

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2024年1月11日(11.01.2024)



(10) 国際公開番号

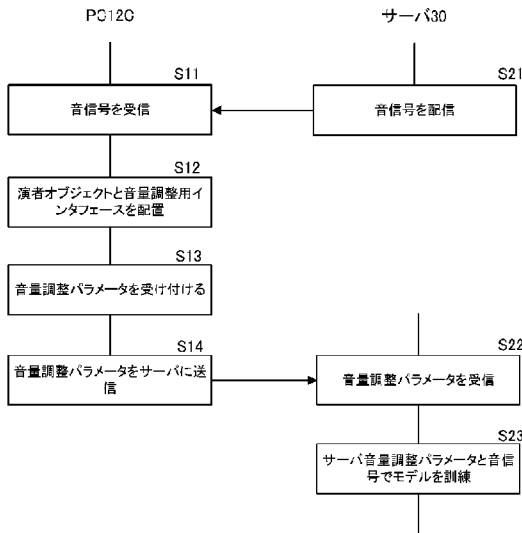
WO 2024/009677 A1

- (51) 国際特許分類:
H04R 3/00 (2006.01)
- (21) 国際出願番号: PCT/JP2023/021288
- (22) 国際出願日: 2023年6月8日(08.06.2023)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2022-107675 2022年7月4日(04.07.2022) JP
- (71) 出願人: ヤマハ株式会社 (YAMAHA CORPORATION) [JP/JP]; 〒4308650 静岡県浜松市中区中沢町10番1号 Shizuoka (JP).
- (72) 発明者: 白木原 太 (SHIRAKIHARA Futoshi); 〒4308650 静岡県浜松市中区中沢町10番

1号 ヤマハ株式会社内 Shizuoka (JP). 松田 遼 (MATSUDA Ryo); 〒4308650 静岡県浜松市中区中沢町10番1号 ヤマハ株式会社内 Shizuoka (JP). 中村 吉就 (NAKAMURA Yoshinari); 〒4308650 静岡県浜松市中区中沢町10番1号 ヤマハ株式会社内 Shizuoka (JP). 竹中 雄耶 (TAKENAKA Yuya); 〒4308650 静岡県浜松市中区中沢町10番1号 ヤマハ株式会社内 Shizuoka (JP). 石川 克己 (ISHIKAWA Katsumi); 〒4308650 静岡県浜松市中区中沢町10番1号 ヤマハ株式会社内 Shizuoka (JP). 大谷 明央 (OHTANI Akio); 〒4308650 静岡県浜松市中区中沢町10番1号 ヤマハ株式会社内 Shizuoka (JP). 山本 和彦 (YAMAMOTO Kazuhiko); 〒4308650 静岡県浜松市中区中沢町10番1号 ヤマハ株式会社内

(54) Title: SOUND PROCESSING METHOD, SOUND PROCESSING DEVICE, AND PROGRAM

(54) 発明の名称: 音処理方法、音処理装置、およびプログラム



30 Server
S11 Receive sound signals
S12 Arrange performer objects and sound volume adjustment interfaces
S13 Accept sound volume adjustment parameters
S14 Transmit sound volume adjustment parameters to server
S21 Distribute sound signals
S22 Receive sound volume adjustment parameters
S23 Train model using server sound volume adjustment parameters and sound signals

(57) Abstract: This sound processing method involves: arranging, in a virtual space, objects of a plurality of performers and a plurality of sound volume adjustment interfaces corresponding to the objects of the plurality of performers; receiving a plurality of sound signals corresponding to the plurality of performers; accepting, from a user, sound volume adjustment parameters that are respectively for the plurality of performers and that correspond to the plurality of sound volume adjustment interfaces; obtaining, using a trained model obtained by learning relationships between sound signals corresponding to the performers of the plurality of sound signals and the sound volume adjustment parameters corresponding to the sound signals, sound volume adjustment parameters corresponding to the plurality of performers; and adjusting and mixing sound volumes of the plurality of sound signals on the basis of the sound volume adjustment parameters obtained by the trained model.

Shizuoka (JP). 藤島 琢哉(FUJISHIMA Takuya);
〒4308650 静岡県浜松市中区中沢町 1 0 番 1
号 ヤマハ株式会社内 Shizuoka (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一 国際調査報告 (条約第21条(3))

(57) 要約: 音処理方法は、仮想空間内に複数の演者のオブジェクトと、前記複数の演者のオブジェクトに対応する複数の音量調整用インタフェースと、を配置し、前記複数の演者にそれぞれ対応する複数の音信号を受信し、利用者から、前記複数の音量調整用インタフェースに対応する、前記複数の演者に対するそれぞれの音量調整パラメータを受け付けて、前記複数の音信号の各演者に対応する音信号と、該音信号に対応する前記音量調整パラメータと、の関係を訓練された訓練済モデルを用いて、前記複数の演者に対するそれぞれの音量調整パラメータを求めて、前記訓練済モデルで求めた該音量調整パラメータに基づいて前記複数の音信号の音量を調整して混合する。

明 細 書

発明の名称：音処理方法、音処理装置、およびプログラム

技術分野

[0001] 本発明の一実施形態は、音処理方法、音処理装置、およびプログラムに関する。

背景技術

[0002] 特許文献1には、オーディオ・ミキサーB2，C3からネットワークを介してパフォーマンスに係る音信号を受信し、オーディオ・ミキサーB2，C3との間の通信遅延時間を測定し、測定された通信遅延時間に応じてオーディオ・ミキサーB2，C3の音信号をミキシングする、オーディオ・ミキサーが開示されている。

[0003] 特許文献2には、プリフェーダとポストフェーダの音量差を補正することで、プリフェーダとポストフェーダの切り替えをスムーズにする構成が開示されている。

[0004] 特許文献3には、スピーカからマイクに至るインパルス応答を測定し、間接音成分を考慮して音量調整を行うことで、間接音成分を考慮した適切な音量に調整する構成が開示されている。

[0005] 特許文献4には、近端側の直接音の音量測定結果とスピーカとマイクの距離を遠端側にフィードバックする構成が開示されている。これにより、遠端側ユーザは、自身の声が正しく拡声されていることを知ることができる。

[0006] 特許文献5には、近端側のマイクで取得した音響特徴量に基づいて遠端側から受信した音信号の音量調整を行う構成が開示されている。これにより特許文献5の発明は、聴取環境を考慮した音量調整を行うことができる。

[0007] 特許文献6には、ステージ上の異なる場所に位置する複数演者に複数のアンプを配置し、ミキサで各アンプの音信号の音量を調整して供給する構成が開示されている。これにより、特許文献6のミキサは、複数のモニタ用スピーカの音量バランスを一括して調整することができる。

先行技術文献

特許文献

- [0008] 特許文献1：特開 2 0 0 5 － 1 2 8 2 9 6 号公報
特許文献2：国際公開第 2 0 1 8 ／ 2 1 4 0 2 号公報
特許文献3：特開 2 0 2 1 － 1 2 9 1 4 5 号公報
特許文献4：特開 2 0 1 0 － 1 0 3 8 5 3 号公報
特許文献5：特開 2 0 2 0 － 2 0 2 4 4 8 号公報
特許文献6：特開 2 0 0 9 － 1 0 0 1 8 5 号公報

発明の概要

発明が解決しようとする課題

- [0009] 先行技術文献に開示された構成は、いずれも仮想空間上の複数の演者の音量バランスを調整するものではない。
- [0010] 本発明の一実施形態は、仮想空間上の複数の演者の音量バランスを適切に調整することができる音処理方法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

- [0011] 本発明の一実施形態に係る音処理方法は、仮想空間内に複数の演者のオブジェクトと、前記複数の演者のオブジェクトに対応する複数の音量調整用インタフェースと、を配置し、前記複数の演者にそれぞれ対応する複数の音信号を受信し、利用者から、前記複数の音量調整用インタフェースに対応する、前記複数の演者に対するそれぞれの音量調整パラメータを受け付けて、前記複数の音信号の各演者に対応する音信号と、該音信号に対応する前記音量調整パラメータと、の関係を訓練された訓練済モデルを用いて、前記複数の演者に対するそれぞれの音量調整パラメータを求めて、前記訓練済モデルで求めた該音量調整パラメータに基づいて前記複数の音信号の音量を調整して混合する。

発明の効果

- [0012] 本発明の一実施形態によれば、仮想空間上の複数の演者の音量バランスを

適切に調整することができる。

図面の簡単な説明

- [0013] [図1]音処理システム1の構成を示すブロック図である。
- [図2]PC12Cの構成を示すブロック図である。
- [図3]ある仮想3次元空間R1の一例を示す斜視図である。
- [図4]訓練段階におけるPC12Cおよびサーバ30の動作を示すフローチャートである。
- [図5]実行段階におけるPC12Cの動作を示すフローチャートである。
- [図6]変形例1に係るPC12A（またはPC12B）の動作を示すフローチャートである。
- [図7]変形例2に係る仮想3次元空間R1の一例を示す斜視図である。
- [図8]変形例4に係る音処理システム1A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

- [0014] 図1は、音処理システム1の構成を示すブロック図である。図1に係る音処理システム1は、第1会場3に設置されたPC（パーソナルコンピュータ）12A、第2会場5に設置されたPC12B、第3会場7に設置されたPC12C、およびサーバ30を含む。PC12A、PC12B、PC12C、およびサーバ30は、ネットワーク9を介して接続される。PC12A、PC12B、およびPC12Cは、本発明の音処理装置の一例である。
- [0015] 第1会場3のPC12Aは、ギターアンプ11Aおよびモーションセンサ13Aに接続される。ギターアンプ11Aは、エレキギター10と接続される。
- [0016] エレキギター10は、音響機器の一例である。ギターアンプ11Aは、オーディオケーブルを介してエレキギター10に接続される。ギターアンプ11Aも音響機器の一例である。また、ギターアンプ11Aは、例えばUSBケーブルによりPC12Aに接続される。無論、ギターアンプ11Aは、無線通信によりPC12Aに接続してもよい。エレキギター10は、演奏音に係るアナログ音信号をギターアンプ11Aに出力する。

