

(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2021-57711

(P2021-57711A)

(43) 公開日 令和3年4月8日(2021.4.8)

(51) Int. Cl. F 1 テーマコード (参考)
 H04R 3/00 (2006.01) H04R 3/00 320 5D220

審査請求 未請求 請求項の数 10 O L (全 20 頁)

(21) 出願番号	特願2019-177967 (P2019-177967)	(71) 出願人	000004075 ヤマハ株式会社 静岡県浜松市中区中沢町10番1号
(22) 出願日	令和1年9月27日(2019.9.27)	(74) 代理人	100125689 弁理士 大林 章
		(74) 代理人	100128598 弁理士 高田 聖一
		(74) 代理人	100121108 弁理士 高橋 太郎
		(72) 発明者	水野 賀文 静岡県浜松市中区中沢町10番1号 ヤマハ株式会社内
		(72) 発明者	近藤 多伸 静岡県浜松市中区中沢町10番1号 ヤマハ株式会社内
		Fターム(参考)	5D220 BA01 BB03

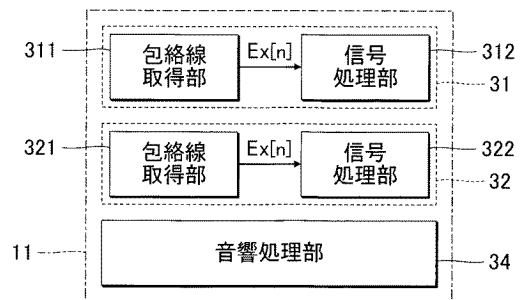
(54) 【発明の名称】 音響処理方法、音響処理装置およびプログラム

(57) 【要約】

【課題】 被り音の影響を低減して適切な音響処理を音信号に対して実行する。

【解決手段】 音響処理装置は、包絡線取得部311と信号処理部312と音響処理部34とを具備する。包絡線取得部311は、音源からの音を收音した音信号の概形を表す観測包絡線を取得する。信号処理部312は、観測包絡線における音源からの音の概形を表す出力包絡線を、観測包絡線から生成する。音響処理部34は、音信号に対して出力包絡線のレベルに応じた音響処理を実行する。

【選択図】 図3



【特許請求の範囲】**【請求項 1】**

音源からの音を収録した音信号の概形を表す観測包絡線を取得し、
前記観測包絡線における前記音源からの音の概形を表す出力包絡線を、前記観測包絡線から生成し、

前記音信号に対して前記出力包絡線のレベルに応じた音響処理を実行する
コンピュータにより実現される音響処理方法。

【請求項 2】

前記音響処理は、前記音信号において前記出力包絡線のレベルに応じた期間の音量を制御するダイナミクス制御を含む

請求項 1 の音響処理方法。

【請求項 3】

前記ダイナミクス制御は、前記音信号において前記出力包絡線のレベルが閾値を下回る期間を消音するゲート処理を含む

請求項 2 の音響処理方法。

【請求項 4】

前記ダイナミクス制御は、前記音信号において前記出力包絡線のレベルが閾値を上回る期間について所定値を上回る音量を低減するコンプレッサ処理を含む

請求項 2 または請求項 3 の音響処理方法。

【請求項 5】

前記観測包絡線の取得においては、当該観測包絡線におけるレベルを単位期間毎に順次に取得し、

前記出力包絡線の生成においては、前記単位期間毎に、前記出力包絡線の 1 個のレベルを生成する

請求項 1 から請求項 4 の何れかの音響処理方法。

【請求項 6】

第 1 音源からの音を収録した第 1 音信号の概形を表す第 1 観測包絡線と、第 2 音源からの音を収録した第 2 音信号の概形を表す第 2 観測包絡線と、を含む複数の観測包絡線を取得し、

前記第 1 観測包絡線における前記第 2 音源からの被り音の混合比と、前記第 2 観測包絡線における前記第 1 音源からの被り音の混合比と、を含む混合行列を利用して、前記複数の観測包絡線から、前記第 1 観測包絡線における前記第 1 音源からの音の概形を表す第 1 出力包絡線と、前記第 2 観測包絡線における前記第 2 音源からの音の概形を表す第 2 出力包絡線と、を含む複数の出力包絡線を生成し、

前記第 1 音信号に対して前記第 1 出力包絡線のレベルに応じた音響処理を実行し、

前記第 2 音信号に対して前記第 2 出力包絡線のレベルに応じた音響処理を実行する

コンピュータにより実現される音響処理方法。

【請求項 7】

音源からの音を収録した音信号の概形を表す観測包絡線を取得する包絡線取得部と、

前記観測包絡線における前記音源からの音の概形を表す出力包絡線を、前記観測包絡線から生成する信号処理部と、

前記音信号に対して前記出力包絡線のレベルに応じた音響処理を実行する音響処理部とを具備する音響処理装置。

【請求項 8】

第 1 音源からの音を収録した第 1 音信号の概形を表す第 1 観測包絡線と、第 2 音源からの音を収録した第 2 音信号の概形を表す第 2 観測包絡線と、を含む複数の観測包絡線を取得する包絡線取得部と、

前記第 1 観測包絡線における前記第 2 音源からの被り音の混合比と、前記第 2 観測包絡線における前記第 1 音源からの被り音の混合比と、を含む混合行列を利用して、前記複数の観測包絡線から、前記第 1 観測包絡線における前記第 1 音源からの音の概形を表す第 1

10

20

30

40

50

出力包絡線と、前記第 2 観測包絡線における前記第 2 音源からの音の概形を表す第 2 出力包絡線と、を含む複数の出力包絡線を生成する信号処理部と、

前記第 1 音信号に対して前記第 1 出力包絡線のレベルに応じた音響処理を実行し、前記第 2 音信号に対して前記第 2 出力包絡線のレベルに応じた音響処理を実行する音響処理部と

を具備する音響処理装置。

【請求項 9】

音源からの音を收音した音信号の概形を表す観測包絡線を取得する包絡線取得部、前記観測包絡線における前記音源からの音の概形を表す出力包絡線を、前記観測包絡線から生成する信号処理部、および、

10

前記音信号に対して前記出力包絡線のレベルに応じた音響処理を実行する音響処理部としてコンピュータを機能させるプログラム。

【請求項 10】

第 1 音源からの音を收音した第 1 音信号の概形を表す第 1 観測包絡線と、第 2 音源からの音を收音した第 2 音信号の概形を表す第 2 観測包絡線と、を含む複数の観測包絡線を取得する包絡線取得部、

前記第 1 観測包絡線における前記第 2 音源からの被り音の混合比と、前記第 2 観測包絡線における前記第 1 音源からの被り音の混合比と、を含む混合行列を利用して、前記複数の観測包絡線から、前記第 1 観測包絡線における前記第 1 音源からの音の概形を表す第 1 出力包絡線と、前記第 2 観測包絡線における前記第 2 音源からの音の概形を表す第 2 出力包絡線と、を含む複数の出力包絡線を生成する信号処理部、および、

20

前記第 1 音信号に対して前記第 1 出力包絡線のレベルに応じた音響処理を実行し、前記第 2 音信号に対して前記第 2 出力包絡線のレベルに応じた音響処理を実行する音響処理部としてコンピュータを機能させるプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、例えば楽器等の音源からの音を收音した音信号を処理する技術に関する。

【背景技術】

【0002】

例えば複数の楽器の演奏音を収録する場面では、楽器毎に別個の收音機器が設置される場合がある。收音機器により收音される音には、当該收音機器が設置された楽器からの音が優勢に含まれるが、当該楽器以外の楽器から到達する音（いわゆる被り音）も含まれる。特許文献 1 には、複数の音源の相互間で発生する被り音の伝達特性を推定し、收音機器により收音される音から他の音源からの被り音を除去する構成が開示されている。

30

【先行技術文献】

【特許文献】

【0003】

【特許文献 1】特開 2013 - 66079 号公報

【発明の概要】

40

【発明が解決しようとする課題】

【0004】

ところで、音信号のレベルに応じて効果付与処理等の各種の音響処理を当該音信号に対して実行する場合がある。例えば、音信号のレベルが閾値を下回る区間を消音するゲート処理、または、音信号のレベルが閾値を上回る区間を抑圧するコンプレッサ処理が想定される。音信号に被り音が含まれる場合、特定の音源からの音に対する音響処理が適切に実行されない可能性がある。以上の事情を考慮して、本開示のひとつの態様は、被り音の影響を低減して適切な音響処理を音信号に対して実行することを目的とする。

【課題を解決するための手段】

【0005】

50

本開示のひとつの態様に係る音響処理方法は、音源からの音を收音した音信号の概形を表す観測包絡線を取得し、前記観測包絡線における前記音源からの音の概形を表す出力包絡線を、前記観測包絡線から生成し、前記音信号に対して前記出力包絡線のレベルに応じた音響処理を実行する。

【0006】

本開示の他の態様に係る音響処理方法は、第1音源からの音を收音した第1音信号の概形を表す第1観測包絡線と、第2音源からの音を收音した第2音信号の概形を表す第2観測包絡線と、を含む複数の観測包絡線を取得し、前記第1観測包絡線における前記第2音源からの被り音の混合比と、前記第2観測包絡線における前記第1音源からの被り音の混合比と、を含む混合行列を利用して、前記複数の観測包絡線から、前記第1観測包絡線における前記第1音源からの音の概形を表す第1出力包絡線と、前記第2観測包絡線における前記第2音源からの音の概形を表す第2出力包絡線と、を含む複数の出力包絡線を生成し、前記第1音信号に対して前記第1出力包絡線のレベルに応じた音響処理を実行し、前記第2音信号に対して前記第2出力包絡線のレベルに応じた音響処理を実行する。

10

【0007】

本開示のひとつの態様に係る音響処理装置は、音源からの音を收音した音信号の概形を表す観測包絡線を取得する包絡線取得部と、前記観測包絡線における前記音源からの音の概形を表す出力包絡線を、前記観測包絡線から生成する信号処理部と、前記音信号に対して前記出力包絡線のレベルに応じた音響処理を実行する音響処理部とを具備する。

