

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7568062号
(P7568062)

(45)発行日 令和6年10月16日(2024.10.16)

(24)登録日 令和6年10月7日(2024.10.7)

(51)国際特許分類

F I

G 1 0 H 1/02 (2006.01)

G 1 0 H 1/02

請求項の数 16 (全14頁)

(21)出願番号	特願2023-508972(P2023-508972)	(73)特許権者	000004075
(86)(22)出願日	令和4年3月9日(2022.3.9)		ヤマハ株式会社
(86)国際出願番号	PCT/JP2022/010400		静岡県浜松市中央区中沢町10番1号
(87)国際公開番号	WO2022/202341	(74)代理人	100108523
(87)国際公開日	令和4年9月29日(2022.9.29)		弁理士 中川 雅博
審査請求日	令和5年9月7日(2023.9.7)	(74)代理人	100125704
(31)優先権主張番号	特願2021-50384(P2021-50384)		弁理士 坂根 剛
(32)優先日	令和3年3月24日(2021.3.24)	(74)代理人	100187931
(33)優先権主張国・地域又は機関	日本国(JP)		弁理士 澤村 英幸
		(72)発明者	須見 康平
			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	浅野 貴裕
			静岡県浜松市中区中沢町10番1号 ヤマハ株式会社内
			最終頁に続く

(54)【発明の名称】 音編集装置、音編集方法および音編集プログラム

(57)【特許請求の範囲】

【請求項1】

第1の音響信号を受け付ける第1の受付部と、
第2の音響信号を受け付ける第2の受付部と、
第1の入力音響信号および第2の入力音響信号と、前記第1の入力音響信号に付与されるべきエフェクトを反映する出力エフェクト情報との間の入出力関係を示す学習済モデルを用いて、前記第1の音響信号および前記第2の音響信号から前記第1の音響信号に付与されるべきエフェクトを反映するエフェクト情報を推定する推定部とを備える、音編集装置。

【請求項2】

前記エフェクト情報は、付与されるべきエフェクトが付与された前記第1の音響信号を生成するためのパラメータを含む、請求項1記載の音編集装置。

【請求項3】

前記学習済モデルは、前記第1の入力音響信号および前記第2の入力音響信号に基づいて、付与されるべきエフェクトが付与された前記第1の入力音響信号を生成するための出力パラメータが前記出力エフェクト情報として学習されることにより生成される、請求項2記載の音編集装置。

【請求項4】

前記エフェクト情報は、付与されるべきエフェクトが付与された前記第1の音響信号を含む、請求項1記載の音編集装置。

【請求項 5】

前記学習済モデルは、前記第 1 の入力音響信号および前記第 2 の入力音響信号に基づいて、付与されるべきエフェクトが付与された前記第 1 の入力音響信号が前記出力エフェクト情報として学習されることにより生成される、請求項 4 記載の音編集装置。

【請求項 6】

前記第 1 の音響信号に付与されるべきエフェクトの程度を調整する調整部をさらに備え、
前記推定部は、前記学習済モデルに基づいて、前記調整部により調整された程度で前記第 1 の音響信号に付与されるべきエフェクトを反映する前記エフェクト情報を推定する、
請求項 1 ～ 5 のいずれか一項に記載の音編集装置。

【請求項 7】

前記学習済モデルは、前記第 1 の入力音響信号に付与されるべきエフェクトの程度に対応して複数生成される、請求項 6 記載の音編集装置。

【請求項 8】

第 1 の音響信号を受け付け、

第 2 の音響信号を受け付け、

第 1 の入力音響信号および第 2 の入力音響信号と、前記第 1 の入力音響信号に付与されるべきエフェクトを反映する出力エフェクト情報との間の入出力関係を示す学習済モデルを用いて、前記第 1 の音響信号および前記第 2 の音響信号から前記第 1 の音響信号に付与されるべきエフェクトを反映するエフェクト情報を推定する、

コンピュータにより実行される、音編集方法。

【請求項 9】

前記エフェクト情報は、付与されるべきエフェクトが付与された前記第 1 の音響信号を生成するためのパラメータを含む、請求項 8 記載の音編集方法。

【請求項 10】

前記学習済モデルは、前記第 1 の入力音響信号および前記第 2 の入力音響信号に基づいて、付与されるべきエフェクトが付与された前記第 1 の入力音響信号を生成するための出力パラメータが前記出力エフェクト情報として学習されることにより生成される、請求項 9 記載の音編集方法。

【請求項 11】

前記エフェクト情報は、付与されるべきエフェクトが付与された前記第 1 の音響信号を含む、請求項 8 記載の音編集方法。

【請求項 12】

前記学習済モデルは、前記第 1 の入力音響信号および前記第 2 の入力音響信号に基づいて、付与されるべきエフェクトが付与された前記第 1 の入力音響信号が前記出力エフェクト情報として学習されることにより生成される、請求項 11 記載の音編集方法。

【請求項 13】

前記コンピュータは、

前記第 1 の音響信号に付与されるべきエフェクトの程度をさらに調整し、

前記エフェクト情報の推定において、前記学習済モデルに基づいて、調整された程度で前記第 1 の音響信号に付与されるべきエフェクトを反映する前記エフェクト情報を推定する、
請求項 8 ～ 12 のいずれか一項に記載の音編集方法。

【請求項 14】

前記学習済モデルは、前記第 1 の入力音響信号に付与されるべきエフェクトの程度に対応して複数生成される、請求項 13 記載の音編集方法。

【請求項 15】

コンピュータに音編集方法を実行させるプログラムであって、

第 1 の音響信号を受け付ける処理と、

第 2 の音響信号を受け付ける処理と、

第 1 の入力音響信号および第 2 の入力音響信号と、前記第 1 の入力音響信号に付与されるべきエフェクトを反映する出力エフェクト情報との間の入出力関係を示す学習済モデル

