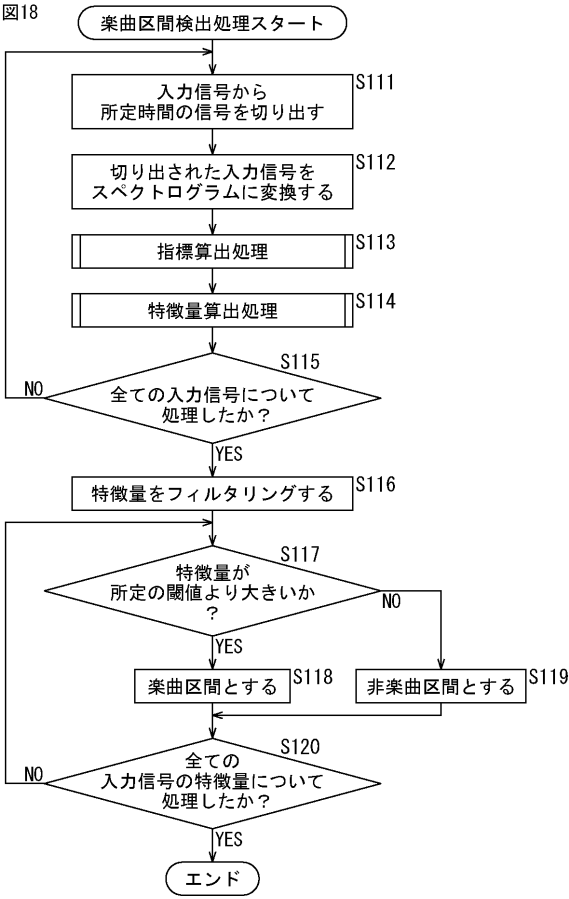


【図 17】

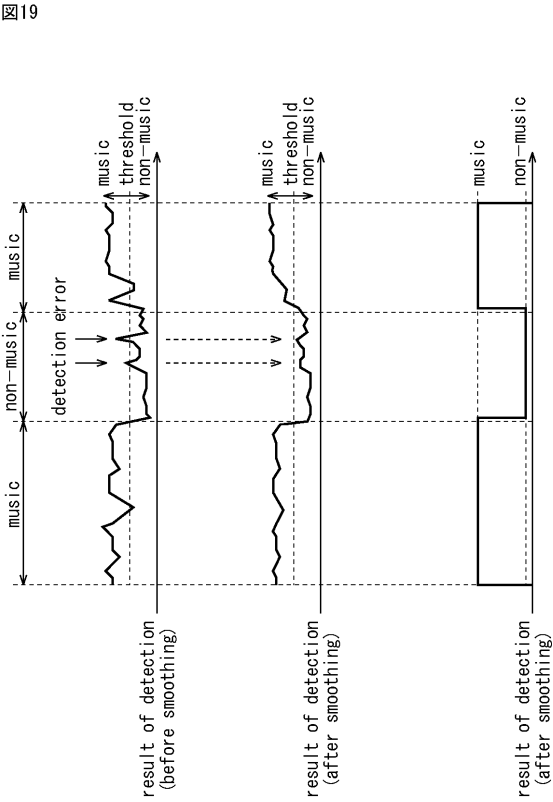
図17



【図 18】



【図 19】



【図 20】

