

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7611614号  
(P7611614)

(45)発行日 令和7年1月10日(2025.1.10)

(24)登録日 令和6年12月26日(2024.12.26)

(51)国際特許分類 F I  
H 0 4 S 7/00 (2006.01) H 0 4 S 7/00 3 4 0  
H 0 4 R 3/00 (2006.01) H 0 4 R 3/00 3 1 0

請求項の数 16 (全30頁)

(21)出願番号	特願2023-516316(P2023-516316)	(73)特許権者	503359821 国立研究開発法人理化学研究所 埼玉県和光市広沢2番1号
(86)(22)出願日	令和4年2月28日(2022.2.28)	(74)代理人	100095407 弁理士 木村 満
(86)国際出願番号	PCT/JP2022/008277	(74)代理人	100110135 弁理士 石井 裕一郎
(87)国際公開番号	WO2022/224586	(74)代理人	100132883 弁理士 森川 泰司
(87)国際公開日	令和4年10月27日(2022.10.27)	(72)発明者	浜中 雅俊 埼玉県和光市広沢2番1号 国立研究開 発法人理化学研究所内
審査請求日	令和5年10月12日(2023.10.12)	審査官	山下 剛史
(31)優先権主張番号	特願2021-70745(P2021-70745)		
(32)優先日	令和3年4月20日(2021.4.20)		
(33)優先権主張国・地域又は機関	日本国(JP)		

最終頁に続く

(54)【発明の名称】 情報処理装置、情報処理方法、プログラム、ならびに、情報記録媒体

(57)【特許請求の範囲】

【請求項1】

カメラを有する情報処理装置であって、

現実世界に固定された第1座標系における前記情報処理装置の第1向きを検知する検知部

、  
前記カメラにより撮影されている撮影画像にユーザの顔画像が含まれていれば、前記撮影画像および前記顔画像から、前記情報処理装置に固定された第2座標系における前記ユーザの顔の第2向きを推定する推定部、

前記検知された第1向きと、前記推定された第2向きと、から、前記第1座標系における前記ユーザの顔の第3向きを算出する算出部、

前記算出された第3向きに応じた情報を出力する出力部  
を備え、

前記出力部は、

仮想音源に対応付けられる仮想方位と、

前記算出された第3向きと、

の角度差に応じた強度で前記仮想音源をミキシングした音声を前記情報として出力することを特徴とする情報処理装置。

【請求項2】

前記情報処理装置は、前記ユーザが装着する音響機器に無線もしくは有線により接続され、

前記出力部は、前記情報を前記音響機器に出力することを特徴とする請求項1に記載の情報処理装置。

【請求項3】

前記音響機器は、外音取り込み可能なヘッドフォン、イヤフォン、ネックスピーカー、骨伝導スピーカー、もしくは、補聴器であることを特徴とする請求項2に記載の情報処理装置。

【請求項4】

前記仮想方位は、あらかじめ定められることを特徴とする請求項1に記載の情報処理装置。

【請求項5】

前記検知部は、前記第1座標系における前記情報処理装置の第1位置をさらに検知し、前記仮想音源に対応付けられる前記仮想方位は、仮想音源が前記第1座標系において配置される仮想位置と、前記検知された第1向きおよび第1位置と、から、算出されることを特徴とする請求項1に記載の情報処理装置。

10

【請求項6】

前記情報処理装置は、前記検知された第1位置および第1向きに応じた映像情報を、前記カメラの撮影方向と同じ方向を表示方向とする画面に表示し、前記顔画像の大きさに応じて、前記仮想音源の波形を補正することを特徴とする請求項5に記載の情報処理装置。

20

【請求項7】

前記撮影画像に前記ユーザの前記顔画像と、前記ユーザの手画像と、が含まれていれば、前記第2座標系における前記ユーザの前記顔と、前記ユーザの手と、の距離に応じて、前記仮想音源の波形を補正することを特徴とする請求項1に記載の情報処理装置。

【請求項8】

前記撮影画像に前記ユーザの手画像が含まれていれば、前記撮影画像の代表点と、前記手画像と、の距離に応じて、前記仮想音源の波形を補正することを特徴とする請求項1に記載の情報処理装置。

30

【請求項9】

前記撮影画像に前記顔画像が含まれていなければ、前記強度を既定値とすることを特徴とする請求項1に記載の情報処理装置。

【請求項10】

前記仮想音源は、仮想空間に配置された発音オブジェクトに対応付けられ、前記情報処理装置は、前記発音オブジェクトが配置され、前記検知された第1位置および第1向きに応じた視点位置および視線方向により観察した前記仮想空間の様子を、前記カメラの撮影方向と同じ方向を表示方向とする画面に表示し、特定条件が満たされると、前記仮想音源に対応付けられる仮想方位と、前記算出された第3向きと、の角度差が最小の発音オブジェクトを、前記ユーザによる注目オブジェクトとして特定し、

40

前記出力部は、前記算出された第3向きに応じた情報を出力するのにかえて、前記特定された注目オブジェクトに対応付けられる仮想音源を、他の仮想音源に優先して出力し、前記特定された注目オブジェクトを他の発音オブジェクトより強調して前記画面に表示し、

解除条件が満たされると、前記注目オブジェクトとしての特定を解除することを特徴とする請求項5に記載の情報処理装置。

【請求項11】

前記情報処理装置は、前記撮影画像に含まれる前記ユーザの手画像に基づくジェスチャ

50

一、もしくは、前記画面に対するタッチ操作に基づいて、前記視点位置を中心に、前記仮想空間に配置された発音オブジェクトの仮想方位を回転させることを特徴とする請求項10に記載の情報処理装置。

【請求項12】

前記発音オブジェクトは、音声とともに再生される動画であり、  
前記注目オブジェクトが特定されている間、前記情報処理装置は、  
前記注目オブジェクトを前記画面内の所定位置に所定拡大率で表示し、  
前記出力部は、前記注目オブジェクトに対応付けられる仮想音源を所定の増幅率で、  
他の仮想音源をミュートして、ミキシングした音声を出力し、  
前記注目オブジェクトとしての特定が解除されると、前記情報処理装置は、前記注目オブジェクトとしての特定が解除された発音オブジェクトの仮想方位が前記算出された第3向きに一致するように、前記視点位置を中心に、前記仮想空間に配置された発音オブジェクトの仮想方位を回転させることを特徴とする請求項10または11に記載の情報処理装置。

