

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5998603号
(P5998603)

(45) 発行日 平成28年9月28日 (2016.9.28)

(24) 登録日 平成28年9月9日 (2016.9.9)

(51) Int. Cl.	F I	
G 1 O L 25/51 (2013.01)	G 1 O L 25/51	
G 1 O L 15/04 (2013.01)	G 1 O L 15/04	3 0 0 B
G 1 O L 25/03 (2013.01)	G 1 O L 25/03	

請求項の数 15 (全 31 頁)

(21) 出願番号	特願2012-94395 (P2012-94395)	(73) 特許権者	000002185
(22) 出願日	平成24年4月18日 (2012.4.18)		ソニー株式会社
(65) 公開番号	特開2013-222113 (P2013-222113A)		東京都港区港南1丁目7番1号
(43) 公開日	平成25年10月28日 (2013.10.28)	(74) 代理人	100093241
審査請求日	平成27年1月6日 (2015.1.6)		弁理士 宮田 正昭
		(74) 代理人	100101801
			弁理士 山田 英治
		(74) 代理人	100095496
			弁理士 佐々木 榮二
		(74) 代理人	100086531
			弁理士 澤田 俊夫
		(74) 代理人	110000763
			特許業務法人大同特許事務所

最終頁に続く

(54) 【発明の名称】 音検出装置、音検出方法、音特徴量検出装置、音特徴量検出方法、音区間検出装置、音区間検出方法およびプログラム

(57) 【特許請求の範囲】

【請求項1】

入力時間信号から所定時間毎の特徴量を抽出する特徴量抽出部と、
 所定数の被検出音の特徴量列を保持する特徴量保持部と、
 上記特徴量抽出部で新たに特徴量が抽出される毎に、該特徴量抽出部で抽出された特徴量の列を、上記保持されている所定数の被検出音の特徴量列とそれぞれ比較して、上記所定数の被検出音の検出結果を得る比較部とを備え、
 上記特徴量抽出部は、
 上記入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得る時間周波数変換部と、
 上記時間周波数分布よりトーンらしさの尤度分布を求める尤度分布検出部とを有し、
 上記求められた尤度分布を周波数方向および時間方向に平滑化して上記所定時間毎の特徴量を抽出する
 音検出装置。

【請求項2】

上記尤度分布検出部は、
 上記時間周波数分布の各時間フレームにおいて周波数方向のピークを検出するピーク検出部と、
 上記検出された各ピークにおいてトーンモデルをフィッティングするフィッティング部と、

上記フィッティング結果に基づき、上記検出された各ピークのトーン成分らしさを示すスコアを得るスコア化部とを備える

請求項 1 に記載の音検出装置。

【請求項 3】

上記特徴量抽出部は、

上記平滑化された尤度分布を周波数方向および / または時間方向に間引く間引き部をさらに備える

請求項 1 または請求項 2 に記載の音検出装置。

【請求項 4】

上記特徴量抽出部は、

上記平滑化された尤度分布を量子化する量子化部をさらに備える

請求項 1 または請求項 2 に記載の音検出装置。

【請求項 5】

上記比較部は、

上記所定数の被検出音のそれぞれについて、上記保持されている被検出音の特徴量列と上記特徴量抽出部で抽出された特徴量列との間の対応する特徴量間の相関演算で類似度を求め、該求められた類似度に基づいて上記被検出音の検出結果を得る

請求項 1 から請求項 4 のいずれかに記載の音検出装置。

【請求項 6】

上記所定数の被検出音の検出結果を時刻情報と共に記録媒体に記録する記録制御部をさらに備える

請求項 1 から請求項 5 のいずれかに記載の音検出装置。

【請求項 7】

入力時間信号から所定時間毎の特徴量を抽出する特徴量抽出ステップと、

上記特徴量抽出ステップで新たに特徴量が抽出される毎に、該特徴量抽出部で抽出された特徴量の列を、保持されている所定数の被検出音の特徴量列とそれぞれ比較して、上記所定数の被検出音の検出結果を得る比較ステップとを備え、

上記特徴量抽出ステップでは、

上記入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得、該時間周波数分布よりトーンらしさの尤度分布を求め、該尤度分布を周波数方向および時間方向に平滑化して上記所定時間毎の特徴量を抽出する

音検出方法。

【請求項 8】

コンピュータに、

入力時間信号から所定時間毎の特徴量を抽出する特徴量抽出ステップと、

上記特徴量抽出ステップで新たに特徴量が抽出される毎に、該特徴量抽出部で抽出された特徴量の列を、保持されている所定数の被検出音の特徴量列とそれぞれ比較して、上記所定数の被検出音の検出結果を得る比較ステップとを備え、

上記特徴量抽出ステップでは、

上記入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得、該時間周波数分布よりトーンらしさの尤度分布を求め、該尤度分布を周波数方向および時間方向に平滑化して上記所定時間毎の特徴量を抽出する

音検出方法を実行させるためのプログラム。

【請求項 9】

入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得る第 1 の時間周波数変換部と、

上記時間周波数分布よりトーンらしさの尤度分布を求める尤度分布検出部と、

上記尤度分布を周波数方向および時間方向に平滑化して所定時間毎の特徴量を抽出する第 1 の特徴量抽出部とを備える

音特徴量抽出装置。

10

20

30

40

50

【請求項 10】

上記尤度分布検出部は、
 上記時間周波数分布の各時間フレームにおいて周波数方向のピークを検出するピーク検出部と、
 上記検出された各ピークにおいてトーンモデルをフィッティングするフィッティング部と、
 上記フィッティング結果に基づき、上記検出された各ピークのトーン成分らしさを示すスコアを得るスコア化部とを備える
 請求項 9 に記載の音特徴量抽出装置。

10

【請求項 11】

上記平滑化された尤度分布を周波数方向および / または時間方向に間引く間引き部をさらに備える
 請求項 9 または請求項 10 に記載の音特徴量抽出装置。

【請求項 12】

上記平滑化された尤度分布を量子化する量子化部をさらに備える
 請求項 9 または請求項 10 に記載の音特徴量抽出装置。

【請求項 13】

上記入力時間信号に基づいて音区間を検出する音区間検出部をさらに備え、
 上記尤度分布検出部は、
 上記検出された音区間の範囲で上記時間周波数分布よりトーンらしさの尤度分布を求め
 る
 請求項 9 から請求項 12 のいずれかに記載の音特徴量抽出装置。

20

【請求項 14】

上記音区間検出部は、
 上記入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得る第 2 の時間周波数変換部と、
 上記時間周波数分布に基づいて、時間フレーム毎の、振幅、トーン成分強度およびスペクトル概形の特徴量を抽出する第 2 の特徴量抽出部と、
 上記抽出された特徴量に基づいて、時間フレーム毎の、音区間らしさを示すスコアを得るスコア化部と、
 上記得られた時間フレーム毎のスコアを時間方向に平滑化する時間平滑化部と、
 上記平滑化された時間フレーム毎のスコアを閾値判定して音区間情報を得る閾値判定部とを有する
 請求項 13 に記載の音特徴量抽出装置。

30

【請求項 15】

入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得る時間周波数変換ステップと、
 上記時間周波数分布よりトーンらしさの尤度分布を求める尤度分布検出ステップと、
 上記尤度分布を周波数方向および時間方向に平滑化する平滑化ステップとを備える
 音特徴量抽出方法。

40

【発明の詳細な説明】

【技術分野】

【0001】

本技術は、音検出装置、音検出方法、音特徴量検出装置、音特徴量検出方法、音区間検出装置、音区間検出方法およびプログラムに関する。

【背景技術】

【0002】

近年、家電（家庭用電気機器）は、操作音、通知音、動作音、警報音など、動作状況に応じて様々な音（以下、「動作状況音」という）を発する。この動作状況音を、家庭内のどこかに設置したマイクロフォン等で観測し、いつどの家電がどのような動作をしている

50

かを検出できれば、いわゆるライフログのような自身の行動履歴の自動収集、あるいは聴覚障害者などへの通知音の可視化、さらには、独居老人の行動見守りなど、様々な応用機能が実現できる。

【0003】

動作状況音は、単純なブザー音、ピープ音の場合や、音楽、音声などの場合もあり、その継続時間長は、短いものでは300ms程度から、長いものでは数十秒程度である。これらが、家電に装備された圧電ブザーや薄型スピーカなど、あまり音質の良くない再生デバイスを通じて再生され、空間に伝播される。

【0004】

例えば、特許文献1には、楽曲の一部断片データを時間周波数分布に変換して特徴量を抽出し、その特徴量を既に登録されている楽曲の特徴量と比較し、楽曲名を同定する技術が記載されている。

10

【先行技術文献】

【特許文献】

【0005】

【特許文献1】特許第4788810号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

特許文献1に記載されると同様の技術を、上述の動作状況音の検出に適用することも考えられる。しかし、家電から発せられる動作状況音に関しては、以下のような、その検出の妨げとなる事項が存在する。

20

【0007】

(1) 数百ミリ秒などの短い動作状況音も認識しなくてはならない。

(2) 再生デバイスの質が悪いため、音が割れていたり、共振が発生して周波数特性が極端に歪んでいたりとすることがある。

(3) 空間伝播により、家庭電化製品自体が発した音と比べて振幅・位相周波数特性が歪むことがある。例えば、図17(a)は、家庭電化製品に近い位置で録音した動作状況音の波形例を示している。これに対して、図17(b)は、家庭電化製品から遠い位置で録音した動作状況音の波形例を示しているが、歪んだものとなっている。

30

【0008】

(4) 空間伝播により、比較的大きな雑音、テレビの出力音、会話音などの非定常な雑音が重畳されることがある。例えば、図17(c)は、雑音原であるテレビの近い位置で録音した動作状況音の波形例を示しているが、動作状況音は雑音に埋もれてしまっている。

(5) 家庭電化製品毎の音の大きさやマイクロフォンまでの距離がそれぞれの家電に依存するため、録音される音の音量がまちまちなになる。

【0009】

本技術の目的は、家電から発生される動作状況音等の被検出音の良好な検出を可能することにある。

40

【課題を解決するための手段】

【0010】

本技術の概念は、

入力時間信号から所定時間毎の特徴量を抽出する特徴量抽出部と、

所定数の被検出音の特徴量列を保持する特徴量保持部と、

上記特徴量抽出部で新たに特徴量が抽出される毎に、該特徴量抽出部で抽出された特徴量の列を、上記保持されている所定数の被検出音の特徴量列とそれぞれ比較して、上記所定数の被検出音の検出結果を得る比較部とを備え、

上記特徴量抽出部は、

上記入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得る時間周

50

波数変換部と、

上記時間周波数分布よりトーンらしさの尤度分布を求める尤度分布検出部と、
上記尤度分布を周波数方向および時間方向に平滑化する平滑化部とを有し、
上記平滑化された尤度分布から上記所定時間毎の特徴量を抽出する
音検出装置にある。

【0011】

本技術において、特徴量抽出部により、入力時間信号から所定時間毎の特徴量が抽出される。この場合、特徴量抽出部では、入力時間信号が時間フレーム毎に時間周波数変換されて時間周波数分布が得られ、この時間周波数分布よりトーンらしさの尤度分布が求められ、この尤度分布が周波数方向および時間方向に平滑化され、この平滑化された尤度分布から所定時間毎の特徴量が抽出される。

10

【0012】

例えば、尤度分布検出部は、時間周波数分布の各時間フレームにおいて周波数方向のピークを検出するピーク検出部と、この検出された各ピークにおいてトーンモデルをフィッティングするフィッティング部と、このフィッティング結果に基づき、検出された各ピークのトーン成分らしさを示すスコアを得るスコア化部とを備える、ようにされてもよい。