【0008】

本開示の他の態様に係る音響処理装置は、第1音源からの音を收音した第1音信号の概形を表す第1観測包絡線と、第2音源からの音を收音した第2音信号の概形を表す第2観測包絡線と、を含む複数の観測包絡線を取得する包絡線取得部と、前記第1観測包絡線における前記第2音源からの被り音の混合比と、前記第2観測包絡線における前記第1音源からの被り音の混合比と、を含む混合行列を利用して、前記複数の観測包絡線から、前記第1観測包絡線における前記第1音源からの音の概形を表す第1出力包絡線と、前記第2観測包絡線における前記第2音源からの音の概形を表す第2出力包絡線と、を含む複数の出力包絡線を生成する信号処理部と、前記第1音信号に対して前記第1出力包絡線のレベルに応じた音響処理を実行し、前記第2音信号に対して前記第2出力包絡線のレベルに応じた音響処理を実行する音響処理部とを具備する。

20

30

【0009】

本開示のひとつの態様に係るプログラムは、音源からの音を收音した音信号の概形を表す観測包絡線を取得する包絡線取得部、前記観測包絡線における前記音源からの音の概形を表す出力包絡線を、前記観測包絡線から生成する信号処理部、および、前記音信号に対して前記出力包絡線のレベルに応じた音響処理を実行する音響処理部としてコンピュータを機能させる。

【0010】

本開示の他の態様に係るプログラムは、第1音源からの音を收音した第1音信号の概形を表す第1観測包絡線と、第2音源からの音を收音した第2音信号の概形を表す第2観測包絡線と、を含む複数の観測包絡線を取得する包絡線取得部、前記第1観測包絡線における前記第2音源からの被り音の混合比と、前記第2観測包絡線における前記第1音源からの被り音の混合比と、を含む混合行列を利用して、前記複数の観測包絡線から、前記第1観測包絡線における前記第1音源からの音の概形を表す第1出力包絡線と、前記第2観測包絡線における前記第2音源からの音の概形を表す第2出力包絡線と、を含む複数の出力包絡線を生成する信号処理部、および、前記第1音信号に対して前記第1出力包絡線のレベルに応じた音響処理を実行し、前記第2音信号に対して前記第2出力包絡線のレベルに応じた音響処理を実行する音響処理部としてコンピュータを機能させる。

40

【図面の簡単な説明】

【0011】

【図1】音響処理システムの構成を例示するブロック図である。

50

【図 2】音響処理装置の構成を例示するブロック図である。

【図 3】制御装置の機能的な構成を例示するブロック図である。

【図 4】観測包絡線の説明図である。

【図 5】推定処理部による推定処理の説明図である。

【図 6】推定処理の具体的な手順を例示するフローチャートである。

【図 7】学習処理の具体的な手順を例示するフローチャートである。

【図 8】音響処理部が実行するゲート処理の説明図である。

【図 9】音響処理部が実行するコンプレッサ処理の説明図である。

【図 10】音響処理装置の全体的な動作の手順を例示するフローチャートである。

【図 11】第 2 実施形態における推定処理の説明図である。

【図 12】第 3 実施形態における推定処理の説明図である。

【発明を実施するための形態】

【0012】

A：第 1 実施形態

図 1 は、本開示の第 1 実施形態に係る音響処理システム 100 の構成を例示するブロック図である。音響処理システム 100 は、N 個（N は 2 以上の自然数）の音源 $S[1] \sim S[N]$ から発生する音響を收音および処理する音楽制作用の録音システムである。各音源 $S[n]$ ($n = 1 \sim N$) は、例えば演奏により発音する楽器である。例えばドラムセットを構成する複数の打楽器（例えばシンバル、キックドラム、スネアドラム、ハイハットおよびフロアタム等）の各々が音源 $S[n]$ に相当する。N 個の音源 $S[1] \sim S[N]$ は、ひとつの音響空間内に相互に近接して設置される。なお、2 個以上の楽器の組合せを音源 $S[n]$ としてもよい。

【0013】

音響処理システム 100 は、N 個の收音装置 $D[1] \sim D[N]$ と音響処理装置 10 と再生装置 20 とを具備する。各收音装置 $D[n]$ は、有線または無線により音響処理装置 10 に接続される。再生装置 20 も同様に、有線または無線により音響処理装置 10 に接続される。なお、音響処理装置 10 と再生装置 20 とを一体に構成してもよい。

【0014】

N 個の收音装置 $D[1] \sim D[N]$ の各々は、N 個の音源 $S[1] \sim S[N]$ の何れかに対応する。すなわち、N 個の收音装置 $D[1] \sim D[N]$ と N 個の音源 $S[1] \sim S[N]$ とは 1 対 1 に対応する。各收音装置 $D[n]$ は、周囲の音を收音するマイクロホンである。例えば、收音装置 $D[n]$ は、音源 $S[n]$ に指向する指向性のマイクロホンである。收音装置 $D[n]$ は、周囲の音の波形を表す音信号 $A[n]$ を生成する。N チャンネルの音信号 $A[1] \sim A[N]$ が音響処理装置 10 に並列に供給される。

【0015】

各收音装置 $D[n]$ は、音源 $S[n]$ から発生した音（以下「目的音」という）の收音を目的として音源 $S[n]$ の近傍に設置される。したがって、收音装置 $D[n]$ には音源 $S[n]$ からの目的音が優勢に到達する。ただし、各音源 $S[n]$ は相互に近接して設置されるから、各收音装置 $D[n]$ には、当該收音装置 $D[n]$ に対応する音源 $S[n]$ 以外の音源 $S[n']$ ($n' = 1 \sim N, n' \neq n$) から発生した音（以下「被り音」という）も到達する。すなわち、收音装置 $D[n]$ が生成する音信号 $A[n]$ は、音源 $S[n]$ から到達する目的音の成分を優勢に含むほか、当該音源 $S[n]$ の周囲に位置する他の音源 $S[n']$ から到達する被り音（spill, bleed, leakage）の成分も含む。なお、各音信号 $A[n]$ をアナログからデジタルに変換する A/D 変換器の図示は便宜的に省略した。

【0016】

音響処理装置 10 は、N チャンネルの音信号 $A[1] \sim A[N]$ を処理するためのコンピュータシステムである。具体的には、音響処理装置 10 は、N チャンネルの音信号 $A[1] \sim A[N]$ に対する音響処理により複数のチャンネルの音信号 B を生成する。再生装置 20 は、音信号 B が表す音を再生する。具体的には、再生装置 20 は、音信号 B をデジタルからアナログに変換する D/A 変換器と、音信号 B を増幅する増幅器と、音信号 B に応じた音響を放音す

10

20

30

40

50

る放音装置とを具備する。

【 0 0 1 7 】

図 2 は、音響処理装置 1 0 の構成を例示するブロック図である。音響処理装置 1 0 は、制御装置 1 1 と記憶装置 1 2 と表示装置 1 3 と操作装置 1 4 と通信装置 1 5 とを具備するコンピュータシステムで実現される。なお、音響処理装置 1 0 は、単体の装置で実現されるほか、相互に別体で構成された複数の装置（すなわちシステム）でも実現される。

【 0 0 1 8 】

制御装置 1 1 は、音響処理装置 1 0 の各要素を制御する単数または複数のプロセッサで構成される。例えば、制御装置 1 1 は、CPU (Central Processing Unit)、SPU (Sound Processing Unit)、DSP (Digital Signal Processor)、FPGA (Field Programmable Gate Array)、またはASIC (Application Specific Integrated Circuit) 等の 1 種類以上のプロセッサで構成される。通信装置 1 5 は、N 個の收音装置 D [1] ~ D [N] および再生装置 2 0 との間で通信する。例えば、通信装置 1 5 は、各收音装置 D [n] が接続される入力ポートと、再生装置 2 0 が接続される出力ポートとを具備する。

【 0 0 1 9 】

表示装置 1 3 は、制御装置 1 1 から指示された画像を表示する。表示装置 1 3 は、例えば液晶表示パネルまたは有機 EL 表示パネルである。操作装置 1 4 は、利用者による操作を受付ける。操作装置 1 4 は、例えば表示装置 1 3 の表示面に対する接触を検知するタッチパネル、または、利用者が操作する操作子である。

【 0 0 2 0 】

記憶装置 1 2 は、制御装置 1 1 が実行するプログラムと制御装置 1 1 が使用するデータとを記憶する単数または複数のメモリである。具体的には、記憶装置 1 2 は、推定処理プログラム P 1 と学習処理プログラム P 2 と音響処理プログラム P 3 とを記憶する。記憶装置 1 2 は、例えば磁気記録媒体または半導体記録媒体等の公知の記録媒体で構成される。なお、複数種の記録媒体の組合せにより記憶装置 1 2 を構成してもよい。また、音響処理装置 1 0 に着脱可能な可搬型の記録媒体、または、音響処理装置 1 0 が通信可能な外部記録媒体（例えばオンラインストレージ）を、記憶装置 1 2 として利用してもよい。

【 0 0 2 1 】

図 3 は、音響処理装置 1 0 の機能的な構成を例示するブロック図である。制御装置 1 1 は、記憶装置 1 2 に記憶されたプログラムを実行することで複数の機能（推定処理部 3 1、学習処理部 3 2、音響処理部 3 4）を実現する。制御装置 1 1 が実現する各機能について以下に詳述する。

【 0 0 2 2 】

[1] 推定処理部 3 1

制御装置 1 1 は、推定処理プログラム P 1 を実行することで推定処理部 3 1 として機能する。推定処理部 3 1 は、N チャネルの音信号 A [1] ~ A [N] を解析する。具体的には、推定処理部 3 1 は、包絡線取得部 3 1 1 と信号処理部 3 1 2 とを具備する。