- [0017] ギターアンプ 11 A は、アナログオーディオ端子を有する。ギターアンプ 11 A は、オーディオケーブルを介してエレキギター 10 からアナログ音信号を受け付ける。ギターアンプ 11 A は、受け付けたアナログ音信号をデジタル音信号に変換する。ギターアンプ 11 A は、当該デジタル音信号にエフェクト等の各種の信号処理を施す。ギターアンプ 11 A は、信号処理後のデジタル音信号をアナログ音信号に変換する。ギターアンプ 11 A は、当該アナログ音信号を増幅する。ギターアンプ 11 A は、内蔵スピーカを介して増幅されたアナログ音信号に基づいて、エレキギター 10 の演奏音を出力する。また、ギターアンプ 11 A は、信号処理後のデジタル音信号を、PC 12 A に送信する。
- [0018] PC 12 A のユーザは、エレキギター 10 の演奏者である。エレキギター 10 の演奏者は、PC 12 A を用いて、自身の演奏音を配信するとともに仮想空間内で仮想的に演奏を行う自身の分身となる 3D モデルのオブジェクトを動作させる。ただし、演奏音の配信を行うユーザと演奏者とは同じ人物である必要はない。PC 12 A は、当該オブジェクトの動作を制御するためのモーションデータを制御する。
- [0019] モーションセンサ 13 A は、演奏者のモーションをキャプチャするためのセンサであり、例えば光学式、慣性式、あるいは画像式等のセンサである。モーションセンサ 13 A は、例えば USB ケーブルにより PC 12 A に接続される。PC 12 A は、モーションセンサ 13 A から受け付けたセンサ情報に基づいてモーションデータを制御する。無論、モーションセンサ 13 A は、無線通信により PC 12 A に接続してもよい。
- [0020] PC 12 A は、ギターアンプ 11 A から受信したギター演奏音に係るデジタル音信号、およびモーションセンサ 13 A のセンサ情報に基づいて制御したモーションデータをサーバ 30 に送信する。
- [0021] 第 2 会場 5 の PC 12 B は、マイク 19 およびモーションセンサ 13 B に接続される。
- [0022] マイク 19 は、音響機器の一例である。マイク 19 は、オーディオケーブ

ルあるいはUSBケーブル等を介してPC12Bに接続される。PC12Bは、オーディオケーブルを介してマイク19からアナログオーディオ信号を受信する。PC12Bは、受信したアナログ音信号をデジタル音信号に変換する。あるいは、マイク19は、USBケーブル等を介してデジタルオーディオ信号をPC12Bに出力してもよい。

[0023] PC12Bのユーザは、歌唱者である。歌唱者は、PC12Bを用いて、自身の歌唱音を配信するとともに仮想空間内で仮想的に歌唱を行う自身の分身となるオブジェクトを動作させる。PC12Bは、当該オブジェクトの動作を制御するためのモーションデータを制御する。ただし、歌唱音の配信を行うユーザと歌唱者とは同じ人物である必要はない。

[0024] モーションセンサ13Bは、歌唱者のモーションをキャプチャするためのセンサであり、例えば光学式、慣性式、あるいは画像式等のセンサである。モーションセンサ13Bは、例えばUSBケーブルによりPC12Bに接続される。PC12Bは、モーションセンサ13Bから受け付けたセンサ情報に基づいてモーションデータを制御する。無論、モーションセンサ13Bは、無線通信によりPC12Bに接続してもよい。

[0025] PC12Bは、マイク19から受信した歌唱音に係るデジタル音信号、およびモーションセンサ13Bのセンサ情報に基づいて制御したモーションデータをサーバ30に送信する。

[0026] 第3会場7のPC12Cは、ヘッドフォン20に接続される。ヘッドフォン20も音響機器の一例である。PC12Cのユーザは、仮想空間内で仮想的に行われる演者のパフォーマンスを視聴する視聴者である。

[0027] 図2は、PC12Cの構成を示すブロック図である。PC12Cは、汎用の情報処理装置である。図2では、PC12Cの構成を示すが、PC12AおよびPC12Bの主要構成も、図2に示す構成と同じである。

[0028] PC12Cは、通信部11、プロセッサ12、RAM13、フラッシュメモリ14、表示器15、ユーザI/F16、およびオーディオI/F17を備えている。

- [0029] 通信部 11 は、例えば B l u e t o o t h（登録商標）または W i - F i（登録商標）等の無線通信機能、U S B または L A N 等の有線通信機能を有する。
- [0030] 表示器 15 は、L C D や O L E D 等からなる。表示器 15 は、プロセッサ 12 の出力した映像を表示する。
- [0031] ユーザ I / F 16 は、操作部の一例である。ユーザ I / F 16 は、マウス、キーボード、あるいはタッチパネル等からなる。ユーザ I / F 16 は、利用者の操作を受け付ける。なお、タッチパネルは、表示器 15 に積層されていてもよい。
- [0032] オーディオ I / F 17 は、アナログオーディオ端子またはデジタルオーディオ端子等を有し、音響機器を接続するためのインタフェースである。本実施形態では、P C 12 C のオーディオ I / F 17 は、音響機器の一例としてヘッドフォン 20 を接続し、ヘッドフォン 20 に音信号を出力する。
- [0033] プロセッサ 12 は、C P U、D S P、または S o C（S y s t e m o n a C h i p）等からなる。プロセッサ 12 は、記憶媒体であるフラッシュメモリ 14 からプログラムを読み出し、R A M 13 に一時記憶することで、種々の動作を行う。なお、プログラムは、フラッシュメモリ 14 に記憶している必要はない。プロセッサ 12 は、例えば、サーバ等の他装置から必要な場合にダウンロードして R A M 13 に一時記憶してもよい。
- [0034] プロセッサ 12 は、通信部 11 を介して、サーバ 30 から音信号およびモーションデータを受信する。サーバ 30 から受信する音信号は、第 1 会場 3 の演奏者の演奏音に係る第 1 音信号および第 2 会場 5 の歌唱者の歌唱音に係る第 2 音信号を含む。サーバ 30 から受信するモーションデータは、第 1 会場 3 の演奏者のモーションおよび第 2 会場 5 の歌唱者のモーションを含む。また、プロセッサ 12 は、通信部 11 を介して、サーバ 30 から空間情報、モデルデータ、および位置情報等も受信する。
- [0035] 空間情報は、例えばライブハウスやコンサートホール等のライブ会場に対応する 3 次元空間の形状を示す情報であり、ある位置を原点とした 3 次元の

座標で表される。空間情報は、実在のコンサートホール等のライブ会場の3DCADデータに基づく座標情報であってもよいし、ある架空のライブ会場の論理的な座標情報（0～1で正規化された情報）であってもよい。

[0036] モデルデータは、3Dモデルのオブジェクトを構成するための3次元CG画像データであり、複数の画像パーツからなる。モデルデータは、演者毎に指定される。例えば第1会場3の演奏者は、自身の分身となるモデルデータを指定する。サーバ30は、指定されたモデルデータを配信する。

[0037] 位置情報は、3次元空間内におけるモデルデータの位置を示す情報である。位置情報は、上記仮想空間内の3次元の座標で表される。位置情報は、スピーカ等の機器の様に位置変化のないモデルデータに対応する位置情報である場合もあるし、演者の様に位置変化するモデルデータに対応する位置情報である場合もある。

[0038] 図3は、ある仮想3次元空間R1の一例を示す斜視図である。図3の仮想3次元空間R1は、一例として直方体形状の空間を示しているが、空間の形状はどのようなものであってもよい。

[0039] プロセッサ12は、サーバ30から受信した空間情報および位置情報に基づいて、図3に示す様な仮想3次元空間R1にオブジェクトを配置する。また、プロセッサ12は、仮想3次元空間R1内に、PC12Cの利用者の位置を設定する。PC12Cの利用者の位置は、仮想3次元空間R1内の視点位置50に対応する。図3では、仮想3次元空間R1を俯瞰して示すが、PC12Cは、プロセッサ12は、空間情報、モデルデータ、位置情報、およびオブジェクトのモーションデータ、および設定した視点位置の情報に基づいて、モデルデータをレンダリングして、設定した視点位置50から仮想3次元空間R1を見た映像を生成する。生成した映像は、表示器15を介して表示する。これにより、PC12Cの視聴者は、設定した視点位置50から仮想3次元空間R1を見た映像を視認することができる。PC12Cのユーザは、ユーザI/F16を介して、仮想3次元空間R1内の視点位置50を変更することができる。プロセッサ12は、変更された視点位置50から仮

想3次元空間R1を見た映像を生成する。これにより、PC12Cのユーザは、仮想3次元空間R1内で自身が移動しているように知覚することができる。

[0040] PC12Cのプロセッサ12は、サーバ30から受信した複数の音信号の音量を調整して混合し、例えばステレオ(L, R)チャンネルの音信号を生成する。この例では、プロセッサ12は、第1会場3の第1音信号および第2会場5の第2音信号を混合する。プロセッサ12は、オーディオI/F17を介してヘッドフォン20にステレオチャンネルの音信号を出力する。

[0041] なお、プロセッサ12は、第1音信号および第2音信号のそれぞれにイコライザやリバーブ処理等のエフェクト処理を行ってもよい。また、プロセッサ12は、第1音信号および第2音信号に対し、それぞれの対応するオブジェクトの位置に音が定位する様な定位処理を行ってもよい。

[0042] PC12Cは、各演者に対応する音信号と、該音信号に対応する音量調整パラメータと、の関係を訓練された訓練済モデルを用いて、複数の演者に対するそれぞれの音量調整パラメータを求めて、複数の音信号の音量を調整して混合する。

[0043] 図4は、訓練段階におけるPC12Cおよびサーバ30の動作を示すフローチャートである。サーバ30は、第1音信号および第2音信号を配信する(S21)。PC12Cのプロセッサ12は、サーバ30から第1音信号および第2音信号を受信する(S11)。