【0013】

特徴量保持部には、所定数の被検出音の特徴量列が保持されている。この被検出音には、家庭電化製品から発せられる動作状況音（操作音、通知音、動作音、警報音など）の他に、人や動物の声音などを含めることができる。比較部により、特徴量抽出部で新たに特徴量が抽出される毎に、この特徴量抽出部で抽出された特徴量の列が、保持されている所定数の被検出音の特徴量列とそれぞれ比較されて、この所定数の被検出音の検出結果が得られる。

20

【0014】

例えば、比較部は、所定数の被検出音のそれぞれについて、保持されている被検出音の特徴量列と特徴量抽出部で抽出された特徴量列との間の対応する特徴量間の相関演算で類似度を求め、この求められた類似度に基づいて被検出音の検出結果を得る、ようにされてもよい。

【0015】

このように本技術においては、入力時間信号の時間周波数分布よりトーンらしさの尤度分布を求め、この尤度分布を周波数方向および時間方向に平滑化したものから所定時間毎の特徴量を抽出して用いるものであり、被検出音（家庭用電化製品から発せられる動作状況音など）の検出を、マイクロフォンの設置位置などに依らずに、精度よく行うことが可能となる。

30

【0016】

なお、本技術において、例えば、特徴量抽出部は、平滑化された尤度分布を周波数方向および/または時間方向に間引く間引き部をさらに備える、ようにされてもよい。また、本技術において、例えば、特徴量抽出部は、平滑化された尤度分布を量子化する量子化部をさらに備える、ようにされてもよい。この場合、特徴量列のデータ量を低減でき、比較演算の負荷を軽減可能となる。

40

【0017】

また、本技術において、例えば、所定数の被検出音の検出結果を時刻情報と共に記録媒体に記録する記録制御部をさらに備える、ようにされてもよい。この場合、例えば、家庭用電化製品の動作履歴など、従って家庭内におけるユーザの行動履歴の取得が可能となる。

【0018】

また、本技術の他の概念は、

入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得る時間周波数変換部と、

上記時間周波数分布よりトーンらしさの尤度分布を求める尤度分布検出部と、

50

上記尤度分布を周波数方向および時間方向に平滑化して所定時間毎の特徴量を抽出する特徴量抽出部とを備える

音特徴量抽出装置にある。

【0019】

本技術において、時間周波数変換部により、入力時間信号が時間フレーム毎に時間周波数変換されて時間周波数分布が得られる。尤度分布検出部により、この時間周波数分布よりトーンらしさの尤度分布が求められる。例えば、尤度分布検出部は、時間周波数分布の各時間フレームにおいて周波数方向のピークを検出するピーク検出部と、この検出された各ピークにおいてトーンモデルをフィッティングするフィッティング部と、このフィッティング結果に基づき、検出された各ピークのトーン成分らしさを示すスコアを得るスコア化部とを備える、ようにされてもよい。そして、特徴量抽出部により、尤度分布が周波数方向および時間方向に平滑化されて所定時間毎の特徴量が抽出される。

10

【0020】

このように本技術においては、入力時間信号の時間周波数分布よりトーンらしさの尤度分布を求め、この尤度分布を周波数方向および時間方向に平滑化したものから所定時間毎の特徴量を抽出するものであり、入力時間信号に含まれる音の特徴量を良好に抽出できる。

【0021】

なお、本技術において、例えば、特徴量抽出部は、平滑化された尤度分布を周波数方向および/または時間方向に間引く間引き部をさらに備える、ようにされてもよい。また、本技術において、例えば、特徴量抽出部は、平滑化された尤度分布を量子化する量子化部をさらに備える、ようにされてもよい。これにより、抽出される特徴量のデータ量の低減が可能となる。

20

【0022】

また、本技術において、例えば、入力時間信号に基づいて音区間を検出する音区間検出部をさらに備え、尤度分布検出部は、検出された音区間の範囲で時間周波数分布よりトーンらしさの尤度分布を求める、ようにされてもよい。これにより、音区間に対応した特徴量を抽出することが可能となる。

【0023】

この場合、音区間検出部は、入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得る時間周波数変換部と、この時間周波数分布に基づいて、時間フレーム毎の、振幅、トーン成分強度およびスペクトル概形の特徴量を抽出する特徴量抽出部と、この抽出された特徴量に基づいて、時間フレーム毎の、音区間らしさを示すスコアを得るスコア化部と、この得られた時間フレーム毎のスコアを時間方向に平滑化する時間平滑化部と、この平滑化された時間フレーム毎のスコアを閾値判定して音区間情報を得る閾値判定部とを有する、ようにされてもよい。

30

【0024】

また、本技術の他の概念は、

入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得る時間周波数変換部と、

40

上記時間周波数分布に基づいて、時間フレーム毎の、振幅、トーン成分強度およびスペクトル概形の特徴量を抽出する特徴量抽出部と、

上記抽出された特徴量に基づいて、時間フレーム毎の、音区間らしさを示すスコアを得るスコア化部とを備える

音区間検出装置にある。

【0025】

本技術において、時間周波数変換部により、入力時間信号が時間フレーム毎に時間周波数変換されて時間周波数分布が得られる。特徴量抽出部により、時間周波数分布に基づいて、時間フレーム毎の、振幅、トーン成分強度およびスペクトル概形の特徴量が抽出される。そして、スコア化部により、抽出された特徴量に基づいて、時間フレーム毎の、音区

50

間らしさを示すスコアが得られる。なお、本技術において、例えば、得られた時間フレーム毎のスコアを時間方向に平滑化する時間平滑化部と、この平滑化された時間フレーム毎のスコアを閾値判定して音区間情報を得る閾値判定部とをさらに備える、ようにされてもよい。

【0026】

このように本技術においては、入力時間信号の時間周波数分布より時間フレーム毎の、振幅、トーン成分強度およびスペクトル概形の特徴量を抽出し、この特徴量から時間フレーム毎の、音区間らしさを示すスコアを得るものであり、音区間情報を精度よく得ることができる。

【発明の効果】

10

【0027】

本技術によれば、家庭電化製品から発せられる動作状況音等の被検出音の検出を良好に行うことができる。

【図面の簡単な説明】

【0028】

【図1】実施の形態としての音検出装置の構成例を示すブロック図である。

【図2】特徴量登録装置の構成例を示すブロック図である。

【図3】音区間とその前後に存在するノイズ区間の一例を示す図である。

【図4】特徴量登録装置を構成する音区間検出部の構成例を示すブロック図である。

【図5】トーン強度特徴量計算部を説明するための図である。

20

【図6】トーン強度特徴量計算部に含まれるトーン性らしさのスコア $S(n,k)$ の分布を得るためのトーン尤度分布検出部の構成例を示すブロック図である。

【図7】2次元多項式関数がトーン性のスペクトルピーク近傍ではよく当てはまるが、ノイズ性のスペクトルピーク近傍ではあまりよく当てはまらないという性質を説明するための模式図である。

【図8】トーン性ピークの時間方向への変化と、スペクトログラム上の小領域内でのフィッティングを模式的に示す図である。

【図9】トーン尤度分布検出部におけるトーン尤度分布検出の処理手順の一例を示すフローチャートである。

【図10】トーン成分検出結果の一例を示す図である。

30

【図11】音声のスペクトログラムの一例を示す図である。

【図12】特徴量抽出部の構成例を示すブロック図である。

【図13】音検出部の構成例を示すブロック図である。

【図14】音検出部の各部の動作を説明するための図である。

【図15】音検出処理をソフトウェアで行うコンピュータ装置の構成例を示すブロック図である。

【図16】CPUによる被検出音の検出処理の手順の一例を示すフローチャートである。

【図17】家庭電化製品自体が発した音の録音状態を説明するための図である。

【発明を実施するための形態】

【0029】

40

以下、発明を実施するための形態（以下、「実施の形態」とする）について説明する。なお、説明を以下の順序で行う。

1. 実施の形態

2. 変形例

【0030】

< 1. 実施の形態 >

[音検出装置]

図1は、実施の形態としての音検出装置100の構成例を示している。この音検出部100は、マイクロフォン101と、音検出部102と、特徴量データベース103と、記録・表示部104を有している。

50

【 0 0 3 1 】

この音検出装置 1 0 0 は、家電から発せられる動作状況音（操作音、通知音、動作音、警報音など）を検出する音検出プロセスを実行し、検出結果の記録および表示を行う。すなわち、この音検出プロセスでは、マイクロフォン 1 0 1 で集音されて得られる時間信号 $f(t)$ から所定時間毎の特徴量が抽出され、特徴量データベースに登録されている所定数の被検出音の特徴量列と比較される。そして、この音検出プロセスでは、所定の被検出音の特徴量列と概ね一致するとの比較結果が得られた場合、その時刻とその所定の被検出音の名が記録および表示される。

【 0 0 3 2 】

マイクロフォン 1 0 1 は、室内の音を集音し、時間信号 $f(t)$ を出力する。この室内の音には、家電 1 ~ 家電 N から発せられる動作状況音（操作音、通知音、動作音、警報音など）も含まれる。音検出部 1 0 2 は、マイクロフォン 1 0 1 から出力される時間信号 $f(t)$ を入力とし、この時間信号から所定時間毎の特徴量を抽出する。この意味で、音検出部 1 0 2 は、特徴量抽出部を構成する。

10

【 0 0 3 3 】

特徴量保持部を構成する特徴量データベース 1 0 3 には、所定数の被検出音の特徴量列が、被検出音名と対応付けられて登録され保持されている。この実施の形態において、この所定数の被検出音は、例えば、家電 1 ~ 家電 N で発生される動作状況音の全部あるいは一部である。音検出部 1 0 2 は、新たな特徴量を抽出する毎に、抽出された特徴量の列を、特徴量データベース 1 0 3 に保持されている所定数の被検出音の特徴量列のそれぞれと比較して、所定数の被検出音の検出結果を得る。この意味で、音検出部 1 0 2 は、比較部を構成している。

20

【 0 0 3 4 】

記録・表示部 1 0 4 は、音検出部 1 0 2 における被検出音の検出結果を、時刻と共に記録媒体に記録し、また、ディスプレイに表示する。例えば、音検出部 1 0 2 における被検出音の検出結果が家電 1 の通知音 A が検出されたことを示している場合、記録・表示部 1 0 4 は、そのときの時刻と家電 1 の通知音 A が鳴った旨を、記録媒体に記録し、また、ディスプレイに表示する。

【 0 0 3 5 】

図 1 に示す音検出装置 1 0 0 の動作を説明する。マイクロフォン 1 0 1 では、室内の音が集音される。このマイクロフォン 1 0 1 から出力される時間信号は音検出部 1 0 2 に供給される。音検出部 1 0 2 では、この時間信号から所定時間毎の特徴量が抽出される。そして、この音検出部 1 0 2 では、新たな特徴量が抽出される毎に、抽出された特徴量の列が、特徴量データベース 1 0 3 に保持されている所定数の被検出音の特徴量列のそれぞれと比較され、所定数の被検出音の検出結果が得られる。この検出結果は、記録・表示部 1 0 4 に供給される。記録・表示部 1 0 4 では、その検出結果が、時刻と共に記録媒体に記録され、また、ディスプレイに表示される。

30

【 0 0 3 6 】

[特徴量登録装置]

図 2 は、特徴量データベース 1 0 3 に、被検出音の特徴量列を登録する特徴量登録装置 2 0 0 の構成例を示している。この特徴量登録装置 2 0 0 は、マイクロフォン 2 0 1 と、音区間検出部 2 0 2 と、特徴量抽出部 2 0 3 と、特徴量登録部 2 0 4 を有している。