【 0 0 2 3 】

包絡線取得部 3 1 1 は、N チャネルの音信号 A [1] ~ A [N] の各々について観測包絡線 $E_x[n]$ ($E_x[1] \sim E_x[N]$) を生成する。各音信号 A [n] の観測包絡線 $E_x[n]$ は、時間軸上における当該音信号 A [n] の波形の概形（エンベロープ）を表す時間領域の信号である。

【 0 0 2 4 】

図 4 は、観測包絡線 $E_x[n]$ の説明図である。時間軸上の所定長の期間（以下「解析期間」という） T_a 毎に N チャネルの観測包絡線 $E_x[1] \sim E_x[N]$ が生成される。各解析期間 T_a は、時間軸上の M 個（M は 2 以上の自然数）の単位期間 $T_u[1] \sim T_u[M]$ で構成される。各単位期間 $T_u[m]$ ($m = 1 \sim M$) は、音信号 A [n] を構成する信号値（サンプル）の U 個分に相当する時間長の期間である。包絡線取得部 3 1 1 は、単位期間 $T_u[m]$ 毎に観測包絡線 $E_x[n]$ のレベル $x[n,m]$ を音信号 A [n] から算定する。1 個の解析期間 T_a における第 n チャネルの観測包絡線 $E_x[n]$ は、当該解析期間 T_a 内の M 個のレベル $x[n,1] \sim x[n,M]$ の時系列で表現される。観測包絡線 $E_x[n]$ における任意の 1 個のレベル $x[n,m]$ は、例えば以

10

20

30

40

50

下の数式(1)で表現される。

【数 1】

$$x[n, m] = \sqrt{\frac{1}{U} \sum_{u=0}^U (a[n, u])^2} \quad (1)$$

【 0 0 2 5 】

数式(1)の記号 $a[n, u]$ は、単位期間 $T_u[m]$ 内における第 n チャンネルの音信号 $A[n]$ を構成する U 個の信号値 $a[n, 1] \sim a[n, U]$ のうち第 u 番目 ($u = 1 \sim U$) の 1 個の信号値を意味する。数式(1)から理解される通り、観測包絡線 $E_x[n]$ の各レベル $x[n, m]$ は、音信号 $A[n]$ の 2 乗平均平方根 (RMS : Root Mean Square) に相当する非負の実効値である。以上の説明から理解される通り、包絡線取得部 3 1 1 は、 N チャンネルの各々について、単位期間 $T_u[m]$ 毎にレベル $x[n, m]$ を生成し、当該レベル $x[n, m]$ の M 個分の時系列 (レベル $x[n, 1] \sim x[n, M]$) を観測包絡線 $E_x[n]$ とする。すなわち、各チャンネルの観測包絡線 $E_x[n]$ は、 M 個のレベル $x[n, 1] \sim x[n, M]$ を要素とする M 次元ベクトルで表現される。

20

【 0 0 2 6 】

図 5 は、推定処理部 3 1 の動作の説明図である。以上に説明した観測包絡線 $E_x[n]$ が、 N チャンネルの音信号 $A[1] \sim A[N]$ の各々について生成される。したがって、 N 個の観測包絡線 $E_x[1] \sim E_x[N]$ を縦方向に配列した N 行 M 列の非負行列 (以下「観測行列」という) X が解析期間 T_a 毎に生成される。観測行列 X における第 n 行第 m 列の要素は、第 n チャンネルの観測包絡線 $E_x[n]$ における第 m 番目のレベル $x[n, m]$ である。なお、以下の各図面においては、音信号 $A[n]$ のチャンネルの総数 N が 3 である場合が例示されている。

【 0 0 2 7 】

図 3 の信号処理部 3 1 2 は、 N チャンネルの観測包絡線 $E_x[1] \sim E_x[N]$ から N チャンネルの出力包絡線 $E_y[1] \sim E_y[N]$ を生成する。図 5 に例示される通り、観測包絡線 $E_x[n]$ に対応する出力包絡線 $E_y[n]$ は、当該観測包絡線 $E_x[n]$ における音源 $S[n]$ からの目的音を強調 (理想的には抽出) した時間領域信号である。すなわち、出力包絡線 $E_y[n]$ においては、音源 $S[n]$ 以外の各音源 $S[n']$ からの被り音のレベルが低減 (理想的には除去) される。以上の説明から理解される通り、出力包絡線 $E_y[n]$ は、音源 $S[n]$ から発生した目的音のレベルの時間的な変化を表す。したがって、第 1 実施形態によれば、各音源 $S[n]$ からの目的音のレベルの時間的な変化を利用者が正確に把握できるという利点がある。

30

【 0 0 2 8 】

信号処理部 3 1 2 は、各解析期間 T_a における N チャンネルの観測包絡線 $E_x[1] \sim E_x[N]$ から当該解析期間 T_a における N チャンネルの出力包絡線 $E_y[1] \sim E_y[N]$ を生成する。すなわち、 N チャンネルの出力包絡線 $E_y[1] \sim E_y[N]$ は解析期間 T_a 毎に生成される。1 個の解析期間 T_a における第 n チャンネルの出力包絡線 $E_y[n]$ は、当該解析期間 T_a 内の相異なる単位期間 $T_u[m]$ に対応する M 個のレベル $y[n, 1] \sim y[n, M]$ の時系列で表現される。すなわち、各出力包絡線 $E_y[n]$ は、 M 個のレベル $y[n, 1] \sim y[n, M]$ を要素とする M 次元ベクトルで表現される。信号処理部 3 1 2 が生成する N チャンネルの出力包絡線 $E_y[1] \sim E_y[N]$ は、 N 行 M 列の非負行列 (以下「係数行列」という) Y を構成する。係数行列 Y (アクティブーション行列) における第 n 行第 m 列の要素は、出力包絡線 $E_y[n]$ における第 m 番目のレベル $y[n, m]$ である。

40

【 0 0 2 9 】

50

1 個の解析期間 T_a において、信号処理部 3 1 2 は、既知の混合行列 Q (基底行列) を利用した非負値行列因子分解 (NMF: Non-negative Matrix Factorization) により、観測行列 X から係数行列 Y を生成する。混合行列 Q は、複数の混合比 $q [n_1, n_2]$ ($n_1 = 1 \sim N$, $n_2 = 1 \sim N$) を配列した N 行 N 列の正方行列である。混合行列 Q は、機械学習により事前に生成されたうえで記憶装置 1 2 に記憶される。混合行列 Q の対角要素である各混合比 $q [n, n]$ ($n_1 = n_2 = n$) は、基準値 (具体的には 1) に設定される。

【 0 0 3 0 】

各観測包絡線 $E x[n]$ は以下の数式 (2) で表現される。

$$E x[n] = q [n, 1] E y[1] + q [n, 2] E y[2] + \dots + q [n, N] E y[N] \quad (2)$$

10

すなわち、観測包絡線 $E x[n]$ に対応する N 個の混合比 $q [n, 1] \sim q [n, N]$ は、当該観測包絡線 $E x[n]$ を N チャンネルの出力包絡線 $E y[1] \sim E y[N]$ の加重和で近似的に表現した場合における各出力包絡線 $E y[n]$ の加重値に相当する。

【 0 0 3 1 】

すなわち、混合行列 Q の各混合比 $q [n_1, n_2]$ は、観測包絡線 $E x[n_1]$ において音源 $S [n_2]$ からの被り音が混合された度合を表す指標である。混合比 $q [n_1, n_2]$ は、收音装置 $D [n_1]$ に対して音源 $S [n_2]$ から到達する被り音の到達率 (ないし減衰率) に関する指標とも換言される。具体的には、混合比 $q [n_1, n_2]$ は、收音装置 $D [n_1]$ が音源 $S [n_1]$ から收音した目的音の音量を 1 (基準値) とした場合に、当該收音装置 $D [n_1]$ が他の音源 $S [n_2]$ から收音した被り音の音量の比率である。したがって、混合比 $q [n_1, n_2]$ と出力包絡線 $E y[n_2]$ のレベル $y [n_2, m]$ との積 $q [n_1, n_2] y [n_2, m]$ は、音源 $S [n_2]$ から收音装置 $D [n_1]$ に到達する被り音の音量に相当する。

20

【 0 0 3 2 】

例えば、図 5 の混合行列 Q における混合比 $q [1, 2]$ は 0.1 であるから、観測包絡線 $E x[1]$ においては、音源 $S [1]$ からの目的音に対して音源 $S [2]$ からの被り音が 0.1 の割合で混合されていることを意味する。また、混合比 $q [1, 3]$ は 0.2 であるから、観測包絡線 $E x[1]$ においては、音源 $S [1]$ からの目的音に対して音源 $S [3]$ からの被り音が 0.2 の割合で混合されていることを意味する。同様に、例えば混合比 $q [3, 1]$ は 0.2 であるから、観測包絡線 $E x[3]$ においては、音源 $S [3]$ からの目的音に対して音源 $S [1]$ からの被り音が 0.2 の割合で混合されていることを意味する。すなわち、混合比 $q [n_1, n_2]$ が大きいほど、音源 $S [n_2]$ から收音装置 $D [n_1]$ に到達する被り音が大きいことを意味する。

30

【 0 0 3 3 】

第 1 実施形態の信号処理部 3 1 2 は、混合行列 Q と係数行列 Y との積 $Q Y$ が観測行列 X に近づくように係数行列 Y を反復的に更新する。例えば、信号処理部 3 1 2 は、観測行列 X と積 $Q Y$ との距離を表す評価関数 $F (X|QY)$ が最小となるように係数行列 Y を算定する。評価関数 $F (X|QY)$ は、例えばユークリッド距離、KL (Kullback-Leibler) ダイバージェンス、板倉齋藤距離、または ダイバージェンス等の任意の距離規範である。

