

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6301891号
(P6301891)

(45) 発行日 平成30年3月28日 (2018.3.28)

(24) 登録日 平成30年3月9日 (2018.3.9)

(51) Int.Cl.	F I
G 1 O L 15/06 (2013.01)	G 1 O L 15/06 3 O O C
G O 6 N 7/00 (2006.01)	G O 6 N 7/00 1 5 O
G O 6 N 99/00 (2010.01)	G O 6 N 99/00 1 5 3
G 1 O L 15/10 (2006.01)	G 1 O L 15/10 5 O O Z
G 1 O L 21/0208 (2013.01)	G 1 O L 21/0208 1 O O B
請求項の数 8 (全 20 頁) 最終頁に続く	

(21) 出願番号	特願2015-191445 (P2015-191445)	(73) 特許権者	000004226
(22) 出願日	平成27年9月29日 (2015.9.29)		日本電信電話株式会社
(65) 公開番号	特開2017-67929 (P2017-67929A)		東京都千代田区大手町一丁目5番1号
(43) 公開日	平成29年4月6日 (2017.4.6)	(74) 代理人	100121706
審査請求日	平成29年2月20日 (2017.2.20)		弁理士 中尾 直樹
		(74) 代理人	100128705
			弁理士 中村 幸雄
		(74) 代理人	100147773
			弁理士 義村 宗洋
		(72) 発明者	小泉 悠馬
			東京都千代田区大手町一丁目5番1号 日
			本電信電話株式会社内
		(72) 発明者	齊藤 翔一郎
			東京都千代田区大手町一丁目5番1号 日
			本電信電話株式会社内
			最終頁に続く

(54) 【発明の名称】 学習装置、識別装置、その方法、およびプログラム

(57) 【特許請求の範囲】

【請求項 1】

雑音は1つ以上の雑音テンプレートの混合で表現されるものとし、周波数領域に変換された雑音の音信号 $x^{(L)}_{j,0}$ を用いて、1つ以上の雑音テンプレートを含み、確率モデルであらわされる雑音モデルを学習する雑音モデル学習部と、

前記音信号 $x^{(L)}_{j,0}$ と、周波数領域に変換された既定の動作音 k と雑音とが混ざった音信号 $x^{(L)}_{i,k}$ とを用いて、前記動作音 k における周波数の重要度 $\alpha_{k,h}$ を計算する周波数重要度計算部と、

$h=1,2,\dots,H_k$ とし、前記重要度 $\alpha_{k,h}$ が高いと判断された周波数 $f_{k,h}$ と、前記周波数 $f_{k,h}$ の総数 H_k とを求める正弦波数及び周波数計算部と、

前記音信号 $x^{(L)}_{j,0}$ と前記音信号 $x^{(L)}_{i,k}$ とを用いて、前記周波数 $f_{k,h}$ における正弦波強度 $A_{k,h}$ を計算する正弦波強度計算部とを含む、

学習装置。

【請求項 2】

請求項1の学習装置で得た雑音モデル及び周波数 $f_{k,h}$ 、総数 H_k 、正弦波強度 $A_{k,h}$ を用いて、識別対象の音響信号に含まれる動作音を識別する識別装置であって、

周波数領域に変換された識別対象の音響信号 X と前記雑音モデルに含まれる雑音テンプレートとを用いて、前記雑音テンプレートの混合比を求め、前記雑音テンプレートと前記混合比とから雑音の推定値 N を計算する雑音計算部と、

前記周波数 $f_{k,h}$ と前記総数 H_k と前記正弦波強度 $A_{k,h}$ とを用いて、正弦波重畳モデルによ

り、動作音 k の近似値を求め、前記音響信号 X と前記近似値と前記推定値 N とから、ある時刻において前記音響信号 X に前記動作音 k が含まれる確率であるフレームワイズ動作音確率を計算するフレームワイズ動作音確率計算部とを含む、
識別装置。

【請求項 3】

請求項 2 の識別装置であって、

識別対象の音響信号に含まれると想定される動作音の種類総数を K とし、 $k=1, 2, \dots, K$ 、 $k'=1, 2, \dots, K$ とし、ある時刻の動作音 k' から次の時刻の動作音 k へ移り変わる遷移確率と前記フレームワイズ動作音確率とを用いて、一つ前の時刻の動作音を考慮した、ある時刻において前記音響信号 X に前記動作音 k が含まれる確率である動作音確率を計算する
動作音確率計算部を含む、
識別装置。

10

【請求項 4】

請求項 2 または請求項 3 の識別装置であって、

前記フレームワイズ動作音確率に基づき得られる動作音の推定結果 z と前記近似値と前記推定値 N とを用いて、前記音響信号 X に含まれる動作音を強調するためのフィルタを設計するフィルタ設計部と、

前記音響信号 X と前記フィルタとを用いて、前記音響信号 X に含まれる動作音を強調するフィルタリング部とを含む、
識別装置。

20

【請求項 5】

請求項 2 から請求項 4 の何れかの識別装置であって、

前記推定結果 z と、推定結果が得られた時刻とを用いて、動作音のログを生成する、
識別装置。

【請求項 6】

雑音は1つ以上の雑音テンプレートの混合で表現されるものとし、周波数領域に変換された雑音の音信号 $x^{(L)}_{j,0}$ を用いて、1つ以上の雑音テンプレートを含み、確率モデルであらわされる雑音モデルを学習する雑音モデル学習ステップと、

前記音信号 $x^{(L)}_{j,0}$ と、周波数領域に変換された既定の動作音 k と雑音とが混ざった音信号 $x^{(L)}_{i,k}$ とを用いて、前記動作音 k における周波数の重要度 $\alpha_{k,h}$ を計算する周波数重要度計算ステップと、

30

$h=1, 2, \dots, H_k$ とし、前記重要度 $\alpha_{k,h}$ が高いと判断された周波数 $f_{k,h}$ と、前記周波数 $f_{k,h}$ の総数 H_k とを求める正弦波数及び周波数計算ステップと、

前記音信号 $x^{(L)}_{j,0}$ と前記音信号 $x^{(L)}_{i,k}$ とを用いて、前記周波数 $f_{k,h}$ における正弦波強度 $A_{k,h}$ を計算する正弦波強度計算ステップとを含む、
学習方法。

【請求項 7】

請求項 6 の学習方法で得た雑音モデル及び周波数 $f_{k,h}$ 、総数 H_k 、正弦波強度 $A_{k,h}$ を用いて、識別対象の音響信号に含まれる動作音を識別する識別方法であって、

周波数領域に変換された識別対象の音響信号 X と前記雑音モデルに含まれる雑音テンプレートとを用いて、前記雑音テンプレートの混合比を求め、前記雑音テンプレートと前記混合比とから雑音の推定値 N を計算する雑音計算ステップと、

40

前記周波数 $f_{k,h}$ と前記総数 H_k と前記正弦波強度 $A_{k,h}$ とを用いて、正弦波重畳モデルにより、動作音 k の近似値を求め、前記音響信号 X と前記近似値と前記推定値 N とから、ある時刻において前記音響信号 X に前記動作音 k が含まれる確率であるフレームワイズ動作音確率を計算するフレームワイズ動作音確率計算ステップとを含む、
識別方法。

【請求項 8】

請求項 1 の学習装置、または、請求項 2 から請求項 5 の何れかの識別装置として、コンピュータを機能させるためのプログラム。

50

【発明の詳細な説明】

【技術分野】

【0001】

音響信号に含まれる所定の音を識別する技術に関する。

【背景技術】

【0002】

音響信号が何の音であるか（足音、水が流れる音など）を分析する技術として、非特許文献1が知られている。非特許文献1では、音響イベント列の生成過程を、大規模かつ複雑な統計モデルでモデル化することで、音響特徴量の生成モデルや識別器を構築している。生成モデルから、長時間の音響信号毎の状況の生成確率を分析したり、新たに入力された音響イベント列と生成モデルの距離を利用して、音響信号が示す状況を推定することを可能とする。

10

【先行技術文献】

【非特許文献】

【0003】

【非特許文献1】井本佳右ほか，“音響イベント列の確率的生成モデルを利用した音響トピックとユーザ行動との関係分析”，日本音響学会，2013年春期音響学会

【発明の概要】

【発明が解決しようとする課題】

【0004】

20

音響信号に含まれる所定の音を識別する技術を、工場内の機器の動作音の識別に利用することを考える。工場などに設置された大型の製造機および成型機などの業務用機器は、故障により稼働がストップするだけで、業務に大きな支障をもたらす。そのため、その動作状況を日常的に監視し、故障の発生を事前に防がなくてはならない。解決案として、業務用機器の管理業者が、定期的に現場へ整備員を派遣し、パーツの摩耗などを確認する方法がある。しかし、多大な人件費や移動費、労力がかかるため、すべての業務用機器や工場でこれを実施するのは難しい。代替案として、その機械を構成するパーツごとに、動作状況ログを残す方法がある。例として、図1のような、ベルトコンベア1、塗料噴射器2、アーム3のパーツで構成される機器の動作ログを残すことを考える。「動作ログ1」のように、「いつ」「どのパーツが」動作したかの“詳細な動作ログ”を残せば、稼働状況や使用頻度から、各パーツの消耗・摩耗度合いなどを推定できる。これにより、故障時期の予測や、パーツ交換の提案などが可能になり、故障率を大幅に落とせる。ところが現在のほとんどの動作ログは、「動作ログ2」のような大まかなログ（作業開始時刻と作業終了時刻だけなど）を残す機能しか備えておらず、詳細な動作ログを残す機能を備えてるものは少ない。