10

【請求項13】

前記発音オブジェクトは、音声を発するアバターであり、  
前記注目オブジェクトが特定されている間、前記情報処理装置は、前記撮影画像に含まれる前記ユーザの手画像に基づくジェスチャー、もしくは、前記画面に対するタッチ操作に基づいて、前記注目オブジェクトの前記仮想空間における位置もしくは向きを変化させることを特徴とする請求項10または11に記載の情報処理装置。

20

【請求項14】

カメラを有する情報処理装置が、  
現実世界に固定された第1座標系における前記情報処理装置の第1向きを検知し、  
前記カメラにより撮影されている撮影画像にユーザの顔画像が含まれていれば、前記撮影画像および前記顔画像から、前記情報処理装置に固定された第2座標系における前記ユーザの顔の第2向きを推定し、  
前記検知された第1向きと、前記推定された第2向きと、から、前記第1座標系における前記ユーザの顔の第3向きを算出し、  
前記算出された第3向きに応じた情報を出力する情報処理方法であって、  
前記情報として、  
仮想音源に対応付けられる仮想方位と、  
前記算出された第3向きと、  
の角度差に応じた強度で前記仮想音源をミキシングした音声を出力する  
ことを特徴とする情報処理方法。

30

【請求項15】

カメラを有する情報処理装置を、  
現実世界に固定された第1座標系における前記情報処理装置の第1向きを検知する検知部、  
前記カメラにより撮影されている撮影画像にユーザの顔画像が含まれていれば、前記撮影画像および前記顔画像から、前記情報処理装置に固定された第2座標系における前記ユーザの顔の第2向きを推定する推定部、  
前記検知された第1向きと、前記推定された第2向きと、から、前記第1座標系における前記ユーザの顔の第3向きを算出する算出部、  
前記算出された第3向きに応じた情報を出力する出力部  
として機能させ、  
前記情報として、  
仮想音源に対応付けられる仮想方位と、  
前記算出された第3向きと、  
の角度差に応じた強度で前記仮想音源をミキシングした音声を出力する  
ことを特徴とするプログラム。

40

50

**【請求項 16】**

請求項15に記載のプログラムが記録されたコンピュータ読取可能な非一時的な情報記録媒体。

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は、現実世界におけるユーザの顔の向きを推定し、これに応じた情報を出力するための情報処理装置、情報処理方法、プログラム、ならびに、情報記録媒体に関する。

**【背景技術】****【0002】**

従来、ユーザの頭部の動きに応じた情報を出力する技術が提案されている。たとえば、特許文献1に開示される音源選択装置は、

ヘッドフォンと、

前記ヘッドフォンを装着した受聴者に対して前記ヘッドフォンを介して定位した複数の仮想音源を提供する仮想音源提供手段と、

前記複数の仮想音源から1つの仮想音源を選択する仮想音源選択手段とを備え、

前記仮想音源提供手段が、

前記受聴者に対して提供する前記複数の仮想音源の複数の定位音源配置パターンを記憶する定位音源配置パターン記憶手段と、

前記受聴者の選択動作に応じて前記複数の定位音源配置パターンから所望のパターンを選択する配置パターン選択手段と、

前記定位音源配置パターンに従って前記複数の仮想音源を提供するミキシング手段とを有し、

前記ヘッドフォンに装着されて前記受聴者の頭部の動きを検出する頭部動作検出センサと、

前記頭部動作検出センサの出力に基づいて前記頭部の動きを判定する頭部動作判定手段とを更に備え、

前記配置パターン選択手段は、前記頭部動作判定手段が前記頭部の動きから予め定めた配置パターン変更動作を検出すると前記定位音源配置パターン記憶手段から別の前記定位音源配置パターンを選択して前記ミキシング手段に出力するように構成されている。

**【0003】**

一方で、近年のスマートフォンやタブレットでは、画面の表示方向と同じ方向を撮影方向とするフロントカメラ(インカメラ、正面カメラ、前面カメラと呼ばれることもある。)により画面を見るユーザの姿をセルフィーとして撮影したり、フロントカメラと逆向きを撮影方向とするリアカメラ(背面カメラと呼ばれることもある。)により、ユーザの前方に広がる世界の様子を、画面で確認しながら撮影できるようにしているものが多い。

**【0004】**

また、近年のスマートフォンやタブレットでは、GPS(Global Positioning System)やWifiアクセスポイント、Bluetooth(登録商標)ビーコン等を利用するジオロケーション検知機能、加速度センサ、地磁気センサなどにより、スマートフォンやタブレットの位置や向きを、現実世界に固定された世界座標系に対して検出できるようにしているものも多い。

**【0005】**

そして、スマートフォンやタブレットの画面に、現実世界の様子を拡張した様子を表示する拡張現実機能を提供する技術も広がりつつある。

**【先行技術文献】****【特許文献】****【0006】**

【文献】特許第4837512号公報

**【発明の概要】**

10

20

30

40

50

**【発明が解決しようとする課題】****【0007】**

ここで、特許文献1に開示される技術では、ヘッドフォンが備える頭部動作検出センサを用いて、ユーザの頭部の動きを検出している。

**【0008】**

しかしながら、スマートフォンやタブレットとともに利用されるヘッドフォンやイヤホン等の音響機器は、ノイズキャンセリング機能や外部音声取り込み機能を有するものが普及しつつあるものの、頭部動作検出センサは有していないことがほとんどである。

**【0009】**

そこで、既に普及しているスマートフォンやタブレット等の機能を利用してユーザの顔の向きを推定する技術が強く求められている。

10

**【0010】**

本発明は、上記の課題を解決するもので、現実世界におけるユーザの顔の向きを推定し、これに応じた情報を出力するための情報処理装置、情報処理方法、プログラム、ならびに、情報記録媒体に関する。

**【課題を解決するための手段】****【0011】**

本発明に係る情報処理装置は、カメラを有し、  
現実世界に固定された第1座標系における前記情報処理装置の第1向きを検知し、  
前記カメラにより撮影されている撮影画像にユーザの顔画像が含まれていれば、前記撮影画像および前記顔画像から、前記情報処理装置に固定された第2座標系における前記ユーザの顔の第2向きを推定し、  
前記検知された第1向きと、前記推定された第2向きと、から、前記第1座標系における前記ユーザの顔の第3向きを算出し、  
前記算出された第3向きに応じた情報を出力する。

20

**【発明の効果】****【0012】**

本発明によれば、現実世界におけるユーザの顔の向きを推定し、これに応じた情報を出力するための情報処理装置、情報処理方法、プログラム、ならびに、情報記録媒体を提供することができる。