40

【 0 0 3 7 】

この特徴量登録装置 2 0 0 は、音登録プロセス（音区間検出プロセスおよび音特徴抽出プロセス）を実行し、被検出音（家電から発せられる動作状況音）の特徴量列を特徴量データベース 1 0 3 に登録する。通例、マイクロフォン 2 0 1 で録音される登録すべき被検出音の前後にはノイズ区間が存在する。そのため、音区間検出プロセスでは、実際に登録すべき有意な音（被検出音）のある音区間が検出される。図 3 は、音区間とその前後に存在するノイズ区間の一例を示している。また、音特徴抽出プロセスでは、マイクロフォン 2 0 1 から得られるその音区間の時間信号 $f(t)$ から、被検出音の検出に有用な特徴量が

50

抽出され、被検出音名と共に特徴量データベース 103 に登録される。

【0038】

マイクログフォン 201 は、被検出音として登録すべき家電の動作状況音を集音する。音区間検出部 202 は、マイクログフォン 201 から出力される時間信号 $f(t)$ を入力とし、この時間信号 $f(t)$ から音区間、すなわち家電から発せられる動作状況音の区間を検出する。特徴量抽出部 203 は、マイクログフォン 201 から出力される時間信号 $f(t)$ を入力とし、この時間信号 $f(t)$ から所定時間毎の特徴量を抽出する。

【0039】

特徴量抽出部 203 は、入力時間信号 $f(t)$ を時間フレーム毎に時間周波数変換して時間周波数分布を得、この時間周波数分布よりトーンらしさの尤度分布を求め、この尤度分布を周波数方向および時間方向に平滑化して所定時間毎の特徴量を抽出する。この場合、特徴量抽出部 203 は、音区間検出部 202 から供給される音区間情報に基づいて音区間の範囲で特徴量を抽出し、家電から発せられる動作状況音の区間に対応した特徴量の列を得る。

10

【0040】

特徴量登録部 204 は、特徴量抽出部 203 で得られた、被検出音としての家電で発せられる動作状況音に対応した特徴量列を、その被検出音名（動作状況音の情報）に対応付けて、特徴量データベース 103 に登録する。図示の例では、特徴量データベース 103 に、I 個の被検出音の特徴量列 $Z1(m)$, $Z2(m)$, \dots , $Zi(m)$, \dots , $Zl(m)$ が登録されている状態を示している。

20

【0041】

「音区間検出部」

図 4 は、音区間検出部 202 の構成例を示している。この音区間検出部 202 の入力は、登録すべき被検出音（家電で発せられる動作状況音）をマイクログフォン 201 で録音して得られる時間信号 $f(t)$ であり、図 3 示すように、前後にノイズ区間も含まれる。また、この音区間検出部 202 の出力は、実際に登録すべき有意な音（被検出音）のある音区間を示す音区間情報である。

【0042】

この音区間検出部 202 は、時間周波数変換部 221 と、振幅特徴量計算部 222 と、トーン強度特徴量計算部 223 と、スペクトル概形特徴量計算部 224 と、スコア計算部 225 と、時間平滑化部 226 と、閾値判定部 227 を有している。

30

【0043】

時間周波数変換部 221 は、入力時間信号 $f(t)$ を時間周波数変換して、時間周波数信号 $F(n,k)$ を得る。ここで、 t は離散時間、 n は時間フレームの番号、 k は離散周波数を表す。時間周波数変換部 221 は、例えば、以下の数式 (1) に示すように、短時間フーリエ変換により、入力時間信号 $f(t)$ を時間周波数変換し、時間周波数信号 $F(n,k)$ を得る。

【0044】

【数 1】

$$F(n,k) = \log \left| \sum_{t=0}^{M-1} W(t)f(t-nR)e^{j2\pi kn} \right| \quad \dots(1)$$

40

ただし、 $W(t)$ は窓関数、 M は窓関数のサイズ、 R はフレーム時間間隔 (= ホップサイズ) を表す。時間周波数信号 $F(n,k)$ は、時間フレーム n 、周波数 k における周波数成分の対数振幅値を表すものであり、いわゆるスペクトログラム (時間周波数分布) である。

【0045】

振幅特徴量計算部 222 は、時間周波数信号 $F(n,k)$ より、振幅特徴量 $x0(n)$, $x1(n)$ を計算する。具体的には、振幅特徴量計算部 222 は、所定の周波数範囲 (下限 KL 、上限 KH) について、以下の数式 (2) で表される、対象フレーム n の近傍時間区間 (前後

50

に長さ L とする) の平均振幅 $A_{ave}(n)$ を求める。

【数 2】

$$A_{ave}(n) = \frac{1}{2L+1} \sum_{n=-L}^L \sum_{k=K_L}^{K_H} \exp(F(n,k)) \quad \cdots(2)$$

【0046】

また、振幅特徴量計算部 222 は、所定の周波数範囲(下限 K_L 、上限 K_H)について、以下の数式(3)で表される、対象フレーム n における絶対振幅 $A_{abs}(n)$ を求める。

10

【数 3】

$$A_{abs}(n) = \sum_{k=K_L}^{K_H} \exp(F(n,k)) \quad \cdots(3)$$

【0047】

さらに、振幅特徴量計算部 222 は、所定の周波数範囲(下限 K_L 、上限 K_H)について、以下の数式(4)で表される、対象フレーム n における相対振幅 $A_{rel}(n)$ を求める。

【数 4】

20

$$A_{rel}(n) = \frac{A_{abs}(n)}{A_{ave}(n)} \quad \cdots(4)$$

【0048】

そして、振幅特徴量計算部 222 は、以下の数式(5)に示すように、絶対振幅 $A_{abs}(n)$ を振幅特徴量 $x_0(n)$ とし、相対振幅 $A_{rel}(n)$ を振幅特徴量 $x_1(n)$ とする。

【数 5】

$$x_0(n) = A_{abs}(n), \quad x_1(n) = A_{rel}(n) \quad \cdots(5)$$

30

【0049】

トーン強度特徴量計算部 223 は、時間周波数信号 $F(n,k)$ より、トーン強度特徴量 $x_2(n)$ を計算する。トーン強度特徴量計算部 223 は、まず、時間周波数信号 $F(n,k)$ の分布(図 5(a) 参照)を、トーン性らしさのスコア $S(n,k)$ の分布(図 5(b) 参照)に変換する。スコア $S(n,k)$ は、 $F(n,k)$ の各時間 n 、各周波数 k にて、その時間周波数成分がどの程度「トーン成分らしいか」を 0 から 1 の間のスコアで表したものである。具体的には、スコア $S(n,k)$ は、 $F(n,k)$ が周波数方向にトーン性のピークを形成する位置では 1 に近く、それ以外の位置では 0 に近い値をとるものである。

【0050】

40

図 6 は、トーン強度特徴量計算部 223 に含まれる、トーン性らしさのスコア $S(n,k)$ の分布を得るためのトーン尤度分布検出部 230 の構成例を示している。このトーン尤度分布検出部 230 は、ピーク検出部 231 と、フィッティング部 232 と、特徴量抽出部 233 と、スコア化部 234 を有している。

【0051】

ピーク検出部 231 で、スペクトログラム(時間周波数信号 $F(n,k)$ の分布)の各時間フレームにおいて、周波数方向のピークが検出される。すなわち、ピーク検出部 231 では、このスペクトログラムに対し、全てのフレーム、全ての周波数で、その位置が周波数方向に関してのピーク(極大値)であるか否かが検出される。

【0052】

50

$F(n,k)$ がピークであるか否かの検出は、例えば、以下の数式(6)を満足するか否かを確認することで行われる。なお、ピークの検出方法として3点を使った方法を示しているが、5点を使った方法であってもよい。

【数6】

$$F(n,k-1) < F(n,k) \text{ かつ } F(n,k) > F(n,k+1) \quad \dots(6)$$

【0053】

フィッティング部232では、ピーク検出部231で検出された各ピークに関し、以下のように、そのピークの近傍領域においてトーンモデルがフィッティングされる。まず、

10

フィッティング部232では、対象とするピークを原点とする座標に座標変換することが行われ、以下の数式(7)に示すように、近傍の時間周波数領域が設定される。ここで、

N は時間方向の近傍領域(例えば3点)、 k は周波数方向の近傍領域(例えば2点)を表す。

【数7】

$$\Gamma = [-\Delta_N \leq n \leq \Delta_N] \times [-\Delta_K \leq k \leq \Delta_K] \quad \dots(7)$$

【0054】

続いて、フィッティング部232では、近傍領域内の時間周波数信号に対し、例えば、

20

以下の数式(8)に示すような2次多項式関数のトーンモデルがフィッティングされる。この場合、フィッティング部232では、例えば、ピーク近傍の時間周波数分布とトーンモデルの二乗誤差最小基準によりフィッティングが行われる。

【数8】

$$Y(k,n) = ak^2 + bk + ckn + dn^2 + en + g \quad \dots(8)$$

【0055】

すなわち、フィッティング部232では、時間周波数信号と多項式関数の近傍領域内における、以下の数式(9)に示すような二乗誤差を最小にする係数が、以下の数式(10)

30

に示すように求められることで、フィッティングが行われる。

【数9】

$$J(a,b,c,d,e,g) = \sum_{\Gamma} (Y(k,n) - F(k,n))^2 \quad \dots(9)$$

$$(\hat{a}, \hat{b}, \hat{c}, \hat{d}, \hat{e}, \hat{g}) = \arg \min J(a,b,c,d,e,g) \quad \dots(10)$$

【0056】

40

この2次多項式関数は、トーン性のスペクトルピーク近傍では、よく当てはまる(誤差が小さい)が、ノイズ性のスペクトルピーク近傍ではあまりよく当てはまらない(誤差が大きい)、という性質をもつ。図7(a)、(b)は、その様子を模式的に示している。図7(a)は、上述の数式(1)で得られる、第 n フレームのトーン性ピーク付近のスペクトルを模式的に示している。

【0057】

図7(b)は、図7(a)のスペクトルに対して、以下の数式(11)で示される2次関数 $f_0(k)$ を当てはめる様子を示している。ただし、 a がピーク曲率、 k_0 が真のピークの周波数、 g_0 が真のピーク位置での対数振幅値である。トーン性の成分のスペクトルピークでは2次関数がよく当てはまるが、ノイズ性のピークでは、ずれが大きい傾向がある

50

【数 1 0】

$$f_0(k) = a(k - k_0)^2 + g_0 \quad \dots(11)$$

【 0 0 5 8】

図 8 (a) は、トーン性ピークの時間方向への変化を模式的に示している。トーン性ピークは、前後の時間フレームで、その概形を保ったまま振幅および周波数が変化をしていく。なお、実際に得られるスペクトルは離散点だが、便宜的に曲線で示している。一点鎖線が前フレーム、実線が現フレーム、点線が次フレームである。 10

【 0 0 5 9】

多くの場合、トーン性の成分はある程度の時間の持続性があり、多少の周波数変化や時間変化を伴うものの、ほぼ同じ形の 2 次関数のシフトで表すことができる。この変化 $Y(k, n)$ は、以下の数式 (1 2) で表される。スペクトルを対数振幅で表しているため、振幅の変化はスペクトルの上下への移動になる。振幅変化項 $f_1(n)$ が加算となるのはそのためである。ただし、 β は周波数の変化率、 $f_1(n)$ はピーク位置における振幅の変化を表す時間関数である。

【数 1 1】

$$Y(k, n) = f_0(k - \beta n) + f_1(n) \quad \dots(12)$$

20

【 0 0 6 0】

この変化 $Y(k, n)$ は、 $f_1(n)$ を時間方向の 2 次関数で近似すると、以下の数式 (1 3) で表される。 a 、 k_0 、 d_1 、 e_1 、 g_0 は定数なので、適切に変数変換をすることで、この数式 (1 3) は、上述の数式 (8) 式と等価となる。