30

【0005】

そこで、動作音から詳細な動作ログを作成する方法を考える。例えば、各パーツの動作音に着目し、その動作音から詳細な動作ログを作成する。機器の動作音を収録し、その動作音がどのパーツから発せられていたかを識別することで、これを実現する。動作音から動作ログを作成するメリットは、“詳細な動作ログ”を容易に作成できることに加え、動作ログと動作音を紐づけることで、人間が遠隔地から各パーツごとの動作音を聞き、異常診断ができる点である（図2）。そのための要素技術として、音響信号から動作音の種類を識別し、所望の動作音だけを個別に強調することが必要である。

40

【0006】

ところが、非特許文献1では、家庭内などの比較的静かな環境を想定し、MFCCやLPCといった一般的な音響特徴量を用いて識別をしている。大型の業務用機械が動いている環境はえてして、他の機械（例えば、冷却ファン）が大騒音のノイズを出していることが多い。また、該当機器に取り付けられた、一部のパーツが大騒音のノイズを出していることもある（コンプレッサーなど）。そのため、各パーツの動作音は信号雑音比（S/N比）が悪く、MFCCやLPCなどの一般的な特徴量では動作音の識別は難しい。また同様に、音響信号

50

から所望の動作音だけを個別に強調することも困難である。

【0007】

本発明は、大騒音のノイズが存在する状況においても、音響信号に含まれる動作音を識別することができる識別装置、識別装置において用いられるパラメータを学習する学習装置、その方法及びプログラムを提供することを目的とする。

【課題を解決するための手段】

【0008】

上記の課題を解決するために、本発明の一態様によれば、学習装置は、雑音は1つ以上の雑音テンプレートの混合で表現されるものとし、周波数領域に変換された雑音の音信号 $x^{(L)}_{j,0}$ を用いて、1つ以上の雑音テンプレートを含み、確率モデルであらわされる雑音モデルを学習する雑音モデル学習部と、音信号 $x^{(L)}_{j,0}$ と、周波数領域に変換された既定の動作音 k と雑音とが混ざった音信号 $x^{(L)}_{i,k}$ とを用いて、動作音 k における周波数の重要度 $W_{k,h}$ を計算する周波数重要度計算部と、 $h=1,2,\dots,H_k$ とし、重要度 $W_{k,h}$ が高いと判断された周波数 $f_{k,h}$ と、周波数 $f_{k,h}$ の総数 H_k とを求める正弦波数及び周波数計算部と、音信号 $x^{(L)}_{j,0}$ と音信号 $x^{(L)}_{i,k}$ とを用いて、周波数 $f_{k,h}$ における正弦波強度 $A_{k,h}$ を計算する正弦波強度計算部とを含む。

【0009】

上記の課題を解決するために、本発明の他の態様によれば、学習方法は、雑音は1つ以上の雑音テンプレートの混合で表現されるものとし、周波数領域に変換された雑音の音信号 $x^{(L)}_{j,0}$ を用いて、1つ以上の雑音テンプレートを含み、確率モデルであらわされる雑音モデルを学習する雑音モデル学習ステップと、音信号 $x^{(L)}_{j,0}$ と、周波数領域に変換された既定の動作音 k と雑音とが混ざった音信号 $x^{(L)}_{i,k}$ とを用いて、動作音 k における周波数の重要度 $W_{k,h}$ を計算する周波数重要度計算ステップと、 $h=1,2,\dots,H_k$ とし、重要度 $W_{k,h}$ が高いと判断された周波数 $f_{k,h}$ と、周波数 $f_{k,h}$ の総数 H_k とを求める正弦波数及び周波数計算ステップと、音信号 $x^{(L)}_{j,0}$ と音信号 $x^{(L)}_{i,k}$ とを用いて、周波数 $f_{k,h}$ における正弦波強度 $A_{k,h}$ を計算する正弦波強度計算ステップとを含む。

【発明の効果】

【0010】

本発明によれば、大騒音のノイズが存在する状況においても、音響信号に含まれる動作音を識別することができるという効果を奏する。

【図面の簡単な説明】

【0011】

【図1】動作ログを説明するための図。

【図2】動作ログと動作音を紐づけ、遠隔地で動作音を人間が聞く構成を説明するための図。

【図3】動作音の時間周波数解析結果の例を示す図。

【図4】機器がとりうる状態遷移確率 $P_{k,k'}$ の一例を図解した図。

【図5】式(4)の意味を説明するための図。

【図6】学習の手順を示す図。

【図7】識別、強調の手順を示す図。

【図8】第一実施形態に係る学習装置の機能ブロック図。

【図9】動作音について周波数重要度を計算した結果の例を示す図。

【図10】第一実施形態に係る識別装置の機能ブロック図。

【発明を実施するための形態】

【0012】

以下、本発明の実施形態について、説明する。なお、以下の説明に用いる図面では、同じ機能を持つ構成部や同じ処理を行うステップには同一の符号を記し、重複説明を省略する。以下の説明において、テキスト中で使用する記号「 \cdot 」「 \sim 」等は、本来直前の文字の真上に記載されるべきものであるが、テキスト記法の制限により、当該文字の直後に記載する。式中においてはこれらの記号は本来の位置に記述している。また、ベクトルや行列の

10

20

30

40

50

各要素単位で行われる処理は、特に断りが無い限り、そのベクトルやその行列の全ての要素に対して適用されるものとする。

【 0 0 1 3 】

< 第一実施形態のポイント >

観測信号（音響信号）から機械音の動作音の種類を識別する。より詳しくは、識別したい動作音毎に与えられた学習データを用いて、大騒音下で動作音の識別と強調を行う技術に関する。

【 0 0 1 4 】

本実施形態では、大騒音下で特定の動作音の種類を識別し、強調する機械学習ベースの技術を提供する。機械の動作音に特化してこれを実現するために、機械の動作音の以下の特徴に着目する。

【 0 0 1 5 】

< 機械の動作音の4つの特徴 >

1. 各動作音は、毎回同じ音が鳴る。
2. 機械の動作順には法則性がある。
3. 工場内の騒音は、ほぼ定常（冷却ファンなど）か、突発音が繰り返して鳴る（打鉄音など）。
4. 観測音は、動作音と騒音が混ざった音である。

【 0 0 1 6 】

以下では、この特徴について詳しく説明する。

< 1. 各動作音は、毎回同じ音が鳴る >

各パーツの動作音は、モーター等に起因する「ウィーン」のような音と、アラームなどに起因する「ピー」という音の2種類に大別できると考える。製造機器は、製品ごとの個体差をなくすために、毎回同じ動きをする。つまり、各パーツでは、動作音として常に同じような音が鳴ると考えられる。また、「ウィーン」や「ピー」といった動作音は、特徴的な周波数を持つことが多い。図3は、比較的静かな環境で収録した動作音の時間周波数解析結果である。この動作音は、「ピー」という音であり、3000Hzと4000Hzのあたりに強いピークを確認できる。このことから動作音 $s_{t,k}$ は、以下の正弦波重畳モデルで近似的に表現できるものとする。

【 0 0 1 7 】

【 数 1 】

$$s_{t,k} \approx \sum_{h=1}^{H_k} A_{k,h} \sin(2\pi f_{k,h} t) \quad (1)$$

【 0 0 1 8 】

ここで ω は円周率、 t は時間領域の時間を表すインデックス、 $k \in \{1, 2, \dots, K\}$ は動作音の種類を表すインデックス（ただし K は音響信号に含まれる動作音の種類の総数）、 H_k は k 番目の動作音を構成する特徴的な周波数の数、 $h \in \{1, 2, \dots, H_k\}$ は k 番目の動作音を構成する H_k 個の特徴的な周波数のうちの h 番目の特徴的な周波数を表すインデックス、 $f_{k,h}$ と $A_{k,h}$ とはそれぞれ k 番目の動作音の h 番目の特徴的な周波数 (Hz) とその強度とを表す。図3を例に挙げれば式(1)は、 $H_k=2$ 、 $f_{k,1}=3000$ 、 $f_{k,2}=4000$ であり、動作音が2つの正弦波の足し算で表現できることを示している。

【 0 0 1 9 】

< 2. 機械の動作には順番に法則性がある >

機械による製品の製造は、アルゴリズムで記述できる決められた「手順」に基づき行われる。図1を例に挙げれば、「ベルトコンベアによる製品の移動」「塗料の噴射」「製品の回収」「休止」、という手順に従って、機械は動作している。これは現在動作しているパーツは、1時刻前の動作しているパーツに応じて推測できることを示している。

時間フレームを z と置き、 $z \in \{1, 2, \dots, K\}$ を時刻 z での動作の種類と置くと、 z は z_{t-1} に依存した1次のマルコフモデルで記述できる。図1を例に挙げると、「ベルトコンベア

10

20

30

40

50

による製品の移動」を $z = 1$ 、「塗料の噴射」を $z = 2$ 、「製品の回収」を $z = 3$ 、「休止」を $z = 4$ と置いたとき、 $z_{t-1} = 1$ ならば、 z_t は1か2しか取りえない。このことを確率を用いて記述すると、以下のように書ける。

【0020】

【数2】

$$p(z_t = k | z_{t-1} = k') = \theta_{k,k'} \quad (2)$$

【0021】

ここで $\theta_{k,k'}$ は、時刻 $t-1$ の状態が k' (ただし、 $k' \in \{1, 2, \dots, K\}$ である)だったとき、時刻 t の状態が k である、動作が移り変わる確率(状態遷移確率)を示す。図4は、図1の機器がとりうる状態遷移確率 $\theta_{k,k'}$ の一例を図解したものである。 $k' = 1$ だったとき、 $\theta_{3,1} = 0$ であることが、 $z_{t-1} = 1$ ならば、 z_t は1か2しか取りえないことを示している。