[0044] プロセッサ12は、仮想空間内に複数の演者のオブジェクトと、複数の演者のオブジェクトに対応する複数の音量調整用インタフェースと、を配置する(S12)。

[0045] 具体的には、プロセッサ12は、図3に示す様に、遠隔地である第1会場3に存在する演奏者31に対応する第1オブジェクト51を配置する。プロセッサ12は、別の遠隔地である第2会場5に存在する歌唱者32に対応する第2オブジェクト52を仮想3次元空間R1内に配置する。さらに、プロセッサ12は、第1オブジェクト51に対応する第1音量調整用インタフェ

ース 7 1 および第 2 オブジェクト 5 2 に対応する第 2 音量調整用インタフェース 7 2 を配置する。なお、本実施形態では、プロセッサ 1 2 は、第 1 会場 3 および第 2 会場 5 の 2 つの会場の演者に対応するオブジェクトおよび音量調整用インタフェースを配置しているが、会場の数は 2 つに限らない。プロセッサ 1 2 は、さらに多数の会場の演者のオブジェクトおよび音量調整用インタフェースを配置してもよい。

[0046] 次に、プロセッサ 1 2 は、P C 1 2 C のユーザから、複数の音量調整用インタフェースに対応する、複数の演者に対するそれぞれの音量調整パラメータを受け付ける (S 1 3)。P C 1 2 C のユーザは、図 3 に示した様に、仮想 3 次元空間 R 1 内に配置された第 1 音量調整用インタフェース 7 1 および第 2 音量調整用インタフェース 7 2 を操作して、音量調整操作を行う。P C 1 2 C のユーザは、例えば、第 1 オブジェクト 5 1 に対応する演奏音が大きすぎると感じた場合に、第 1 音量調整用インタフェース 7 1 を操作して音量を下げる操作を行う。図 3 の例では、第 1 音量調整用インタフェース 7 1 および第 2 音量調整用インタフェース 7 2 は、スライダの操作子になっている。したがって、P C 1 2 C のユーザは、第 1 オブジェクト 5 1 に対応する演奏音が大きすぎると感じた場合に、第 1 音量調整用インタフェース 7 1 を下方向に移動させる。また、P C 1 2 C のユーザは、例えば、第 2 オブジェクト 5 2 に対応する歌唱音が小さすぎると感じた場合に、第 2 音量調整用インタフェース 7 2 を上方向に移動させ、音量を上げる操作を行う。

[0047] P C 1 2 C は、受け付けた音量調整パラメータをサーバ 3 0 に送信する (S 1 4)。サーバ 3 0 は、P C 1 2 C から音量調整パラメータを受信する (S 2 2)。この例では、サーバ 3 0 は、P C 1 2 C から音量調整パラメータを受信しているが、他にも多数の情報処理装置から音量調整パラメータを受信する。サーバ 3 0 は、受信した多数の音量調整パラメータを用いて、所定のモデルに、所定のアルゴリズムを用いて配信した複数の演者に対応する音信号と音量調整パラメータとの関係を訓練させる (S 2 3)。

[0048] 本実施形態において、モデルを訓練させるためのアルゴリズムは限定され

ず、CNN (Convolutional Neural Network) やRNN (Recurrent Neural Network) 等の任意の機械訓練アルゴリズムを用いることができる。機械訓練アルゴリズムは、教師あり訓練、教師なし訓練、半教師訓練、強化訓練、逆強化訓練、能動訓練、あるいは転移訓練等であってもよい。また、サーバ30は、HMM (Hidden Markov Model: 隠れマルコフモデル) やSVM (Support Vector Machine) 等の機械訓練モデルを用いてモデルを訓練させてもよい。

[0049] 例えば、ある特定の演者（例えば第1会場3の演奏者31）の演奏音が、多数の視聴者で音量が大きすぎると感じられた場合、多数の視聴者が音量を下げる操作を行う。この場合、所定のモデルは、第1会場3の演奏音の音信号に対して、音量を下げる様な音量調整パラメータを出力するように訓練される。この様に、複数の演者のそれぞれの音信号と、それぞれの音量調整パラメータは、相関関係を有する。したがって、サーバ30は、所定のモデルに、複数の演者のそれぞれの音信号と、それぞれの音量調整パラメータと、の関係を訓練させ、訓練済モデルを生成することができる。

[0050] 図5は、実行段階におけるPC12Cの動作を示すフローチャートである。PC12Cのプロセッサ12は、サーバ30から第1音信号、第2音信号、および訓練済モデルを受信する(S31)。なお、訓練済モデルは、第1音信号および第2音信号とは別に事前に受信してもよい。

[0051] プロセッサ12は、仮想空間内に複数の演者のオブジェクトを配置する(S32)。具体的には、プロセッサ12は、第1オブジェクト51および第2オブジェクト52を仮想3次元空間R1内に配置する。この例では、実行段階においてプロセッサ12は第1音量調整用インタフェース71および第2音量調整用インタフェース72を配置しない。

[0052] プロセッサ12は、訓練済モデルを用いて、複数の演者に対するそれぞれの音量調整パラメータを求める(S33)。上述の様に、訓練済モデルは、複数の演者のそれぞれの音信号と、それぞれの音量調整パラメータと、の関

係を訓練されている。したがって、プロセッサ 12 は、訓練済モデルを用いて、第 1 オブジェクト 51 に対応する第 1 音信号および第 2 オブジェクト 52 に対応する第 2 音信号にそれぞれ対応する第 1 音量調整パラメータおよび第 2 音量調整パラメータを求める。

[0053] プロセッサ 12 は、訓練済モデルで求めた音量調整パラメータに基づいて複数の音信号の音量を調整して混合する (S34)。具体的には、プロセッサ 12 は、第 1 音信号の音量を第 1 音量調整パラメータで調整し、第 2 音信号の音量を第 2 音量調整パラメータで調整し、音量調整後の第 1 音信号および第 2 音信号を混合する。

[0054] この様に、PC12C は、複数の利用者から受け付けた音量調整パラメータで訓練された訓練済モデルを用いて複数の演者の音信号を適切な音量バランスに調整して混合することで、仮想 3 次元空間 R1 内で仮想的に歌唱または演奏を行う複数の演者の音量バランスを適切に調整することができる。これにより、仮想 3 次元空間 R1 で仮想的に行われる演者のパフォーマンスを視聴する視聴者は、音量バランスの調整操作を行う必要なく、仮想空間内における仮想的な演奏をより良い音量バランスで簡単に視聴できるという顧客体験を享受できる。

[0055] なお、この例では、実行段階においてプロセッサ 12 は、第 1 音量調整用インタフェース 71 および第 2 音量調整用インタフェース 72 を配置しなかったが、第 1 音量調整用インタフェース 71 および第 2 音量調整用インタフェース 72 を配置してもよい。この場合、PC12C のユーザは、プロセッサ 12 が訓練済モデルを用いて求めた第 1 音量調整パラメータおよび第 2 音量調整パラメータをさらに微調整することができる。また、PC12C は、微調整を行った音量調整パラメータもサーバ 30 に送信してもよい。サーバ 30 は、微調整を行った音量調整パラメータも受信して、訓練済モデルをさらに再訓練してもよい。これにより、仮想 3 次元空間 R1 内で行われるパフォーマンスの進行に伴って音量調整パラメータが更新される。そのため、パフォーマンスを視聴する視聴者は、仮想 3 次元空間 R1 内で行われるパフォ

ーマンスの進行に合わせて常に適切な音量バランスで視聴できるという顧客体験を享受できる。

[0056] (変形例1)

図6は、変形例1に係るPC12A（またはPC12B）の動作を示すフローチャートである。

[0057] 上記実施形態では、PC12Cがサーバ30から訓練済モデルを受信し、第1音信号の音量を第1音量調整パラメータで調整し、第2音信号の音量を第2音量調整パラメータで調整し、音量調整後の第1音信号および第2音信号を混合する例を示した。つまり、訓練済モデルで求めた音量調整パラメータは、受信した複数の音信号を混合する受信側機器で用いられる音量調整パラメータであった。

[0058] 変形例1の音処理システム1では、送信側の機器であるPC12AおよびPC12Bがそれぞれ訓練済モデルを受信し、第1音信号の音量を第1音量調整パラメータで調整し、第2音信号の音量を第2音量調整パラメータで調整する。

[0059] 具体的には、PC12Aは、まずサーバ30から訓練済モデルを受信する(S41)。PC12Aは、受信した訓練済モデルを用いて、送信する音信号の音量調整パラメータを求める(S42)。上述の様に、訓練済モデルは、複数の演者のそれぞれの音信号と、それぞれの音量調整パラメータと、の関係性を訓練されている。したがって、PC12Aは、訓練済モデルを用いて、第1音信号に対応する第1音量調整パラメータを求めることができる。PC12Aは、訓練済モデルで求めた第1音量調整パラメータに基づいて第1音信号の音量を調整する(S43)。PC12Aは、調整後の第1音信号をサーバ30に送信する(S44)。PC12Bも同様に、訓練済モデルに基づいて第2音信号の音量を第2音量調整パラメータで調整する。

[0060] つまり、変形例1において、訓練済モデルで求めた音量調整パラメータは、複数の演者でそれぞれ利用される複数の機器で用いられる音量調整パラメータであり、複数の機器は、それぞれ、音量調整パラメータに基づいて複数

の音信号の音量を調整し、受信側の機器は、複数の機器で音量を調整された後の複数の音信号を受信して混合する。

[0061] これにより、変形例 1 の音処理システム 1 でも、仮想 3 次元空間 R 1 で仮想的に行われる演者のパフォーマンスを視聴する視聴者は、音量バランスの調整操作を行う必要なく、仮想空間内における仮想的な演奏をより良い音量バランスで簡単に視聴できるという顧客体験を享受できる。