10

20

30

40

50

を用いて、前記第 1 の音響信号および前記第 2 の音響信号から前記第 1 の音響信号に付与されるべきエフェクトを反映するエフェクト情報を推定する処理とを、

前記コンピュータにより実行させる、音編集プログラム。

【請求項 16】

コンピュータに音編集方法を実行させるプログラムであって、

第 1 の音響信号を受け付ける処理と、

第 2 の音響信号を受け付ける処理と、

第 1 の入力音響信号および第 2 の入力音響信号と、前記第 1 の入力音響信号に付与されるべきエフェクトを反映する出力エフェクト情報との間の入出力関係を示す学習済モデルを用いて、前記第 1 の音響信号および前記第 2 の音響信号から前記第 1 の音響信号に付与されるべきエフェクトを反映するエフェクト情報を出力する処理とを、

前記コンピュータにより実行させる、音編集プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音を編集する音編集装置、音編集方法および音編集プログラムに関する。

【背景技術】

【0002】

合奏においては、複数の演奏者により同時に楽器が演奏される。したがって、各演奏者は、周囲の演奏者による楽器の音量とのバランスが維持されるように、自己の音量を調整することが好ましい。しかしながら、演奏者は、自己の出音が聴こえにくいため、自己の音量を大きくする傾向がある。この場合、他の演奏者も自己の音量を大きくすることとなるため、音量のバランスを維持することは容易ではない。特に、演奏会場が狭い場合には、音が飽和して会場内を回ることとなるので、音量のバランスを維持することはより困難になる。

【0003】

【文献】特開 2020 - 160139 号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

音の明瞭度が大きくなるように音響信号にエフェクトを付加することにより、演奏者は、楽器の音量を大きくすることなく自己の出音を認識可能になると考えられる。例えば、特許文献 1 には、様々なエフェクトを音響信号に付加する効果付加装置が記載されている。しかしながら、各演奏者による音の明瞭度は周囲の演奏者による音に応じて変化するため、音の明瞭度が大きくなるように音響信号にエフェクトを付加することは容易ではない。

【0005】

本発明の目的は、音の明瞭度を容易に大きくすることが可能な音編集装置、音編集方法および音編集プログラムを提供することである。

【課題を解決するための手段】

【0006】

本発明の一局面に従う音編集装置は、第 1 の音響信号を受け付ける第 1 の受付部と、第 2 の音響信号を受け付ける第 2 の受付部と、第 1 の入力音響信号および第 2 の入力音響信号と、第 1 の入力音響信号に付与されるべきエフェクトを反映する出力エフェクト情報との間の入出力関係を示す学習済モデルを用いて、第 1 の音響信号および第 2 の音響信号から第 1 の音響信号に付与されるべきエフェクトを反映するエフェクト情報を推定する推定部とを備える。

【0007】

本発明の他の局面に従う音編集方法は、第 1 の音響信号を受け付け、第 2 の音響信号を受け付け、第 1 の入力音響信号および第 2 の入力音響信号と、第 1 の入力音響信号に付与されるべきエフェクトを反映する出力エフェクト情報との間の入出力関係を示す学習済モ

10

20

30

40

50

デルを用いて、第１の音響信号および第２の音響信号から第１の音響信号に付与されるべきエフェクトを反映するエフェクト情報を推定し、コンピュータにより実行される。

【０００８】

本発明のさらに他の局面に従う音編集プログラムは、コンピュータに音編集方法を実行させるプログラムであって、第１の音響信号を受け付ける処理と、第２の音響信号を受け付ける処理と、第１の入力音響信号および第２の入力音響信号と、第１の入力音響信号に付与されるべきエフェクトを反映する出力エフェクト情報との間の入出力関係を示す学習済モデルを用いて、第１の音響信号および第２の音響信号から第１の音響信号に付与されるべきエフェクトを反映するエフェクト情報を推定する処理とを、コンピュータにより実行させる。

10

本発明のさらに他の局面に従う音編集プログラムは、コンピュータに音編集方法を実行させるプログラムであって、第１の音響信号を受け付ける処理と、第２の音響信号を受け付ける処理と、第１の入力音響信号および第２の入力音響信号と、第１の入力音響信号に付与されるべきエフェクトを反映する出力エフェクト情報との間の入出力関係を示す学習済モデルを用いて、第１の音響信号および第２の音響信号から第１の音響信号に付与されるべきエフェクトを反映するエフェクト情報を出力する処理とを、コンピュータにより実行させる。

【発明の効果】

【０００９】

本発明によれば、音の明瞭度を容易に大きくすることが可能になる。

20

【図面の簡単な説明】

【００１０】

【図１】図１は本発明の第１の実施の形態に係る音編集装置を含む処理システムの構成を示すブロック図である。

【図２】図２は図１の音学習装置および音編集装置の構成を示すブロック図である。

【図３】図３は第１の音響信号および第３の音響信号の一例を示す図である。

【図４】図４は図２の音学習装置による音学習処理の一例を示すフローチャートである。

【図５】図５は図２の音編集装置による音編集処理の一例を示すフローチャートである。

【図６】図６は本発明の第２の実施の形態に係る音編集装置を含む処理システムの構成を示すブロック図である。

30

【図７】図７は図６の音学習装置および音編集装置の構成を示すブロック図である。

【図８】図８は図７の音学習装置による音学習処理の一例を示すフローチャートである。

【図９】図９は図７の音編集装置による音編集処理の一例を示すフローチャートである。

【図１０】図１０は他の実施の形態に係る音編集装置の構成を示すブロック図である。

【発明を実施するための形態】

【００１１】

[１] 第１の実施の形態

(１) 処理システムの構成

以下、本発明の実施の形態に係る音編集装置、音編集方法および音編集プログラムについて図面を用いて詳細に説明する。図１は、本発明の第１の実施の形態に係る音編集装置を含む処理システムの構成を示すブロック図である。図１に示すように、処理システム１００は、ＲＡＭ（ランダムアクセスメモリ）１１０、ＲＯＭ（リードオンリメモリ）１２０、ＣＰＵ（中央演算処理装置）１３０および記憶部１４０を備える。