【0022】

<3.工場内の騒音は、ほぼ定常か、突発音が繰り返して鳴る>

工場内の騒音(背景雑音、単に雑音ともいう)は、ほぼ定常(冷却ファンなど)か数個のスペクトル状態で記述できる突発音(打鋸音など)の2種類に大別できると考える。ここで定常な雑音とは、送風機のような「ゴー」という雑音である。また数個のスペクトル状態で記述できる突発音とは、ドリルやハンマーのような「ドドドドド」など一定のリズムで大きな打撃音が鳴るものである。また、実際の騒音(背景雑音)は、これらの音、つまり M 種類の雑音のテンプレート $\mu_{\omega,m}$ が、混合比率 w_m (ただし $\sum_{m=1}^M w_m = 1$)で混ぜ合わさったものと考え、時刻 t での雑音の音響信号 $N_{\omega,t}$ を以下のように記述する。

【0023】

【数3】

$$N_{\omega,t} = \sum_{m=1}^M w_{m,t} \mu_{\omega,m} \quad (3)$$

【0024】

ここで ω は周波数領域の周波数インデックスであり、 $\omega \in \{1, 2, \dots, W\}$ である。つまり、本実施形態では雑音を1つ以上の雑音テンプレートの混合で表現する。

【0025】

<4.観測音は、動作音と騒音が混ざった音である>

実際にマイクロホンで観測される音(振幅スペクトル) $X_{\omega,t}$ は、1.から3.で説明される音が足しあわされて観測されるものとする。

【0026】

【数4】

$$X_{\omega,t} = S_{\omega,z_t} + N_{\omega,t} \quad (4)$$

【0027】

ここで S_{ω,z_t} (ただし、下付き添え字の z_t は z を意味する。)は式(1)の $s_{t,k}$ をフーリエ変換し、絶対値を取った、動作音の振幅スペクトルである。式(4)の意味を、図5を使って説明する。まず S_{ω,z_t} は、時刻 t では z_t 番目のパーツが動作しており、その動作音は S_{ω,z_t} であることを示している。つまり、式(1)(2)にしたがって動作音が生成されていることを表す。続いて $N_{\omega,t}$ は時刻 t では、それぞれの雑音源 m が w_m の混合比で混ぜ合わさっている(式(3))。そして、上述の手順で生成された S_{ω,z_t} と $N_{\omega,t}$ とが重畳され、観測信号 $X_{\omega,t}$ が生成される。

【0028】

本実施形態は、1.から4.の説明に基づき、式(4)の逆問題を解いて z_t (時刻 t で動作しているパーツの種類)を識別するものである。また、 z_t を解くことにより動作音の振幅スペクトル S_{ω,z_t} が求まり、動作音だけの強調が可能になる。この逆問題を解くために必要なパラメータを、学習データから学習する。本実施形態の学習と識別、強調の手順の

10

20

30

40

50

例を、それぞれ図 6 と図 7 に示す。以降、まず図 6 に沿って学習の手順を説明し、次に図 7 に沿って識別/強調の手順を説明する。

【 0 0 2 9 】

< 学習手順 >

本実施形態の学習手順を実現する学習装置 1 0 0 について説明する。

【 0 0 3 0 】

学習装置 1 0 0 は、学習データ(音響信号 $x^{(L)}_t$ 、ラベルファイルlabel)を入力とし、雑音モデルのパラメータ $\theta = \{\mu_m, \mu_{1,m}, \dots, \mu_{M,m}\}_{m=1}^M$ 、各動作音 k の正弦波重畳モデルパラメータ(K 個の H_k と、 $(\sum_{k=1}^K H_k)$ 個の周波数 $f_{k,h}$ と、 $(\sum_{k=1}^K H_k)$ 個の正弦波強度 $A_{k,h}$)を学習し、出力する。

10

【 0 0 3 1 】

学習装置 1 0 0 は、周波数領域変換部 1 0 2 と、データ切り出し部 1 0 3 と、雑音モデル学習部 1 0 4 と、 K 個の特徴周波数計算部 1 0 5 - k とを含む。図 8 は、学習装置 1 0 0 の機能ブロック図を示す。

【 0 0 3 2 】

< 学習データ >

識別対象の機器の近くに(内部でも可能)マイクロホンを設置し、音響信号(以下、マイクロホンで収音された音響信号を「観測信号」ともいう) $x^{(L)}_t$ を収集する。なお、上付き添え字(L)は学習データ自体、または、学習データから得られるデータを表すインデックスである。収集した音響信号 $x^{(L)}_t$ に、生成したい詳細な動作ログと同じ粒度でラベルファイル(動作ログ)labelを人手で作成する。ラベルファイルlabelは、時刻 t に対応する音響信号 $x^{(L)}_t$ がどの動作音に対応するかを示す情報であり、例えば、各時刻 t に対応する各音響信号 $x^{(L)}_t$ に各動作音を示すラベルを付与してもよいし、対象となる学習データ(音響信号 $x^{(L)}_1, x^{(L)}_2, \dots$)全てに対して、各動作音を示すラベルとその動作音の開始時刻及び終了時刻を付与してもよい。

20

【 0 0 3 3 】

このときのサンプリング周波数は、動作音の特徴的な周波数を記録できるように設定すればよい。例えば16kHzなどに設定する。

【 0 0 3 4 】

さらに可能であれば、識別対象の機器の近くに複数のマイクロホンを設置したり、雑音発生源の近くにマイクロホンを設置してもよい。複数のマイクロホンを設置した場合、ビームフォーミングやスペクトルサブトラクションなどの技術が使用可能になるため、動作音、雑音の推定精度が向上する。

30

【 0 0 3 5 】

< 周波数領域変換部 1 0 2 >

周波数領域変換部 1 0 2 は、音響信号 $x^{(L)}_t$ を受け取り、音響信号 $x^{(L)}_t$ を短時間フーリエ変換(STFT)などの手法で周波数領域に変換する。その後、複素スペクトルの絶対値を取り、振幅スペクトル(以下、この振幅スペクトルを単に「周波数領域に変換された音響信号」または「音響信号」ともいう) $X^{(L)} = (X^{(L)}_1, \dots, X^{(L)}_2, \dots, X^{(L)}_1, \dots, X^{(L)}_1, \dots)^T$ を得(S 1 0 2)、出力する。ただし、 T は転置を表す。ここで、フーリエ変換長は32ms、シフト幅は16msなどに設定できる。

40

【 0 0 3 6 】

< データ切り出し部 1 0 3 >

データ切り出し部 1 0 3 は、周波数領域に変換された音響信号 $X^{(L)}$ と、ラベルファイルlabelとを受け取り、ラベルファイルlabelを元に、周波数領域に変換された音響信号 $X^{(L)}$ を、雑音と、動作音の種類ごとに分割し(S 1 0 3)、雑音の音響信号 $x^{(L)}_{j,0}$ と動作音 k の音響信号 $x^{(L)}_{i,k}$ と出力する。ここで、雑音の音響信号 $x^{(L)}_{j,0}$ には雑音のみが、動作音 k の音響信号 $x^{(L)}_{i,k}$ には動作音と雑音が重畳された状態となっている。なお、音響信号 $x^{(L)}_{j,0}$ に含まれる雑音と音響信号 $x^{(L)}_{i,k}$ に含まれる雑音は、同じ種類の雑音である。ただし、そのレベルは異なってもよい。

50

【 0 0 3 7 】

なお、音響信号 $X^{(L)}$ から切り出された k 番目の動作音の l_k フレーム分の音響信号を $x^{(L)}_{i,k}=(X^{(L)}_{1,i,k}, X^{(L)}_{2,i,k}, \dots, X^{(L)}_{j,i,k}, \dots, X^{(L)}_{J,i,k})^T$ とおき、音響信号 $X^{(L)}$ から切り出された雑音の J フレーム分の音響信号を $x^{(L)}_{j,0}=(X^{(L)}_{1,j,0}, X^{(L)}_{2,j,0}, \dots, X^{(L)}_{j,j,0}, \dots, X^{(L)}_{J,j,0})^T$ とおく。ただし、 l_k は動作音 k 毎に異なり、 $i=1,2,\dots,l_k$ である。

【 0 0 3 8 】

なお、ある時刻 の音響信号 $X^{(L)}$ に2つ以上の動作音が含まれる(2つ以上の動作音が重畳された状態の場合)には、推定精度を高めるために、その部分を排除し、以下の学習には用いない構成としてもよい。

【 0 0 3 9 】

< 雑音モデル学習部 1 0 4 >

雑音モデル学習部 1 0 4 は、音響信号 $X^{(L)}$ から切り出された雑音の J フレーム分の音響信号 $x^{(L)}_{j,0}$ を受け取り、この値を用いて、雑音モデルを学習し(S 1 0 4)、学習装置 1 0 0 の出力値として雑音モデルパラメータ を出力する。ここで、雑音モデルは、 M 個(M は1以上の整数)の雑音テンプレート $\mu_{1,m}, \mu_{2,m}, \dots, \mu_{J,m}$ (ただし、 $m=1,2,\dots,M$)を含み、確率モデルであらわされる。言い換えると、学習時には、雑音のみの学習データ(雑音の J フレーム分の音響信号 $x^{(L)}_{j,0}$)を用いて、何らかの確率モデルにより雑音モデルを事前学習する。GMMの場合は、何らかの確率モデルとして「混合ガウス分布(GMM)」を用いる。GMMの場合、EMアルゴリズムなどの手法を用いて雑音モデルを学習する(例えば、