[0062] なお、変形例 1 では、P C 1 2 A（または P C 1 2 B）が音信号の音量を調整する例を示したが、例えばギターアンプ 1 1 A が訓練済モデルに基づいて音信号の音量を調整してもよいし、エレキギター 1 0 が訓練済モデルに基づいて音信号の音量を調整してもよい。あるいは、P C 1 2 A が訓練済モデルに基づいてギターアンプ 1 1 A における音量調整パラメータを求めて、該音量調整パラメータをギターアンプ 1 1 A に入力し、ギターアンプ 1 1 A が音信号の音量を調整してもよい。あるいは、P C 1 2 A が訓練済モデルに基づいてエレキギター 1 0 における音量調整パラメータを求めて、該音量調整パラメータをエレキギター 1 0 に入力し、エレキギター 1 0 が音信号の音量を調整してもよい。

[0063] （変形例 2）

訓練済モデルは、音量調整パラメータだけでなく、複数の音信号の各演者に対応する音信号と、該音信号に施すエフェクト処理のエフェクトパラメータとの関係を訓練されてもよい。

[0064] 図 7 は、変形例 2 に係る仮想 3 次元空間 R 1 の一例を示す斜視図である。図 3 と共通する構成については同一の符号を付し、説明を省略する。

[0065] P C 1 2 C のプロセッサ 1 2 は、訓練段階として、仮想空間内に複数の演者のオブジェクトと、複数の演者のオブジェクトに対応する複数のエフェクト調整用インタフェースと、を配置する。具体的には、プロセッサ 1 2 は、図 7 に示す様に、第 1 オブジェクト 5 1 に対応する第 1 エフェクト調整用インタフェース 7 1 A および第 2 オブジェクト 5 2 に対応する第 2 エフェクト調整用インタフェース 7 2 A と、を配置する。

- [0066] この例では、第1エフェクト調整用インタフェース71Aおよび第2エフェクト調整用インタフェース72Aは、それぞれイコライザのエフェクトパラメータを調整するための操作子である。第1エフェクト調整用インタフェース71Aおよび第2エフェクト調整用インタフェース72Aは、それぞれ高音域（High）、中音域（Mid）、および低音域（Low）のレベルを調整する操作子を含む。
- [0067] PC12Cのユーザは、第1エフェクト調整用インタフェース71Aおよび第2エフェクト調整用インタフェース72Aを操作して、エフェクトパラメータの調整操作を行う。
- [0068] PC12Cは、受け付けたエフェクトパラメータをサーバ30に送信する。サーバ30は、PC12Cを含む多数の情報処理装置からエフェクトパラメータを受信する。サーバ30は、受信した多数のエフェクトパラメータを用いて、所定のモデルに、所定のアルゴリズムを用いて配信した複数の演者に対応する音信号とエフェクトパラメータとの関係を訓練させる。
- [0069] 実行段階において、PC12Cのプロセッサ12は、サーバ30から第1音信号、第2音信号、および訓練済モデルを受信する。プロセッサ12は、訓練済モデルを用いて、複数の演者の音信号に施すそれぞれのエフェクトパラメータを求める。プロセッサ12は、訓練済モデルで求めたエフェクトパラメータに基づいて複数の音信号にエフェクト処理を施す。また、プロセッサ12は、エフェクト処理後の複数の音信号の音量を調整して混合する。
- [0070] この様に、変形例2のPC12Cは、複数の利用者から受け付けたエフェクトパラメータで訓練された訓練済モデルを用いて複数の演者の音信号に適切なエフェクト処理を施して混合することで、仮想3次元空間R1内で仮想的に歌唱または演奏を行う複数の演者の音質を適切に調整することができる。これにより、仮想3次元空間R1で仮想的に行われる演者のパフォーマンスを視聴する視聴者は、エフェクトパラメータの調整操作を行う必要なく、仮想空間内における仮想的な演奏をより良い音質で簡単に視聴できるという顧客体験を享受できる。

[0071] なお、エフェクトは、上記の例で示したイコライザに限らない。エフェクトは、コンプレッサ、あるいはリバーブ等その他のエフェクトであってもよい。例えば、P C 1 2 Cのユーザは、第1会場3の演奏音に響きがないと感じた場合に、第1会場3の演奏音に強いリバーブ処理をかけるように、エフェクトパラメータを調整する。サーバ30は、P C 1 2 Cを含む多数の情報処理装置からエフェクトパラメータを受信し、第1会場3の演奏音に強いリバーブ処理をかける様な訓練済モデルを生成する。これにより、第1会場3の演奏音には、自動的に強いリバーブ処理が施されるため、視聴者は、改めて第1会場3の演奏音に強いリバーブ処理をかけるエフェクトパラメータを調整する必要なく、仮想空間内における仮想的な演奏をより良い音質で簡単に視聴できるという顧客体験を享受できる。

[0072] なお、エフェクト処理は、受信側のP C 1 2 Cではなく、送信側のP C 1 2 A、P C 1 2 B、あるいはギターアンプ11Aやエレキギター10、マイク19等で行ってもよい。この場合、送信側のP C 1 2 AおよびP C 1 2 Bが、訓練済モデルをサーバ30から受信して、該訓練済モデルに基づいてエフェクトパラメータを求めて、エフェクト処理を施す。また、ギターアンプ11Aが訓練済モデルに基づいてエフェクトパラメータを求めてエフェクト処理を行ってもよいし、エレキギター10が訓練済モデルに基づいてエフェクトパラメータを求めてエフェクト処理を行ってもよい。あるいは、P C 1 2 Aが訓練済モデルに基づいてギターアンプ11Aにおけるエフェクト処理のエフェクトパラメータを求めて、該エフェクトパラメータをギターアンプ11Aに入力し、ギターアンプ11Aが入力したエフェクトパラメータに基づいてエフェクト処理を行ってもよい。あるいは、P C 1 2 Aが訓練済モデルに基づいてエレキギター10におけるエフェクト処理のエフェクトパラメータを求めて、該エフェクトパラメータをエレキギター10に入力し、エレキギター10がエフェクト処理を行ってもよい。

[0073] (変形例3)

変形例3に係る音処理システム1は、複数の演者でそれぞれ利用される複

数の音響機器の情報を取得し、取得した複数の音響機器の情報に基づいて、複数の音信号に施すエフェクト処理のエフェクトパラメータを調整する。

[0074] 例えば、P C 1 2 Aは、エレキギター 1 0およびギターアンプ 1 1 Aの情報をサーバ 3 0に送信する。エレキギター 1 0およびギターアンプ 1 1 Aの情報とは、例えばエレキギター 1 0およびギターアンプ 1 1 Aのそれぞれの機種名、あるいは製造番号等の情報を含む。同様に、P C 1 2 Bは、マイク 1 9の情報をサーバ 3 0に送信し、P C 1 2 Cは、ヘッドフォン 2 0の情報をサーバ 3 0に送信する。

[0075] サーバ 3 0は、複数の音響機器の情報とそれぞれの音響機器に対応する適切なイコライザ等のエフェクトパラメータをテーブルとして記憶している。サーバ 3 0は、P C 1 2 A、P C 1 2 B、またはP C 1 2 Cから受信した音響機器の情報に対応するエフェクトパラメータをテーブルから読み出して、読み出したエフェクトパラメータをP C 1 2 A、P C 1 2 B、またはP C 1 2 Cに送信する。

[0076] P C 1 2 A、P C 1 2 B、またはP C 1 2 Cは、サーバ 3 0からエフェクトパラメータを受信して、対応する音響機器の音信号に施すエフェクト処理のエフェクトパラメータを調整する。例えば、P C 1 2 Cは、サーバ 3 0から受信したエフェクトパラメータに基づいて、ヘッドフォン 2 0に出力する音信号のイコライザのパラメータを調整する。

[0077] これにより、各会場の利用者は、利用する音響機器のイコライザ等のエフェクトパラメータを手動で調整する必要なく、適切なエフェクトパラメータに簡単に調整できるという顧客体験を享受できる。例えば、ある演者がある音響機器（例えばあるマイク）を用いて歌唱音を配信する場合と、別のある音響機器（別のあるマイク）を用いて歌唱音を配信する場合と、で音質が異なる場合がある。この様に、音響機器の違いによる収録環境の差により、配信される歌唱音の音質が異なる場合がある。しかし、変形例 3 の音処理システム 1 では、このような音響機器の違いによる収録環境の差を補正することができる。

[0078] なお、サーバは、複数の音響機器の情報とそれぞれの音響機器に対応する適切なエフェクトパラメータとの関係を訓練した訓練済モデルを用いて、対応する音響機器のエフェクトパラメータを求めてもよい。

[0079] (変形例4)

図8は、変形例4に係る音処理システム1Aの構成を示すブロック図である。図1と同じ構成については同じ符号を付し、説明を省略する。音処理システム1Aは、第1会場3の演者および第2会場5の演者が、互いに演奏音または歌唱音に係る音信号を送信し、リモートセッションを行う。

[0080] PC12Aは、第2会場5の演者の歌唱音に係る音信号を受信し、音量を調整してヘッドフォン20Aに出力する。第1会場3の演者は、ヘッドフォン20Aを介して第2会場5の演者の歌唱音を聴く。また、第1会場3の演者は、PC12Aを用いて第2会場5の演者の歌唱音の音量を調整して、当該歌唱音に合わせた演奏を行う。

[0081] PC12Bは、第1会場3の演者の演奏音に係る音信号を受信し、音量を調整してヘッドフォン20Bに出力する。第2会場5の演者は、ヘッドフォン20Bを介して第1会場3の演者の演奏音を聴く。また、第2会場5の演者は、PC12Bを用いて第1会場3の演者の演奏音の音量を調整して、当該演奏音に合わせた演奏を行う。

[0082] サーバ30は、PC12AおよびPC12Bでそれぞれ調整された音量調整パラメータを受信し、所定のモデルを訓練する。これにより、サーバ30は、当該バンド用に訓練した訓練済モデルを生成することができる。バンドメンバーは、次にリモートセッションを行う場合、情報処理装置を用いてサーバ30から訓練済モデルを受信し、訓練済モデルを用いて音量調整を行う。

[0083] これにより、PC12AおよびPC12Bの演者は、音量の調整操作を行う必要なく、過去に調整したより良い音量でリモートセッションを行うことができるという顧客体験を享受できる。