40

【００１２】

処理システム１００は、例えばエフェクタまたはスピーカに設けられる。また、処理システム１００は、パーソナルコンピュータ等の情報処理装置により実現されてもよいし、演奏機能を備えた電子楽器により実現されてもよい。ＲＡＭ１１０、ＲＯＭ１２０、ＣＰＵ１３０および記憶部１４０は、バス１５０に接続される。ＲＡＭ１１０、ＲＯＭ１２０およびＣＰＵ１３０により音学習装置１０および音編集装置２０が構成される。本実施の形態では、音学習装置１０と音編集装置２０とは共通の処理システム１００により構成さ

50

れるが、別個の処理システムにより構成されてもよい。

【0013】

R A M 1 1 0 は、例えば揮発性メモリからなり、C P U 1 3 0 の作業領域として用いられるとともに、各種データを一時的に記憶する。R O M 1 2 0 は、例えば不揮発性メモリからなり、音学習プログラムおよび音編集プログラムを記憶する。C P U 1 3 0 は、R O M 1 2 0 に記憶された音学習プログラムを R A M 1 1 0 上で実行することにより音学習処理を行う。また、C P U 1 3 0 は、R O M 1 2 0 に記憶された音編集プログラムを R A M 1 1 0 上で実行することにより音編集処理を行う。音学習処理および音編集処理の詳細については後述する。

【0014】

音学習プログラムまたは音編集プログラムは、R O M 1 2 0 ではなく記憶部 1 4 0 に記憶されてもよい。あるいは、音学習プログラムまたは音編集プログラムは、コンピュータが読み取り可能な記憶媒体に記憶された形態で提供され、R O M 1 2 0 または記憶部 1 4 0 にインストールされてもよい。あるいは、処理システム 1 0 0 がインターネット等のネットワークに接続されている場合には、当該ネットワーク上のサーバ（クラウドサーバを含む。）から配信された音学習プログラムまたは音編集プログラムが R O M 1 2 0 または記憶部 1 4 0 にインストールされてもよい。

【0015】

記憶部 1 4 0 は、ハードディスク、光学ディスク、磁気ディスクまたはメモリカード等の記憶媒体を含み、学習済モデル M および複数の学習データ D 1 を記憶する。学習済モデル M または複数の学習データ D 1 は、記憶部 1 4 0 に記憶されず、コンピュータが読み取り可能な記憶媒体に記憶されていてもよい。あるいは、処理システム 1 0 0 がネットワークに接続されている場合には、学習済モデル M または複数の学習データ D 1 は、当該ネットワーク上のサーバに記憶されていてもよい。学習済モデル M は、複数の学習データ D 1 に基づいて構築される。学習済モデル M の詳細については後述する。

【0016】

本実施の形態では、各学習データ D 1 は、第 1 の入力音響信号、第 2 の入力音響信号および出力音響信号をそれぞれ示す複数（マルチトラック）の波形データを含む。第 1 の入力音響信号は、第 1 の使用者が使用する楽器と同じ種類の楽器を用いて演奏された音等、第 1 の使用者により演奏されると仮定された音に対応する。第 2 の入力音響信号は、第 2 の使用者が使用する楽器と同じ種類の楽器を用いて演奏された音等、第 2 の使用者により演奏されると仮定された音に対応する。

【0017】

出力音響信号は、本実施の形態における出力エフェクト情報の例であり、第 1 の入力音響信号および第 2 の入力音響信号に基づいて、付与されるべきエフェクトが第 1 の入力音響信号に付与された音響信号である。第 2 の入力音響信号が同時に入力される状況において、出力音響信号に対応する音の明瞭度は、第 1 の入力音響信号に対応する音の明瞭度よりも大きい。出力音響信号を示す波形データは、エフェクトのパラメータが調整されることにより、第 1 の入力音響信号を示す波形データから生成されてもよい。

【0018】

（2）音学習装置および音編集装置

図 2 は、図 1 の音学習装置 1 0 および音編集装置 2 0 の構成を示すブロック図である。図 2 に示すように、音学習装置 1 0 は、機能部として、第 1 の取得部 1 1、第 2 の取得部 1 2、第 3 の取得部 1 3 および構築部 1 4 を含む。図 1 の C P U 1 3 0 が音学習プログラムを実行することにより、音学習装置 1 0 の機能部が実現される。音学習装置 1 0 の機能部の少なくとも一部は、電子回路等のハードウェアにより実現されてもよい。

【0019】

第 1 の取得部 1 1 は、記憶部 1 4 0 等に記憶された各学習データ D 1 から第 1 の入力音響信号を取得する。第 2 の取得部 1 2 は、各学習データ D 1 から第 2 の入力音響信号を取得する。第 3 の取得部 1 3 は、各学習データ D 1 から出力音響信号を取得する。

10

20

30

40

50

【 0 0 2 0 】

構築部 1 4 は、各学習データ D 1 について、第 1 の取得部 1 1 および第 2 の取得部 1 2 によりそれぞれ取得された第 1 の入力音響信号および第 2 の入力音響信号に基づいて、第 3 の取得部 1 3 により取得された出力音響信号を機械学習する。複数の学習データ D 1 について機械学習を繰り返すことにより、構築部 1 4 は、第 1 の入力音響信号および第 2 の入力音響信号と、出力音響信号との間の入出力関係を示す学習済モデル M を構築する。