[非特許文献2]小西貞則、“多変量解析入門、付録C EMアルゴリズム”、岩波書店、pp.294-298、2010年

この場合、雑音データ(ここでは、雑音の J フレーム分の音響信号 $x^{(L)}_{j,0}=(X^{(L)}_{1,j,0}, X^{(L)}_{2,j,0}, \dots, X^{(L)}_{j,j,0})^T$ であり、 $X^{(L)}_{j,j,0}$ を $N^{(L)}_{j,j,0}$ とも記載する($X^{(L)}_{j,j,0}=N^{(L)}_{j,j,0}$))。ここで j は、時刻フレーム の何れかに対応する)は以下の確率分布に従って生成されたものとしてモデル化する。

【 0 0 4 0 】

【 数 5 】

$$N^{(L)}_{\omega,\tau} \sim \sum_{m=1}^M \alpha_m N(\mu_{\omega,m}, \Sigma_m) \quad (5)$$

【 0 0 4 1 】

ここで $\mu_{\omega,m}$ 、 $\mu_{\omega,m}$ はそれぞれ、 m 番目の混合分布の混合比、周波数 における平均ベクトル(雑音テンプレート)、共分散行列である。この場合、雑音モデルパラメータは、雑音GMMパラメータであり、 $\theta = \{\mu_{\omega,m}, \mu_{\omega,m}, \dots, \mu_{\omega,m}, \Sigma_m\}_{m=1}^M$ で表される。なお、共分散行列 Σ_m は対角行列に制限してもよいし、 m 番目の混合分布の分散 Σ_m を用いて、 Σ_m と制限してもよい。ここで I は単位行列を表す。

【 0 0 4 2 】

GMMの学習は、雑音の音響信号 $N^{(L)}$ を構成する雑音を M 種類にクラスタリングしていることに相当する。つまり、工場の中で鳴っている M 種類の雑音を推定しており、 $\mu_{\omega,m}$ は m 番目の雑音テンプレートとみなすこともできる。そのため、GMMの混合数 M は、工場内で想定される雑音源の数を設定すべきであり、例えば8などに設定できる。

【 0 0 4 3 】

なお、学習データを得る際に、複数のマイクロホンを配置した場合、GMMの学習と平均ベクトル(雑音テンプレート) $\mu_{\omega,m}$ の作成手順が異なる。複数チャネルを用いてビームフォーミングやスペクトラルサブトラクションを施した音響特徴量ベクトルを $N^{(add)}$ とする。この音響特徴量ベクトル $N^{(add)}$ を、音響信号 $X^{(L)}$ から切り出された雑音の J フレーム分の音響信号 $x^{(L)}_{j,0}=N^{(L)}$ と結合させて、以下のGMMを学習する。

【 0 0 4 4 】

10

20

30

40

50

【数 6】

$$\begin{bmatrix} N_{1,\tau}^{(L)} \\ \vdots \\ N_{\Omega,\tau}^{(L)} \\ N_{\tau}^{(add)} \end{bmatrix} \sim \sum_{m=1}^M \alpha_m N(\bar{\mu}_m, \Sigma_m), \quad \bar{\mu}_m = \begin{bmatrix} \mu_{1,m} \\ \vdots \\ \mu_{\Omega,m} \\ \mu_m^{(add)} \end{bmatrix} \quad (6)$$

【 0 0 4 5 】

つまり雑音GMMパラメータは、 $\theta = \{\mu_m, \Sigma_m\}_{m=1}^M$ であり、雑音テンプレートは μ_m の上位次元となる。

10

【 0 0 4 6 】

なお、識別・強調時に雑音を推定するには、事前学習した雑音モデルを用いて、観測信号から雑音を推定する。推定法には、最小平均二乗誤差(MMSE)推定や最大事後確率(MAP)推定などを用いることができる。どの推定法を用いるかは任意であるが、雑音モデルに用いた確率分布の種類に応じて、推定精度と計算速度のトレードオフを考慮しながら決めることが望ましい。なお、GMMの場合は、このトレードオフを考慮して、MMSE推定を使うことが一般的である。

【 0 0 4 7 】

<特徴周波数計算部 1 0 5 - k >

K個の特徴周波数計算部 1 0 5 - k はそれぞれ以下の処理を行う。特徴周波数計算部 1 0 5 - k は、雑音の音響信号 $x^{(L)}_{j,0}$ と動作音kの音響信号 $x^{(L)}_{i,k}$ を受け取り、動作音kの正弦波重畳モデルパラメータ(ここでは、正弦波数 H_k , その周波数 $f_{k,h}$, 正弦波強度 $A_{k,h}$)を学習し(S 1 0 5)、出力する。例えば、特徴周波数計算部 1 0 5 - k は、周波数重要度計算部 1 0 5 A と、正弦波数及び周波数計算部 1 0 5 B と正弦波強度計算部 1 0 5 C とを含む(図 8 参照)。

20

【 0 0 4 8 】

(周波数重要度計算部 1 0 5 A)

周波数重要度計算部 1 0 5 A は、雑音の音響信号 $x^{(L)}_{j,0}$ と動作音kの音響信号 $x^{(L)}_{i,k}$ を受け取り、これらの値を用いて、動作音kにおける周波数の重要度 $w_{k,h}$ を計算し(S 1 0 5 A)、出力する。なお、重要度 $w_{k,h}$ とは、周波数の成分が前記動作音kを識別する際に重要な成分か否かを示す指標である。例えば、以下のようにして重要度 $w_{k,h}$ を計算する。

30

【 0 0 4 9 】

まず、雑音の音響信号 $x^{(L)}_{j,0}$ と動作音kの音響信号 $x^{(L)}_{i,k}$ を用いて、ロジスティック回帰モデルを学習する(例えば参考文献 2 参照)。

(参考文献 2)小西貞則、“多変量解析入門、4章ロジスティック回帰モデル”、岩波書店、pp.79-96、2010年。

【 0 0 5 0 】

説明変数行列 X_k と応答変数ベクトル Y_k をそれぞれ

【 0 0 5 1 】

40

【数 7】

$$X_k = (\underbrace{x_{1,k}, x_{2,k}, \dots, x_{I_k,k}}_{I_k}, \underbrace{x_{1,0}, x_{2,0}, \dots, x_{J,0}}_J) \quad (9)$$

$$Y_k = (\underbrace{1, 1, \dots, 1}_{I_k}, \underbrace{0, 0, \dots, 0}_J) \quad (10)$$

と置き、以下のロジスティック回帰モデルを学習する。

【 0 0 5 2 】

50

【数 8】

$$Y_k = \frac{1}{1 + \exp(-X_k^T \beta_k + \beta_0)} \quad (11)$$

【0053】

ここで $X_k = (x_{1,k}, x_{2,k}, \dots, x_{n,k})^T$ は回帰ベクトルである。

【0054】

ロジスティック回帰モデルの回帰ベクトルの特徴として、識別に有効な次元に対応する回帰係数の絶対値が大きくなる。すなわち、 $|\beta_{\omega,k}|$ が大きな周波数ほど、動作音kの識別に有効な周波数、すなわち重要な周波数とみなすことができる。そこで本実施形態では周波数 ω に対する重要度 $\gamma_{\omega,k}$ を以下のように定義する。

10

【0055】

【数 9】

$$\gamma_{\omega,k} = \frac{|\beta_{\omega,k}|}{\max_{\omega} \{|\beta_{\omega,k}|\}} \quad (12)$$

【0056】

この値が大きな周波数ほど、その動作音の特徴をよく表現する周波数と考える。2つの異なる動作音について周波数重要度を計算した結果の例を図9に示す。いくつかの周波数において、周波数重要度のピークが確認できる。なお、回帰ベクトルの過学習を避け、安定して周波数重要度を計算するために、ロジスティック回帰の学習時に正則化項を付与した方がよい。この正則化項には、回帰ベクトルのL1ノルムを用いるLASSOや回帰ベクトルのL2ノルムを用いるリッジ回帰を使うことができる。

20

【0057】

(正弦波数及び周波数計算部 105B)

正弦波数及び周波数計算部 105B は、重要度 $\gamma_{\omega,k}$ を受け取り、重要度 $\gamma_{\omega,k}$ が高いと判断された周波数 $f_{k,h}$ と、周波数 $f_{k,h}$ の総数 H_k とを求め (S105B)、出力する。なお、重要度 $\gamma_{\omega,k}$ が高いとは、周波数 ω の成分が動作音kを識別する際に重要な成分であることを意味する。例えば、周波数重要度に対する閾値 γ_{TH} を設定する。重要度 $\gamma_{\omega,k}$ と閾値 γ_{TH} との大小関係により、重要度 $\gamma_{\omega,k}$ が高いか否かを判断する。重要度 $\gamma_{\omega,k}$ が式 (12) で表される場合には、重要度 $\gamma_{\omega,k}$ が大きければ大きいほど、動作音kに対して周波数 ω が重要であると考えられるので、 $\gamma_{\omega,k} > \gamma_{TH}$ となる周波数が動作音kを識別する際に重要な成分であると考ええる。そこで、 $\gamma_{\omega,k} > \gamma_{TH}$ となる周波数を $f_{k,h}$ とする。ただし、 $f_{k,h}$ は周波数の値自体であってもよいし、それと等価な値 (例えば周波数のインデックス) であってもよい。さらに、 $\gamma_{\omega,k} > \gamma_{TH}$ となる周波数 $f_{k,h}$ の個数を数え、その総数を H_k とする。総数 H_k と周波数 $f_{k,h}$ の例を図9に示す。動作音1を例に挙げれば、 $H_k=2$ であり、 $f_{k,1}$ は2100Hz付近、 $f_{k,2}$ は4200Hz付近となる。なお閾値 γ_{TH} は総数 H_k が大きくなりすぎないように経験的に決定すべきであり、たとえば0.5に設定できる。