30

**【図面の簡単な説明】****【0013】**

【図1】本発明の実施形態に係る情報処理装置の概要構成を示す説明図である。

【図2】本発明の実施形態に係る情報処理装置が実行する情報処理方法の制御を示すフローチャートである。

【図3】本発明の実施形態に係る情報処理装置による表示例をグレースケールで示す図面代用写真である。

【図4】本発明の実施形態に係る情報処理装置による表示例をモノクロ2値で示す図面代用写真である。

【図5】本発明の実施形態に係る情報処理装置による表示例をグレースケールで示す図面代用写真である。

40

【図6】本発明の実施形態に係る情報処理装置による表示例をモノクロ2値で示す図面代用写真である。

【図7】本発明の実施形態に係る情報処理装置による表示例をグレースケールで示す図面代用写真である。

【図8】本発明の実施形態に係る情報処理装置による表示例をモノクロ2値で示す図面代用写真である。

【図9】本発明の実施形態に係る情報処理装置による表示例をグレースケールで示す図面代用写真である。

【図10】本発明の実施形態に係る情報処理装置による表示例をモノクロ2値で示す図面

50

代用写真である。

【図 1 1】本発明の実施形態に係る情報処理装置による表示例をグレースケールで示す図面代用写真である。

【図 1 2】本発明の実施形態に係る情報処理装置による表示例をモノクロ2値で示す図面代用写真である。

【図 1 3】本発明の実施形態に係る情報処理装置による表示例をグレースケールで示す図面代用写真である。

【図 1 4】本発明の実施形態に係る情報処理装置による表示例をモノクロ2値で示す図面代用写真である。

【図 1 5】本発明の実施形態に係る情報処理装置による表示例をグレースケールで示す図面代用写真である。

10

【図 1 6】本発明の実施形態に係る情報処理装置による表示例をモノクロ2値で示す図面代用写真である。

【図 1 7】本発明の実施形態に係る情報処理装置による表示例をグレースケールで示す図面代用写真である。

【図 1 8】本発明の実施形態に係る情報処理装置による表示例をモノクロ2値で示す図面代用写真である。

【図 1 9】本発明の実施形態に係る情報処理装置による仮想コンサート会場の舞台の表示例をグレースケールで示す図面代用写真である。

【図 2 0】本発明の実施形態に係る情報処理装置による仮想コンサート会場の舞台の表示例をモノクロ2値で示す図面代用写真である。

20

【図 2 1】本発明の実施形態に係る情報処理装置による複数のディスプレイが配置された仮想部屋の表示例をグレースケールで示す図面代用写真である。

【図 2 2】本発明の実施形態に係る情報処理装置による複数のディスプレイが配置された仮想部屋の表示例をモノクロ2値で示す図面代用写真である。

【図 2 3】本発明の実施形態に係る情報処理装置による複数の動画コンテンツが配置された仮想部屋の表示例をグレースケールで示す図面代用写真である。

【図 2 4】本発明の実施形態に係る情報処理装置による複数の動画コンテンツが配置された仮想部屋の表示例をモノクロ2値で示す図面代用写真である。

【図 2 5】本発明の実施形態に係る情報処理装置による仮想部屋において1つの動画コンテンツが注目されたときの表示例をグレースケールで示す図面代用写真である。

30

【図 2 6】本発明の実施形態に係る情報処理装置による仮想部屋において1つの動画コンテンツが注目されたときの表示例をモノクロ2値で示す図面代用写真である。

【図 2 7】本発明の実施形態に係る情報処理装置による仮想部屋において1つの動画コンテンツが注目されたときの表示例をグレースケールで示す図面代用写真である。

【図 2 8】本発明の実施形態に係る情報処理装置による仮想部屋において1つの動画コンテンツが注目されたときの表示例をモノクロ2値で示す図面代用写真である。

【図 2 9】本発明の実施形態に係る情報処理装置による仮想部屋において1つの動画コンテンツが注目されたときの表示例をグレースケールで示す図面代用写真である。

【図 3 0】本発明の実施形態に係る情報処理装置による仮想部屋において1つの動画コンテンツが注目されたときの表示例をモノクロ2値で示す図面代用写真である。

40

【図 3 1】本発明の実施形態に係る注目オブジェクトを処理する情報処理装置の概要構成を示す説明図である。

【発明を実施するための形態】

【0014】

以下に、本発明の実施形態を説明する。なお、本実施形態は、説明のためのものであり、本発明の範囲を制限するものではない。したがって、当業者であれば、本実施形態の各要素もしくは全要素を、これと均等なものに置換した実施形態を採用することが可能である。また、各実施例にて説明する要素は、用途に応じて適宜省略することも可能である。このように、本発明の原理にしたがって構成された実施形態は、いずれも本発明の範囲に

50

含まれる。

【0015】

(構成)

図1は、本発明の実施形態に係る情報処理装置の概要構成を示す説明図である。以下、本図を参照して概要を説明する。

【0016】

本図に示すように、本実施形態に係る情報処理装置101は、カメラ151を有する。検知部111、推定部112、算出部113、出力部114を備える。また、情報の出力先として、音響機器152やディスプレイの画面153等を採用することができる。

【0017】

本実施形態に係る情報処理装置101は、典型的には、プログラムをスマートフォンやタブレット等の可搬型のコンピュータが実行することによって実現される。当該コンピュータは、各種の出力装置や入力装置に接続され、これらの機器と情報を送受する。

【0018】

コンピュータにて実行されるプログラムは、当該コンピュータが通信可能に接続されたサーバにより配布、販売することができるほか、CD-ROM(Compact Disk Read Only Memory)やフラッシュメモリ、EEPROM(Electrically Erasable Programmable ROM)などの非一時的(non-transitory)な情報記録媒体に記録した上で、当該情報記録媒体を配布、販売等することも可能である。

【0019】

プログラムは、コンピュータが有するハードディスク、ソリッドステートドライブ、フラッシュメモリ、EEPROM等の非一時的な情報記録媒体にインストールされる。すると、当該コンピュータにより、本実施形態における情報処理装置が実現されることになる。一般的には、コンピュータのCPU(Central Processing Unit)は、コンピュータのOS(Operating System)による管理の下、情報記録媒体からRAM(Random Access Memory)へプログラムを読み出してから、当該プログラムに含まれるコードを解釈、実行する。ただし、CPUがアクセス可能なメモリ空間内に情報記録媒体をマッピングできるようなアーキテクチャでは、RAMへの明示的なプログラムのロードは不要なこともある。なお、プログラムの実行の過程で必要とされる各種情報は、RAM内に一時的(temporary)に記録しておくことができる。