【 0 0 3 4 】

N 個の音源 $S [1] \sim S [N]$ のうち任意の 2 個の音源 $S [k_1]$ および音源 $S [k_2]$ に着目する ($k_1 = 1 \sim N$, $k_2 = 1 \sim N$, $k_1 \neq k_2$)。 N チャンネルの観測包絡線 $E x[1] \sim E x[N]$ は、観測包絡線 $E x[k_1]$ と観測包絡線 $E x[k_2]$ とを含む。観測包絡線 $E x[k_1]$ は、音源 $S [k_1]$ からの目的音を收音した音信号 $A [k_1]$ の概形である。観測包絡線 $E x[k_1]$ は「第 1 観測包絡線」の一例であり、音源 $S [k_1]$ は「第 1 音源」の一例であり、音信号 $A [k_1]$ は「第 1 音信号」の一例である。他方、観測包絡線 $E x[k_2]$ は、音源 $S [k_2]$ からの目的音を收音した音信号 $A [k_2]$ の概形である。観測包絡線 $E x[k_2]$ は「第 2 観測包絡線」の一例であり、音源 $S [k_2]$ は「第 2 音源」の一例であり、音信号 $A [k_2]$ は「第 2 音信号」の一例である。

40

【 0 0 3 5 】

混合行列 Q は、混合比 $q [k_1, k_2]$ と混合比 $q [k_2, k_1]$ とを含む。混合比 $q [k_1, k_2]$ は、観測包絡線 $E x[k_1]$ における音源 $S [k_2]$ からの被り音の混合比であり、混合比 $q [k_2, k_1]$ は、

50

観測包絡線 $E_x[k2]$ における音源 $S[k1]$ からの被り音の混合比である。Nチャンネルの出力包絡線 $E_y[1] \sim E_y[N]$ は、出力包絡線 $E_y[k1]$ と出力包絡線 $E_y[k2]$ とを含む。出力包絡線 $E_y[k1]$ は、「第1出力包絡線」の一例であり、観測包絡線 $E_x[k1]$ における音源 $S[k1]$ からの目的音の概形を表す信号を意味する。他方、出力包絡線 $E_y[k2]$ は、「第2出力包絡線」の一例であり、観測包絡線 $E_x[k2]$ における音源 $S[k2]$ からの目的音の概形を表す信号を意味する。

【0036】

図6は、制御装置11が係数行列Yを生成する処理（以下「推定処理」という） S_a の具体的な手順を例示するフローチャートである。推定処理 S_a は、操作装置14に対する利用者からの指示を契機として開始され、N個の音源 $S[1] \sim S[N]$ による発音に並行して実行される。例えば、音響処理システム100の利用者は、音源 $S[n]$ としての楽器を演奏する。複数の利用者による演奏に並行して推定処理 S_a が実行される。推定処理 S_a は、解析期間 T_a 毎に実行される。

10

【0037】

推定処理 S_a を開始すると、包絡線取得部311は、Nチャンネルの音信号 $A[1] \sim A[N]$ からNチャンネルの観測包絡線 $E_x[1] \sim E_x[N]$ （すなわち観測行列X）を生成する（ S_{a1} ）。具体的には、包絡線取得部311は、前掲の数式(1)の演算により各観測包絡線 $E_x[n]$ におけるレベル $x[n,m]$ を算定する。

【0038】

信号処理部312は、係数行列Yを初期化する（ S_{a2} ）。例えば、信号処理部312は、直前の解析期間 T_a における観測行列Xを現在の解析期間 T_a における係数行列Yの初期値として設定する。なお、係数行列Yの初期化の方法は以上の例示に限定されない。例えば、信号処理部312は、現在の解析期間 T_a について生成した観測行列Xを、現在の解析期間 T_a における係数行列Yの初期値として設定してもよい。また、信号処理部312は、直前の解析期間 T_a における観測行列Xまたは係数行列Yの各要素に乱数を加算した行列を、現在の解析期間 T_a における係数行列Yの初期値として設定してもよい。

20

【0039】

信号処理部312は、既知の混合行列Qと現在の係数行列Yとの積 QY と、現在の解析期間 T_a の観測行列Xとの距離を表す評価関数 $F(X|QY)$ を算定する（ S_{a3} ）。信号処理部312は、所定の終了条件が成立したか否かを判定する（ S_{a4} ）。終了条件は、例えば評価関数 $F(X|QY)$ が所定の閾値を下回ること、または、係数行列Yを更新した回数が所定の閾値に到達したことである。

30

【0040】

終了条件が成立していない場合（ $S_{a4} : NO$ ）、信号処理部312は、評価関数 $F(X|QY)$ が減少するように係数行列Yを更新する（ S_{a5} ）。終了条件が成立するまで（ $S_{a4} : YES$ ）、評価関数 $F(X|QY)$ の算定（ S_{a3} ）と係数行列Yの更新（ S_{a5} ）とが反復される。係数行列Yは、終了条件が成立した段階（ $S_{a4} : YES$ ）の数値で確定される。

【0041】

Nチャンネルの観測包絡線 $E_x[1] \sim E_x[N]$ の生成（ S_{a1} ）と複数の出力包絡線 $E_y[1] \sim E_y[N]$ の生成（ $S_{a2} \sim S_{a5}$ ）とは、N個の音源 $S[1] \sim S[N]$ からの収音に並行して解析期間 T_a 毎に実行される。

40

【0042】

以上の説明から理解される通り、第1実施形態においては、各音信号 $A[n]$ の概形を表す観測包絡線 $E_x[n]$ に対する処理で出力包絡線 $E_y[n]$ が生成されるから、各音信号 $A[n]$ を解析する構成と比較して、音源 $S[n]$ 毎の目的音（出力包絡線 $E_y[n]$ ）のレベルを推定する推定処理 S_a の負荷を軽減することが可能である。

【0043】

[2] 学習処理部32

図3に例示される通り、制御装置11は、学習処理プログラムP2を実行することで学習処理部32として機能する。学習処理部32は、推定処理 S_a に使用される混合行列Q

50

を生成する。混合行列 Q は、推定処理 S_a の実行前の任意の時点において生成（ないし訓練）される。具体的には、初期的な混合行列 Q が新規に生成されるほか、生成済の混合行列 Q が訓練（再訓練）される。学習処理部 3 2 は、包絡線取得部 3 2 1 と信号処理部 3 2 2 とを具備する。

【 0 0 4 4 】

包絡線取得部 3 2 1 は、訓練用に準備された N チャンネルの音信号 $A[1] \sim A[N]$ の各々について観測包絡線 $E_x[n]$ ($E_x[1] \sim E_x[N]$) を生成する。訓練用の音信号 $A[n]$ の時間長は、 M 個の単位期間 $T_u[1] \sim T_u[M]$ の時間長（すなわち解析期間 T_a の時間長）に相当する。すなわち、 N チャンネルの観測包絡線 $E_x[1] \sim E_x[N]$ を含む N 行 M 列の観測行列 X が生成される。包絡線取得部 3 2 1 による動作は包絡線取得部 3 1 1 による動作と同様である。

10

【 0 0 4 5 】

信号処理部 3 2 2 は、解析期間 T_a における N チャンネルの観測包絡線 $E_x[1] \sim E_x[N]$ から混合行列 Q と N チャンネルの出力包絡線 $E_y[1] \sim E_y[N]$ とを生成する。すなわち、観測行列 X から混合行列 Q と係数行列 Y とが生成される。 N チャンネルの観測包絡線 $E_x[1] \sim E_x[N]$ を利用して混合行列 Q を更新する処理を 1 エポックとして、所定の終了条件が成立するまで当該エポックを複数回にわたり反復することで、推定処理 S_a に使用される混合行列 Q が確定される。終了条件は、前述の推定処理 S_a の終了条件とは相違してよい。信号処理部 3 2 2 が生成した混合行列 Q は記憶装置 1 2 に記憶される。

【 0 0 4 6 】

信号処理部 3 2 2 は、非負値行列因子分解により観測行列 X から混合行列 Q と係数行列 Y とを生成する。すなわち、信号処理部 3 2 2 は、エポック毎に、混合行列 Q と係数行列 Y との積 QY が観測行列 X に近付くように係数行列 Y を更新する。信号処理部 3 2 2 は、係数行列 Y の更新を複数のエポックにわたり反復し、観測行列 X と積 QY との距離を表す評価関数 $F(X|QY)$ が徐々に減少するように係数行列 Y を算定する。

20

【 0 0 4 7 】

図 7 は、制御装置 1 1 が混合行列 Q を生成（すなわち訓練）する処理（以下「学習処理」という） S_b の具体的な手順を例示するフローチャートである。学習処理 S_b は、操作装置 1 4 に対する利用者からの指示を契機として開始される。例えば、推定処理 S_a が実行される正式な演奏の開始前（例えばリハーサル）に音源 $S[n]$ としての楽器を演奏者が演奏する。音響処理システム 1 0 0 の利用者は、演奏音を收音することで訓練用の N チャンネルの音信号 $A[1] \sim A[N]$ を取得する。

30

【 0 0 4 8 】

なお、音源 $S[n]$ の位置、收音装置 $D[n]$ の位置、または音源 $S[n]$ と收音装置 $D[n]$ の相対的な位置関係等の收音条件が変化すると、各收音装置 $D[n]$ に他の音源 $S[n]$ から到達する被り音の程度も変化する。したがって、收音条件が変更されるたびに、利用者からの指示に応じて学習処理 S_b が実行されることで混合行列 Q が更新される。

【 0 0 4 9 】

なお、各楽器の演奏に並行した推定処理 S_a の実行中に收音条件の変化または推定結果の誤差に気付いた場合、利用者は、音響処理システム 1 0 0 に対して混合行列 Q の再訓練を指示する。音響処理システム 1 0 0 は、利用者からの指示に応じて、現時点の混合行列 Q を利用した推定処理 S_a を実行しながら、現在の演奏を収録することで訓練用の音信号 $A[n]$ を取得する。学習処理部 3 2 は、訓練用の音信号 $A[n]$ を利用した学習処理 S_b により混合行列 Q を再訓練する。推定処理部 3 1 は、再訓練後の混合行列 Q を、以後の演奏に対する推定処理 S_a に利用する。すなわち、演奏の途中で混合行列 Q が更新される。