30

【0058】

(正弦波強度計算部 105C)

正弦波強度計算部 105C は、総数 H_k と、 H_k 個の周波数 $f_{k,h}$ と、雑音の音響信号 $x^{(L)}_{j,0}$ と動作音kの音響信号 $x^{(L)}_{i,k}$ とを受け取り、周波数 $f_{k,h}$ における音響信号の正弦波強度 $A_{k,h}$ を計算し (S105C)、出力する。例えば、正弦波強度 $A_{k,h}$ は以下のように計算する。

40

【0059】

【数 10】

$$A_{k,h} = \frac{1}{I_k} \sum_{i=1}^{I_k} X_{f_{k,h},i,k}^{(L)} - \frac{1}{J} \sum_{j=1}^J X_{f_{k,h},j,0}^{(L)} \quad (13)$$

【0060】

50

$X_{f_{k,h},i,k}^{(L)}$ (ただし、下付添え字 $f_{k,h}$ は $f_{k,h}$ を意味する。) はフレーム i の音響信号 $x_{i,k}^{(L)} = (X_{1,i,k}^{(L)}, X_{2,i,k}^{(L)}, \dots, X_{f_{k,h},i,k}^{(L)})^T$ のうちの周波数 $f_{k,h}$ に対応する周波数成分を意味し、 $X_{f_{k,h},j,0}^{(L)}$ は、フレーム j の $x_{j,0}^{(L)} = (X_{1,j,0}^{(L)}, X_{2,j,0}^{(L)}, \dots, X_{f_{k,h},j,0}^{(L)})^T$ のうちの周波数 $f_{k,h}$ に対応する周波数成分を意味する。また万が一、 $A_{k,h}$ が 0 を下回った場合、 $A_{k,h} = 0$ とする。ここで、式 (13) の第一項は動作音 k があるときの周波数 $f_{k,h}$ の周波数成分 (振幅スペクトル値) の平均値、第二項は雑音の周波数 $f_{k,h}$ の周波数成分 (振幅スペクトル値) の平均値である。

【 0 0 6 1 】

学習装置 1 0 0 は、以上の処理により、学習データから雑音モデルパラメータ $\theta = \{\mu_m, \mu_{1,m}, \dots, \mu_{f_{k,h},m}\}_{m=1}^M$ と正弦波重畳モデルパラメータ (K 個の H_k と、 $(\sum_{k=1}^K H_k)$ 個の周波数 $f_{k,h}$ と、 $(\sum_{k=1}^K H_k)$ 個の正弦波強度 $A_{k,h}$) とを求め、出力する。

10

【 0 0 6 2 】

< 識別及び強調手順 >

図 7 を用いて、本実施形態の識別及び強調手順を実現する識別装置 2 0 0 について説明する。

【 0 0 6 3 】

識別装置 2 0 0 は、雑音モデルパラメータ $\theta = \{\mu_m, \mu_{1,m}, \dots, \mu_{f_{k,h},m}\}_{m=1}^M$ と正弦波重畳モデルパラメータ (K 個の H_k と、 $(\sum_{k=1}^K H_k)$ 個の周波数 $f_{k,h}$ と、 $(\sum_{k=1}^K H_k)$ 個の正弦波強度 $A_{k,h}$) とを受け取り、識別及び強調処理を行う前に予め、設定しておく。

【 0 0 6 4 】

20

識別装置 2 0 0 は、識別対象の音響信号 x_t を受け取り、音響信号 x_t に含まれる動作音を識別し (例えば、時刻 t の音響信号 x_t にどの動作音 k が含まれるかを識別する)、動作音の種類 z と、動作音を強調した音響信号 $x_t^{(E)}$ を出力する。

【 0 0 6 5 】

なお、識別対象の音響信号 x_t は、学習データと同じ状況で収集される。

【 0 0 6 6 】

図 8 は、識別装置 2 0 0 の機能ブロック図の例を示す。識別装置 2 0 0 は、周波数領域変換部 2 0 2、雑音計算部 2 0 3、フレームワイズ動作音確率計算部 2 0 4、動作音確率計算部 2 0 5、動作音推定部 2 0 6、ウィナーフィルタ設計部 2 0 7、ウィナーフィルタリング部 2 0 8 及び時間領域変換部 2 0 9 を含む。

30

【 0 0 6 7 】

< 周波数領域変換部 2 0 2 >

周波数領域変換部 2 0 2 は、識別対象の音響信号 x_t を受け取り、音響信号 x_t を短時間フーリエ変換 (STFT) などの手法で周波数領域に変換する。その後、複素スペクトルの絶対値を取り、振幅スペクトル (以下、この振幅スペクトルを単に「周波数領域に変換された音響信号」または「音響信号」ともいう) $X = (X_1, X_2, \dots, X_{f_{k,h}}, \dots, X_{f_{k,h}})^T$ を得 (S 2 0 2)、出力する。ここで、変換パラメータは、学習時と同じものを用いる。

【 0 0 6 8 】

< 雑音計算部 2 0 3 >

雑音計算部 2 0 3 は、音響信号 X と雑音モデルパラメータ θ を受け取り、雑音の推定値 $N = (N_1, N_2, \dots, N_{f_{k,h}}, \dots, N_{f_{k,h}})^T$ を式 (3) で計算し (S 2 0 3)、出力する。例えば、音響信号 X と雑音モデルパラメータ θ に含まれる雑音テンプレート $\mu_{1,m}, \dots, \mu_{f_{k,h},m}$ とを用いて、雑音テンプレート $\mu_{1,m}, \dots, \mu_{f_{k,h},m}$ の混合比 w_m を求め、雑音テンプレート $\mu_{1,m}, \dots, \mu_{f_{k,h},m}$ と混合比 w_m とから雑音の推定値 N を計算する。雑音 GMM を用いた場合、まず、混合比 w_m を以下の式で計算する。

40

【 0 0 6 9 】

【数 1 1】

$$w_{m,\tau} = \frac{\alpha_m N(X_\tau | \mu_m, \Sigma_m)}{\sum_{l=1}^M \alpha_l N(X_\tau | \mu_l, \Sigma_l)} \quad (14)$$

$$N(X_\tau | \mu_m, \Sigma_m) = \frac{1}{(2\pi)^{\Omega/2} |\Sigma_m|^{1/2}} \exp\left\{-\frac{1}{2} D_\tau^T \Sigma_m^{-1} D_\tau\right\} \quad (15)$$

$$D_\tau = X_\tau - \mu_m \quad (16)$$

$$X_\tau = (X_{1,\tau}, X_{2,\tau}, \dots, X_{\Omega,\tau})$$

$$\mu_m = (\mu_{1,m}, \mu_{2,m}, \dots, \mu_{\Omega,m})$$

【0 0 7 0】

その後、推定値 $N = (N_1, N_2, \dots, N, \dots, N)^T$ を式(3)で計算し、出力する。

【0 0 7 1】

【数 1 2】

$$N_{\omega,\tau} = \sum_{m=1}^M w_{m,\tau} \mu_{\omega,m} \quad (3)$$

【0 0 7 2】

なお、複数のマイクロホンを用いて音響信号を収録している際は、複数チャネルを用いて
ビームフォーミングやスペクトラルサブトラクションを施した音響特徴量ベクトル $X^{(add)}$
を、観測データと結合させて、式(14)を計算すればよい。例えば、式(14)～(16)におい
て、

【0 0 7 3】

【数 1 3】

$$X_\tau = (X_{1,\tau}, X_{2,\tau}, \dots, X_{\Omega,\tau}, X_\tau^{(add)})$$

$$\mu_m = (\mu_{1,m}, \mu_{2,m}, \dots, \mu_{\Omega,m}, \mu_m^{(add)})$$

とする。

【0 0 7 4】

< フレームワイズ動作音確率計算部 2 0 4 >

フレームワイズ動作音確率計算部 2 0 4 は、音響信号 X と、雑音の推定値 N と、正弦波重畳モデルパラメータ(ここでは、 K 個の H_k と、 $(\sum_{k=1}^K H_k)$ 個の周波数 $f_{k,h}$ と、 $(\sum_{k=1}^K H_k)$ 個の正弦波強度 $A_{k,h}$) とを受け取り、まず、正弦波重畳モデルにより、動作音 k の近似値を求める。ここでは、正弦波重畳モデルパラメータを用いて、式(1)により時間領域での動作音 k の近似値 $s_{t,k}$ を計算する。

【0 0 7 5】

【数 1 4】

$$s_{t,k} = \sum_{h=1}^{H_k} A_{k,h} \sin(2\pi f_{k,h} t) \quad (1)$$

【0 0 7 6】

計算した近似値 $s_{t,k}$ を窓関数で切り出し、短時間フーリエ変換(STFT)などの手法で周波数領域に変換する。その後、複素スペクトルの絶対値を取り、振幅スペクトル(以下、この振幅スペクトルを単に「周波数領域に変換された動作音の近似値」または「動作音の近似値」ともいう) $S = (S_1, S_2, \dots, S, \dots, S)^T$ を得る。ここで、変換パラメータは、学習時と同じものを用いる。なお、動作音の近似値 S は時刻ごとに計算するのではなく、正弦波重畳モデルパラメータを取得した段階で、識別及び強調処理を行う前に計算しておく。このため、上述の通り、サンプリング周波数は、動作音の特徴的な周波

10

20

30

40

50

数を記録できるように設定すればよい。動作音の近似値 S は、フレームワイズ動作音確率計算部 204 において以降の処理で用いられるとともに、ウィナーフィルタ設計部 207 に出力される。