【0020】

さらに、コンピュータは、各種画像処理計算を高速に行うためのGPU(Graphics Processing Unit)を備えることが望ましい。GPUならびにTensorFlow等のライブラリを使うことで、CPUの制御の下、各種の人工知能処理における学習機能や分類機能を利用することができるようになる。

【0021】

なお、ソフトウェアをインストールするようなコンピュータにより本実施形態の情報処理装置を実現するのではなく、専用の電子回路を用いて本実施形態の情報処理装置101を構成することも可能である。たとえば、ポータブルカメラやポータブル電子ゲーム装置などを情報処理装置101として利用することができる。

【0022】

この態様では、プログラムを電子回路の配線図やタイミングチャート等を生成するための素材として利用することもできる。このような態様では、プログラムに定められる仕様を満たすような電子回路がFPGA(Field Programmable Gate Array)やASIC(Application Specific Integrated Circuit)により構成され、当該電子回路は、当該プログラムに定められた機能を果たす専用機器として機能して、本実施形態の情報処理装置を実現する。

【0023】

以下では、理解を容易にするため、情報処理装置101は、コンピュータがプログラムを実行することによって実現される態様を想定して説明する。

【0024】

10

20

30

40

50

なお、情報処理装置101には、情報の出力先として、ヘッドホン、イヤホン、ネックスピーカー、骨伝導スピーカー、補聴器等の音響機器152を、無線または有線にて接続することができる。これらの音響機器152には、外部音声取り込み機能を有することが望ましい。

【0025】

また、上述のように、情報の出力先として、情報処理装置101が有する液晶ディスプレイ、有機EL(Organic Electro-Luminescence)ディスプレイ、電子インクを用いたペーパーディスプレイ等の画面153を採用することができる。これらのディスプレイをタッチスクリーンとすることで、情報処理装置101の入力装置として機能させることもできる。

【0026】

さて、本実施形態の情報処理装置101において、検知部111は、現実世界に固定された第1座標系における情報処理装置101の第1向きを検知する。

【0027】

現実世界に固定された第1座標系における情報処理装置101の向き(第1向き)は、情報処理装置101が有する地磁気センサや重力を検知する慣性センサ、加速度センサ、ジャイロセンサ等を介して検知することができる。

【0028】

また、第1座標系における情報処理装置101の位置(第1位置)は、GPS、Wifiアクセスポイント、Bluetoothビーコン等を用いたジオロケーション検知機能により検知することも可能である。

【0029】

一方、推定部112は、カメラ151により撮影されている撮影画像にユーザの顔画像が含まれていれば、撮影画像および顔画像から、情報処理装置101に固定された第2座標系におけるユーザの顔の第2向きを推定する。

【0030】

すなわち、情報処理装置101は、撮影画像に描画されている顔画像を画像認識により抽出し、目、鼻、口などの特徴部位を認識した上で、顔画像に基いて、情報処理装置101に対する相対的なユーザの顔の向き(第2向き)を推定する。この処理には、一般的なフェーストラッキング技術を適用することができる。

【0031】

なお、撮影画像内における顔画像の位置や大きさに基づいて、情報処理装置101に対する相対的なユーザの顔の位置(第2位置)をさらに推定することとしても良い。

【0032】

さらに、算出部113は、検知された第1向きと、推定された第2向きと、から、現実世界(第1座標系)におけるユーザの顔の第3向きを算出する。

【0033】

第1座標系と第2座標系の間での方向に関する変換は、第1向きに基いて、一意に定めることができる。また、第1位置が検知されている場合は、第1向きと第1位置に基いて、第1座標系と第2座標系の間での座標値の座標変換を、一意に定めることができる。

【0034】

そこで、撮影画像に基いて推定された第2向きの第2座標系における成分を、第1座標系における成分に変換することで、世界(地球)に対してユーザの顔がどちらを向いているかを表す第3向きを算出することができる。

【0035】

そして、出力部114は、算出された第3向きに応じた情報を出力する。情報の出力先としては、ユーザが装着する音響機器152やディスプレイの画面153を採用することができる。

【0036】

出力される情報としては、1つもしくは複数の仮想音源を現実世界内に設定し、第3向きに応じて各仮想音源に対応付けられる波形の強度、音色、位相等を変化させてミキシング

10

20

30

40

50

した音声情報を採用することができる。

【0037】

仮想音源は、現実世界内で十分遠方に仮想的に配置したと想定して、聴取点からの仮想的な方向(仮想方位)のみをあらかじめ決めて対応付けても良い。また、仮想音源を現実世界内の位置に仮想的に配置しても良い。

【0038】

前者の場合、出力部114は、

仮想音源に対応付けられる仮想方位と、

算出された第3向きと、

の角度差に応じた強度(増幅率)で、仮想音源をミキシングする。角度差が小さければ、ユーザの正面に仮想音源があることになるから、ミキシング時の波形の強度を大きくすることで、顔の向きに応じて変化する音声拡張現実をユーザに提供することができるようになる。

10

【0039】

なお、仮想音源が複数ある場合には、顔の向きを一周させたと仮定した場合の平均音圧が大きく変化しないように、すなわち、各仮想音源についての角度差に基づいた増幅率の比は維持したまま、仮想音源のパワーの総和がほぼ一定となるように、ミキシング時の増幅率を調整することで、仮想音源全体の迫力を維持したまま、特定の仮想音源を強調することができるようになる。これをパワー補正という。

【0040】

また、角度差に応じてステレオ出力の左右の増幅率や時間差を変化させることで、仮想音源の方向をユーザに知得させることも可能である。たとえば、ユーザの右側に仮想音源がある場合には、右側の増幅率を左側の増幅率よりも大きくしたり、右側が左側に先行するように時間差を設定したりすれば、簡易的なバイノーラル再生を実現することができる。仮想音源の方向をユーザに感じさせることができる。

20

【0041】

後者の場合は、仮想音源に対応付けられる仮想方位を、

仮想音源が第1座標系において配置される仮想位置と、

検知部111により検知された第1向きおよび第1位置と、

から、算出した上で、前者と同様の処理を行えば良い。

30

【0042】

なお、第1向きおよび第1位置の測定精度、ならびに、第2位置の推定精度が十分に高い場合には、仮想方位の算出にあたって、第1位置ではなく、現実世界におけるユーザの顔の位置(第3位置)を利用して良い。第3位置は、フェーストラッキングによって求められた情報処理装置101に対する相対的な顔の位置(第2位置)を、情報処理装置101に固定された座標系に座標変換すれば得られる。