40

【 0 0 5 0 】

学習処理 S_b を開始すると、包絡線取得部 3 2 1 は、訓練用の N チャンネルの音信号 $A[1] \sim A[N]$ から N チャンネルの観測包絡線 $E_x[1] \sim E_x[N]$ を生成する (S_{b1})。具体的には、包絡線取得部 3 2 1 は、前掲の数式(1)の演算により各観測包絡線 $E_x[n]$ におけるレベル $x[n,m]$ を算定する。

50

【 0 0 5 1 】

信号処理部 3 2 2 は、混合行列 Q および係数行列 Y を初期化する (S b2)。例えば、信号処理部 3 2 2 は、対角要素 ($q [n,n]$) を 1 に設定し、対角要素以外の各要素を乱数に設定する。なお、混合行列 Q の初期化の方法は以上の例示に限定されない。例えば、過去の学習処理 S_b で生成された混合行列 Q を、今回の学習処理 S_b における初期的な混合行列 Q として再訓練してもよい。また、信号処理部 3 2 2 は、例えば観測行列 X を係数行列 Y の初期値として設定する。なお、係数行列 Y の初期化の方法は以上の例示に限定されない。例えば、過去の学習処理 S_b において今回と同じ音信号 $A [n]$ が利用される場合、信号処理部 3 2 2 は、その学習処理 S_b により生成した係数行列 Y を、今回の学習処理 S_b における係数行列 Y の初期値としてもよい。また、信号処理部 3 2 2 は、以上に例示した観測行列 X または係数行列 Y の各要素に乱数を加算した行列を、現在の解析期間 T_a における係数行列 Y の初期値として設定してもよい。

10

【 0 0 5 2 】

信号処理部 3 2 2 は、混合行列 Q と係数行列 Y との積 QY と、現在の解析期間 T_a の観測行列 X との距離を表す評価関数 $F (X|QY)$ を算定する (S b3)。信号処理部 3 2 2 は、所定の終了条件が成立したか否かを判定する (S b4)。学習処理 S_b の終了条件は、例えば、評価関数 $F (X|QY)$ が所定の閾値を下回ること、または、係数行列 Y を更新した回数が所定の閾値に到達したことである。

【 0 0 5 3 】

終了条件が成立していない場合 (S b4 : NO)、信号処理部 3 2 2 は、評価関数 $F (X|QY)$ が減少するように混合行列 Q および係数行列 Y を更新する (S b5)。混合行列 Q および係数行列 Y の更新 (S b5) と評価関数 $F (X|QY)$ の算定 (S b3) とを 1 エポックとして、終了条件が成立するまで (S b4 : YES)、当該エポックが反復される。混合行列 Q は、終了条件が成立した段階 (S b4 : YES) の数値で確定される。

20

【 0 0 5 4 】

以上の説明から理解される通り、第 1 実施形態においては、各観測包絡線 $E_x[n]$ における他の音源 $S [n']$ からの被り音の混合比 $q [n,n']$ を含む混合行列 Q が、訓練用の N チャンネルの観測包絡線 $E_x[1] \sim E_x[N]$ から事前に生成される。混合行列 Q は、各音源 $S [n]$ に対応する音信号 $A [n]$ に他の音源 $S [n']$ からの被り音が含まれる度合 (音被りの度合) を表す。ここでは、音信号 $A [n]$ の概形を表す観測包絡線 $E_x[n]$ が処理されるから、音信号 $A [n]$ を処理する構成と比較して、混合行列 Q を生成する学習処理 S_b の負荷を軽減することが可能である。

30

【 0 0 5 5 】

なお、推定処理 S_a と学習処理 S_b との相違点は、推定処理 S_a では混合行列 Q が固定されるのに対して学習処理 S_b では混合行列 Q が係数行列 Y とともに更新される点である。すなわち、混合行列 Q の更新の有無以外の点については推定処理 S_a と学習処理 S_b とは共通する。したがって、学習処理部 3 2 の機能を推定処理部 3 1 として利用してもよい。すなわち、学習処理部 3 2 による学習処理 S_b において混合行列 Q を固定し、かつ、単位期間 $T_u[m]$ の M 個分にわたる観測包絡線 $E_x[n]$ を纏めて処理することで推定処理 S_a が実現される。前述の例示では、推定処理部 3 1 と学習処理部 3 2 とを別個の要素として説明したが、推定処理部 3 1 と学習処理部 3 2 とを 1 個の要素として音響処理装置 1 0 に搭載してもよい。

40

【 0 0 5 6 】

[3] 音響処理部 3 4

図 3 に例示される通り、制御装置 1 1 は、音響処理プログラム P_4 を実行することで音響処理部 3 4 として機能する。音響処理部 3 4 は、 N チャンネルの音信号 $A [1] \sim A [N]$ の各々に対して音響処理を実行することで音信号 $B [n]$ ($B [1] \sim B [N]$) を生成する。具体的には、音響処理部 3 4 は、推定処理部 3 1 が生成した出力包絡線 $E_y[n]$ のレベル $y [n,m]$ に応じた音響処理を、音信号 $A [n]$ に対して実行する。出力包絡線 $E_y[n]$ は、前述の通り、音信号 $A [n]$ における音源 $S [n]$ からの目的音の概形を表す包絡線である。具体的には、

50

音響処理部 3 4 は、出力包絡線 $E y[n]$ のレベル $y [n,m]$ に応じて音信号 $A [n]$ に設定された複数の処理期間 H の各々について音響処理を実行する。

【 0 0 5 7 】

例えば、 N 個の音源 $S [1] \sim S [N]$ のうち任意の 2 個の音源 $S [k1]$ および音源 $S [k2]$ に着目する。音響処理部 3 4 は、音信号 $A [k1]$ に対して出力包絡線 $E y[k1]$ のレベル $y [k1,m]$ に応じた音響処理を実行し、音信号 $A [k2]$ に対して出力包絡線 $E y[k2]$ のレベル $y [k2,m]$ に応じた音響処理を実行する。

【 0 0 5 8 】

音響処理部 3 4 は、 N チャネルの音信号 $B [1] \sim B [N]$ から音信号 B を生成する。具体的には、音響処理部 3 4 は、 N チャネルの音信号 $B [1] \sim B [N]$ の各々に係数を乗算したうえで N チャネル分を混合することで音信号 B を生成する。各音信号 $B [n]$ の係数（すなわち加重値）は、例えば操作装置 1 4 に対する利用者からの指示に応じて設定される。

10

【 0 0 5 9 】

音響処理部 3 4 は、音信号 $A [n]$ の音量を制御するダイナミクス制御を含む音響処理を実行する。ダイナミクス制御は、例えばゲート処理およびコンプレッサ処理等のエフェクタ処理を含む。利用者は、操作装置 1 4 を適宜に操作することで音響処理の種類を選択することが可能である。音響処理の種類は、 N チャネルの音信号 $A [1] \sim A [N]$ の各々について個別に選択されてもよいし、 N チャネルの音信号 $A [1] \sim A [N]$ について一括的に選択されてもよい。

【 0 0 6 0 】

20

[3 A] ゲート処理

図 8 は、音響処理のうちゲート処理の説明図である。利用者がゲート処理を選択した場合、音響処理部 3 4 は、出力包絡線 $E y[n]$ のレベル $y [n,m]$ が所定の閾値 y_{TH1} を下回る可変長の期間を処理期間 H として設定する。閾値 y_{TH1} は、例えば操作装置 1 4 に対する利用者からの指示に応じた可変値である。ただし、閾値 y_{TH1} を所定値に固定してもよい。

【 0 0 6 1 】

音響処理部 3 4 は、音信号 $A [n]$ における各処理期間 H の音量を低減する。具体的には、音響処理部 3 4 は、処理期間 H 内における音信号 $A [n]$ のレベルをゼロに設定（すなわち消音）する。以上に例示したゲート処理によれば、音信号 $A [n]$ における他の音源 $S [n']$ からの被り音を有効に低減できる。

30

【 0 0 6 2 】

[3 B] コンプレッサ処理

図 9 は、音響処理のうちコンプレッサ処理の説明図である。利用者がコンプレッサ処理を選択した場合、音響処理部 3 4 は、第 n チャネルの出力包絡線 $E y[n]$ のレベル $y [n,m]$ が所定の閾値 y_{TH2} を上回る処理期間 H において、当該第 n チャネルの音信号 $A [n]$ のゲインを低下させる。閾値 y_{TH2} は、例えば操作装置 1 4 に対する利用者からの指示に応じた可変値である。ただし、閾値 y_{TH2} を所定値に固定してもよい。

【 0 0 6 3 】

音響処理部 3 4 は、音信号 $A [n]$ における各処理期間 H の音量を低減する。具体的には、音響処理部 3 4 は、音信号 $A [n]$ の各処理期間 H についてゲインを低下させることで信号値を低減する。音信号 $A [n]$ のゲインを低減する度合（レシオ）は、例えば操作装置 1 4 に対する利用者からの指示に応じて設定される。前述の通り、出力包絡線 $E y[n]$ は音源 $S [n]$ からの目的音の概形を表す信号である。したがって、出力包絡線 $E y[n]$ のレベル $y [n,m]$ が閾値 y_{TH2} を上回る処理期間 H について音信号 $A [n]$ の音量を低減することで、音信号 $A [n]$ の目的音における音量の変化を有効に制御できる。

40

【 0 0 6 4 】

図 10 は、音響処理装置 1 0 の制御装置 1 1 が実行する全体的な動作を例示するフローチャートである。例えば N 個の音源 $S [1] \sim S [N]$ の発音に並行して、解析期間 T_a 毎に図 1 0 の処理が実行される。