【数 1 2】

$$\begin{aligned} Y(k, n) &= a(k - k_0 - \beta n)^2 + g_0 + d_1 n^2 + e_1 n \\ &= ak^2 - 2ak_0k - 2ak_0kn + a\beta^2 n^2 + d_1 n^2 + 2ak_0\beta n + e_1 n + ak_0^2 + g_0 \end{aligned}$$

30

... (13)

【 0 0 6 1】

図 8 (b) は、スペクトログラム上の小領域内でのフィッティングを模式的に示している。トーン性ピークでは、類似した形状が緩やかに時間変化するため、数式 (8) がよく適合する傾向にある。しかし、ノイズ性のピーク近傍に関しては、ピークの形状やピークの周波数がばらつくため、数式 (8) はあまりよく適合しない、つまり、最適に当てはめても誤差が大きいものとなる。 40

【 0 0 6 2】

なお、上述の数式 (1 0) では、 a 、 b 、 c 、 d 、 e 、 g の全ての係数に関するフィッティングを行う計算を示した。しかし、いくつかの係数についてはあらかじめ定数に固定した上でのフィッティングを行ってもよい。また、2 次以上の多項式関数でフィッティングしてもよい。

【 0 0 6 3】

図 6 に戻って、特徴量抽出部 2 3 3 では、フィッティング部 2 3 2 で得られる各ピークにおけるフィッティング結果 (上述の数式 (1 0) 参照) に基づいて、以下の数式 (1 4) に示すような特徴量 (x_0 、 x_1 、 x_2 、 x_3 、 x_4 、 x_5) が抽出される。各特徴量は、各 50

ピークにおける周波数成分の性質を表す特徴量であり、それ自体を音声や楽音などの分析に用いることができる。

【数 1 3】

$$\begin{aligned}
 \text{[ピークの曲率]} \quad x_0 &= \hat{a} \\
 \text{[ピークの周波数]} \quad x_1 &= -\frac{\hat{b}}{2\hat{a}} \\
 \text{[ピーク対数振幅値]} \quad x_2 &= \hat{g} \\
 \text{[周波数の変化率]} \quad x_3 &= -\frac{\hat{c}}{2\hat{a}} \\
 \text{[振幅の変化率]} \quad x_4 &= \hat{e} \\
 \text{[フィッティングの正規化誤差]} \quad x_5 &= \frac{J(\hat{a}, \hat{b}, \hat{c}, \hat{d}, \hat{e}, \hat{g})}{\sum_{\Gamma} (F(n, k) - \hat{g})^2}
 \end{aligned} \quad \dots(14)$$

【0064】

スコア化部 234 では、各ピークのトーン成分らしさを定量化するために、ピーク毎に特徴量抽出部 233 で抽出された特徴量が用いられて、各ピークのトーン成分らしさを示すスコア $S(n, k)$ が得られる。スコア化部 234 では、特徴量 ($x_0, x_1, x_2, x_3, x_4, x_5$) のうち、一つまたは複数の特徴量が用いられて、以下の数式 (15) に示すように、スコア $S(n, k)$ が求められる。この場合、少なくとも、フィッティングの正規化誤差 x_5 、あるいは周波数方向のピークの曲率 x_0 が使用される。

【数 1 4】

$$S(n, k) = \text{Sigm} \left(\sum_{i=0}^5 w_i H_i(x_i) + w_6 \right) \quad \dots(15)$$

【0065】

ただし、 $\text{Sigm}(x)$ はシグモイド関数であり、 w_i は予め定める荷重係数であり、 $H_i(x_i)$ は、 i 番目の特徴量 x_i に対して施すあらかじめ定める非線形関数である。非線形関数 $H_i(x_i)$ には、例えば、以下の数式 (16) に示すような関数を用いることができる。ただし、 u_i, v_i は、あらかじめ定める荷重係数である。 w_i, u_i, v_i は、なんらかの適切な定数をあらかじめ定めてもよいが、例えば、多数のデータを用いて最急降下学習などを行うことで、自動的に決定することもできる。

【数 1 5】

$$H_i(x_i) = \text{Sigm}(u_i x_i + v_i) \quad \dots(16)$$

【0066】

スコア化部 234 では、上述したように、ピーク毎に、数式 (15) によって、トーン成分らしさを示すスコア $S(n, k)$ が求められる。なお、スコア化部 234 では、ピークではない位置 (n, k) におけるスコア $S(n, k)$ は 0 とされる。スコア化部 234 では、時間周波数信号 $f(n, k)$ の各時刻、各周波数において、0 から 1 の間の値を取るトーン成分ら

10

20

30

40

50

しさのスコア $S(n, k)$ が得られる。

【0067】

図9のフローチャートは、トーン尤度分布検出部230におけるトーン尤度分布検出の処理手順の一例を示している。トーン尤度分布検出部230は、ステップST1において、処理を開始し、その後、ステップST2の処理に移る。このステップST2において、トーン尤度分布検出部230は、フレーム（時間フレーム）の番号 n を0に設定する。

【0068】

次に、トーン尤度分布検出部230は、ステップST3において、 $n < N$ であるか否かを判断する。なお、スペクトログラム（時間周波数分布）のフレームは0から $N - 1$ まで存在するものとする。 $n < N$ でないとき、トーン尤度分布検出部230は、全てのフレームの処理が終了したものと判断し、ステップST4において、処理を終了する。

10

【0069】

$n < N$ であるとき、トーン尤度分布検出部230は、ステップST5において、離散周波数 k を0に設定する。そして、トーン尤度分布検出部230は、ステップST6において、 $k < K$ であるか否かを判断する。なお、スペクトログラム（時間周波数分布）の離散周波数 k は0から $K - 1$ まで存在するものとする。 $k < K$ でないとき、トーン尤度分布検出部230は、全ての離散周波数の処理が終了したものと判断し、ステップST7において、 n をインクリメントし、その後、ステップST3に戻り、次のフレームの処理に移る。

【0070】

20

ステップST6で $k < K$ であるとき、トーン尤度分布検出部230は、ステップST8において、 $F(n, k)$ がピークであるか否かを判断する。ピークでないとき、トーン尤度分布検出部230は、ステップST9において、スコア $S(n, k)$ を0とし、ステップST10において、 k をインクリメントし、その後、ステップST6に戻り、次の離散周波数の処理に移る。

【0071】

ステップST8でピークであるとき、トーン尤度分布検出部230は、ステップST11の処理に移る。このステップST11において、トーン尤度分布検出部230は、そのピークの近傍領域においてトーンモデルをフィッティングする。そして、トーン尤度分布検出部230は、ステップST12において、フィッティング結果に基づいて、種々の特徴量（ $x_0, x_1, x_2, x_3, x_4, x_5$ ）を抽出する。

30

【0072】

次に、トーン尤度分布検出部230は、ステップST13において、ステップST12で抽出された特徴量を用いて、そのピークのトーン成分らしさを示す、0から1の間の値をとるスコア $S(n, k)$ を求める。トーン尤度分布検出部230は、このステップST14の処理の後、ステップST10において、 k をインクリメントし、その後、ステップST6に戻り、次の離散周波数の処理に移る。

【0073】

図10は、図11に示すような時間周波数分布（スペクトログラム） $F(n, k)$ から、図6に示すトーン尤度分布検出部230で得られたトーン成分らしさのスコア $S(n, k)$ の分布の一例を示している。スコア $S(n, k)$ の値が大きいほど黒く表示されているが、ノイズ性のピークは概ね検出されていないのに対し、トーン性の成分（図11で黒い太横線を形成している成分）のピークは概ね検出されていることが分かる。

40

【0074】

図4に戻って、トーン強度特徴量計算部223は、続いて、スコア $S(n, k)$ が所定の閾値 S_{thsd} より大きい位置（図5（b）参照）について、その近傍周波数位置の成分のみを抽出するトーン成分抽出フィルタ $H(n, k)$ （図5（c）参照）を作成する。以下の数式（17）は、このトーン成分抽出フィルタ $H(n, k)$ を表している。

【数 16】

$$H(n, k) = \begin{cases} 1 & k_T - \Delta k < k < k_T + \Delta k \\ 0 & \text{otherwise} \end{cases} \quad \dots(17)$$

【0075】

ただし、 k_T はトーン成分が検出された周波数であり、 k は所定の周波数幅である。ここで、上述したように時間周波数信号 $F(n, k)$ を得るための短時間フーリエ変換（数式（1）参照）における窓関数 $W(t)$ のサイズが M であるとき、 k は $2/M$ とされることが望ましい。

10

【0076】

トーン強度特徴量計算部223は、続いて、このトーン成分抽出フィルタ $H(n, k)$ を、元の時間周波数信号時間周波数信号 $F(n, k)$ に乗算して、図5(d)に示すように、トーン成分のみを残したスペクトル（トーン成分スペクトル） $F_T(n, k)$ を得る。以下の数式（18）は、このトーン成分スペクトル $F_T(n, k)$ を表している。

【数 17】

$$F_T(n, k) = H(n, k)F(n, k) \quad \dots(18)$$

20

【0077】

トーン強度特徴量計算部223は、最後に、所定の周波数範囲（下限 K_L 、上限 K_H ）について総和をとり、以下の数式（19）で表される、対象フレーム n におけるトーン成分強度 $A_{\text{tone}}(n)$ を求める。

【数 18】

$$A_{\text{tone}}(n) = \sum_{k=K_L}^{K_H} \exp(F_T(n, k)) \quad \dots(19)$$

30

【0078】

そして、トーン強度特徴量計算部223は、以下の数式（20）に示すように、トーン成分強度 $A_{\text{tone}}(n)$ をトーン強度特徴量 $x_2(n)$ とする。

【数 19】

$$x_2(n) = A_{\text{tone}}(n) \quad \dots(20)$$

40

【0079】

スペクトル概形特徴量計算部224は、スペクトル概形特徴量 $x_3(n)$ 、 $x_4(n)$ 、 $x_5(n)$ 、 $x_6(n)$ を、以下の数式（21）に示すように、求める。ただし、 L は、特徴量の次元数であり、ここでは、 $L = 7$ の場合を示している。

【数 2 0】

$$\begin{aligned}
 x_3(n) &= \sum_{k=0}^{N/2-1} F(k, n) \cos(2\pi k / N) \\
 x_4(n) &= \sum_{k=0}^{N/2-1} F(k, n) \cos(4\pi k / N) \\
 x_5(n) &= \sum_{k=0}^{N/2-1} F(k, n) \cos(6\pi k / N) \\
 x_6(n) &= \sum_{k=0}^{N/2-1} F(k, n) \cos(8\pi k / N)
 \end{aligned}
 \quad \left. \vphantom{\begin{aligned} x_3(n) \\ x_4(n) \\ x_5(n) \\ x_6(n) \end{aligned}} \right\} \dots(21)$$

10

【0080】

このスペクトル概形特徴量は、対数スペクトルを離散コサイン変換により展開した低次ケプストラムである。ここでは、4次までを示したが、より高次の係数まで使用してもよい。また、いわゆるMFCC(Mel-Frequency Cepstral Coefficients)のように、周波数軸を歪曲させてから離散コサイン変換を施したものをを用いてもよい。

20

【0081】

上述の振幅特徴量 $x_0(n)$ 、 $x_1(n)$ 、トーン強度特徴量 $x_2(n)$ 、スペクトル概形特徴量 $x_3(n)$ 、 $x_4(n)$ 、 $x_5(n)$ 、 $x_6(n)$ は、フレーム n における L 次元(ここでは7次元)の特徴量ベクトル $x(n)$ を構成する。因みに、「音の大きさ、音の高さ、音色」を音の三要素と言い、音の性質を表す基本的な属性である。特徴量ベクトル $x(n)$ は、振幅(音の大きさに関係)、トーン成分強度(音の高さに関係)、スペクトル概形(音色の関係)により構成されることで、音の三要素の全てに関する特徴量を構成している。