【0043】

このほか、ユーザの顔の向きを、羅針盤のように、ディスプレイの画面153に表示しても良い。ユーザが顔の向きを変えた場合に、向きの変化に応じて、「羅針盤」の「針」の向きが変化するような表示態様を採用すれば、ユーザの視界にディスプレイの画面153が入る範囲内であれば、本実施形態が適切に稼働していることをユーザが確認できるようになる。

40

【0044】

また、仮想音源に対応付けられる仮想方位と、算出された第3向きと、の角度差に応じて、仮想音源の波形を、強度以外についても演出的に補正することで、正面側にある仮想音源を強調してユーザに聞かせることもできる。

【0045】

たとえば、仮想音源の波形にもとづいて、直接音とリバーブ音を生成し、角度差に応じて両者の混合比を変化させることとしても良い。角度差が小さければ、直接音の割合を増やすことで、仮想音源が正面側で大きな音で聞こえているようにユーザに感じさせること

50

ができる。これを反響補正という。

【0046】

また、正面側にある仮想音源の中心音域を求め、他の方向にある仮想音源については、求められた中心音域をイコライザーにより弱めることで、周波数的なかぶりを減らし、正面側にある仮想音源を浮き上がらせてユーザに聴取させることもできる。これを中心音域補正と呼ぶ。

【0047】

このほか、正面側にある仮想音源については、倍音成分を強化するサチレーションを加えてきらびやかな音にして、正面側にある仮想音源を浮き上がらせてユーザに聴取させることもできる。これをサチレーション補正という。

10

【0048】

さて、情報処理装置101が有するカメラ151が、いわゆるフロントカメラである場合、その撮影方向は、ディスプレイの画面153の表示方向と一致し、ユーザが位置すると想定する方向に向かう。

【0049】

したがって、ユーザが情報処理装置101の画面153を正面から見ていれば、ユーザの顔はカメラ151によって撮影されているはずである。

【0050】

そこで、ユーザの顔がカメラ151によって撮影されていない場合は、ユーザは特定の仮想音源に集中して聴取してはいない、と想定して、仮想音源の強度・ステレオ音声の時間差の調整や演出的な補正を平均的な既定値としても良い。

20

【0051】

なお、演出的な補正は、ユーザのジェスチャーによって調整することも可能である。

【0052】

たとえば、ユーザがディスプレイの画面153に顔を近付けた場合は、ユーザが正面に集中しようとしていると想定して、正面にある仮想音源を強調するような演出的な補正をしても良い。この態様では、撮影画像に描画されている顔画像の大きさに基づいて、演出的な補正の強さを変化させれば良いことになる。

【0053】

このほか、ユーザのジェスチャーにより、演出的な補正の強さを変化させることもできる。たとえば、耳をすますジェスチャーにより、ユーザが正面に集中しようとしていると想定して、正面にある仮想音源を強調するような演出的な補正をしても良い。

30

【0054】

たとえば、撮影画像から、ユーザの顔画像と、ユーザの手画像と、を画像認識し、ユーザの顔の位置(たとえば、顔の中心位置)と、ユーザの手の位置(たとえば、小指先端の位置)と、を推定した上で、両者の距離(近さ)に応じて、演出的な補正の強さを変化させることで、耳をすますジェスチャーに簡易に対応することができる。

【0055】

このほか、より簡易的に、ユーザの手画像の撮影画像内における位置(たとえば、小指先端の位置)と、撮影画像の代表点(たとえば、撮影画像の中心位置や顔画像の中心位置等。)と、の距離(近さ)に応じて、演出的な補正の強さを変化させることとしても良い。

40

【0056】

ここで、代表点として、撮影画像の中心位置を採用した場合には、フェーストラッキングに失敗して顔画像が認識できなかったときでも、手画像が認識されていれば、補正の強さを調整することができるようになる。

【0057】

なお、強度・時間差の調整や演出上の補正においては、算出された角度差や距離をそのまま直ちに反映させるのではなく、直近一定時間(たとえば、100ms程度)内の平均や減衰平均を使うことにより、算出された値に近付けるようにして、値の変化を滑らかにすることとしても良い。

50

## 【 0 0 5 8 】

上記の説明では、音声により拡張現実をユーザに体感させることとしており、現実世界においてユーザが顔の向きを変えると(ユーザの顔が情報処理装置101のカメラ151で撮影されていれば)その向きに応じて仮想音源の出力が変化する。

## 【 0 0 5 9 】

したがって、現実世界における環境音に指向性が存在する場合であっても、仮想音源の出力は、これと連動した指向性を有することになり、音響機器152が外部音声取り込み可能な場合や、ネックスピーカーのように外部音声もそのままユーザの耳に入る場合であっても、環境音と仮想音は、顔の向きに応じて矛盾なく混合されてユーザに提供されることになり、ユーザに、リアルな音声拡張現実の提供をすることができる。

10

## 【 0 0 6 0 】

なお、情報処理装置101がリアカメラを有する場合には、リアカメラで撮影した現実世界の撮影画像と、仮想音源と同じ位置に配置された仮想オブジェクトの外観と、を合成した拡張現実画像を生成して、情報処理装置101のディスプレイの画面153に表示することで、視覚と聴覚の両方の拡張現実をユーザに提供することもできる。

## 【 0 0 6 1 】

(制御の流れ)

図2は、本発明の実施形態に係る情報処理装置が実行する情報処理方法の制御を示すフローチャートである。以下、本図を参照して説明する。なお、以下の処理の各工程は、適用される態様に依りて、適宜省略が可能である。

20

## 【 0 0 6 2 】

本処理が開始されると、情報処理装置101は、まず、各仮想音源の再生用のパラメータを既定値で初期化して(ステップS200)、ミキシング再生を開始する(ステップS201)。このパラメータには、増幅率(左右等、各チャンネル毎の増幅率としても良いし、全体のものとしても良い。)、直接音とリバーブ音の混合比、サチレーションの強度等、種々のものを採用することができ、処理開始の当初は、これらのパラメータに既定値が設定されることになる。

## 【 0 0 6 3 】

この後、仮想音源の再生はバックグラウンド処理として並行して実行されるが、ミキシング用のパラメータは、以下の処理によって、ユーザの顔の向き等に応じて変更される。

30

## 【 0 0 6 4 】

ついで、情報処理装置101は、地磁気センサ、ジャイロセンサ、加速度センサ等を介して、情報処理装置101の現実世界(第1座標系)における第1向き(や第1位置)を検知する(ステップS202)。