【 0 0 6 5 】

50

制御装置 11 (推定処理部 31) は、Nチャネルの観測包絡線 $E_x[1] \sim E_x[N]$ と、混合行列 Q と、Nチャネルの出力包絡線 $E_y[1] \sim E_y[N]$ とを取得する (S1)。具体的には、制御装置 11 は、Nチャネルの観測包絡線 $E_x[1] \sim E_x[N]$ とNチャネルの出力包絡線 $E_y[1] \sim E_y[N]$ とを図 6 の推定処理 S_a により生成し、事前の学習処理 S_b により生成された混

制御装置 11 (推定処理部 31) は、前述の推定処理 S_a により、Nチャネルの観測包絡線 $E_x[1] \sim E_x[N]$ と、混合行列 Q とから、Nチャネルの出力包絡線 $E_y[1] \sim E_y[N]$ を生成する (S1)。具体的には、制御装置 11 は、第 1 に、Nチャネルの音信号 $A[1] \sim A[N]$ から観測包絡線 $E_x[1] \sim E_x[N]$ を生成する。第 2 に、制御装置 11 は、Nチャネルの出力包絡線 $E_y[1] \sim E_y[N]$ を図 6 の推定処理 S_a により生成する。

【 0066 】

制御装置 11 (音響処理部 34) は、Nチャネルの音信号 $A[1] \sim A[N]$ の各々に対して出力包絡線 $E_y[n]$ のレベル $y[n,m]$ に応じた音響処理を実行する (S2)。具体的には、制御装置 11 は、出力包絡線 $E_y[n]$ のレベル $y[n,m]$ に応じて音信号 $A[n]$ に設定される各処理期間 H について音響処理を実行する。

【 0067 】

以上に説明した通り、第 1 実施形態においては、観測包絡線 $E_x[n]$ における音源 $S[n]$ からの目的音の概形を表す出力包絡線 $E_y[n]$ のレベル $y[n,m]$ に応じた音響処理が音信号 $A[n]$ に対して実行されるから、音信号 $A[n]$ に含まれる被り音の影響を低減して適切な音響処理を音信号 $A[n]$ に対して実行することが可能である。

【 0068 】

B : 第 2 実施形態

第 2 実施形態について説明する。なお、以下に例示する各形態において機能が第 1 実施形態と同様である要素については、第 1 実施形態の説明で使用した符号を流用して各々の詳細な説明を適宜に省略する。

【 0069 】

第 1 実施形態においては、複数の単位期間 $T_u[m]$ ($T_u[1] \sim T_u[M]$) を含む解析期間 T_a 毎に推定処理 S_a が実行される。第 2 実施形態においては、単位期間 $T_u[m]$ 毎に推定処理 S_a が実行される。すなわち、第 2 実施形態は、第 1 実施形態における 1 個の解析期間 T_a に含まれる単位期間 $T_u[m]$ の個数 M を 1 に限定した形態である。

【 0070 】

図 11 は、第 2 実施形態における推定処理 S_a の説明図である。第 2 実施形態においては、時間軸上の単位期間 $T_u[i]$ 毎 (i は自然数) 毎にNチャネルのレベル $x[1,i] \sim x[N,i]$ が生成される。観測行列 X は、1 個の単位期間 $T_u[i]$ に対応するNチャネルのレベル $x[1,i] \sim x[N,i]$ を縦方向に配列したN行1列の非負行列である。したがって、複数の単位期間 $T_u[i]$ にわたる観測行列 X の時系列が、Nチャネルの観測包絡線 $E_x[1] \sim E_x[N]$ に相当する。すなわち、第 n チャネルの観測包絡線 $E_x[n]$ は、複数の単位期間 $T_u[i]$ にわたるレベル $x[n,i]$ の時系列で表現される。同様に、係数行列 Y は、1 個の単位期間 $T_u[i]$ に対応するNチャネルのレベル $y[1,i] \sim y[N,i]$ を縦方向に配列したN行1列の非負行列である。したがって、複数の単位期間 $T_u[i]$ にわたる係数行列 Y の時系列が、Nチャネルの出力包絡線 $E_y[1] \sim E_y[N]$ に相当する。混合行列 Q は、第 1 実施形態と同様に、複数の混

【 0071 】

第 1 実施形態においては、M 個の単位期間 $T_u[1] \sim T_u[M]$ を含む解析期間 T_a 毎に図 6 の推定処理 S_a が実行される。第 2 実施形態においては、単位期間 $T_u[i]$ 毎に推定処理 S_a が実行される。すなわち、N 個の音源 $S[1] \sim S[N]$ による発音に並行して実時間的に推定処理 S_a が実行される。なお、推定処理 S_a の内容は第 1 実施形態と同様である。他方、学習処理 S_b は、第 1 実施形態と同様に、M 個の単位期間 $T_u[1] \sim T_u[m]$ を含む 1 個の解析期間 T_a について実行される。すなわち、第 2 実施形態においては、推定処理 S_a が単位期間 $T_u[i]$ 毎のレベル $y[n,i]$ を算定するリアルタイム処理であるのに対し、学習処理 S_b は、複数の単位期間 $T_u[1] \sim T_u[M]$ にわたる出力包絡線 $E_y[n]$ を算定するノンリアルタイ

10

20

30

40

50

△処理である。

【 0 0 7 2 】

以上の説明から理解される通り、第 2 実施形態によれば、N 個の音源 $S[1] \sim S[N]$ による発音に対する出力包絡線 $E_y[n]$ の遅延が低減される。すなわち、N 個の音源 $S[1] \sim S[N]$ による発音に並行して実時間的に各出力包絡線 $E_y[n]$ を生成できる。

【 0 0 7 3 】

図 1 0 に例示した処理 (S_1, S_2) は、単位期間 $T_u[i]$ 毎に実行される。具体的には、制御装置 1 1 (音響処理部 3 4) は、音信号 $A[n]$ に対する音響処理を単位期間 $T_u[i]$ 毎に実行する (S_3)。したがって、N 個の音源 $S[1] \sim S[N]$ の発音に対して遅延なく各音信号 $A[n]$ を処理できる。

【 0 0 7 4 】

C : 第 3 実施形態

図 1 2 は、第 3 実施形態における推定処理 S_a の説明図である。第 1 実施形態の推定処理部 3 1 における包絡線取得部 3 1 1 は、相異なる音源 $S[n]$ に対応する N チャネルの観測包絡線 $E_x[1] \sim E_x[N]$ を生成する。第 3 実施形態の包絡線取得部 3 1 1 は、相異なる周波数帯域に対応する 3 系統の観測包絡線 $E_x[n]$ ($E_x[n]_L, E_x[n]_M, E_x[n]_H$) をチャネル毎に生成する。観測包絡線 $E_x[n]_L$ は低周波数帯域に対応し、観測包絡線 $E_x[n]_M$ は中周波数帯域に対応し、観測包絡線 $E_x[n]_H$ は高周波数帯域に対応する。低周波数帯域は中周波数帯域の低域側に位置し、高周波数帯域は中周波数帯域の高域側に位置する。具体的には、低周波数帯域は中周波数帯域の下端値を下回る周波数帯域であり、高周波数帯域は中周波数帯域の上端値を上回る周波数帯域である。なお、観測包絡線 $E_x[n]$ が算定される周波数帯域の総数は 3 に限定されず任意である。なお、低周波数帯域と中周波数帯域と高周波数帯域とは、部分的に相互に重複してもよい。

【 0 0 7 5 】

包絡線取得部 3 1 1 は、各音信号 $A[n]$ を低周波数帯域と中周波数帯域と高周波数帯域の 3 個の周波数帯域に分割し、第 1 実施形態と同様の方法により周波数帯域毎に観測包絡線 $E_x[n]$ ($E_x[n]_L, E_x[n]_M, E_x[n]_H$) を生成する。以上の説明から理解される通り、観測行列 X は、3 系統の観測包絡線 $E_x[n]$ ($E_x[n]_L, E_x[n]_M, E_x[n]_H$) を N チャネルにわたり配列した 3 N 行 M 列の非負行列である。また、混合行列 Q は、相異なる周波数帯域に対応する 3 個の要素を N チャネルにわたり配列した 3 N 行 3 N 列の正方行列である。

【 0 0 7 6 】

信号処理部 3 1 2 は、相異なる周波数帯域に対応する 3 系統の出力包絡線 $E_y[n]$ ($E_y[n]_L, E_y[n]_M, E_y[n]_H$) を N チャネルの各々について生成する。出力包絡線 $E_y[n]_L$ は低周波数帯域に対応し、出力包絡線 $E_y[n]_M$ は中周波数帯域に対応し、出力包絡線 $E_y[n]_H$ は高周波数帯域に対応する。したがって、係数行列 Y は、3 系統の出力包絡線 $E_y[n]$ ($E_y[n]_L, E_y[n]_M, E_y[n]_H$) を N チャネルにわたり配列した 3 N 行 M 列の非負行列である。信号処理部 3 1 2 は、既知の混合行列 Q を利用した非負値行列因子分解により、観測行列 X から係数行列 Y を生成する。

【 0 0 7 7 】

以上の説明では推定処理 S_a に着目したが、学習処理 S_b についても同様である。具体的には、学習処理部 3 2 の包絡線取得部 3 2 1 は、相異なる周波数帯域に対応する 3 系統の観測包絡線 $E_x[n]$ ($E_x[n]_L, E_x[n]_M, E_x[n]_H$) を N チャネルの各々の音信号 $A[n]$ から生成する。すなわち、包絡線取得部 3 2 1 は、3 系統の観測包絡線 $E_x[n]$ ($E_x[n]_L, E_x[n]_M, E_x[n]_H$) を N チャネルにわたり配列した 3 N 行 N 列の観測行列 X を生成する。混合行列 Q は、相異なる周波数帯域に対応する 3 個の要素を N チャネルにわたり配列した 9 行 9 列の正方行列である。係数行列 Y は、相異なる周波数帯域に対応する 3 系統の出力包絡線 $E_y[n]$ ($E_y[n]_L, E_y[n]_M, E_y[n]_H$) を N チャネルにわたり配列した 3 N 行 N 列の非負行列である。信号処理部 3 2 2 は、非負値行列因子分解により観測行列 X から混合行列 Q と係数行列 Y とを生成する。