【0082】

スコア計算部225は、特徴量ベクトル $x(n)$ の要素を合成し、フレーム n が実際に登録すべき有意な音(被検出音)のある音区間であるかどうかを、0から1の間のスコア $S(n)$ で表現する。これは、例えば、以下の数式(22)により求められる。ただし、 $\text{Sigm}()$ はシグモイド関数であり、 $u_i, v_i, w_i (i = 0, \dots, L-1)$ はサンプルデータより経験的に決める定数である。

30

【数 2 1】

$$\begin{aligned}
 S(n) &= \text{Sigm}\left(\sum_{i=0}^{L-1} w_i \xi_i(x_i(n)) + w_L\right) \\
 \xi_i(x_i) &= \text{Sigm}(u_i x_i(n) + v_i)
 \end{aligned}
 \quad \dots(22)$$

40

【0083】

時間平滑化部226は、スコア計算部225で求められたスコア $S(n)$ を時間方向に平滑化する。この平滑化の処理では、単純に移動平均をとってもよいし、例えばメジアンフィルタのように中央値を取るようなフィルタを用いてもよい。以下の数式(23)は、平滑化スコア $S_a(n)$ を、平均処理で得る例を示している。ただし、 n は、フィルタのサイズであり、経験的に決める定数である。

【数 2 2】

$$S_a(n) = \frac{1}{2\Delta n + 1} \sum_{\tau=n-\Delta n}^{n+\Delta n} S(n) \quad \dots(23)$$

【0084】

閾値判定部 2 2 7 は、時間平滑化部 2 2 6 で得られた各フレーム n の平滑化スコア $S_a(n)$ を閾値と比較し、閾値以上となるフレーム区間を音区間と判定し、そのフレーム区間を示す音区間情報を出力する。

10

【0085】

図 4 に示す音区間検出部 2 0 2 の動作を説明する。登録すべき被検出音（家電で発せられる動作状況音）をマイクロフォン 2 0 1 で録音して得られる時間信号 $f(t)$ は、時間周波数変換部 2 2 1 に供給される。この時間周波数変換部 2 2 1 では、入力時間信号 $f(t)$ が時間周波数変換されて、時間周波数信号 $F(n, k)$ が得られる。この時間周波数信号 $F(n, k)$ は、振幅特徴量計算部 2 2 2、トーン強度特徴量計算部 2 2 3 およびスペクトル概形特徴量計算部 2 2 4 に供給される。

【0086】

振幅特徴量計算部 2 2 2 では、時間周波数信号 $F(n, k)$ より、振幅特徴量 $x_0(n)$ 、 $x_1(n)$ が計算される（数式（5）参照）。また、トーン強度特徴量計算部 2 2 3 では、時間周波数信号 $F(n, k)$ より、トーン強度特徴量 $x_2(n)$ が計算される（数式（20）参照）。さらに、スペクトル概形特徴量計算部 2 2 4 では、スペクトル概形特徴量 $x_3(n)$ 、 $x_4(n)$ 、 $x_5(n)$ 、 $x_6(n)$ が計算される（数式（21）参照）。

20

【0087】

振幅特徴量 $x_0(n)$ 、 $x_1(n)$ 、トーン強度特徴量 $x_2(n)$ 、スペクトル概形特徴量 $x_3(n)$ 、 $x_4(n)$ 、 $x_5(n)$ 、 $x_6(n)$ は、フレーム n における L 次元（ここでは 7 次元）の特徴量ベクトル $x(n)$ として、スコア計算部 2 2 5 に供給される。スコア計算部 2 2 5 では、特徴量ベクトル $x(n)$ の要素が合成されて、フレーム n が実際に登録すべき有意な音（被検出音）のある音区間であるかどうかを表現する、0 から 1 の間のスコア $S(n)$ が計算される（数式（22）参照）。このスコア $S(n)$ は、時間平滑化部 2 2 6 に供給される。

30

【0088】

時間平滑化部 2 2 6 では、スコア $S(n)$ が時間方向に平滑化され（数式（23）参照）、平滑化スコア $S_a(n)$ は閾値判定部 2 2 7 に供給される。閾値判定部 2 2 7 では、各フレーム n の平滑化スコア $S_a(n)$ が閾値と比較され、閾値以上となるフレーム区間が音区間と判定され、そのフレーム区間を示す音区間情報が出力される。

【0089】

図 4 に示す音区間検出部 2 0 2 は、入力時間信号 $f(t)$ の時間周波数分布 $F(n, k)$ より時間フレーム毎の、振幅、トーン成分強度およびスペクトル概形の特徴量を抽出し、この特徴量から時間フレーム毎の、音区間らしさを示すスコア $S(n)$ を得るものである。そのため、登録すべき検出音がノイズ環境下で録音される場合であっても、この検出音の区間を示す音区間情報を精度よく得ることができる。

40

【0090】

「特徴量抽出部」

図 1 2 は、特徴量抽出部 2 0 3 の構成例を示している。この特徴量抽出部 2 0 3 の入力には、登録すべき被検出音（家電で発せられる動作状況音）をマイクロフォン 2 0 1 で録音して得られる時間信号 $f(t)$ であり、図 3 示すように、前後にノイズ区間も含まれる。また、この特徴量抽出部 2 0 3 の出力は、登録すべき被検出音の区間で所定時間毎に抽出された特徴量の列である。

【0091】

この特徴量抽出部 2 0 3 は、時間周波数変換部 2 4 1 と、トーン尤度分布検出部 2 4 2

50

と、時間周波数平滑化部 2 4 3 と、間引き・量子化部 2 4 4 を有している。時間周波数変換部 2 4 1 は、上述の音区間検出部 2 0 2 の時間周波数変換部 2 2 1 と同様に、入力時間信号 $f(t)$ を時間周波数変換して、時間周波数信号 $F(n, k)$ を得る。なお、特徴量抽出部 2 0 3 は、音区間検出部 2 0 2 の時間周波数変換部 2 2 1 で得られた時間周波数信号 $F(n, k)$ を利用してもよく、その場合には、この時間周波数変換部 2 4 1 を不要とできる。

【 0 0 9 2 】

トーン尤度分布検出部 2 4 2 は、音区間検出部 2 0 2 からの音区間情報に基づいて、音区間のトーン尤度分布を検出する。すなわち、トーン尤度分布検出部 2 4 2 は、まず、上述した音区間検出部 2 0 2 のトーン強度特徴量計算部 2 2 3 におけると同様にして、時間周波数信号 $F(n, k)$ の分布 (図 5 (a) 参照) を、トーン性らしさのスコア $S(n, k)$ の分布 (図 5 (b) 参照) に変換する。

10

【 0 0 9 3 】

トーン尤度分布検出部 2 4 2 は、続いて、音区間情報を用いて、以下の数式 (2 4) に示すように、登録すべき有意な音 (被検出音) のある音区間のトーン尤度分布 $Y(n, k)$ を求める。

【 数 2 3 】

$$Y(n, k) = \begin{cases} S(n, k) & n \text{ が音区間内の場合} \\ 0 & n \text{ が音区間外の場合} \end{cases} \quad \dots (24)$$

20

【 0 0 9 4 】

時間周波数平滑化部 2 4 3 は、トーン尤度分布検出部 2 4 2 で求められた音区間のトーン尤度分布 $Y(n, k)$ を、時間方向および周波数方向に平滑化し、以下の数式 (2 5) に示すように、平滑化されたトーン尤度分布 $Y_a(n, k)$ を得る。

【 数 2 4 】

$$Y_a(n, k) = \sum_{\tau=-\Delta_n}^{\Delta_n} \sum_{\lambda=-\Delta_k}^{\Delta_k} Y(n-\tau, k-\lambda) H(\tau, \lambda) \quad \dots (25)$$

30

【 0 0 9 5 】

ただし、 k は平滑化フィルタの周波数方向の片側サイズ、 n は時間方向の片側サイズ、 $H(n, k)$ は平滑化フィルタの 2 次元インパルス応答である。なお、上述では表記を簡単にするため、周波数方向に歪みのないフィルタを用いて説明した。しかし、例えば、メル周波数のように、周波数軸を歪曲するフィルタを用いて平滑化を行ってもよい。

【 0 0 9 6 】

間引き・量子化部 3 4 4 は、時間周波数平滑化部 2 4 3 で得られた平滑化されたトーン尤度分布 $Y_a(n, k)$ を間引きし、さらに、量子化して、以下の数式 (2 6) に示すように、登録すべき有意な音 (被検出音) の特徴量 $Z(m, l)$ を生成する。

40

【 数 2 5 】

$$z(m, l) = \text{Quant} \left[Y_a(mT, lK) \right] \quad (0 \leq m \leq M-1, 0 \leq l \leq L-1) \quad \dots (26)$$

【 0 0 9 7 】

50

ただし、 T は時間方向の離散化ステップ、 K は周波数方向の離散化ステップ、 m は間引きされた離散時間、 l は間引きされた離散周波数である。また、 M は時間方向のフレーム数 (= 登録すべき有意な音 (被検出音) の時間長に相当する)、 L は周波数方向の次元数、 $\text{Quant}[\]$ は量子化の関数である。

【0098】

上述の特徴量 $z(m, l)$ は、周波数方向にまとめて、以下の数式 (27) に示すように、ベクトル表記して、 $Z(m)$ で表すことができる。

【数26】

$$Z(m) = [z(m, 0), \dots, z(m, L-1)] \quad (0 \leq m \leq M-1) \quad \dots (27)$$

10

【0099】

この場合、上述の特徴量 $Z(m, l)$ は、時間方向に T 毎に抽出された M 個のベクトル $Z(0)$ 、 \dots 、 $Z(M-1)$ 、 $Z(M)$ により構成されていることになる。したがって、間引き・量子化部 244 からは、登録すべき被検出音の区間で所定時間毎に抽出された特徴量 (ベクトル) の列 $Z(m)$ が得られる。

【0100】

なお、時間周波数平滑化部 243 で得られた平滑化されたトーン尤度分布 $Y_a(n, k)$ をそのまま特徴量抽出部 203 の出力、つまり特徴量列として用いることも考えられる。しかし、平滑化されているので全ての時間、周波数のデータを持っている必要はない。時間方向および周波数方向に間引きすることで、情報量を減らすことができる。また、量子化により、例えば、8ビットや16ビットのデータを2ビットや3ビットのデータに変換できる。このように間引きおよび量子化が行われることで、特徴量 (ベクトル) 列 $Z(m)$ の情報量を低減でき、後述する音検出装置 100 におけるマッチング計算の処理負荷を軽減することが可能となる。

20

【0101】

図12に示す特徴量抽出部 203 の動作を説明する。登録すべき被検出音 (家電で発せられる動作状況音) をマイクロフォン 201 で録音して得られる時間信号 $f(t)$ は、時間周波数変換部 241 に供給される。この時間周波数変換部 241 では、入力時間信号 $f(t)$ が時間周波数変換されて、時間周波数信号 $F(n, k)$ が得られる。この時間周波数信号 $F(n, k)$ は、トーン尤度分布検出部 242 に供給される。また、このトーン尤度分布検出部 242 には、音区間検出 202 で得られた音区間情報も供給される。

30

【0102】

このトーン尤度分布検出部 242 では、時間周波数信号 $F(n, k)$ の分布がトーン性らしさのスコア $S(n, k)$ の分布に変換され、さらに、音区間情報が用いられて、登録すべき有意な音 (被検出音) のある音区間のトーン尤度分布 $Y(n, k)$ が求められる (数式 (24) 参照)。このトーン尤度分布 $Y(n, k)$ は、時間周波数平滑化部 243 に供給される。