【 0 0 2 1 】

本例では、構築部 1 4 は例えば U - N e t を用いて機械学習するが、実施の形態はこれに限定されない。構築部 1 4 は、C N N (Convolutional Neural Network) または F C N (Fully Convolutional Network) 等の他の方式を用いて機械学習してもよい。構築部 1 4 により構築された学習済モデル M は、例えば記憶部 1 4 0 に記憶される。構築部 1 4 により構築された学習済モデル M は、ネットワーク上のサーバ等に記憶されてもよい。

10

【 0 0 2 2 】

音編集装置 2 0 は、機能部として、第 1 の受付部 2 1 、第 2 の受付部 2 2 および推定部 2 3 を含む。図 1 の C P U 1 3 0 が音編集プログラムを実行することにより、音編集装置 2 0 の機能部が実現される。音編集装置 2 0 の機能部の少なくとも一部は、電子回路等のハードウェアにより実現されてもよい。

【 0 0 2 3 】

本実施の形態では、第 1 の受付部 2 1 および第 2 の受付部 2 2 は、楽曲データ D 2 を取得する。楽曲データ D 2 は、第 1 の音響信号および第 2 の音響信号をそれぞれ示す複数の波形データを含み、例えば使用者を含む複数の演奏者が合奏することにより生成される。第 1 の音響信号は、使用者により演奏された音に対応する。第 2 の音響信号は、他の演奏者により演奏された音または使用者の周囲で発生した音に対応する。第 1 の受付部 2 1 は、楽曲データ D 2 から第 1 の音響信号を受け付ける。第 2 の受付部 2 2 は、楽曲データ D 2 から第 2 の音響信号を受け付ける。

20

【 0 0 2 4 】

推定部 2 3 は、記憶部 1 4 0 等に記憶された学習済モデル M を用いて、付与されるべきエフェクトが第 1 の音響信号に付与された第 3 の音響信号を楽曲データ D 2 に含まれる第 1 の音響信号および第 2 の音響信号から推定する。また、推定部 2 3 は、推定された第 3 の音響信号を出力する。本実施の形態では、第 3 の音響信号がエフェクト情報の例である。

30

【 0 0 2 5 】

図 3 は、第 1 の音響信号および第 3 の音響信号の一例を示す図である。図 3 の左欄には、楽曲データ D 2 に含まれる第 1 の音響信号と、第 1 の音響信号を周波数解析することにより得られたスペクトルとが示される。図 3 の右欄には、推定部 2 3 により出力された第 3 の音響信号と、第 3 の音響信号を周波数解析することにより得られたスペクトルとが示される。

【 0 0 2 6 】

図 3 の例では、一点鎖線で囲まれた A 部に示すように、比較的低い周波帯域においては、第 3 の音響信号の強度が第 1 の音響信号の強度よりも低減される。一方、二点鎖線で囲まれた B 部に示すように、比較的高い周波帯域においては、第 3 の音響信号の強度が第 1 の音響信号の強度よりも増強される。これにより、第 2 の音響信号が同時に発生した状況において、第 3 の音響信号に対応する音の明瞭度は、第 1 の音響信号に対応する音の明瞭度よりも大きくなる。

40

【 0 0 2 7 】

したがって、使用者は、推定部 2 3 により出力された第 3 の音響信号を用いることにより、楽器の音量を大きくすることなく自己の出音を容易に認識することができる。そのため、使用者は、合奏において、周囲の演奏者による楽器の音量とのバランスが維持されるように、適した音量で自己の楽器を演奏することができる。あるいは、ミキシングエンジニアは、複数の楽器の音量のバランスが維持されるように容易にミキシングを行うことができる。

50

【 0 0 2 8 】

(3) 音学習処理および音編集処理

図 4 は、図 2 の音学習装置 1 0 による音学習処理の一例を示すフローチャートである。図 4 の音学習処理は、図 1 の C P U 1 3 0 が音学習プログラムを実行することにより行われる。

【 0 0 2 9 】

第 1 の取得部 1 1 は、記憶部 1 4 0 等に記憶されたいずれかの学習データ D 1 から第 1 の入力音響信号を取得する (ステップ S 1) 。第 2 の取得部 1 2 は、ステップ S 1 の学習データ D 1 から第 2 の入力音響信号を取得する (ステップ S 2) 。第 3 の取得部 1 3 は、ステップ S 1 の学習データ D 1 から出力音響信号を取得する (ステップ S 3) 。ステップ S 1 ~ S 3 は、いずれが先に実行されてもよいし、同時に実行されてもよい。

10

【 0 0 3 0 】

次に、構築部 1 4 は、ステップ S 1 で取得された第 1 の入力音響信号およびステップ S 2 で取得された第 2 の入力音響信号と、ステップ S 3 で取得された出力音響信号との間の入出力関係を機械学習する (ステップ S 4) 。続いて、構築部 1 4 は、所定回数の機械学習が実行されたか否かを判定する (ステップ S 5) 。所定回数の機械学習が実行されていない場合、構築部 1 4 はステップ S 1 に戻る。

【 0 0 3 1 】

所定回数の機械学習が実行されるまで、学習データ D 1 または学習パラメータが変更されつつステップ S 1 ~ S 5 が繰り返される。機械学習の繰り返し回数は、構築される学習済モデルの精度に応じて予め設定される。所定回数の機械学習が実行された場合、構築部 1 4 は、機械学習の結果に基づいて、第 1 の入力音響信号および第 2 の入力音響信号と、出力音響信号との間の入出力関係を示す学習済モデル M を構築し (ステップ S 6) 、音学習処理を終了する。