## 【 0 0 6 5 】

さらに、情報処理装置101は、カメラ151にて撮影されている撮影画像から、ユーザの顔画像を画像認識により抽出する試行をする(ステップS203)。

## 【 0 0 6 6 】

当該試行により、ユーザの顔画像の抽出に成功すれば(ステップS204;Yes)、情報処理装置101は、当該顔画像に基づいて、情報処理装置101に対する相対的な(第2座標系における)ユーザの顔の第2向き(や第2位置)を推定する(ステップS205)。

40

## 【 0 0 6 7 】

そして、情報処理装置101は、推定された第2向き(や第2位置)を、検知された第1向き(や第1位置)に基いて、座標変換により、第1座標系におけるユーザの顔の第3向き(や第3位置)を算出する(ステップS206)。

## 【 0 0 6 8 】

ついで、情報処理装置101は、仮想音源のそれぞれについて、以下の処理を繰り返す(ステップS207)。

## 【 0 0 6 9 】

すなわち、情報処理装置101は、当該仮想音源の第1座標系における仮想方位を取得す

50

る(ステップS208)。この仮想方位は、あらかじめ定めたものとしても良いし、第1座標系における当該仮想音源の仮想位置と情報処理装置101の第1位置(あるいは、ユーザの顔の第3位置)に基づいて算定しても良い。

【0070】

ついで、情報処理装置101は、当該仮想方位と、第3向きと、の角度差に基づいて、当該仮想音源に対する再生用の新たなパラメータを算出する(ステップS209)。最も単純には、角度差に基づいて新たな増幅率を算出することとなるが、反響補正やサチレーション補正を加えても良い。

【0071】

このほか、当該仮想音源の仮想位置と第1位置(あるいは第3位置)との距離(近さ)に応じて、増幅率をさらに補正することとしても良い。すなわち、距離が小さければ小さいほど、増幅率を大きくする等である。

10

【0072】

すべての仮想音源について処理を繰り返した(ステップS210)後、情報処理装置101は、全仮想音源の再生用の新たなパラメータを、互いの関係に基づいてさらに補正する(ステップS211)。この補正には、たとえば、正面側の仮想音源を他の仮想音源に比べて強調する中心音域補正や、仮想音源全体の迫力をそのまま維持できるようにするパワー補正等が含まれる。

【0073】

そして、全仮想音源の再生用のパラメータを、新たなパラメータに滑らかに近付けるように(あるいは、そのまま新たなパラメータとするように)、新たなパラメータに基づく更新の処理を各仮想音源について繰り返してから(ステップS212-S214)、処理をステップS202に戻す。

20

【0074】

一方、ユーザの顔画像の抽出に失敗すれば(ステップS204;No)、全仮想音源のパラメータを既定値に近付けるように(あるいは、そのまま既定値とするように)、更新する既定値に基づく更新の処理を各仮想音源について繰り返してから(ステップS215-S217)、処理をステップS202に戻す。

【0075】

なお、上記の制御の流れでは省略したが、撮影画像からユーザの手画像を認識し、ユーザのジェスチャーに基づいて、正面側の仮想音声の増幅率を変更したり、サチレーション補正や中心音域補正の強度を変更したりしても良い。

30

【0076】

(ディスプレイの画面への出力)

図3乃至18は、本発明の実施形態に係る情報処理装置による表示例をグレイスケールもしくはモノクロ2値で示す図面代用写真である。以下、これらの図を参照して説明する。

【0077】

図3、4では、スマートフォンからなる情報処理装置101のディスプレイの画面153に種々の情報が表示されている。本図下方中央にある三角形の再生ボタンをタップすると、仮想音源からなる曲の再生が開始される。

40

【0078】

画面153の上方には、ユーザの顔の向きが検出された結果と、小指の先端が検出された検出の結果と、がウィンドウ内に表示されている。ユーザが操作に慣れるまではこの検出結果を見て、カメラ151でユーザの顔が撮影されるような位置を保持しつつ、ジェスチャーを確認・練習することができる。

【0079】

再生ボタンの左側にあるオンオフボタンをタップあるいはスライドすることで、図5、6に示すようにウィンドウを閉じることができる。再度同じオンオフボタンをタップあるいはスライドすることで、ウィンドウを再度表示することもできる。

【0080】

50

画面153の中央には、円形に並んだ楽器のアイコンが並べられている。これは、仮想空間上に配置された仮想音源のパートの方位を表している。

【0081】

本図では、楽器が等間隔に並べられているが、必ずしも等間隔で円形である必要はなく、任意の配置が可能である。

【0082】

この円の上をスワイプすると、楽器は円形の中心を中心として回転し、図7、8に示すように、自分の好きな楽器を好きな方向に配置することができる。

【0083】

円形の中央にあるのは、操作しているユーザの-avatarであり、白矢印の方向がユーザの顔の向きを表している。白矢印の先にある楽器のアイコンが、ユーザの正面側に位置する仮想音源に相当する。

10

【0084】

本処理の開始時は、白矢印は既定の方向(たとえば上)を向いており、ユーザが顔の向きをかえたり、スマートフォンの位置を動かしたりすると、それに応じて白矢印の向きが変化する。

【0085】

avatarの上をタップすると、白矢印の方向および楽器の配置(距離)がリセットされる。

【0086】

図3、4では、白矢印の方向に、扇形が表示されている。これは、増幅率が0.5倍以上の範囲を表している。耳をすますジェスチャーをすることによって、扇形の角度が変化し、どの仮想音源が強調されているかがユーザに伝えられる。

20

【0087】

画面153の下部には2本のスライダーが並んでいる。上のスライダーは、仮想空間上で円形に並んだ楽器との距離を表しており、スライダーを移動することで距離を変化させることができる。図3、4に示す配置では、距離は20メートルであるが、図9、10では、10メートルになっており、図11、12では、30メートルになっている。そして、画面153に示されるavatarから楽器までの距離も、この距離に応じて変化する。

【0088】

下のスライダーは、フォーカスの利き具合、すなわち、扇形の角度に連動している。上記のように、ジェスチャーによってフォーカスの利き具合を変化させることもできるが、スライダーを直接移動させることで、調整することもできる。

30

【0089】

再生ボタンの右側にある歯車型の設定ボタンを押すと、図13、14に示すように、設定フォームに遷移する。

【0090】

設定フォームでは、各楽器のマスターボリューム(ミキサー増幅率の既定値)が設定できる。情報処理装置101は、角度差に応じた乗数をマスターボリュームに掛け合わせることで、ミキシングに用いる増幅率を一旦計算した上で、全体のパワーがほぼ一定になるように補正を行う。