【0103】

時間周波数平滑化部 243 では、トーン尤度分布 $Y(n, k)$ が時間方向および周波数方向に平滑化され、平滑化されたトーン尤度分布 $Y_a(n, k)$ が得られる (数式 (25) 参照)。このトーン尤度分布 $Y_a(n, k)$ は間引き・量子化部 244 に供給される。間引き・量子化部 244 では、トーン尤度分布 $Y_a(n, k)$ が間引きされ、さらに、量子化されて、登録すべき有意な音 (被検出音) の特徴量 $z(m, l)$ 、従って特徴量列 $Z(m)$ が得られる (数式 (26)、数式 (27) 参照)。

40

【0104】

図2に戻って、特徴量登録部 204 は、特徴量登録部 204 で生成された登録すべき被検出音の特徴量列 $Z(m)$ を、被検出音名 (動作状況音の情報) と対応付けて、特徴量データベース 103 に登録する。

【0105】

50

図 2 に示す特徴登録装置 200 の動作を説明する。マイクロフォン 201 では、被検出音として登録すべき家電の動作状況音が集音される。このマイクロフォン 201 から出力される時間信号 $f(t)$ は、音区間検出部 202 および特徴量抽出部 203 に供給される。音区間検出部 202 では、入力時間信号 $f(t)$ から、音区間、すなわち家電から発せられる動作状況音の区間が検出されて、音区間情報が出力される。この音区間情報は特徴量抽出部 203 に供給される。

【0106】

特徴量抽出部 203 では、入力時間信号 $f(t)$ が時間フレーム毎に時間周波数変換されて時間周波数信号 $F(n,k)$ の分布が得られ、さらに、この時間周波数分布からトーンらしさの尤度分布、つまりスコア $S(n,k)$ の分布が求められる。そして、特徴量抽出部 203 10
では、音区間情報に基づいて、スコア $S(n,k)$ の分布から音区間のトーン尤度分布 $Y(n,k)$ が得られ、それが時間方向および周波数方向に平滑化され、さらに、間引き・量子化の処理が施されて、特徴量列 $Z(m)$ が生成される。

【0107】

特徴量抽出部 203 で生成された、登録すべき被検出音（家電の動作状況音）の特徴量列 $Z(m)$ は、特徴量登録部 204 に供給される。特徴量登録部 204 では、特徴量列 $Z(m)$ を、その被検出音名（動作状況音の情報）に対応付けて、特徴量データベース 103 に登録することが行われる。以下では、 I 個の被検出音が登録されたものとし、それらの特徴量列を $Z1(m)$, $Z2(m)$, \dots , $Zi(m)$, \dots , $ZI(m)$ と表記し、また、各特徴量列の時間フレーム数（時間方向に並ぶベクトルの個数）を $M1$, $M2$, \dots , Mi , \dots , MI と記述する。 20

【0108】

「音検出部」

図 13 は、音検出部 102 の構成例を示している。この音検出部 102 は、信号バッファ部 121 と、特徴量抽出部 122 と、特徴量バッファ部 123 と、比較部 124 を有している。信号バッファ部 121 は、マイクロフォン 101 で集音されて得られる時間信号 $f(t)$ の信号サンプルを所定数バッファリングする。所定数とは、特徴量抽出部 122 が新たに 1 フレーム分の特徴量列を計算できるだけのサンプル数である。

【0109】

特徴量抽出部 122 は、信号バッファ部 121 にバッファリングされた時間信号 $f(t)$ 30
の信号サンプルに基づいて、所定時間毎の特徴量を抽出する。詳細説明は省略するが、この特徴量抽出部 203 は、上述した特徴登録装置 200 の特徴量抽出部 203（図 12 参照）と同様に構成される。

【0110】

ただし、特徴量抽出部 122 においては、トーン尤度分布検出部 242 は全区間のトーン尤度分布 $Y(n,k)$ を求める。つまり、トーン尤度分布検出部 242 は、時間周波数信号 $F(n,k)$ の分布から得られたスコア $S(n,k)$ の分布をそのまま出力する。そして、間引き・量子化部 244 は、入力時間信号 $f(t)$ の全区間において、 T （時間方向の離散化ステップ）毎に、新たに抽出された特徴量（ベクトル） $X(n)$ を出力する。ここで、 n は現在抽出された特徴量のフレーム番号（現在の離散時間に相当する）である。 40

【0111】

特徴量バッファ部 123 は、特徴量抽出部 122 から出力される特徴量（ベクトル） $X(n)$ を、図 14 に示すように、最新から N 個保存する。ここで、 N は、少なくとも、特徴量データベース 103 に登録（保持）されている特徴量列 $Z1(m)$, $Z2(m)$, \dots , $Zi(m)$, \dots , $ZI(m)$ のうち、最も長い特徴量列のフレーム数（時間方向に並ぶベクトルの個数）と同じかそれ以上の数である。

【0112】

比較部 124 は、特徴量抽出部 122 で新たな特徴量 $X(n)$ が抽出される毎に、信号バッファ部 123 に保存されている特徴量の列を、特徴量データベース 103 に登録されている I 個の被検出音の特徴量列と順次比較し、 I 個の被検出音の検出結果を得る。ここで 50

、 i を被検出音の番号とすると、被検出音の長さ（フレーム数 M_i ）はそれぞれの被検出音で異なる。

【0113】

比較部124は、図14に示すように、特徴量バッファ部123の最新フレーム n に、被検出音の特徴量列の最終フレーム $Z_i(M_i - 1)$ を合わせ、特徴量バッファ部123に保存されている N 個の特徴量のうち、被検出音の特徴量列の長さ分のフレームを用いて類似度を算出する。この類似度 $Sim(n, i)$ は、例えば、以下の数式(28)で示すように、特徴量間の相関演算により計算できる。ただし、 $Sim(n, i)$ は、第 n フレームにおける第 i 番目の被検出音の特徴量列との間の類似度を意味する。比較部124は、類似度が所定の閾値より大きい場合には、「時刻 n において第 i 番目の被検出音が鳴っている」と判定し、その判定結果を出力する。

10

【0114】

【数27】

$$Sim(n, i) = \frac{\sum_{m=0}^{M_i-1} X(n - M_i - 1 + m) Z_i(m)}{\sqrt{\sum_{m=0}^{M_i-1} X^2(n - M_i - 1 + m) \sum_{m=0}^{M_i-1} Z_i^2(m)}} \quad \dots(28)$$

20

【0115】

図13に示す音検出部102の動作を説明する。マイクロフォン101で集音されて得られる時間信号 $f(t)$ は信号バッファ部121に供給され、その信号サンプルが所定数バッファリングされる。特徴量抽出部122では、信号バッファ部121にバッファリングされた時間信号 $f(t)$ の信号サンプルに基づいて、所定時間毎に特徴量が抽出される。そして、この特徴量抽出部122からは、 T （時間方向の離散化ステップ）毎に、新たに抽出された特徴量（ベクトル） $X(n)$ が順次出力される。

30

【0116】

特徴量バッファ部123には、特徴量抽出部122で抽出された特徴量 $X(n)$ が供給され、最新から N 個保存される。比較部124では、特徴量抽出部122で新たな特徴量 $X(n)$ が抽出される毎に、信号バッファ部123に保存されている特徴量の列が、特徴量データベース103に登録されている I 個の被検出音の特徴量列と順次順次比較され、 I 個の被検出音の検出結果が得られる。

【0117】

この場合、比較部124では、特徴量バッファ部123の最新フレーム n に、被検出音の特徴量列の最終フレーム $Z_i(M_i - 1)$ を合わせ、被検出音の特徴量列の長さ分のフレームが用いられて類似度が算出される（図14参照）。そして、比較部124では、類似度が所定の閾値より大きい場合には、「時刻 n において第 i 番目の被検出音が鳴っている」と判定され、その判定結果が出力される。

40

【0118】

なお、図1に示す音検出装置100は、ハードウェアで構成できる他、ソフトウェアで構成することもできる。例えば、図15に示すコンピュータ装置300に、図1に示す音検出装置100の一部または全部の機能を持たせ、上述したと同様の被検出音の検出処理を行わせることができる。

【0119】

コンピュータ装置300は、CPU(Central Processing Unit)301、ROM(Read O

50

nlyMemory) 302、RAM(Random Access Memory) 303、データ入出力部(データI/O) 304およびHDD(Hard Disk Drive) 305により構成されている。ROM 302には、CPU 301の処理プログラムなどが格納されている。RAM 303は、CPU 301のワークエリアとして機能する。CPU 301は、ROM 302に格納されている処理プログラムを必要に応じて読み出し、読み出した処理プログラムをRAM 303に転送して展開し、当該展開された処理プログラムを読み出して、トーン成分検出処理を実行する。

【0120】

このコンピュータ装置300においては、入力時間信号 $f(t)$ は、データI/O 304を介して入力され、HDD 305に蓄積される。このようにHDD 305に蓄積される入力時間信号 $f(t)$ に対して、CPU 301により、被検出音の検出処理が行われる。そして、検出結果がデータI/O 304を介して外部に出力される。なお、HDD 305には、I個の被検出音の特徴量列が予め登録されて保持されている。

10

【0121】

図16のフローチャートは、CPU 301による被検出音の検出処理の手順の一例を示している。CPU 301は、ステップST 21において、処理を開始し、その後、ステップST 22の処理に移る。このステップST 22において、CPU 181は、入力時間信号 $f(t)$ を、例えばHDD 305に構成される信号バッファ部に入力する。そして、CPU 301は、ステップST 23において、1フレーム分の特徴量列を計算できるだけのサンプル数が溜まったか否かを判断する。

20

【0122】

1フレーム分のサンプル数が溜まったとき、CPU 301は、ステップST 24において、特徴量 $X(n)$ を抽出する処理を行う。CPU 301は、ステップST 25において、抽出された特徴量 $X(n)$ を、例えばHDD 305に構成される特徴量バッファ部に入力する。そして、CPU 301は、ステップST 26において、被検出音番号 i が0にセットされる。

【0123】

次に、CPU 301は、ステップST 27において、 $i < I$ であるか否かを判定する。 $i < I$ であるとき、CPU 301は、ステップST 28において、信号バッファ部に保存されている特徴量の列とHDD 305に登録されている i 番目の被検出音の特徴量列 $Z_i(m)$ との間の類似度を算出する。そして、CPU 301は、ステップST 29において、類似度 $>$ 閾値を満足するか否かを判定する。

30

【0124】

類似度 $>$ 閾値を満足するとき、CPU 301は、ステップST 30において、一致結果を出力する。すなわち、「時刻 n において第 i 番目の被検出音が鳴っている」との判定結果を、検出出力として出力する。その後、CPU 301は、ステップST 31において、 i をインクリメントし、ステップST 27の処理に戻る。なお、ステップST 29において、類似度 $>$ 閾値を満足しないとき、CPU 301は、直ちに、ステップST 31において、 i をインクリメントし、ステップST 27の処理に戻る。また、ステップST 27で $i > I$ でないとき、現在のフレームの処理を終了したものと判断し、ステップST 22の処理に戻り、次のフレームの処理に移る。

40

【0125】

次に、CPU 181は、ステップST 3において、フレーム(時間フレーム)の番号 n を0に設定する。そして、CPU 181は、ステップST 4において、 $n < N$ であるか否かを判断する。なお、スペクトログラム(時間周波数分布)のフレームは0から $N - 1$ まで存在するものとする。 $n < N$ でないとき、CPU 181は、全てのフレームの処理が終了したものと判断し、ステップST 5において、処理を終了する。

【0126】

$n < N$ であるとき、CPU 181は、ステップST 6において、離散周波数 k を0に設定する。そして、CPU 181は、ステップST 7において、 $k < K$ であるか否かを判断