20

【 0 0 3 2 】

図 5 は、図 2 の音編集装置 2 0 による音編集処理の一例を示すフローチャートである。図 5 の音編集処理は、図 1 の C P U 1 3 0 が音編集プログラムを実行することにより行われる。

【 0 0 3 3 】

第 1 の受付部 2 1 は、楽曲データ D 2 から第 1 の音響信号を受け付ける (ステップ S 1 1) 。第 2 の受付部 2 2 は、ステップ S 1 1 の楽曲データ D 2 から第 2 の音響信号を受け付ける (ステップ S 1 2) 。ステップ S 1 1 , S 1 2 は、いずれが先に実行されてもよいし、同時に実行されてもよい。推定部 2 3 は、音学習処理のステップ S 6 で構築された学習済モデル M を用いて、ステップ S 1 1 , S 1 2 でそれぞれ受け付けられた第 1 の音響信号および第 2 の音響信号から第 3 の音響信号を推定し (ステップ S 1 3) 、音編集処理を終了する。

30

【 0 0 3 4 】

(4) 実施の形態の効果

以上説明したように、本実施の形態に係る音編集装置 2 0 は、第 1 の音響信号を受け付ける第 1 の受付部 2 1 と、第 2 の音響信号を受け付ける第 2 の受付部 2 2 と、第 1 の入力音響信号および第 2 の入力音響信号と、第 1 の入力音響信号に付与されるべきエフェクトを反映する出力エフェクト情報との間の入出力関係を示す学習済モデル M を用いて、第 1 の音響信号および第 2 の音響信号から第 1 の音響信号に付与されるべきエフェクトを反映するエフェクト情報を推定する推定部 2 3 とを備える。

40

【 0 0 3 5 】

この構成によれば、第 2 の音響信号が変化する場合でも、学習済モデル M を用いて、音の明瞭度が大きくなるように第 1 の音響信号に付与されるべきエフェクトを反映するエフェクト情報を得ることが可能になる。これにより、音の明瞭度を容易に大きくすることができる。

【 0 0 3 6 】

50

エフェクト情報は、付与されるべきエフェクトが付与された第１の音響信号（第３の音響信号）を含んでもよい。この場合、推定された第３の音響信号を用いることにより、明瞭度が大きくされた音を容易に得ることができる。

【００３７】

学習済モデルＭは、第１の入力音響信号および第２の入力音響信号に基づいて、付与されるべきエフェクトが付与された第１の入力音響信号（出力音響信号）が出力エフェクト情報として学習されることにより生成されてもよい。この場合、第１の音響信号および第２の音響信号から第３の音響信号を推定するための学習済モデルＭを容易に生成することができる。

【００３８】

[２] 第２の実施の形態

(１) 処理システムの構成

第２の実施の形態に係る音編集装置２０、音編集方法および音編集プログラムについて、第１の実施の形態に係る音編集装置２０、音編集方法および音編集プログラムと異なる点を説明する。図６は、本発明の第２の実施の形態に係る音編集装置２０を含む処理システム１００の構成を示すブロック図である。図６に示すように、処理システム１００は、エフェクト付与部１６０をさらに備える。エフェクト付与部１６０は、例えばイコライザまたはコンプレッサを含み、バス１５０に接続される。エフェクト付与部１６０は、入力されたパラメータに基づいて、音響信号にエフェクトを付与する。

【００３９】

本実施の形態では、記憶部１４０等に記憶される各学習データＤ１は、第１の入力音響信号および第２の入力音響信号をそれぞれ示す複数の波形データを含む。また、各学習データＤ１は、出力音響信号を示す波形データに代えて、出力音響信号を生成するために第１の入力音響信号に付与されるべきエフェクトを反映するパラメータ（以下、出力パラメータと呼ぶ。）を含む。出力パラメータは、本実施の形態における出力エフェクト情報の例である。

【００４０】

(２) 音学習装置および音編集装置

図７は、図６の音学習装置１０および音編集装置２０の構成を示すブロック図である。本実施の形態においては、音学習装置１０の第３の取得部１３は、各学習データＤ１から出力パラメータを取得する。第１の取得部１１および第２の取得部１２の動作は、第１の実施の形態における第１の取得部１１および第２の取得部１２の動作とそれぞれ同様である。

【００４１】

構築部１４は、各学習データＤ１について、第１の取得部１１および第２の取得部１２によりそれぞれ取得された第１の入力音響信号および第２の入力音響信号に基づいて、第３の取得部１３により取得された出力パラメータを機械学習する。複数の学習データＤ１について機械学習を繰り返すことにより、構築部１４は、第１の入力音響信号および第２の入力音響信号と、出力パラメータとの間の入出力関係を示す学習済モデルＭを構築する。

【００４２】

本例では、構築部１４は例えばＣＮＮを用いて機械学習するが、実施の形態はこれに限定されない。構築部１４は、ＲＮＮ（Recurrent Neural Network）またはAttention等の他の方式を用いて機械学習してもよい。構築部１４により構築された学習済モデルＭは、例えば記憶部１４０に記憶される。構築部１４により構築された学習済モデルＭは、ネットワーク上のサーバ等に記憶されてもよい。

【００４３】

音編集装置２０において、第１の受付部２１および第２の受付部２２は、合奏により生成された第１の音響信号および第２の音響信号をそれぞれリアルタイムで取得する。推定部２３は、記憶部１４０等に記憶された学習済モデルＭを用いて、付与されるべきエフェクトが付与された第１の音響信号を生成するためのパラメータを第１の音響信号および第