【0077】

次に、雑音の推定値 N と、動作音 k の近似値 $S_{\omega,k}$ とを重畳し、以下の変数を計算する。

【0078】

【数15】

$$\tilde{X}_{\omega,\tau,k} = S_{\omega,k} + N_{\omega,\tau} \quad (17)$$

10

【0079】

この変数 $\tilde{X}_{\omega,\tau,k}$ は、時刻 τ で動作音 k が鳴っていた場合の、音響信号 X の推定値である。

【0080】

そして、音響信号 X と、推定値 $\tilde{X}_{\omega,\tau,k}$ とを用いて、時刻 $\tau-1$ の状態と関係なく時刻 τ だけの情報から、ある時刻 τ において音響信号 X に動作音 k が含まれる確率であるフレームワイズ動作音確率 $p(z_{\tau}|X_{\omega,\tau}, N_{\omega,\tau})$ を以下のように計算し(204)、出力する。

【0081】

【数16】

$$\zeta_k = \exp\{-(\tilde{X}_{\omega,\tau,k} - X_{\omega,\tau})^T (\tilde{X}_{\omega,\tau,k} - X_{\omega,\tau})\} \quad (18)$$

20

$$p(z_{\tau} = k | X_{\omega,\tau}, N_{\omega,\tau}) = \frac{\zeta_k}{\sum_{k=1}^K \zeta_k} \quad (19)$$

【0082】

<動作音確率計算部 205>

動作音確率計算部 205 は、フレームワイズ動作音確率 $p(z_{\tau}|X_{\omega,\tau}, N_{\omega,\tau})$ と遷移確率 $\theta_{k,k'}$ を受け取り、これらの値を用いて、時刻 $\tau-1$ の動作音 k を考慮した、ある時刻 τ において音響信号 X に動作音 k が含まれる確率である動作音確率 $p(z_{\tau}|X_{\omega,\tau}, N_{\omega,\tau}, z_{\tau-1})$ を計算し(205)、出力する。動作の状態を逐次的に推定するために、隠れマルコフモデルの前向き状態推定アルゴリズムに基づき、以下のように動作音確率 $p(z_{\tau}|X_{\omega,\tau}, N_{\omega,\tau}, z_{\tau-1})$ を求める。

30

【0083】

【数17】

$$\eta_k = p(z_{\tau} = k | X_{\omega,\tau}, N_{\omega,\tau}) \sum_{k'=1}^K \theta_{k,k'} p(z_{\tau-1} = k' | X_{\omega,\tau}, N_{\omega,\tau}, z_{\tau-2}) \quad (20)$$

$$p(z_{\tau} = k | X_{\omega,\tau}, N_{\omega,\tau}, z_{\tau-1}) = \frac{\eta_k}{\sum_{k=1}^K \eta_k} \quad (21)$$

40

【0084】

ここで遷移確率 $\theta_{k,k'}$ は、学習時に最尤推定で求めてもよいし、機械の動作法則に従って人手で決めてもよい。なお、実用上は、動作音の学習データがほとんど集まらないことが多いので、遷移確率 $\theta_{k,k'}$ は人手で与えることが多い。より詳しくは、遷移確率 $\theta_{k,k'}$ を学習するのに十分なラベル量が、雑音GMMパラメータや正弦波重畳モデルパラメータを学習するのに必要なラベル量よりも圧倒的に多いため、費用対効果や人件費を考慮すると、人手で遷移確率 $\theta_{k,k'}$ を与えた方(チューニングした方が)が効率的な場合が多い。

【0085】

<動作音推定部 206>

50

動作音推定部 206 は、動作音確率 $p(z_t | X_{\omega, \tau}, N_{\omega, \tau}, z_{t-1})$ を受け取り、時刻 t での動作音の種類 z_t を、以下の式に従って推定し (S 206)、推定結果である種類 z_t を出力する。

【0086】

【数 18】

$$z_t = \arg \max_k p(z_t = k | X_{\omega, \tau}, N_{\omega, \tau}, z_{t-1}) \quad (22)$$

【0087】

式(22)による状態推定は、最大事後確率 (MAP) 推定と呼ばれ、時刻 t までのデータを得た下で、最も事後確率 (動作音確率 $p(z_t | X_{\omega, \tau}, N_{\omega, \tau}, z_{t-1})$) の高い状態 k を時刻 t での動作音の種類 z_t と推定することを示している。推定結果 z_t は、フレームワイズ動作音確率 $p(z_t | X_{\omega, \tau}, N_{\omega, \tau})$ に基づき得られる動作音確率 $p(z_t | X_{\omega, \tau}, N_{\omega, \tau}, z_{t-1})$ を利用するため、推定結果 z_t は、間接的に、フレームワイズ動作音確率 $p(z_t | X_{\omega, \tau}, N_{\omega, \tau})$ に基づき得られる値と言える。

【0088】

なお、図 2 のように動作ログと動作音を紐づける必要がなければ、以降の処理を省略し、推定結果である種類 z_t のみを識別装置 200 の出力値として出力すればよい。

【0089】

<ウィナーフィルタ設計部 207>

ウィナーフィルタ設計部 207 は、推定結果である種類 z_t と雑音の推定値 $N_{\omega, \tau}$ (雑音計算部 203 の出力値) と動作音の近似値 S_k (フレームワイズ動作音確率計算部 204 においてフレームワイズ動作音確率の計算過程で得られる値) とを受け取り、次式により、音響信号 $X_{\omega, \tau}$ に含まれる動作音 $k=z_t$ だけを強調するためのウィナーフィルタ $G_{\omega, \tau} = (G_{1, \tau}, G_{2, \tau}, \dots, G_{z_t, \tau}, \dots, G_{K, \tau})^T$ を設計し (S 207)、出力する。

【0090】

【数 19】

$$G_{\omega, \tau} = \frac{S_{\omega, z_t}^2}{S_{\omega, z_t}^2 + N_{\omega, \tau}^2} \quad (23)$$

【0091】

<ウィナーフィルタリング部 208>

ウィナーフィルタリング部 208 は、音響信号 $X_{\omega, \tau}$ とウィナーフィルタ $G_{\omega, \tau}$ とを受け取り、音響信号 $X_{\omega, \tau}$ をウィナーフィルタ $G_{\omega, \tau}$ でフィルタリングし (S 208)、音響信号 $X_{\omega, \tau}$ に含まれる動作音 $k=z_t$ を強調し、強調した音響信号 $X^{(E)}_{\omega, \tau} = (X^{(E)}_{1, \tau}, X^{(E)}_{2, \tau}, \dots, X^{(E)}_{z_t, \tau}, \dots, X^{(E)}_{K, \tau})^T$ を出力する。例えば、次式のように、音響信号 $X_{\omega, \tau}$ にウィナーフィルタ $G_{\omega, \tau}$ を乗算することでフィルタリングする。

$$X^{(E)}_{\omega, \tau} = G_{\omega, \tau} X_{\omega, \tau}$$

【0092】

<時間領域変換部 209>

時間領域変換部 209 は、音響信号 $X^{(E)}_{\omega, \tau}$ を受け取り、周波数領域変換部 202 で行った周波数領域への変換方法に対応する時間領域への変換方法 (例えば IFFT) を用いて、音響信号 $X^{(E)}_{\omega, \tau}$ を時間領域の音響信号 $x^{(E)}_t$ に変換し (S 209)、識別装置 200 の出力値として出力する。

【0093】

上述の方法によって、識別装置 200 は、動作音の種類 z_t と強調した音響信号 $x^{(E)}_t$ を出力する。なお、強調した音響信号 $x^{(E)}_t$ を必要としない場合には、ウィナーフィルタ設計部 207、ウィナーフィルタリング部 208、時間領域変換部 209 を設けず、その処理 (S 207 ~ S 209) を省略してもよい。

【0094】

<効果>

10

20

30

40

50

以上の構成により、大騒音のノイズが存在する状況においても、音響信号に含まれる動作音を識別することができる。

【0095】

本実施形態では、機械音に特化して大騒音下で特定の動作音の種類を識別し強調するために、上述の機械の動作音の4つの特徴に着目した、識別/強調アルゴリズムとその学習法を提供する。なお、識別対象の動作法則に着目した、隠れマルコフモデルの前向き状態推定アルゴリズム自体は、古くからある技術である(非特許文献1等)。しかし、1.動作音を正弦波重畳モデルで近似計算(式(1))し、2.雑音を、雑音テンプレートの混合で表現することで(式(3)(14))、3.音響信号からの動作音の種類を、少量の学習データから効率よく、学習・識別/強調するアルゴリズム(システム全体)については従来技術ではなく、これにより、大騒音のノイズが存在する状況においても、音響信号に含まれる動作音を識別することができる。

10

【0096】

また、正弦波重畳モデルのパラメータ推定で用いたロジスティック回帰のパラメータ推定は公知の技術であるが、本実施形態では、この公知技術を利用して、回帰係数の絶対値比率(式(12))を周波数重要度とみなし、正弦波重畳モデルのパラメータ推定に用いている(S105A~S105C)。

【0097】

また、統計的パターン認識に基づく状態推定は、大量の学習データ必要である。例えば、音声認識などは、人間の発話だけで100時間以上の学習データが必要であった。本実施形態では、機械の動作音の動作特性に着目し、学習法を工夫することで、少ない学習データ量でパラメータ推定が可能である。具体的には、各パーツの動作が50回程度(機械の種類にもよるが、15分程度)のデータで学習が可能である。また、動作音を識別するだけでなく、強調する方法も同時に提供することで、遠隔地から動作音に基づく異常診断を行えるようになり、移動時間・移動経費の削減など、業務効率化にも寄与する。