50

する。なお、スペクトログラム（時間周波数分布）の離散周波数 k は 0 から $K - 1$ まで存在するものとする。 $k < K$ でないとき、CPU 181 は、全ての離散周波数の処理が終了したものと判断し、ステップ ST 8 において、 n をインクリメントし、その後に、ステップ ST 4 に戻り、次のフレームの処理に移る。

【0127】

ステップ ST 7 で $k < K$ であるとき、CPU 181 は、ステップ ST 9 において、 $F(n, k)$ がピークであるか否かを判断する。ピークでないとき、CPU 181 は、ステップ ST 10 において、スコア $S(n, k)$ を 0 とし、ステップ ST 11 において、 k をインクリメントし、その後に、ステップ ST 7 に戻り、次の離散周波数の処理に移る。

【0128】

ステップ ST 9 でピークであるとき、CPU 181 は、ステップ ST 12 の処理に移る。このステップ ST 12 において、CPU 181 は、そのピークの近傍領域においてトーンモデルをフィッティングする。そして、CPU 181 は、ステップ ST 13 において、フィッティング結果に基づいて、種々の特徴量 ($x_0, x_1, x_2, x_3, x_4, x_5$) を抽出する。

【0129】

次に、CPU 181 は、ステップ ST 14 において、ステップ ST 13 で抽出された特徴量を用いて、そのピークのトーン成分らしさを示す、 0 から 1 の間の値をとるスコア $S(n, k)$ を求める。CPU 181 は、このステップ ST 14 の処理の後、ステップ ST 11 において、 k をインクリメントし、その後に、ステップ ST 7 に戻り、次の離散周波数の処理に移る。

【0130】

上述したように、図 1 に示す音検出装置 100 においては、マイクロフォン 101 で集音されて得られる入力時間信号 $f(t)$ の時間周波数分布よりトーンらしさの尤度分布を求め、この尤度分布を周波数方向および時間方向に平滑化したものから所定時間毎の特徴量を抽出して用いるものである。従って、被検出音（家庭用電化製品から発せられる動作状況音など）の検出を、マイクロフォン 101 の設置位置などに依らずに、精度よく行うことができる。

【0131】

また、図 1 に示す音検出装置 100 においては、音検出部 102 で得られた被検出音の検出結果を、時刻と共に記録媒体に記録し、また、ディスプレイに表示するものである。従って、家庭内における家電等の動作状況を自動的に記録でき、自らの行動履歴（いわゆるライフログ）の取得が可能になる。また、聴覚障害者などに、音による通知を自動的に視覚化することが可能になる。

【0132】

< 2 . 変形例 >

なお、上述実施の形態においては、家庭内において、家電から発せられる動作状況音（操作音、通知音、動作音、警報音など）を検出する例を示した。しかし、本技術は、家庭内の用途に限らず、生産工場などで製造された製品の音機能に関する検査の自動化にも用いることができる。また、動作状況音の検出だけに限られるものではなく、特定の人や動物の音声、さらにはその他の環境音の検出にも、本技術を適用できることは勿論である。

【0133】

また、上述実施の形態においては、短時間フーリエ変換により時間周波数変換を行うように説明したが、ウェーブレット変換など、その他の変換手法を使用して、入力時間信号を時間周波数変換することも考えられる。また、上述実施の形態においては、検出された各ピーク近傍の時間周波数分布とトーンモデルの二乗誤差最小基準によりフィッティングを行うように説明したが、4 乗誤差最小基準やエントロピー最小基準などによりフィッティングを行うことも考えられる。

【0134】

また、本技術は、以下のような構成を取ることもできる。

10

20

30

40

50

(1) 入力時間信号から所定時間毎の特徴量を抽出する特徴量抽出部と、
 所定数の被検出音の特徴量列を保持する特徴量保持部と、
 上記特徴量抽出部で新たに特徴量が抽出される毎に、該特徴量抽出部で抽出された特徴量の列を、上記保持されている所定数の被検出音の特徴量列とそれぞれ比較して、上記所定数の被検出音の検出結果を得る比較部とを備え、

上記特徴量抽出部は、
 上記入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得る時間周波数変換部と、
 上記時間周波数分布よりトーンらしさの尤度分布を求める尤度分布検出部とを有し、
 上記求められた尤度分布を周波数方向および時間方向に平滑化して上記所定時間毎の特徴量を抽出する
 音検出装置。

10

(2) 上記尤度分布検出部は、
 上記時間周波数分布の各時間フレームにおいて周波数方向のピークを検出するピーク検出部と、
 上記検出された各ピークにおいてトーンモデルをフィッティングするフィッティング部と、
 上記フィッティング結果に基づき、上記検出された各ピークのトーン成分らしさを示すスコアを得るスコア化部とを備える

前記(1)に記載の音検出装置。

20

(3) 上記特徴量抽出部は、
 上記平滑化された尤度分布を周波数方向および/または時間方向に間引く間引き部をさらに備える
 前記(1)または(2)に記載の音検出装置。

(4) 上記特徴量抽出部は、
 上記平滑化された尤度分布を量子化する量子化部をさらに備える
 前記(1)または(2)に記載の音検出装置。

(5) 上記比較部は、
 上記所定数の被検出音のそれぞれについて、上記保持されている被検出音の特徴量列と上記特徴量抽出部で抽出された特徴量列との間の対応する特徴量間の相関演算で類似度を求め、該求められた類似度に基づいて上記被検出音の検出結果を得る

30

前記(1)から(4)のいずれかに記載の音検出装置。

(6) 上記所定数の被検出音の検出結果を時刻情報と共に記録媒体に記録する記録制御部をさらに備える
 前記(1)から(5)のいずれかに記載の音検出装置。

(7) 入力時間信号から所定時間毎の特徴量を抽出する特徴量抽出ステップと、
 上記特徴量抽出ステップで新たに特徴量が抽出される毎に、該特徴量抽出部で抽出された特徴量の列を、保持されている所定数の被検出音の特徴量列とそれぞれ比較して、上記所定数の被検出音の検出結果を得る比較ステップとを備え、

上記特徴量抽出ステップでは、
 上記入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得、該時間周波数分布よりトーンらしさの尤度分布を求め、該尤度分布を周波数方向および時間方向に平滑化して上記所定時間毎の特徴量を抽出する
 音検出方法。

40

(8) コンピュータに、
 入力時間信号から所定時間毎の特徴量を抽出する特徴量抽出ステップと、
 上記特徴量抽出ステップで新たに特徴量が抽出される毎に、該特徴量抽出部で抽出された特徴量の列を、保持されている所定数の被検出音の特徴量列とそれぞれ比較して、上記所定数の被検出音の検出結果を得る比較ステップとを備え、

上記特徴量抽出ステップでは、

50

上記入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得、該時間周波数分布よりトーンらしさの尤度分布を求め、該尤度分布を周波数方向および時間方向に平滑化して上記所定時間毎の特徴量を抽出する

音検出方法を実行させるためのプログラム。

(9) 入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得る時間周波数変換部と、

上記時間周波数分布よりトーンらしさの尤度分布を求める尤度分布検出部と、

上記尤度分布を周波数方向および時間方向に平滑化して所定時間毎の特徴量を抽出する特徴量抽出部とを備える

音特徴量抽出装置。

10

(10) 上記尤度分布検出部は、

上記時間周波数分布の各時間フレームにおいて周波数方向のピークを検出するピーク検出部と、

上記検出された各ピークにおいてトーンモデルをフィッティングするフィッティング部と、

上記フィッティング結果に基づき、上記検出された各ピークのトーン成分らしさを示すスコアを得るスコア化部とを備える

前記(9)に記載の音特徴量抽出装置。

(11) 上記平滑化された尤度分布を周波数方向および/または時間方向に間引く間引き部をさらに備える

20

前記(9)または(10)に記載の音特徴量抽出装置。

(12) 上記平滑化された尤度分布を量子化する量子化部をさらに備える

前記(9)または(10)に記載の音特徴量抽出装置。

(13) 上記入力時間信号に基づいて音区間を検出する音区間検出部をさらに備え、

上記尤度分布検出部は、

上記検出された音区間の範囲で上記時間周波数分布よりトーンらしさの尤度分布を求める

前記(9)から(12)のいずれかに記載の音特徴量抽出装置。

(14) 上記音区間検出部は、

上記入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得る時間周波数変換部と、

30

上記時間周波数分布に基づいて、時間フレーム毎の、振幅、トーン成分強度およびスペクトル概形の特徴量を抽出する特徴量抽出部と、

上記抽出された特徴量に基づいて、時間フレーム毎の、音区間らしさを示すスコアを得るスコア化部と、

上記得られた時間フレーム毎のスコアを時間方向に平滑化する時間平滑化部と、

上記平滑化された時間フレーム毎のスコアを閾値判定して音区間情報を得る閾値判定部とを有する

前記(13)に記載の音特徴量抽出装置。

(15) 入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得る時間周波数変換ステップと、

40

上記時間周波数分布よりトーンらしさの尤度分布を求める尤度分布検出ステップと、

上記尤度分布を周波数方向および時間方向に平滑化する平滑化ステップとを備える

音特徴量抽出方法。

(16) 入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得る時間周波数変換部と、

上記時間周波数分布に基づいて、時間フレーム毎の、振幅、トーン成分強度およびスペクトル概形の特徴量を抽出する特徴量抽出部と、

上記抽出された特徴量に基づいて、時間フレーム毎の、音区間らしさを示すスコアを得るスコア化部とを備える

50

音区間検出装置。

(17) 上記得られた時間フレーム毎のスコアを時間方向に平滑化する時間平滑化部と

、
上記平滑化された時間フレーム毎のスコアを閾値判定して音区間情報を得る閾値判定部とをさらに備える

前記(16)に記載の音区間検出装置。

(18) 入力時間信号を時間フレーム毎に時間周波数変換して時間周波数分布を得る時間周波数変換ステップと、

上記時間周波数分布に基づいて、時間フレーム毎の、振幅、トーン成分強度およびスペクトル概形の特徴量を抽出する特徴量抽出ステップと、

上記抽出された特徴量に基づいて、時間フレーム毎の、音区間らしさを示すスコアを得るスコア化ステップとを備える

音区間検出方法。

【符号の説明】

【0135】

- 100・・・音検出装置
- 101・・・マイクロフォン
- 102・・・音検出部
- 103・・・特徴量データベース
- 104・・・記録・表示部
- 121・・・信号バッファ部
- 122・・・特徴量抽出部
- 123・・・特徴量バッファ部
- 124・・・比較部
- 200・・・特徴量登録装置
- 201・・・マイクロフォン
- 202・・・音区間検出部
- 203・・・特徴量抽出部
- 204・・・特徴量登録部
- 221・・・時間周波数変換部
- 222・・・振幅特徴量計算部
- 223・・・トーン強度特徴量計算部
- 224・・・スペクトル概形特徴量計算部
- 225・・・スコア計算部
- 226・・・時間平滑化部
- 227・・・閾値判定部
- 230・・・トーン尤度分布検出部
- 231・・・ピーク検出部
- 232・・・フィッティング部
- 233・・・特徴量抽出部
- 234・・・スコア化部
- 241・・・時間周波数変換部
- 242・・・トーン尤度分布検出部
- 243・・・時間周波数変換部
- 244・・・真引き・量子化部