10

20

30

40

50

２の音響信号から順次推定する。また、推定部２３は、推定されたパラメータを順次出力する。本実施の形態では、パラメータがエフェクト情報の例である。

【００４４】

エフェクト付与部１６０は、推定部２３により出力されたパラメータに基づいて、第１の受付部２１により取得された第１の音響信号にエフェクトを付与する。これにより、図３の右欄に示された第３の音響信号と同様の第４の音響信号が生成される。したがって、第２の音響信号が同時に発生した状況において、第４の音響信号に対応する音の明瞭度は、第１の音響信号に対応する音の明瞭度よりも大きくなる。

【００４５】

（３）音学習処理および音編集処理

図８は、図７の音学習装置１０による音学習処理の一例を示すフローチャートである。図８の例では、音学習処理は、ステップＳ２１～Ｓ２６を含む。ステップＳ２１，Ｓ２２は、図４の音学習処理のステップＳ１，Ｓ２とそれぞれ同様である。ステップＳ２３において、第３の取得部１３は、学習データＤ１から出力パラメータを取得する（ステップＳ２３）。ステップＳ２１～Ｓ２３は、いずれが先に実行されてもよいし、同時に実行されてもよい。

【００４６】

構築部１４は、ステップＳ２１で取得された第１の入力音響信号およびステップＳ２２で取得された第２の入力音響信号と、ステップＳ２３で取得された出力パラメータとの間の入出力関係を機械学習する（ステップＳ２４）。ステップＳ２５，Ｓ２６は、図４の音学習処理のステップＳ５，Ｓ６とそれぞれ同様である。これにより、ステップＳ２６において、第１の入力音響信号および第２の入力音響信号と、出力パラメータとの間の入出力関係を示す学習済モデルＭが構築される。

【００４７】

図９は、図７の音編集装置２０による音編集処理の一例を示すフローチャートである。第１の受付部２１は、合奏により生成された第１の音響信号を受け付ける（ステップＳ３１）。第２の受付部２２は、合奏により生成された第２の音響信号を受け付ける（ステップＳ３２）。ステップＳ３１，Ｓ３２は、略同時に実行される。

【００４８】

推定部２３は、音学習処理のステップＳ２６で構築された学習済モデルＭを用いて、ステップＳ３１，Ｓ３２でそれぞれ受け付けられた第１の音響信号および第２の音響信号からパラメータを推定する（ステップＳ３３）。その後、推定部２３は、ステップＳ３３で推定されたパラメータを図７のエフェクト付与部１６０に出力し（ステップＳ３４）、ステップＳ３１に戻る。合奏が終了するまでステップＳ３１～Ｓ３４が繰り返される。

【００４９】

（４）実施の形態の効果

本実施の形態においては、第１の実施の形態と同様に、第２の音響信号が変化する場合でも、学習済モデルＭを用いて、音の明瞭度が大きくなるように第１の音響信号に付与されるべきエフェクトを反映するエフェクト情報を得ることが可能になる。これにより、音の明瞭度を容易に大きくすることができる。

【００５０】

エフェクト情報は、付与されるべきエフェクトが付与された第１の音響信号を生成するためのパラメータを含んでもよい。この場合、エフェクト情報を高速に得ることができる。また、エフェクト情報に基づいて第１の音響信号にパラメータが付与された第４の音響信号を用いることにより、明瞭度が大きくされた音を容易に得ることができる。

【００５１】

学習済モデルＭは、第１の入力音響信号および第２の入力音響信号に基づいて、付与されるべきエフェクトが付与された第１の入力音響信号を生成するための出力パラメータが出力エフェクト情報として学習されることにより生成されてもよい。この場合、第１の音響信号および第２の音響信号からパラメータを推定するための学習済モデルＭを容易に生

10

20

30

40

50

成することができる。

【 0 0 5 2 】

[3] 他の実施の形態

(1) 第 1 の実施の形態において、第 1 の入力音響信号および第 2 の入力音響信号と、出力音響信号との間の入出力関係を示す学習済モデル M が音学習装置 1 0 により構築されるが、実施の形態はこれに限定されない。第 2 の実施の形態と同様に、第 1 の入力音響信号および第 2 の入力音響信号と、出力パラメータとの間の入出力関係を示す学習済モデル M が音学習装置 1 0 により構築されてもよい。

【 0 0 5 3 】

この場合、構築された学習済モデル M を用いて、付与されるべきエフェクトが付与された第 1 の音響信号を生成するためのパラメータが第 1 の音響信号および第 2 の音響信号から音編集装置 2 0 により推定される。この構成においては、音学習装置 1 0 または音編集装置 2 0 を実現するための CPU 1 3 0 の処理速度が比較的低くてもよい。また、処理システム 1 0 0 は、エフェクト付与部 1 6 0 を含んでもよい。音編集装置 2 0 により推定されたパラメータがエフェクト付与部 1 6 0 に出力されることにより、第 4 の音響信号が生成される。

10

【 0 0 5 4 】

(2) 第 2 の実施の形態において、第 1 の入力音響信号および第 2 の入力音響信号と、出力パラメータとの間の入出力関係を示す学習済モデル M が音学習装置 1 0 により構築されるが、実施の形態はこれに限定されない。第 1 の実施の形態と同様に、第 1 の入力音響信号および第 2 の入力音響信号と、出力音響信号との間の入出力関係を示す学習済モデル M が音学習装置 1 0 により構築されてもよい。

20

【 0 0 5 5 】

この場合、構築された学習済モデル M を用いて、付与されるべきエフェクトが第 1 の音響信号に付与された第 3 の音響信号が第 1 の音響信号および第 2 の音響信号から音編集装置 2 0 により推定される。したがって、処理システム 1 0 0 はエフェクト付与部 1 6 0 を含まなくてもよい。この構成においては、音学習装置 1 0 または音編集装置 2 0 を実現するための CPU 1 3 0 の処理速度が比較的高いことが好ましい。