10

20

30

40

50

【 0 0 7 8 】

第3実施形態においても第1実施形態と同様の効果が実現される。また、第3実施形態においては、各チャンネルの観測包絡線 $E_x[n]$ および出力包絡線 $E_y[n]$ が複数の周波数帯域に分離されるから、音源 $S[n]$ の目的音を高精度に反映した観測包絡線 $E_x[n]$ および出力包絡線 $E_y[n]$ を生成できるという利点がある。なお、図12においては第1実施形態を基礎とした構成を例示したが、単位期間 $T_u[i]$ 毎に推定処理 S_a を実行する第2実施形態にも、第3実施形態の構成は同様に適用される。

【 0 0 7 9 】

D：変形例

以上に例示した各態様に付加される具体的な変形の態様を以下に例示する。以下の例示から任意に選択された2以上の態様を、相互に矛盾しない範囲で適宜に併合してもよい。

【 0 0 8 0 】

(1) 前述の各形態では、前掲の数式(1)の演算により各音信号 $A[n]$ の観測包絡線 $E_x[n]$ を生成したが、包絡線取得部311または包絡線取得部321が観測包絡線 $E_x[n]$ を生成する方法は以上の例示に限定されない。例えば、音信号 $A[n]$ の正側の各ピークから経時的に減衰する曲線または直線により観測包絡線 $E_x[n]$ を構成してもよい。また、音信号 $A[n]$ の正側の成分を平滑化することで観測包絡線 $E_x[n]$ を生成してもよい。

【 0 0 8 1 】

(2) 前述の各形態では、音響処理装置10の包絡線取得部311および包絡線取得部321が各音信号 $A[n]$ から観測包絡線 $E_x[n]$ を生成したが、外部装置で生成された観測包絡線 $E_x[n]$ を包絡線取得部311または包絡線取得部321が受信してもよい。すなわち、包絡線取得部311または包絡線取得部321は、音信号 $A[n]$ に対する処理で観測包絡線 $E_x[n]$ を生成する要素と、外部装置により生成された観測包絡線 $E_x[n]$ を受信する要素との双方を含む。

【 0 0 8 2 】

(3) 前述の各形態では非負値行列因子分解を例示したが、 N チャンネルの観測包絡線 $E_x[1] \sim E_x[N]$ から N チャンネルの出力包絡線 $E_y[1] \sim E_y[N]$ を生成するための方法は以上の例示に限定されない。例えば、非負拘束最小2乗法 ($NNLS$: Non-Negative Least Squares) を利用して各出力包絡線 $E_y[n]$ を生成してもよい。すなわち、混合行列 Q と係数行列 Y とにより観測行列 X を近似する任意の最適化手法が利用される。

【 0 0 8 3 】

(4) 前述の各形態では、音響処理部34が音信号 $A[n]$ に対してゲート処理またはコンプレッサ処理を実行する構成を例示したが、音響処理部34が実行する音響処理の内容は以上の例示に限定されない。ゲート処理またはコンプレッサ処理のほか、例えばリミッタ処理、エクスパンダ処理またはマキシマイザ処理等のダイナミクス制御を、音響処理部34が実行してもよい。リミッタ処理は、例えば、音信号 $A[n]$ において出力包絡線 $E_y[n]$ のレベル $y_{n,m}$ が閾値を上回る各処理期間 H について、所定値を上回る音量を当該所定値に設定する処理である。エクスパンダ処理は、音信号 $A[n]$ における各処理期間 H の音量を減少させる処理である。また、マキシマイザ処理は、音信号 $A[n]$ における各処理期間 H の音量を増加させる処理である。また、音響処理は、音信号 $A[n]$ の音量を制御するダイナミクス制御に限定されない。例えば、音信号 $A[n]$ の各処理期間 H に波形の歪を発生させるディストーション処理、または、音信号 $A[n]$ の各処理期間 H に残響を付与するリバース処理、等の各種の音響処理が、音響処理部34により実行される。

【 0 0 8 4 】

(5) 携帯電話機またはスマートフォン等の端末装置との間で通信するサーバ装置により音響処理装置10を実現してもよい。例えば、音響処理装置10は、端末装置から受信した N チャンネルの音信号 $A[1] \sim A[N]$ に対する推定処理 S_a または学習処理 S_b により、 N チャンネルの出力包絡線 $E_y[1] \sim E_y[N]$ を生成する。なお、 N チャンネルの観測包絡線 $E_x[1] \sim E_x[N]$ が端末装置から送信される構成では、包絡線取得部311または包絡線取得部321が N チャンネルの観測包絡線 $E_x[1] \sim E_x[N]$ を端末装置から受信する。音響処理装置10

10

20

30

40

50

の音響処理部 3 4 は、各音信号 A [n] に対する音響処理で生成した音信号 B を端末装置に送信する。

【 0 0 8 5 】

(6) 前述の各形態においては、推定処理部 3 1 と学習処理部 3 2 と音響処理部 3 4 とを具備する音響処理装置 1 0 を例示したが、音響処理装置 1 0 の一部の要素を省略してもよい。例えば、外部装置により生成された混合行列 Q が音響処理装置 1 0 に供給される構成では学習処理部 3 2 が省略される。また、混合行列 Q を生成する学習処理部 3 2 を具備する装置は、機械学習装置とも換言される。

【 0 0 8 6 】

(7) 以上に例示した音響処理装置 1 0 の機能は、前述の通り、制御装置 1 1 を構成する単数または複数のプロセッサと記憶装置 1 2 に記憶されたプログラム (P 1 ~ P 4) との協働により実現される。本開示に係るプログラムは、コンピュータが読取可能な記録媒体に格納された形態で提供されてコンピュータにインストールされ得る。記録媒体は、例えば非一過性 (non-transitory) の記録媒体であり、C D - R O M 等の光学式記録媒体 (光ディスク) が好例であるが、半導体記録媒体または磁気記録媒体等の公知の任意の形式の記録媒体も包含される。なお、非一過性の記録媒体とは、一過性の伝搬信号 (transitory, propagating signal) を除く任意の記録媒体を含み、揮発性の記録媒体も除外されない。また、配信装置が通信網を介してプログラムを配信する構成では、当該配信装置においてプログラムを記憶する記憶装置が、前述の非一過性の記録媒体に相当する。

【 0 0 8 7 】

E : 付記

以上に例示した形態から、例えば以下の構成が把握される。

【 0 0 8 8 】

本開示のひとつの態様 (態様 1) に係る音響処理方法は、音源からの音を收音した音信号の概形を表す観測包絡線を取得し、前記観測包絡線における前記音源からの音の概形を表す出力包絡線を、前記観測包絡線から生成し、前記音信号に対して前記出力包絡線のレベルに応じた音響処理を実行する。

【 0 0 8 9 】

以上の態様によれば、観測包絡線における音源からの音の概形を表す出力包絡線のレベルに応じた音響処理が音信号に対して実行されるから、音信号に含まれる被り音の影響を低減して適切な音響処理を音信号に対して実行することが可能である。

【 0 0 9 0 】

なお、「観測包絡線の取得」は、音信号に対する信号処理で観測包絡線を生成する動作と、他装置により生成された観測包絡線を受信する動作との双方を含む。また、「観測包絡線における音源からの音の概形を表す出力包絡線」とは、観測包絡線における音源以外の音源からの被り音が抑圧 (理想的には除去) された包絡線を意味する。

【 0 0 9 1 】

態様 1 の具体例 (態様 2) において、前記音響処理は、前記音信号において前記出力包絡線のレベルに応じた期間の音量を制御するダイナミクス制御を含む。態様 2 の具体例 (態様 3) において、前記ダイナミクス制御は、前記音信号において前記出力包絡線のレベルが閾値を下回る期間を消音するゲート処理を含む。以上の態様によれば、音信号において音以外の被り音の音量を有効に低減できる。また、態様 2 または態様 3 の具体例 (態様 4) において、前記ダイナミクス制御は、前記音信号において前記出力包絡線のレベルが閾値を上回る期間について所定値を上回る音量を低減するコンプレッサ処理を含む。以上の態様によれば、音信号における音の音量を有効に低減できる。

【 0 0 9 2 】

態様 1 から態様 4 の何れかの具体例 (態様 5) において、前記観測包絡線の取得においては、当該観測包絡線におけるレベルを単位期間毎に順次に取得し、前記出力包絡線の生成においては、前記単位期間毎に、前記出力包絡線の 1 個のレベルを生成する。以上の態様によれば、音源による発音に対する出力包絡線の遅延を十分に低減できる。

【 0 0 9 3 】

本開示のひとつの態様（態様 6）に係る音響処理方法は、第 1 音源からの音を収録した第 1 音信号の概形を表す第 1 観測包絡線と、第 2 音源からの音を収録した第 2 音信号の概形を表す第 2 観測包絡線と、を含む複数の観測包絡線を取得し、前記第 1 観測包絡線における前記第 2 音源からの被り音の混合比と、前記第 2 観測包絡線における前記第 1 音源からの被り音の混合比と、を含む混合行列を利用して、前記複数の観測包絡線から、前記第 1 観測包絡線における前記第 1 音源からの音の概形を表す第 1 出力包絡線と、前記第 2 観測包絡線における前記第 2 音源からの音の概形を表す第 2 出力包絡線と、を含む複数の出力包絡線を生成し、前記第 1 音信号に対して前記第 1 出力包絡線のレベルに応じた音響処理を実行し、前記第 2 音信号に対して前記第 2 出力包絡線のレベルに応じた音響処理を実行する。

10

【 0 0 9 4 】

以上の態様によれば、第 1 観測包絡線における第 1 音源からの音の概形を表す第 1 出力包絡線のレベルに応じた音響処理が第 1 音信号に対して実行され、第 2 観測包絡線における第 2 音源からの音の概形を表す第 2 出力包絡線のレベルに応じた音響処理が第 2 音信号に対して実行される。したがって、第 1 音信号および第 2 音信号の各々に含まれる被り音の影響を低減して適切な音響処理を実行することが可能である。