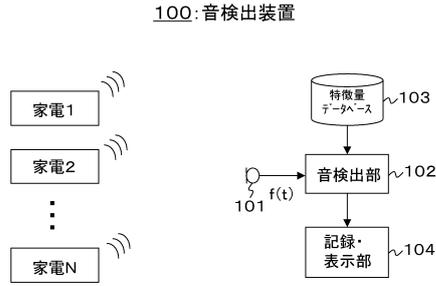
10

20

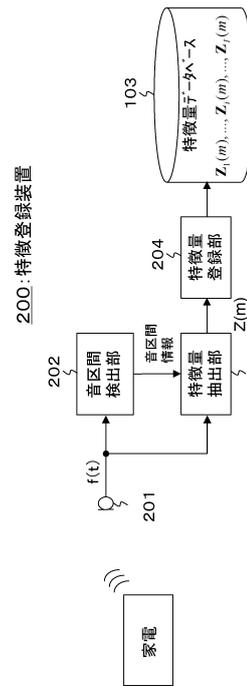
30

40

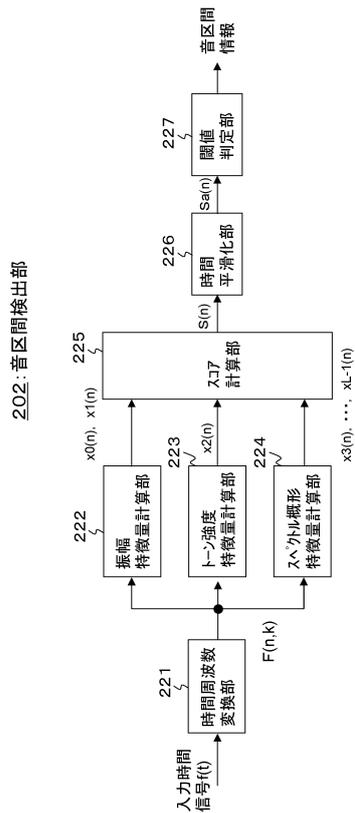
【図1】



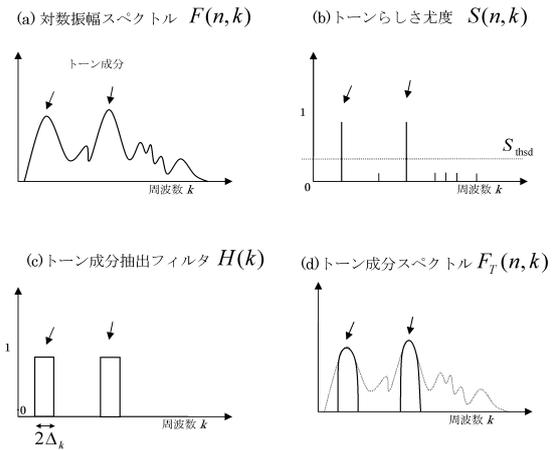
【図2】



【図4】



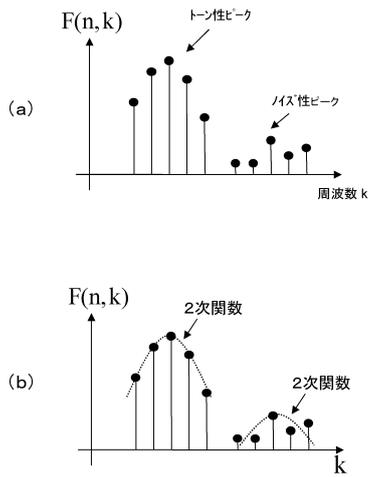
【図5】



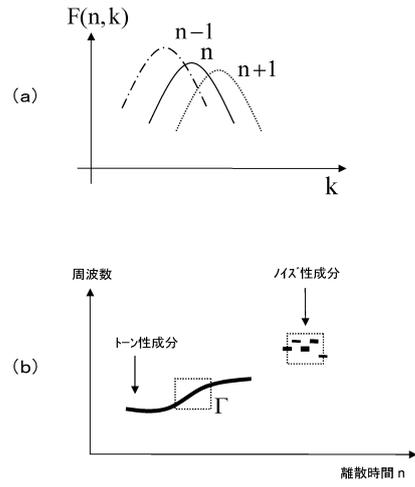
【図6】



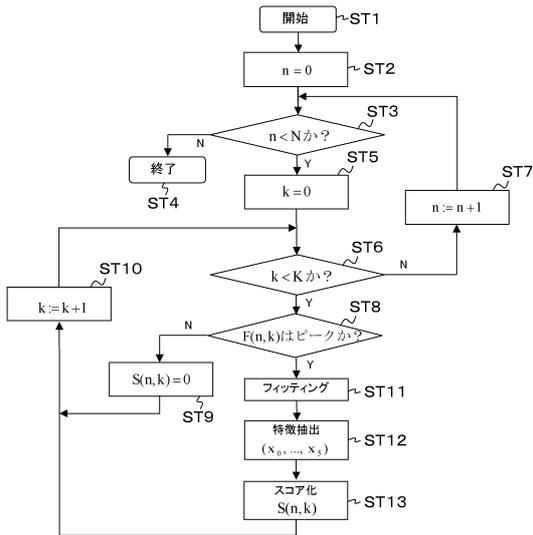
【図7】



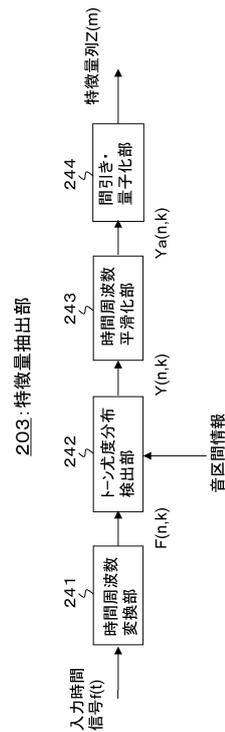
【図8】



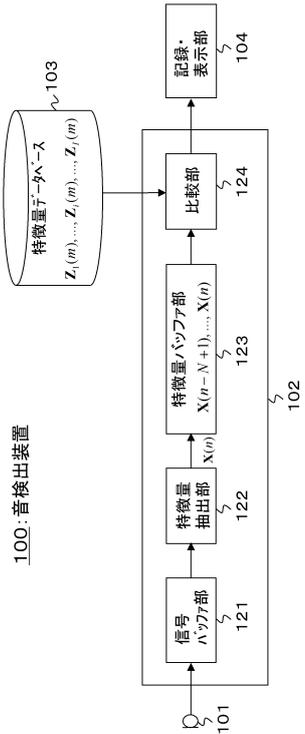
【図9】



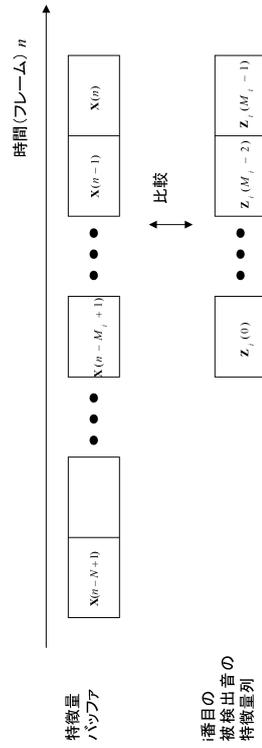
【図12】



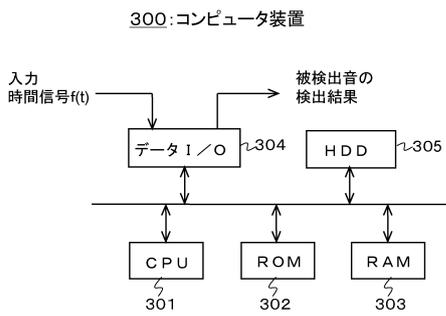
【図13】



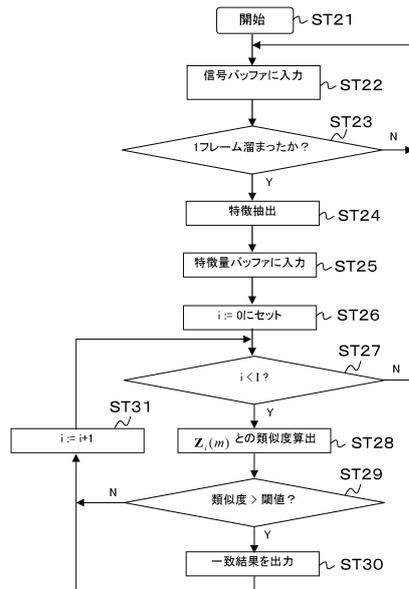
【図14】



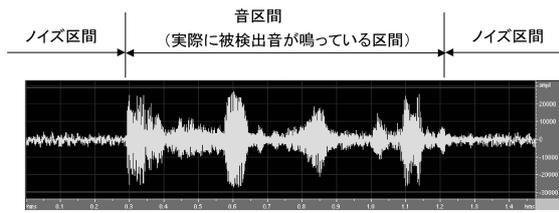
【図15】



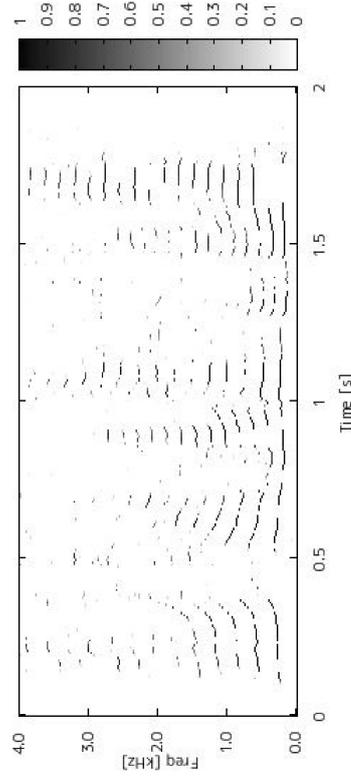
【図16】



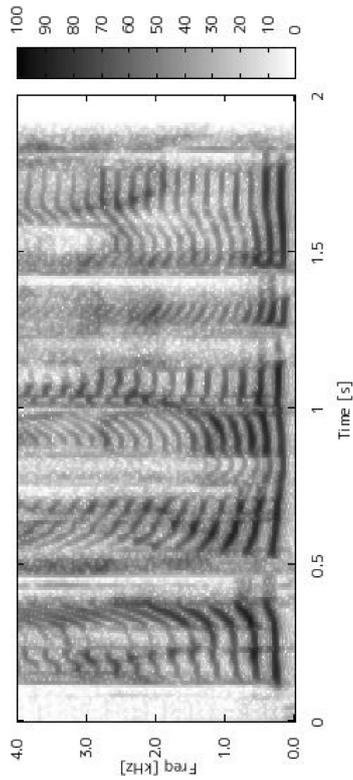
【図 3】



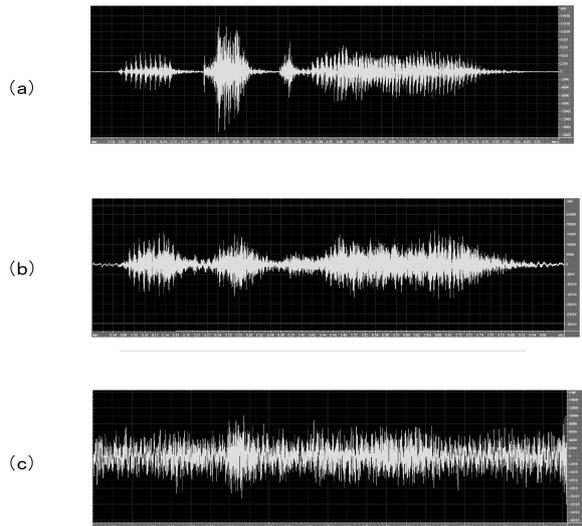
【図 10】



【図 11】



【図 17】



フロントページの続き

- (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
- (72)発明者 西口 正之
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 倉田 宜典
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 大野 弘

(56)参考文献 特開平09-026354(JP,A)

(58)調査した分野(Int.Cl., DB名)

G10L 25/51
G10L 15/04
G10L 25/03