20

【0098】

<変形例>

また、推定結果 z と、推定結果が得られた時刻 t を用いて、動作音のログ(図1参照)を生成してもよい。このような構成により詳細な動作ログを容易に残すことができる。工場などに設置された大型の製造機・成型機などの、詳細な動作ログを音響信号から作成することで、機械の稼働状況や使用頻度から、各パーツの消耗・摩耗度合いなどを推定でき、故障時期の予測や交換時期の提案などが可能になり、故障率を大幅に落とせる。産業、特に製造業の効率化に寄与する技術である。

30

【0099】

本実施形態では、動作音推定部206において、動作音確率 $p(z_t | X_{\omega, \tau}, N_{\omega, \tau}, z_{1:t-1})$ を用いて、時刻 t での動作音の種類 z_t を推定しているが、フレームワイズ動作音確率 $p(z_t | X_{\omega, \tau}, N_{\omega, \tau})$ を用いて、次式により、時刻 t での動作音の種類 z_t を推定してもよい。

【0100】

【数20】

$$z_t = \arg \max_k p(z_t = k | X_{\omega, \tau}, N_{\omega, \tau}) \quad (22)$$

40

【0101】

ただし、この場合、機械の動作音の4つの特徴のうちの1つである、「2. 機械の動作順には法則性がある。」を利用していないため、推定精度は落ちる。

【0102】

<その他の変形例>

本発明は上記の実施形態及び変形例に限定されるものではない。例えば、上述の各種の処理は、記載に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的にあるいは個別に実行されてもよい。その他、本発明の趣旨を

50

逸脱しない範囲で適宜変更が可能である。

【0103】

< プログラム及び記録媒体 >

また、上記の実施形態及び変形例で説明した各装置における各種の処理機能をコンピュータによって実現してもよい。その場合、各装置が有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、上記各装置における各種の処理機能がコンピュータ上で実現される。

【0104】

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等どのようなものでもよい。

10

【0105】

また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させてもよい。

【0106】

このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶部に格納する。そして、処理の実行時、このコンピュータは、自己の記憶部に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。また、このプログラムの別の実施形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよい。さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆるASP(Application Service Provider)型のサービスによって、上述の処理を実行する構成としてもよい。なお、プログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの(コンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等)を含むものとする。

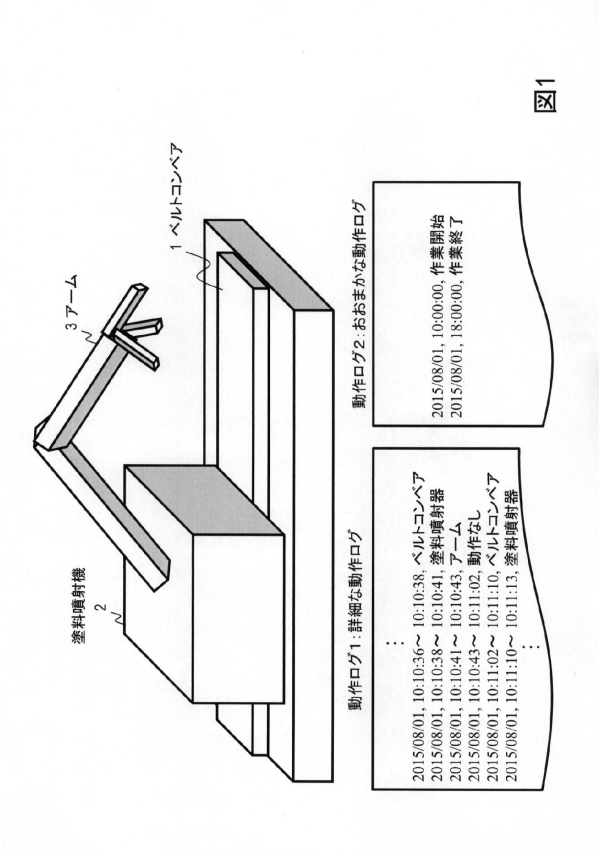
20

【0107】

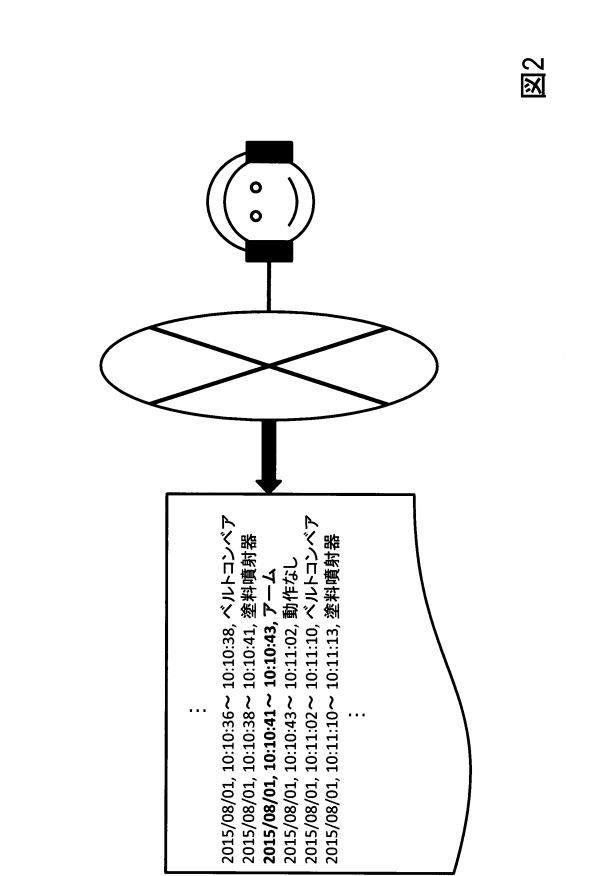
また、コンピュータ上で所定のプログラムを実行させることにより、各装置を構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