[0084] なお、上記の音処理システム1Aは、第1会場3および第2会場5でリモ

ートセッションを行う例である。しかし、音処理システム 1 A は、さらに多数の会場で演奏音または歌唱音に係る音信号を送受信し、それぞれの演者が他の演者に合わせて演奏を行う、リモート合奏を行うこともできる。

[0085] (変形例 5)

変形例 5 の P C 1 2 C は、複数の演者のオブジェクトの第 1 位置情報と、視聴者の第 2 位置情報とに基づいて受信した複数の音信号の音量を調整して混合する。

[0086] 例えば、P C 1 2 C のプロセッサ 1 2 は、仮想 3 次元空間 R 1 内のユーザの視点位置 5 0 と第 1 オブジェクト 5 1 の距離、および視点位置 5 0 と第 2 オブジェクト 5 2 の距離に基づいて、第 1 音信号および第 2 音信号の音量を調整する。P C 1 2 C のプロセッサ 1 2 は、視点位置 5 0 との距離の近いオブジェクトに対応する音信号の音量を大きくし、視点位置 5 0 との距離の遠いオブジェクトに対応する音信号の音量を小さくする。

[0087] これにより、仮想 3 次元空間 R 1 で仮想的に行われる演者のパフォーマンスを視聴する視聴者は、仮想 3 次元空間 R 1 における自身と演者との距離感を認識しながら視聴できるという顧客体験を享受できる。

[0088] (変形例 6)

変形例 6 の P C 1 2 C は、視点位置 5 0、第 1 オブジェクト 5 1 の位置、および第 2 オブジェクト 5 2 の位置に基づいて視点位置 5 0 を受聴点とした音響処理を施す。視点位置 5 0 を受聴点とした音響処理とは、例えば定位処理である。

[0089] P C 1 2 C のプロセッサ 1 2 は、例えば、視点位置 5 0 から見て第 1 オブジェクト 5 1 および第 2 オブジェクト 5 2 の位置に第 1 オブジェクト 5 1 および第 2 オブジェクト 5 2 の音が定位する様な定位処理を行う。

[0090] プロセッサ 1 2 は、例えば H R T F (H e a d R e l a t e d T r a n s f e r F u n c t i o n) に基づく定位処理を行う。H R T F は、ある仮想の音源位置から利用者の右耳および左耳に至る伝達関数を表す。例えば、図 3 に示す様に、第 1 オブジェクト 5 1 の位置は、視点位置 5 0 から見

て前方左側である。プロセッサ 12 は、第 1 オブジェクト 51 に対応する音信号に、ユーザの前方左側の位置に定位する様な H R T F を畳み込むバイノーラル処理を行う。これにより、P C 1 2 C のユーザは、仮想 3 次元空間 R 1 内の視点位置 50 に居て、自身の前方左側の第 1 オブジェクト 51 の音を聴いている様に知覚することができる。

[0091] (変形例 7)

上記実施形態では、いずれも訓練段階において、サーバ 30 が多数の情報処理装置から音量調整パラメータを受信し、所定のモデルを訓練する例を示した。しかし、サーバ 30 は、ある 1 つの情報処理装置から音量調整パラメータを受信し、所定のモデルを訓練してもよい。例えば熟練のオペレータが音量バランスの調整操作を行った場合、サーバ 30 は、当該熟練のオペレータの音量調整操作を訓練した訓練済モデルを生成し、当該訓練済モデルを配信する。これにより、当該訓練済モデルは、多数の情報処理装置で共有される。他の情報処理装置は、配信された訓練済モデルを用いて音量調整を行う。

[0092] これにより、仮想 3 次元空間 R 1 で仮想的に行われる演者のパフォーマンスを視聴する視聴者は、音量バランスの調整操作を行う必要なく、熟練のオペレータにより調整されたより良い音量バランスで仮想的な演奏をより簡単に視聴できるという顧客体験を享受できる。

[0093] また、サーバ 30 は、あるバンドのメンバーでリモートセッションを行う場合に、当該バンドのメンバーで調整された音量調整パラメータを受信し、所定のモデルを訓練してもよい。これにより、サーバ 30 は、当該バンド用に訓練した訓練済モデルを生成することができる。バンドメンバーは、次にリモートセッションを行う場合、情報処理装置を用いてサーバ 30 から訓練済モデルを受信し、訓練済モデルを用いて音量調整を行う。

[0094] これにより、リモートセッションを行うバンドのメンバーは、音量バランスの調整操作を行う必要なく、過去に調整されたより良い音量バランスでリモートセッションを行うことができるという顧客体験を享受できる。

[0095] なお、サーバ30は、ある1つの情報処理装置から受信した音量調整パラメータで訓練した訓練済モデルを再訓練してもよい。サーバ30は、当該1つの情報処理装置から再度、音量調整パラメータを受信して訓練済モデルを再訓練してもよいし、別の情報処理装置から音量調整パラメータを受信して、訓練済モデルを再訓練してもよい。

[0096] (その他の例)

ユーザは、スライダ等の操作子ではなく、「第1演者の音量を大きくする」等の音声入力により音量調整パラメータを入力してもよい。

[0097] サーバ30を含む音処理システムの運営者は、仮想3次元空間R1内のパフォーマンスの環境を提供するとともに、訓練済モデルを販売してもよい。例えば、サーバ30は、特定の訓練済モデルに課金処理を行い、課金確認後に訓練済モデルをダウンロードするように構成する。課金処理は、サーバ30ではなく、別の課金用のサーバで行ってもよい。サーバ30は、あるユーザに所定の金額を課金した後、例えばある熟練のオペレータの音量調整操作により訓練された訓練済モデルを当該ユーザにダウンロードさせる。この場合、サーバ30は、当該熟練のオペレータに、訓練済モデルがダウンロードされる毎に報酬を支払う支払い処理を行ってもよい。この様に、サーバ30を含む音処理システムの運営者は、オペレータにインセンティブを与えてもよい。これにより、オペレータは、自身の音量調整の技術を販売することができる。したがって、運営者は、多数のユーザに対して音処理システムを利用するモチベーションを高めることができる。

[0098] なお、サーバ30は、複数のオペレータの音量調整パラメータで訓練した複数の訓練済モデルを蓄積してもよい。サーバ30は、複数の訓練済モデルのうち、視聴者から指定された任意の訓練済モデルを該視聴者の利用する情報処理装置にダウンロードさせる。この場合、サーバ30は、ダウンロードされた訓練済モデルを訓練したオペレータに対して報酬を支払う処理を行ってもよい。これにより、運営者は、多数のオペレータに対して音処理システムを利用するモチベーションを高めることができる。

[0099] なお、課金処理は、ダウンロード毎の課金ではなく、1ヶ月あるいは1年単位の課金（サブスクリプション）であってもよい。

[0100] 本実施形態の説明は、すべての点で例示であって、制限的なものではないと考えられるべきである。本発明の範囲は、上述の実施形態ではなく、請求の範囲によって示される。さらに、本発明の範囲は、請求の範囲と均等の範囲を含む。

符号の説明

[0101] 1, 1A : 音処理システム、3 : 第1会場、5 : 第2会場、7 : 第3会場、9 : ネットワーク、10 : エレキギター、11 : 通信部、11A : ギターアンプ、12 : プロセッサ、13 : RAM、13A, 13B : モーションセンサ、14 : フラッシュメモリ、15 : 表示器、16 : ユーザI/F、17 : オーディオI/F、19 : マイク、20, 20A, 20B : ヘッドフォン、30 : サーバ、31 : 演奏者、32 : 歌唱者、50 : 視点位置、51 : 第1オブジェクト、52 : 第2オブジェクト、71 : 第1音量調整用インタフェース、71A : 第1エフェクト調整用インタフェース、72 : 第2音量調整用インタフェース、72A : 第2エフェクト調整用インタフェース

請求の範囲

- [請求項1] 仮想空間内に複数の演者のオブジェクトと、前記複数の演者のオブジェクトに対応する複数の音量調整用インタフェースと、を配置し、
前記複数の演者にそれぞれ対応する複数の音信号を受信し、
利用者から、前記複数の音量調整用インタフェースに対応する、前記複数の演者に対するそれぞれの音量調整パラメータを受け付けて、
前記複数の音信号の各演者に対応する音信号と、該音信号に対応する前記音量調整パラメータと、の関係を訓練された訓練済モデルを用いて、前記複数の演者に対するそれぞれの音量調整パラメータを求めて、
前記訓練済モデルで求めた該音量調整パラメータに基づいて前記複数の音信号の音量を調整して混合する、
音処理方法。
- [請求項2] 前記訓練済モデルは、前記複数の音信号の各演者に対応する音信号と、該音信号に施すエフェクト処理のエフェクトパラメータと、の関係を訓練され、
前記訓練済モデルを用いて、前記複数の演者に対するそれぞれのエフェクトパラメータを求めて、
前記訓練済モデルで求めた該エフェクトパラメータに基づいて前記複数の音信号にエフェクト処理を施す、
請求項 1 に記載の音処理方法。
- [請求項3] 前記複数の演者でそれぞれ利用される複数の音響機器の情報を受信し、
受信した前記複数の音響機器の情報に基づいて前記複数の演者に対するそれぞれのエフェクトパラメータを求める、
請求項 2 に記載の音処理方法。
- [請求項4] 前記複数の演者でそれぞれ利用される複数の音響機器の情報を取得し、