【 0 0 5 6 】

(3) 上記実施の形態において、学習済モデル M を用いて第 1 の音響信号および第 2 の音響信号からエフェクト情報が推定されるが、実施の形態はこれに限定されない。第 1 の音響信号および第 2 の音響信号とエフェクト情報との対応関係を示すテーブル等の対応情報が記憶部 1 4 0 等に記憶されている場合には、当該対応情報を用いて第 1 の音響信号および第 2 の音響信号からエフェクト情報が推定されてもよい。

30

【 0 0 5 7 】

(4) 図 1 0 は、他の実施の形態に係る音編集装置 2 0 の構成を示すブロック図である。図 1 0 に示すように、他の実施に係る音編集装置 2 0 は、機能部として調整部 2 4 をさらに含む。調整部 2 4 は、例えば図示しない表示装置に表示される GUI (Graphical User Interface) であり、使用者により操作される。調整部 2 4 は、GUI ではなく、物理的なダイヤル、スイッチまたはボタンであってもよい。

40

【 0 0 5 8 】

使用者は、音楽性を低下させてでも音の明瞭度を増加させたい場合等には、エフェクトの程度が大きくなるように調整部 2 4 を操作することができる。一方、使用者は、リラックスしたい場合等には、エフェクトの程度が小さくなるように調整部 2 4 を操作することができる。調整部 2 4 は、使用者による操作に基づいて、第 1 の音響信号に付与されるべきエフェクトの程度を調整する。推定部 2 3 は、学習済モデル M に基づいて、調整部 2 4 により調整された程度で第 1 の音響信号に付与されるべきエフェクトを反映するエフェクト情報を推定する。

【 0 0 5 9 】

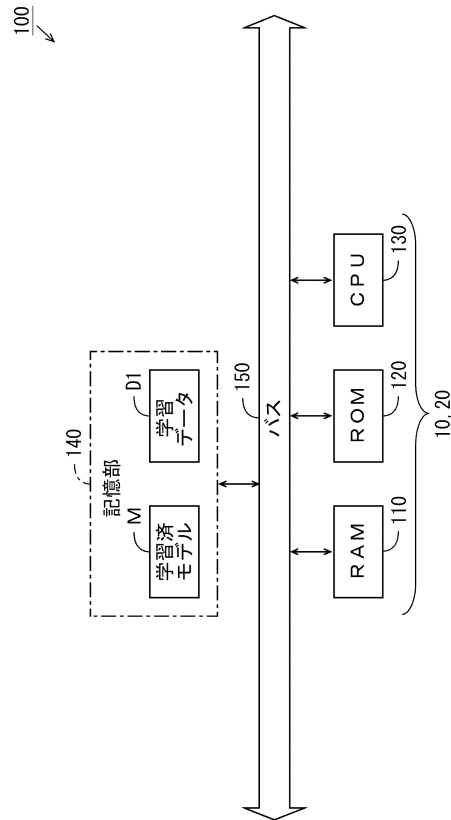
この構成においては、学習データ D 1 は、エフェクトの程度に対応して複数用意される

50

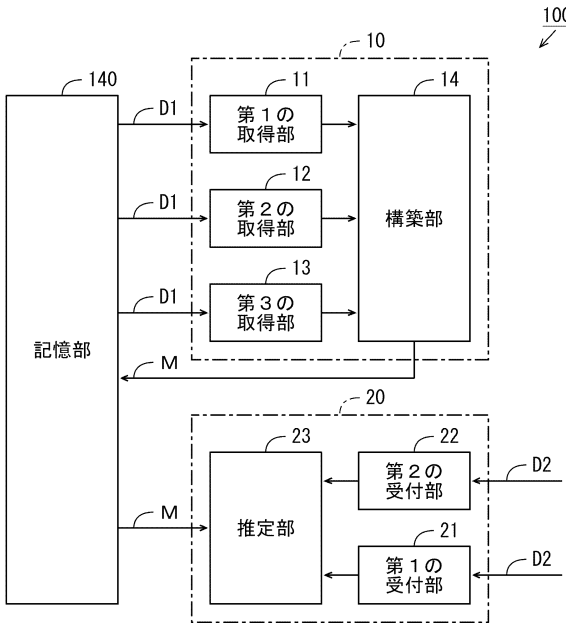
。また、音学習装置 1 0 の構築部 1 4 は、第 1 の入力音響信号に付与されるべきエフェクトの程度に対応して、学習済モデル M を複数生成する。