【 0 0 9 5 】

本開示のひとつの態様に係る音響処理装置は、以上に例示した各態様の音響処理方法を実行する。また、本開示のひとつの態様に係るプログラムは、以上に例示した各態様の音響処理方法をコンピュータに実行させる。

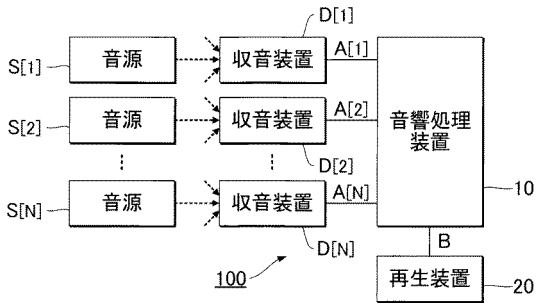
20

【 符号の説明 】

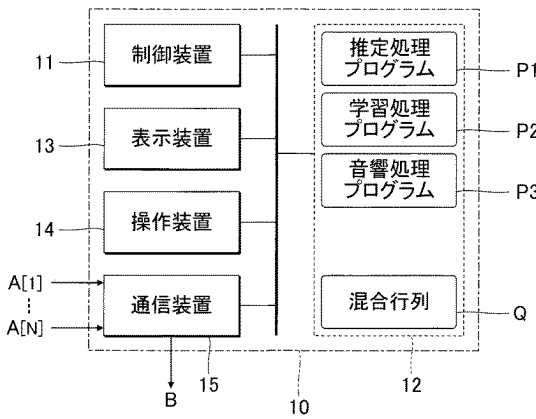
【 0 0 9 6 】

1 0 0 ... 音響処理システム、 1 0 ... 音響処理装置、 2 0 ... 再生装置、 D [n] (D [1] ~ D [N]) ... 収録装置、 1 1 ... 制御装置、 1 2 ... 記憶装置、 1 3 ... 表示装置、 1 4 ... 操作装置、 1 5 ... 通信装置、 3 1 ... 推定処理部、 3 1 1 ... 包絡線取得部、 3 1 2 ... 信号処理部、 3 2 ... 学習処理部、 3 2 1 ... 包絡線取得部、 3 2 2 ... 信号処理部、 3 4 ... 音響処理部。

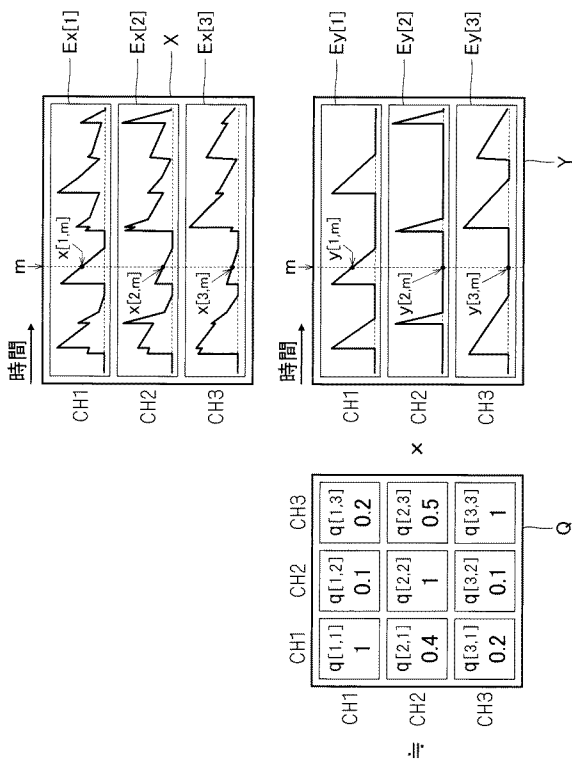
【図1】



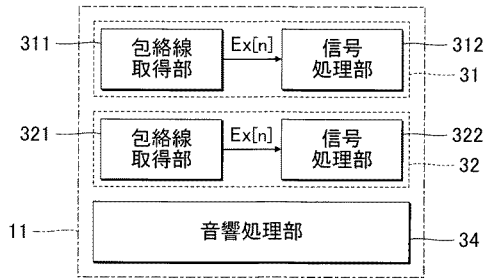
【図2】



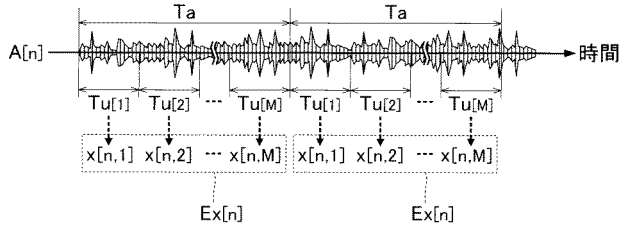
【図5】



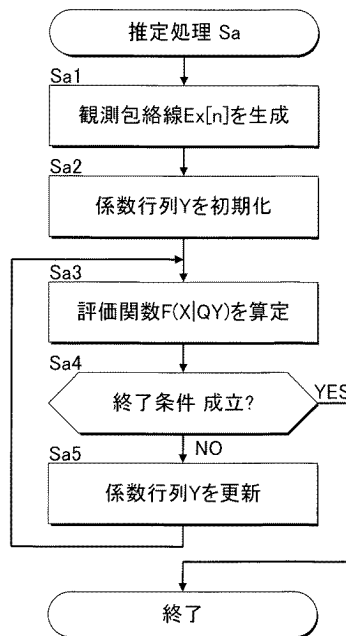
【図3】



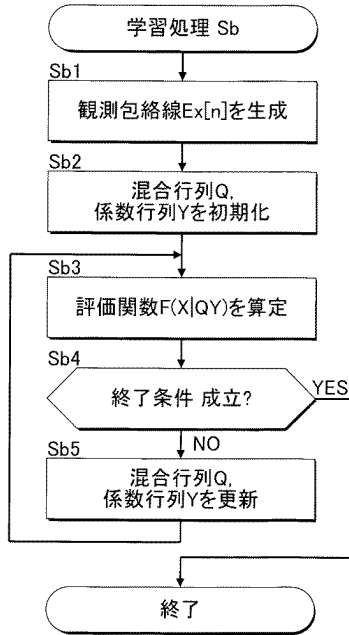
【図4】



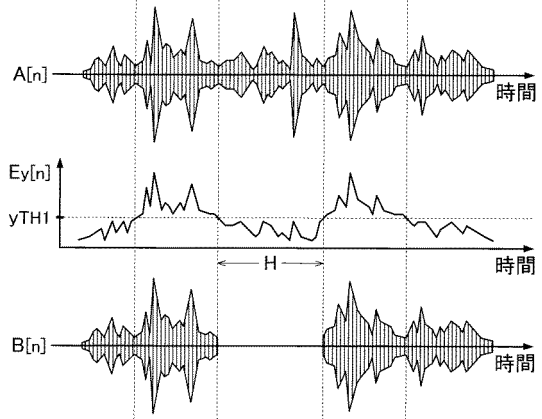
【図6】



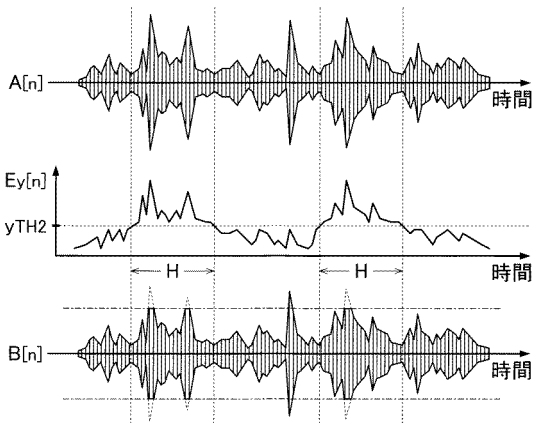
【 図 7 】



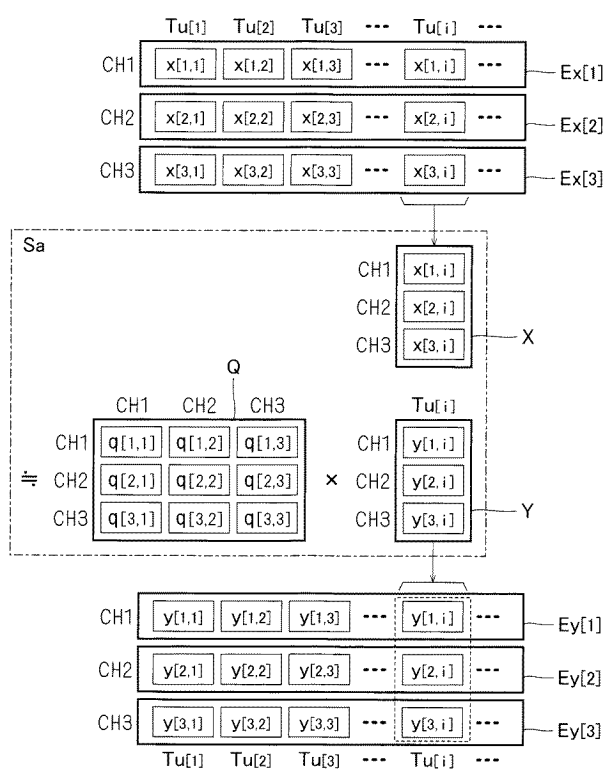
【 図 8 】



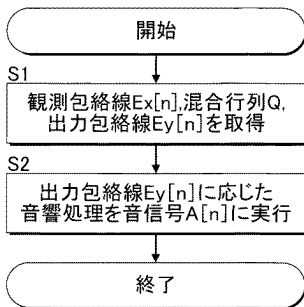
【 図 9 】



【 図 1 1 】



【 図 1 0 】



【 図 1 2 】