取得した前記複数の音響機器の情報に基づいて、前記複数の音信号の音量を調整する、

請求項 1 乃至請求項 3 のいずれか 1 項に記載の音処理方法。

[請求項5] 前記複数の演者のオブジェクトの第 1 位置情報と、視聴者の第 2 位置情報と、を取得し、

前記第 1 位置情報および前記第 2 位置情報に基づいて前記複数の音信号の音量を調整する、

請求項 1 乃至請求項 3 のいずれか 1 項に記載の音処理方法。

[請求項6] 前記複数の利用者は、演者を含む、

請求項 1 乃至請求項 3 のいずれか 1 項に記載の音処理方法。

[請求項7] 前記訓練済モデルで求めた前記音量調整パラメータは、受信した前記複数の音信号を混合する受信側機器で用いられる音量調整パラメータである、

請求項 1 乃至請求項 3 のいずれか 1 項に記載の音処理方法。

[請求項8] 前記訓練済モデルで求めた前記音量調整パラメータは、前記複数の演者でそれぞれ利用される複数の機器で用いられる音量調整パラメータであり、

前記複数の機器は、それぞれ、前記音量調整パラメータに基づいて前記複数の音信号の音量を調整し、

受信側の機器は、前記複数の機器で音量を調整された後の前記複数の音信号を受信して混合する、

請求項 1 乃至請求項 3 のいずれか 1 項に記載の音処理方法。

[請求項9] 前記複数の演者にそれぞれ対応する複数の音信号を、ネットワークを介して受信する、

請求項 1 乃至請求項 3 のいずれか 1 項に記載の音処理方法。

[請求項10] 第 1 の利用者の第 1 の情報処理装置から、前記音量調整パラメータを受け付けて所定のモデルを訓練して前記訓練済モデルを生成し、

前記訓練済モデルを第 2 の利用者の第 2 の情報処理装置に送信し、

前記第2の情報処理装置が、前記訓練済モデルを用いて、前記複数の演者に対するそれぞれの音量調整パラメータを求める、

請求項1乃至請求項3のいずれか1項に記載の音処理方法。

[請求項11] 前記第1の情報処理装置または前記第2の情報処理装置から前記音量調整パラメータを受け付けて、前記訓練済モデルを再訓練する、
請求項10に記載の音処理方法。

[請求項12] サーバが、前記第2の利用者に対して課金処理を行い、前記第1の利用者に対して報酬の支払い処理を行う、
請求項10に記載の音処理方法。

[請求項13] 仮想空間内に複数の演者のオブジェクトと、前記複数の演者のオブジェクトに対応する複数の音量調整用インタフェースと、を配置し、
前記複数の演者にそれぞれ対応する複数の音信号を受信し、
利用者から、前記複数の音量調整用インタフェースに対応する、前記複数の演者に対するそれぞれの音量調整パラメータを受け付けて、
前記複数の音信号の各演者に対応する音信号と、該音信号に対応する前記音量調整パラメータと、の関係を訓練された訓練済モデルを用いて、前記複数の演者に対するそれぞれの音量調整パラメータを求めて、
前記訓練済モデルで求めた該音量調整パラメータに基づいて前記複数の音信号の音量を調整して混合する、
プロセッサを備えた、
音処理装置。

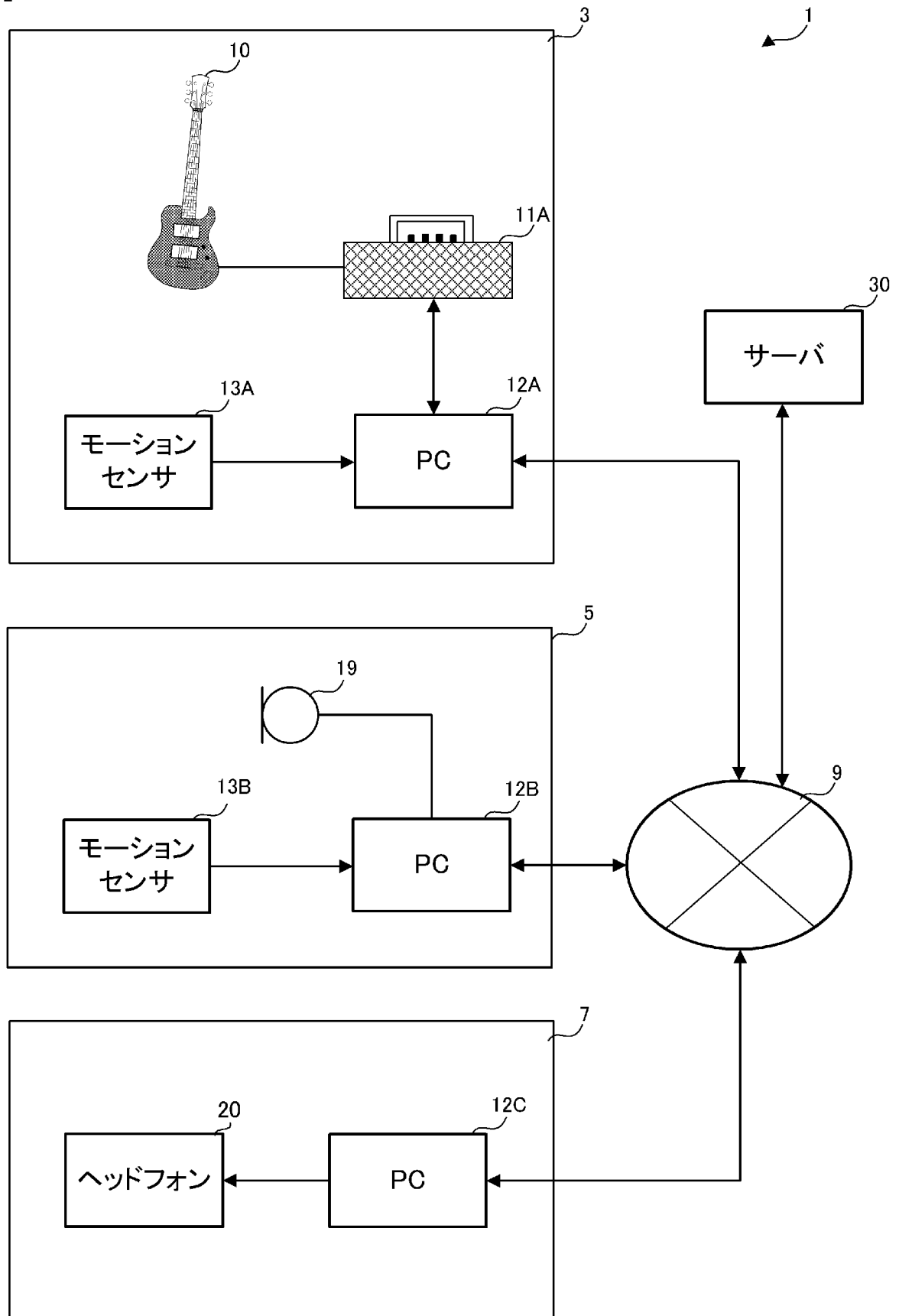
[請求項14] 仮想空間内に複数の演者のオブジェクトと、前記複数の演者のオブジェクトに対応する複数の音量調整用インタフェースと、を配置し、
前記複数の演者にそれぞれ対応する複数の音信号を受信し、
利用者から、前記複数の音量調整用インタフェースに対応する、前記複数の演者に対するそれぞれの音量調整パラメータを受け付けて、
前記複数の音信号の各演者に対応する音信号と、該音信号に対応す

る前記音量調整パラメータと、の関係を訓練された訓練済モデルを用いて、前記複数の演者に対するそれぞれの音量調整パラメータを求めて、

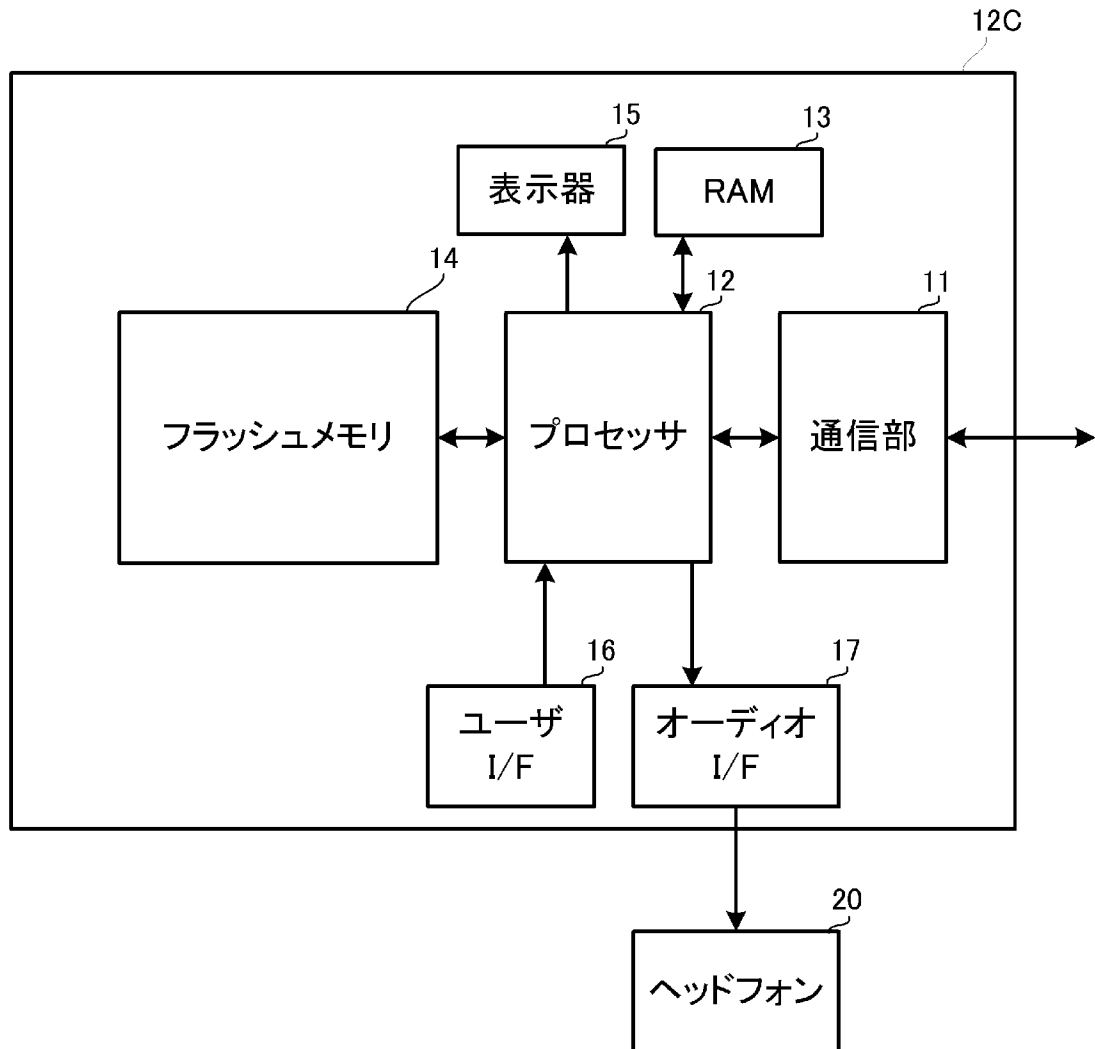
前記訓練済モデルで求めた該音量調整パラメータに基づいて前記複数の音信号の音量を調整して混合する、