【図面】

【図 1】



【図 2】



10

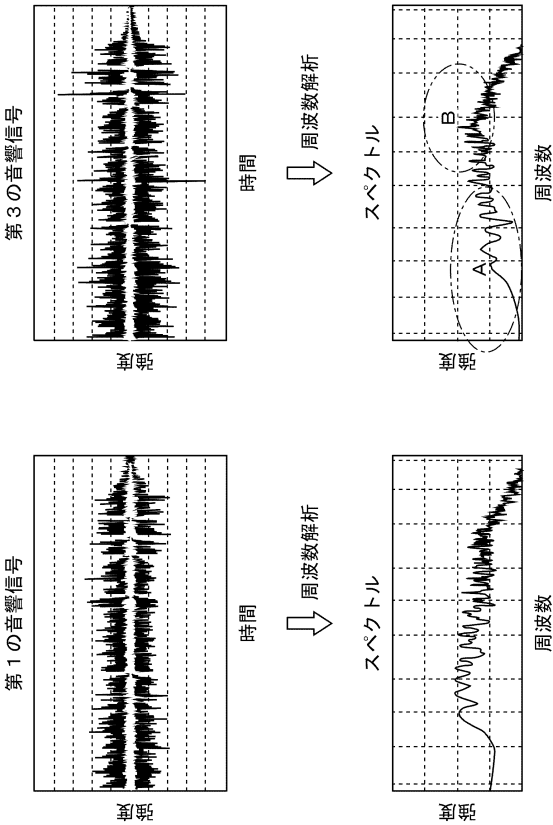
20

30

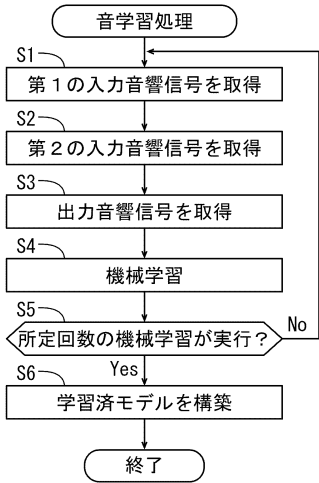
40

50

【図 3】



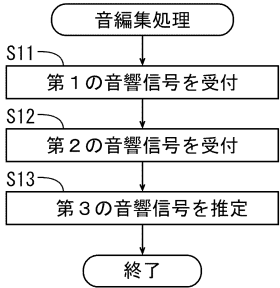
【図 4】



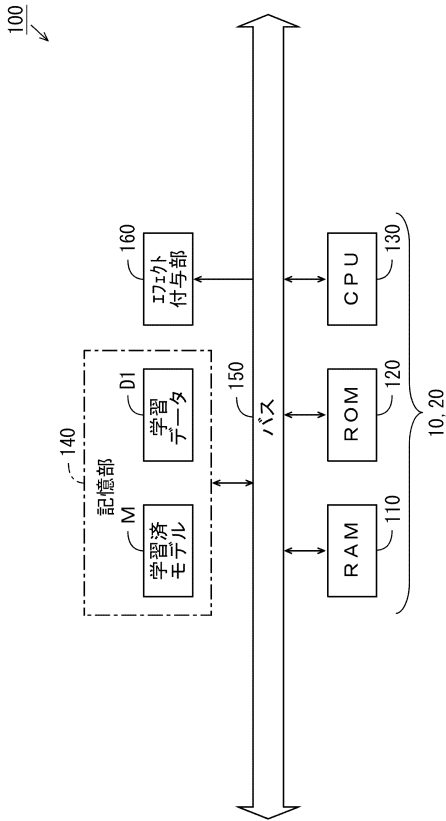
10

20

【図 5】



【図 6】

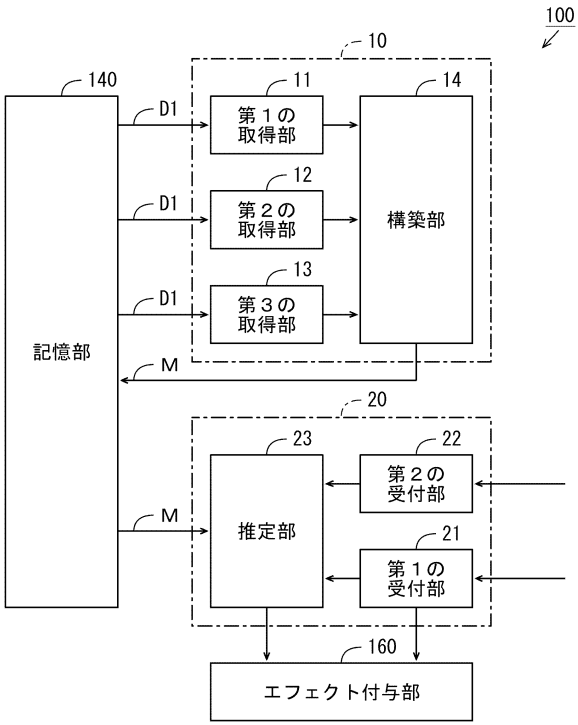


30

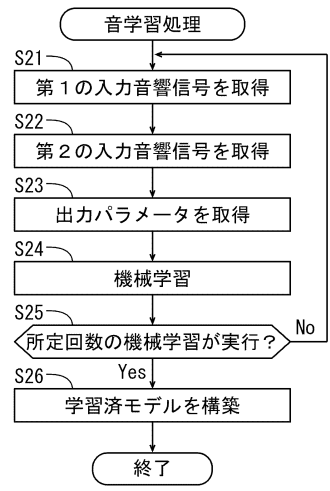
40

50

【図 7】



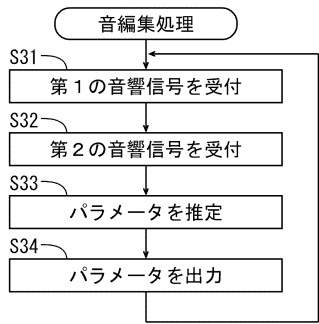
【図 8】



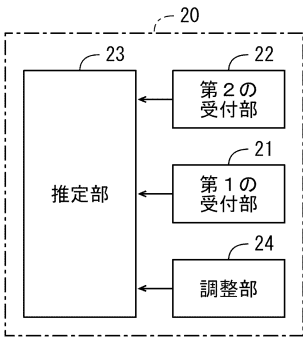
10

20

【図 9】



【図 10】



30

40

50

フロントページの続き

(72)発明者 大寄 郁弥
静岡県浜松市中区中沢町 1 0 番 1 号 ヤマハ株式会社内
審査官 大野 弘
(56)参考文献 特開 2 0 2 0 - 0 1 3 1 7 0 (J P , A)
(58)調査した分野 (Int.Cl. , D B 名)
G 1 0 H 1 / 0 2