40

【0091】

図15、16では、ブーストモードの設定がされている。ブーストモードでは、全体のパワーが一定になるように増幅率を調整する際に、正面側の仮想音源の強度を倍増させることで、正面の楽器を強調することができる。

【0092】

図17、18は、上記のスマートフォンと同様の機能をタブレットにて実現した場合の出力例である。

【0093】

これらの図では、リアカメラで撮影されている無人の公園に、仮想楽器を演奏する仮想人物の動画が重畳された拡張現実画像が表示されている。

50

## 【0094】

一方で、拡張現実ではなく、仮想現実の本実施形態を提供することもできる。図19、20は、仮想のコンサート会場の舞台に仮想楽器の演奏者を円状に配置し、その中央にユーザを配置したかのような仮想現実をユーザに提供する。

## 【0095】

本表示例では、仮想楽器の演奏者が10人、舞台の上に配置されており、演奏者のアバター(本図では、10人の演奏者のうちの3人のアバター)が舞台の上で楽器を演奏する映像を構成することで、仮想オブジェクトとしている。各仮想オブジェクトには、楽器の演奏音が仮想音源として対応付けられ、仮想音源は、上記実施形態と同様に、ミキシングされて出力される。

10

## 【0096】

この態様では、ユーザは仮想コンサートの指揮者のような体験をすることができる。

## 【0097】

上記実施形態と同様に、ユーザは、耳をすますジェスチャー等により、複数の演奏者のアバターのうち、ユーザが向いているアバター、すなわち、ユーザの正面に位置するアバターを、注目の対象を表す注目オブジェクトとして特定することができる。

## 【0098】

ユーザが情報処理装置101を自身の正面にて把持しており、顔を画面153の中央に向けている場合には、画面153の中央に表示されている仮想オブジェクトが注目オブジェクトとなる。

20

## 【0099】

一方で、ユーザが情報処理装置101を自身の正面にて把持していても、ユーザが顔を画面153の中央ではなく、右側や左側など、その他の方向に向けている場合には、画面中央に表示されている仮想オブジェクトではなく、顔を向けている方向に表示されている仮想オブジェクトが注目オブジェクトになる。すなわち、仮想音源に対応付けられる仮想方位と、第3向きとの、の角度差が閾角以下で最小の発音オブジェクトを、注目オブジェクトとして特定する。

## 【0100】

ジェスチャーではなく、所望の仮想オブジェクトに顔を向け続け、その向け続けた時間が所定の閾時間を超えると、当該仮想オブジェクトが注目オブジェクトとして特定されるようにしても良い。

30

## 【0101】

一旦注目オブジェクトとして演奏者のアバターが選択された後は、ユーザは、注目オブジェクトの位置や向きを変化させることができるようにしても良い。

## 【0102】

たとえば、画面153がタッチスクリーンとして構成されている場合には、タッチスクリーンに触れてなぞる操作をすると、注目オブジェクトが、なぞり操作の軌跡を平行移動した同じ形状の軌跡に沿って移動するようにしても良い。この態様では、注目オブジェクトが特定されているので、画面153に表示された注目オブジェクトそのものをタッチする必要はなく、画面153において注目オブジェクトが表示されている場所以外においてなぞり操作をすることができ、注目オブジェクトを指で隠さずに、注目オブジェクトの位置を変更することができる。

40

## 【0103】

また、画面153に対して、二本指や三本指でタッチして回転させる操作をすると、注目オブジェクトが、タッチする指の数に応じた軸周りに、タッチした角度だけ回転するようにしても良い。この態様においても、画面153に表示された注目オブジェクトそのものをタッチする必要はない。

## 【0104】

なお、タッチスクリーンで構成された画面153に対して行うこれらの種々の操作は、ジェスチャーによって代替することも可能である。

50

## 【0105】

上述の例では、仮想空間内の仮想オブジェクトとして、楽器を演奏する演奏者のアバターを配置したが、音声とともに再生される動画を仮想オブジェクトとすることもできる。図21, 22, 23, 24は、仮想部屋の中に複数の仮想ディスプレイが配置され、当該仮想ディスプレイにて仮想動画が再生される様子を示す表示例である。

## 【0106】

これらの表示例では、仮想音源として機能する音声とともに各仮想ディスプレイにて再生される仮想動画が、仮想オブジェクトに相当する。

## 【0107】

これらの図では、仮想空間内では、10個の仮想動画がユーザの周りに配置されており、ユーザは、情報処理装置101を把持して自身の向きを変更することによって、ユーザは複数の仮想動画を並べて見比べることができる。ユーザの頭が向いている仮想動画の仮想音源が、他の仮想動画の仮想音源よりも優先されて出力される。

10

## 【0108】

この態様では、10個より多くの仮想動画をユーザが順に見比べることも可能である。すなわち、仮想空間においてユーザから見えない位置に配置された仮想ディスプレイにおいて、仮想動画を交換すれば良い。

## 【0109】

この態様では、ユーザが情報処理装置101を把持したまま、自身の体を現実空間で回転させることで、仮想動画を順に並べて見ることができる。

20

## 【0110】

また、指の向きを短時間で水平方向に、右から左へ、あるいは、左から右へ変化させるようなジェスチャー、あるいは、画面153を構成するタッチスクリーンを左スライドあるいは右スライドさせる動作等によって、仮想空間内において、ユーザを中心に仮想ディスプレイを回転させても良い。

## 【0111】

これにより、複数の仮想動画を、カルーセル表示あるいはカバーフロー表示に類似した態様でユーザに提供することができる。

## 【0112】

上記実施形態と同様に、ユーザは、画面153に表示されているいずれかの仮想動画に顔を向けて、耳をすます等のジェスチャーをしたり、閲覧時間の時間長だけ顔を向け続けたり、等によって、当該仮想動画を注目オブジェクトとして特定することができる。

30

## 【0113】

仮想動画が注目オブジェクトとして特定されると、情報処理装置101は、注目オブジェクトとして特定された仮想動画(注目動画)を、画面153内の中央等、所定位置に、所定拡大率で表示して再生するとともに、当該注目オブジェクトに対応付けられる仮想音源、すなわち、当該仮想動画とともに再生させるべき音声(注目音声)を、他の仮想音源に優先して再生する。このとき、出力される音声のミキシングは、注目音声は所定の増幅率で、他の仮想音源はミュートする、すなわち、注目音声のみが出力され、他の仮想音声は出力されないようにしても良い。