音処理を情報処理装置に実行させるプログラム。

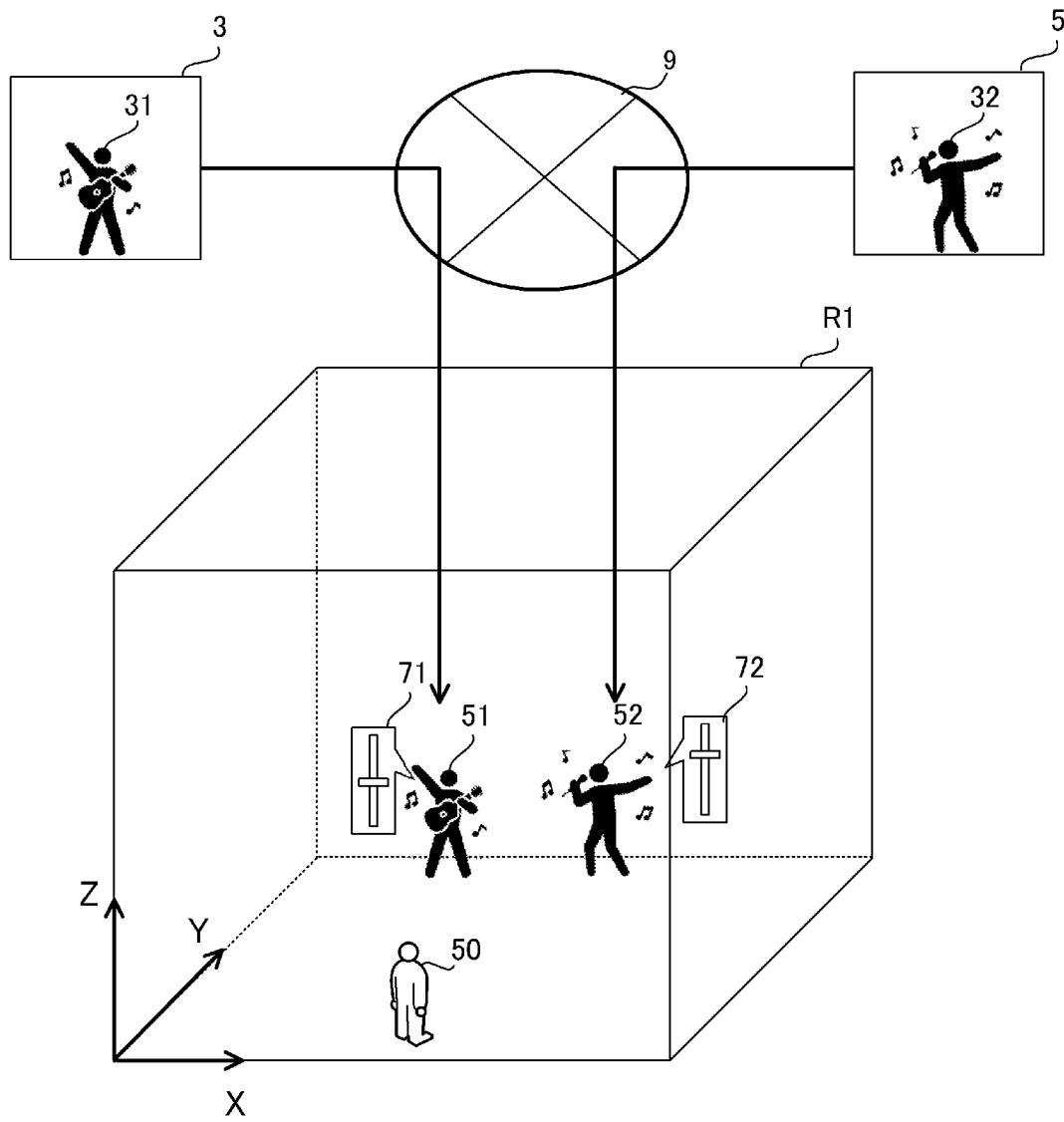
[図1]



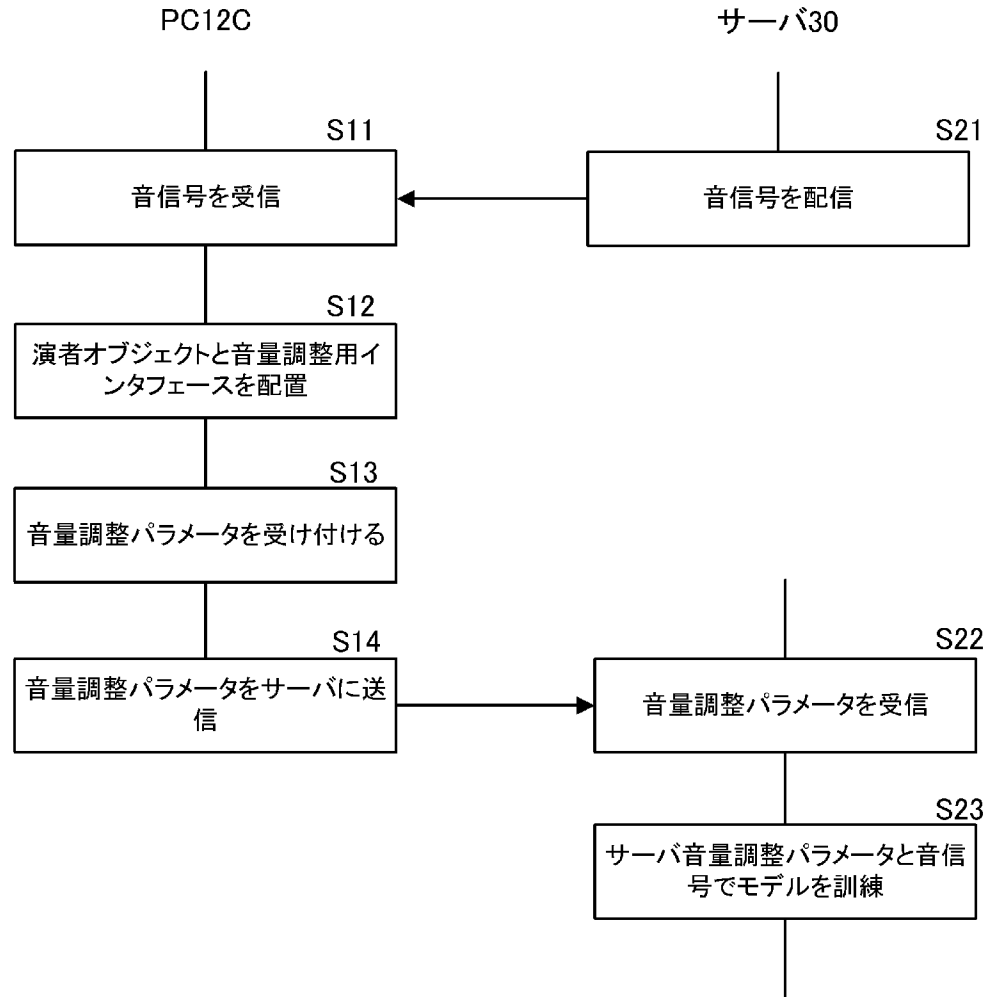
[図2]



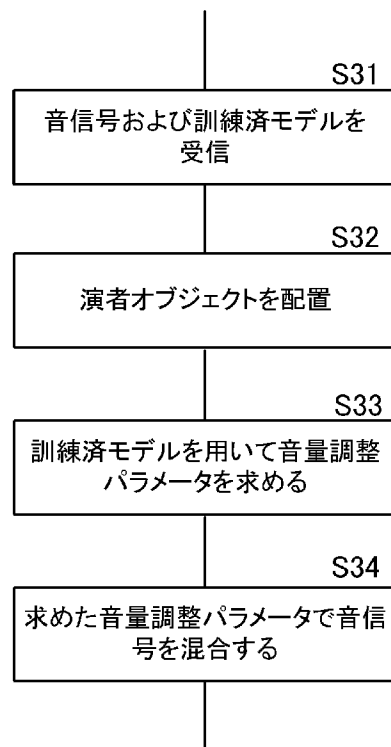
[図3]



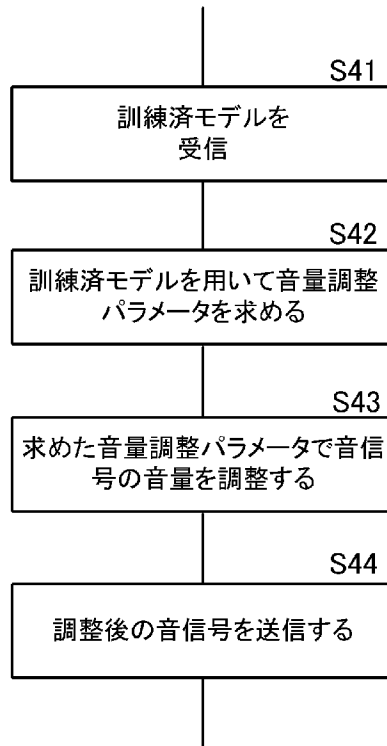
[図4]