30

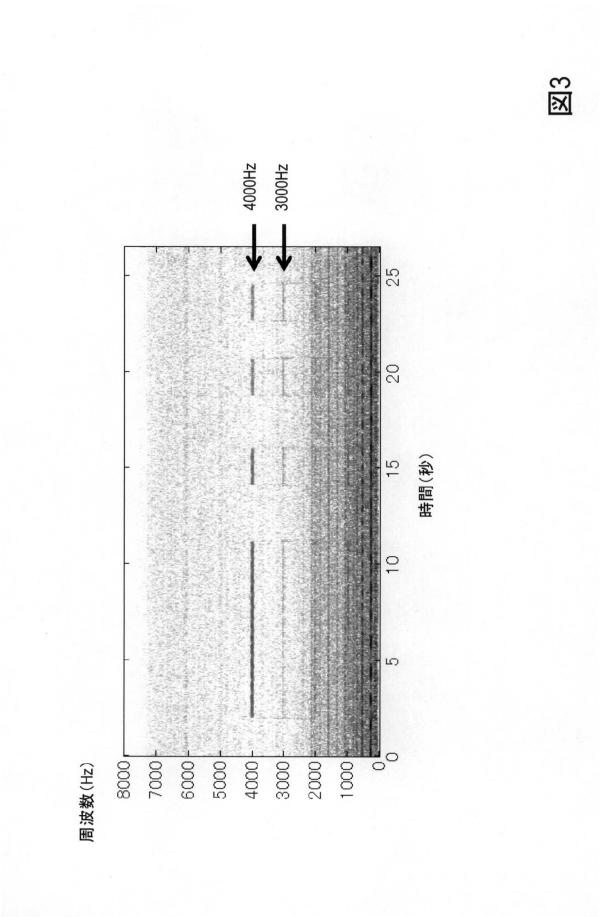
【図 1】



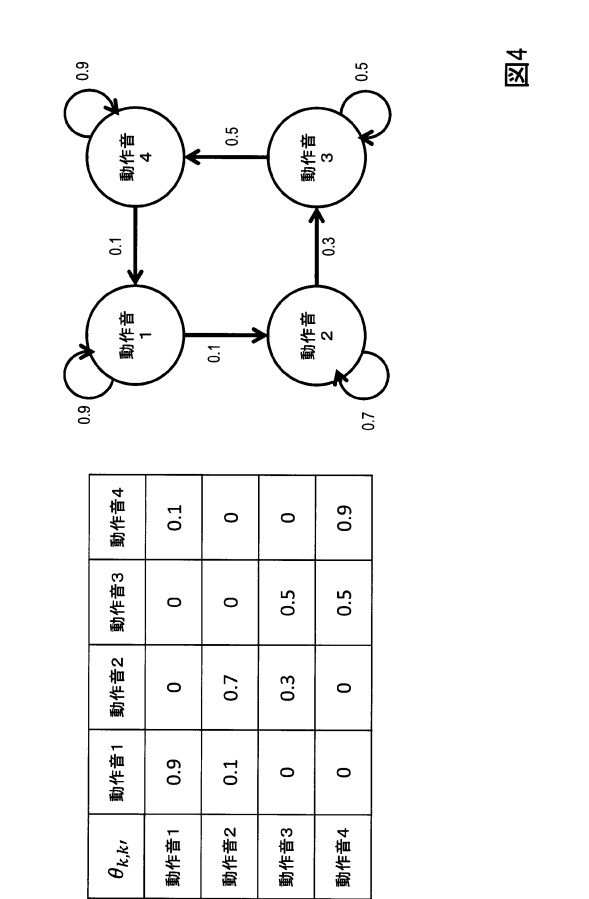
【図 2】



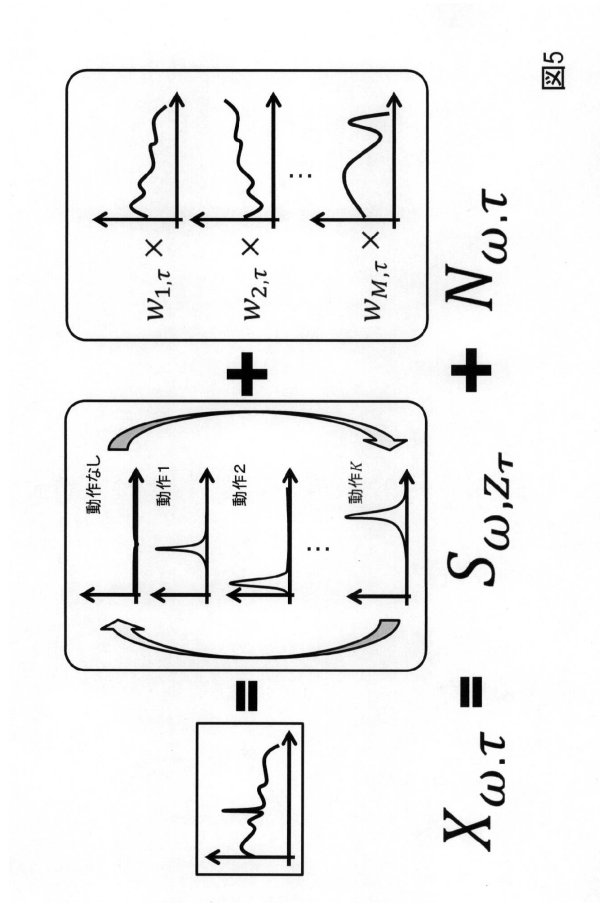
【図 3】



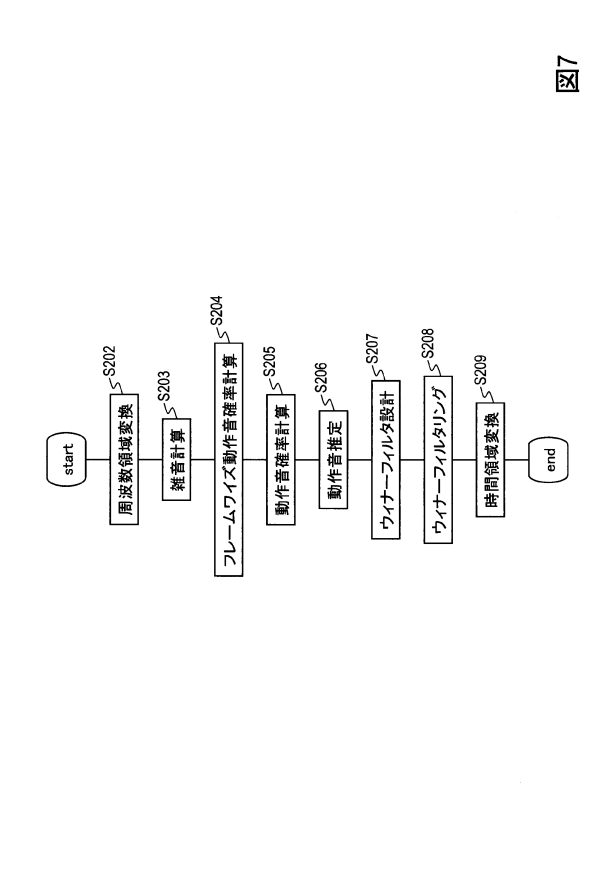
【図 4】



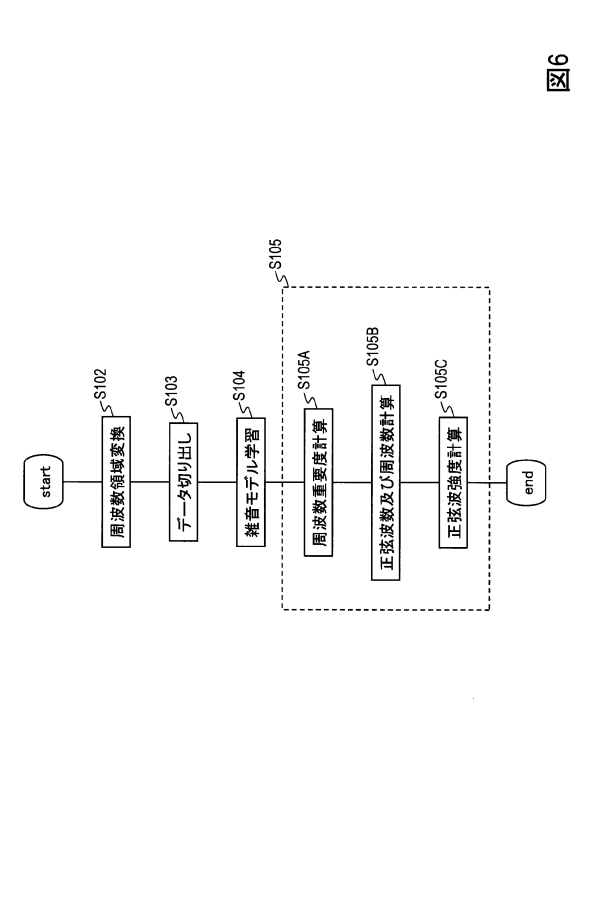
【図5】



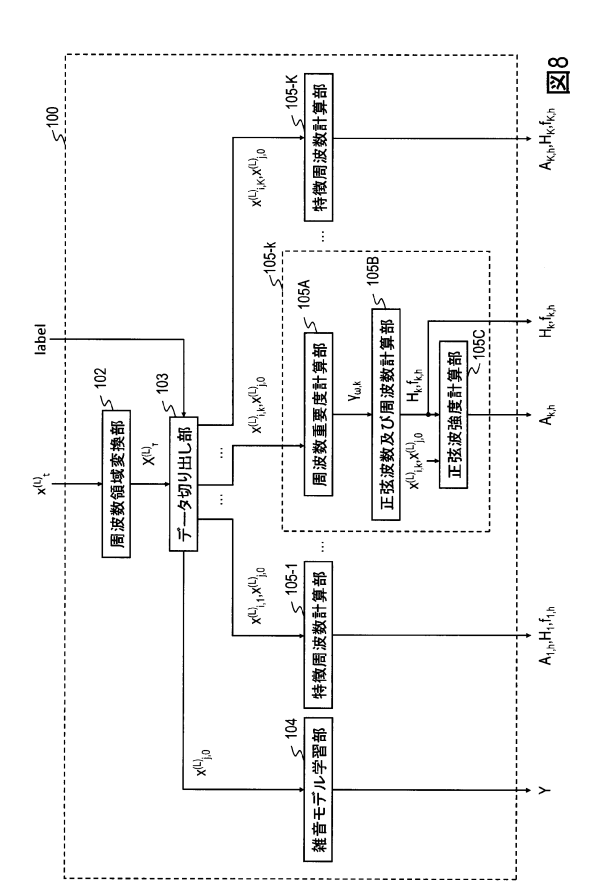
【図7】



【図6】



【図8】



【図 9】

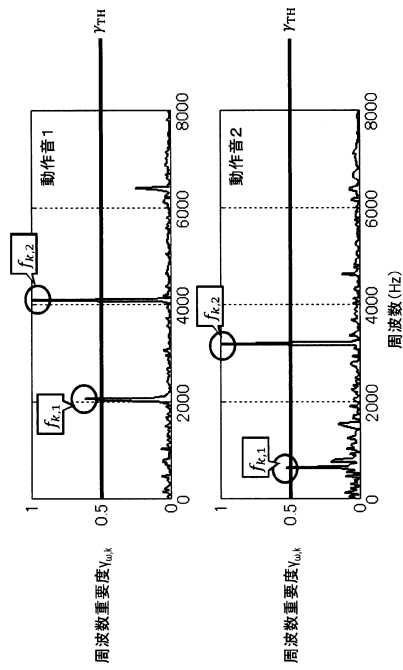


図9

【図 10】

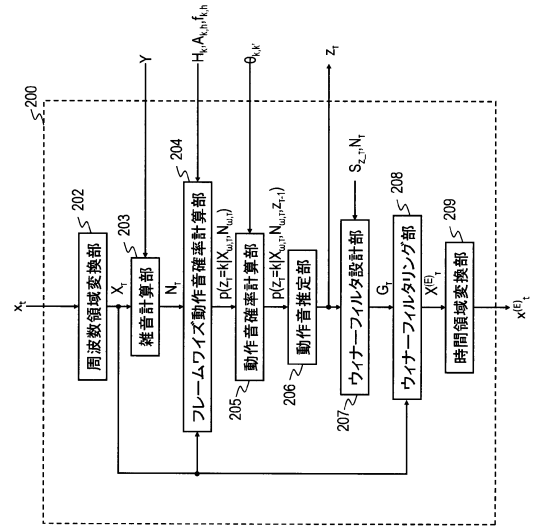


図10

フロントページの続き

(51)Int.Cl. F I
G 0 1 H 3/00 (2006.01) G 1 0 L 15/06 5 0 0 Z
G 0 1 H 3/00 A

(72)発明者 植松 尚
東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
(72)発明者 丹羽 健太
東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
(72)発明者 小林 和則
東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内

審査官 鈴木 圭一郎

(56)参考文献 特開平10-282985(JP,A)
特開2013-020252(JP,A)
国際公開第2013/032025(WO,A1)
小泉悠馬, 競技音を抽出するための特徴選択と音源強調の統合的アプローチの検討, 日本音響学会2015年秋季研究発表会講演論文集CD-ROM, 日本, 一般社団法人日本音響学会, 2015年9月18日, p579-582

(58)調査した分野(Int.Cl., DB名)
G 1 0 L 1 5 / 0 0 - 1 5 / 3 4
G 1 0 L 2 1 / 0 0 - 2 1 / 1 8