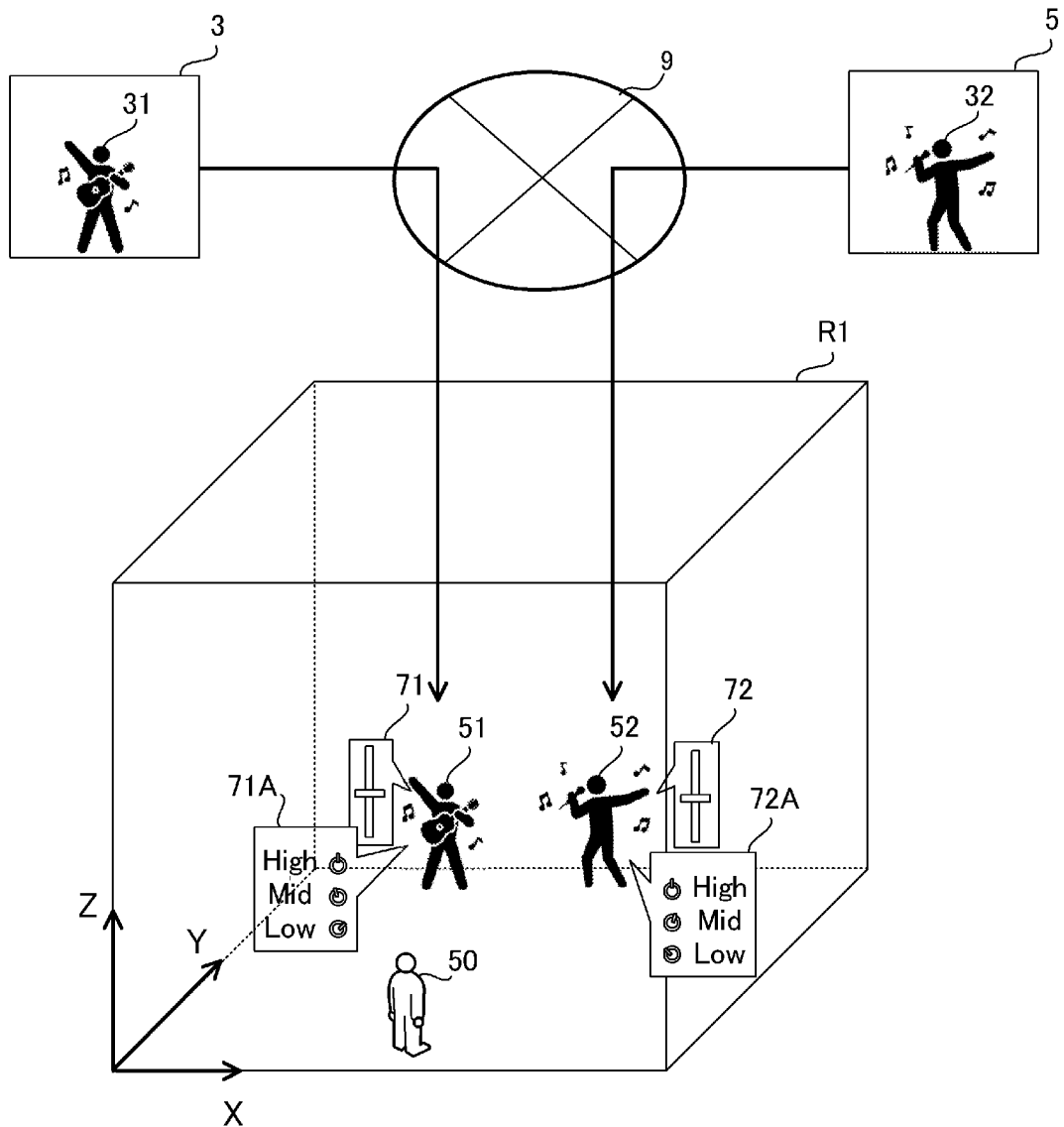
[図5]



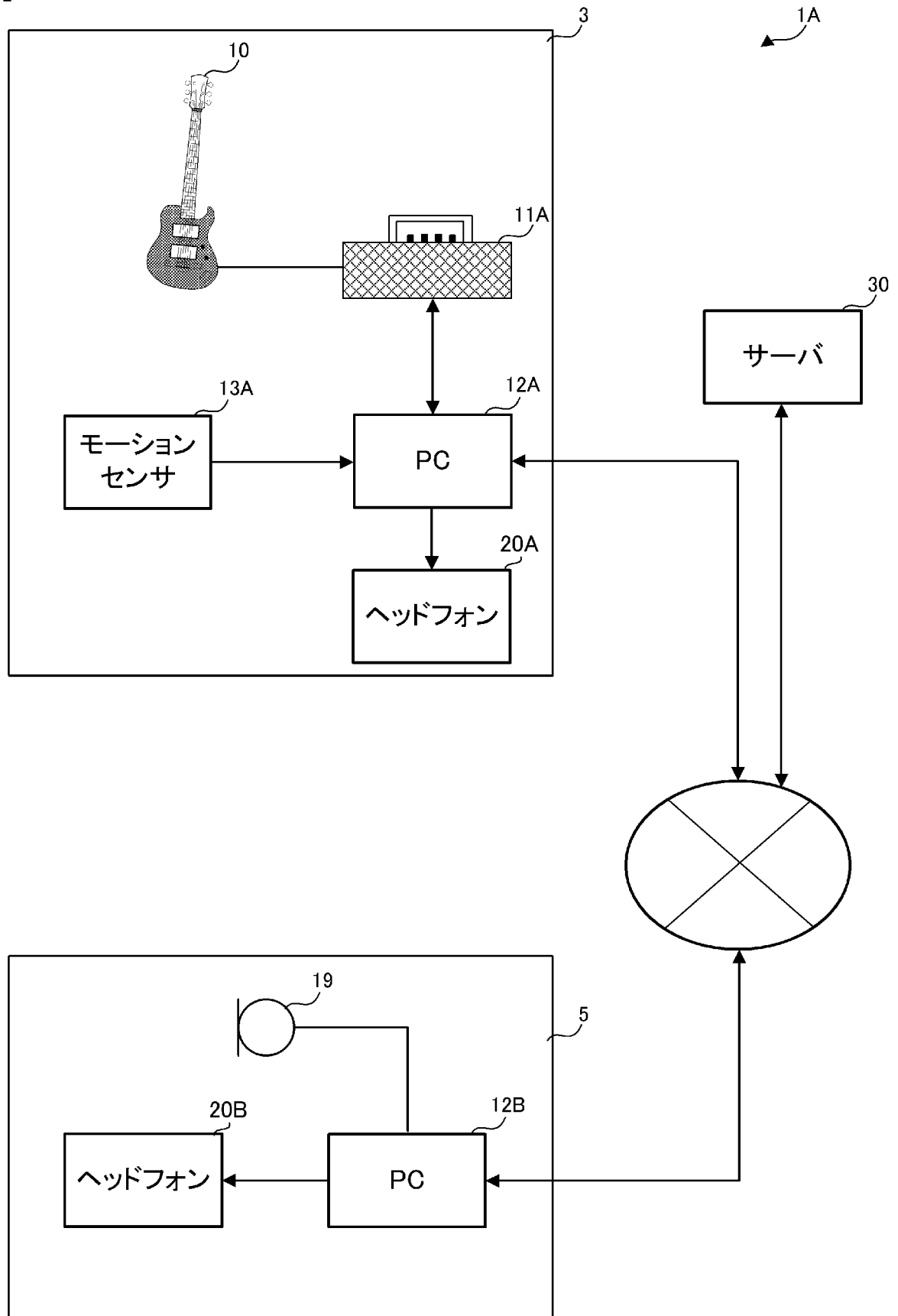
[図6]



[図7]



[図8]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2023/021288

A. CLASSIFICATION OF SUBJECT MATTER**H04R 3/00**(2006.01)i

FI: H04R3/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04R3/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996

Published unexamined utility model applications of Japan 1971-2023

Registered utility model specifications of Japan 1996-2023

Published registered utility model applications of Japan 1994-2023

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2018/055860 A1 (SONY CORP) 29 March 2018 (2018-03-29) paragraphs [0017]-[0055], [0091]-[0093], fig. 1-4, 11	1-14
Y	WO 2021/192072 A1 (YAMAHA CORP) 30 September 2021 (2021-09-30) paragraphs [0035]-[0038]	1-14



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

04 September 2023

Date of mailing of the international search report

12 September 2023

Name and mailing address of the ISA/JP

Japan Patent Office (ISA/JP)

3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915

Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2023/021288

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
WO	2018/055860	A1	29 March 2018	US 2019/0174247 A1 paragraphs [0041]-[0079], [0116]-[0121], fig. 1-4, 11 CN 109716794 A	
WO	2021/192072	A1	30 September 2021	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） H04R 3/00(2006.01)i FI: H04R3/00		
B. 調査を行った分野 調査を行った最小限資料（国際特許分類（IPC）） H04R3/00 最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2023年 日本国実用新案登録公報 1996-2023年 日本国登録実用新案公報 1994-2023年 国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y	WO 2018/055860 A1（ソニー株式会社）29.03.2018（2018-03-29） 段落0017-0055、0091-0093、図1-4、11	1-14
Y	WO 2021/192072 A1（ヤマハ株式会社）30.09.2021（2021-09-30） 段落0035-0038	1-14
☐ C欄の続きにも文献が列挙されている。 ☒ パテントファミリーに関する別紙を参照。		
* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的技術水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献 “T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献		
国際調査を完了した日 04.09.2023	国際調査報告の発送日 12.09.2023	
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 佐久 聖子 5Z 3142 電話番号 03-3581-1101 内線 3591	

国際調査報告
パテントファミリーに関する情報

国際出願番号

PCT/JP2023/021288

引用文献			公表日	パテントファミリー文献	公表日
WO	2018/055860	A1	29.03.2018	US 2019/0174247 A1 [0041]-[0079], [0116]- [0121], FIGS. 1-4, 11 CN 109716794 A	
WO	2021/192072	A1	30.09.2021	(ファミリーなし)	