40

## 【0114】

図25, 26は、図23, 24において画面中央に描画されていた仮想動画が注目オブジェクトとして特定され、画面中央に拡大されて、当該注目オブジェクト動画および音声再生されている様子を示している。

## 【0115】

図27, 28, 29, 30は、ここで再生されている音声に合わせてユーザがダンスをし、現実空間における情報処理装置101やユーザの向きは変化しているが、画面中央にて再生される動画は注目オブジェクトのままであることを示している。

## 【0116】

この態様では、ユーザが手を広げて情報処理装置101のカメラ151に近付けるジェスチ

50

ャーをしたり、タッチスクリーンに対して短時間だけタップする等によって、注目オブジェクトとしての特定を解除することができる。

【0117】

特定が解除されたときは、当該特定が解除された仮想動画がユーザの頭が向いている先に位置するように、仮想空間内においてユーザの周りを囲む仮想動画を、ユーザを中心に回転させることとしても良い。すなわち、特定が解除された仮想オブジェクトの仮想方位が、算出された第3向きに一致するように、仮想空間内における仮想始点を中心に、仮想空間内に配置された仮想オブジェクト(の仮想方位)を回転させることになる。

【0118】

この態様によれば、特定が解除された直後は、その仮想動画がユーザの顔が向く方に配置されており、仮想動画の列が、以前と同じ順序でほぼ同じ位置に配置されることとなるので、ユーザは、仮想動画を直感的に、順に見比べることができるようになる。

10

【0119】

これらの実施形態に係る情報処理装置101について、以下に整理して説明する。図31は、本発明の実施形態に係る注目オブジェクトを処理する情報処理装置の概要構成を示す説明図である。

【0120】

本実施形態に係る情報処理装置101は、図1に開示する構成に加えて、特定部301と、解除部302と、を有する。特定部301、解除部302は、検知部111、推定部112、算出部113から各種の情報を取得し、それに応じて、出力部114を制御する。

20

【0121】

上記のように、本実施形態に係る情報処理装置101では、複数の発音オブジェクトが仮想空間に配置されている。各発音オブジェクトは、たとえば、上記実施形態における仮想オブジェクトとすることができ、上記実施形態では、仮想楽器を演奏する演奏者のアバター、あるいは、仮想動画を再生する仮想ディスプレイに相当する。

【0122】

各発音オブジェクトは、仮想音源に対応付けられている。上記実施形態では、仮想音源は、仮想楽器により出力される演奏音や、仮想動画とともに再生される音声に相当する。

【0123】

そして、情報処理装置101は、仮想空間の様子を画面153に表示する。具体的には、第1位置および第1向きに応じた視点位置および視線方向により観察した仮想空間の様子を、カメラ151の撮影方向と同じ方向を表示方向とする画面153に表示する。

30

【0124】

情報処理装置101(の画面153)の位置や向きを変化させたり、ユーザの頭の位置や向きを変化させると、それに応じて、画面153に表示される仮想世界の様子が増える。これにより、情報処理装置101の画面153が、仮想空間を覗き込むための「窓」として機能することとなる。

【0125】

ここで、情報処理装置101の特定部301は、特定条件が満たされたか否かを判定し、それに応じた処理を行う。

40

【0126】

また、情報処理装置101の解除部302は、解除条件が満たされたか否か、を判定し、それに応じた処理を行う。

【0127】

特定条件とは、複数の発音オブジェクトのうち、いずれかをユーザによる注目オブジェクトとして特定するための条件であり、解除条件とは、注目オブジェクトとしての特定を解除するための条件である。

【0128】

上記実施形態では、特定条件として、耳をすますジェスチャーがされていること、所定時間以上特定の発音オブジェクトの方を向き続けること、等を採用しており、解除条件と

50

して、手を広げてカメラ151に近付けるジェスチャーがされたことや、画面153を構成するタッチスクリーンをタップすること等を採用しているが、その他の条件を採用することも可能である。

【0129】

情報処理装置101において、特定部301は、特定条件が満たされる、と判定すると、仮想音源に対応付けられる仮想方位と、算出された第3向きと、の角度差が最小の発音オブジェクトを、ユーザによる注目オブジェクトとして特定する。

【0130】

注目オブジェクトは、仮想空間に配置された発音オブジェクトのうちユーザが注目したい、あるいは、注目していると推定されるオブジェクトである。上記実施形態では、画面153に表示されている発音オブジェクトであって、ユーザが向いている発音オブジェクトが注目オブジェクトとなりうる。すなわち、ユーザが画面153の中央を向いていれば、画面153中央に表示されている発音オブジェクトが、ユーザが画面153の右端を向いていれば、画面153の右端に表示されている発音オブジェクトが、ユーザが画面153の左端を向いていれば、画面153の左端に表示されている発音オブジェクトが、それぞれ注目オブジェクトとなりうる。

10

【0131】

さて、注目オブジェクトが特定されていないときは、出力部114は、仮想音源に対応付けられる仮想方位と、算出された第3向きと、の角度差に応じた強度で、仮想音源をミキシングするが、注目オブジェクトが特定されているときは、出力部114は、算出された第3向きに応じた情報を出力するのにかえて、特定された注目オブジェクトに対応付けられる仮想音源を、他の仮想音源に優先して出力する。

20

【0132】

上記実施形態では、注目オブジェクトに相当する演奏者の仮想楽器の演奏音や、仮想動画に伴う音声が、他の音声に優先して出力されることになる。ここで、「優先」とは、たとえば、注目オブジェクトの仮想音源の増幅率を所定の定数とし、他の仮想音源の増幅率をゼロ(ミュート)あるいは小さい値とする、等が含まれる。

【0133】

また、情報処理装置101は、画面153において、特定された注目オブジェクトを他の発音オブジェクトより強調して画面153に表示しても良い。

30

【0134】

上記実施形態では、注目オブジェクトに相当する演奏者の色を明るくしたり、演奏者にマークをつけたり、等の態様を採用することもできる。また、注目オブジェクトに相当する仮想動画を画面中央に所定サイズで表示することによって、強調表示をすることとしている。

【0135】

情報処理装置101において、解除部302は、解除条件が満たされると、注目オブジェクトとしての特定を解除する。これにより、仮想音源の優先出力や画面153における強調表示は終了し、最先に説明した出力手法が採用されることになる。

【0136】

仮想空間内に配置された仮想オブジェクトの位置は、仮想空間内の視点を中心に回転させることもできる。すなわち、情報処理装置101は、撮影画像に含まれるユーザの手画像に基づくジェスチャー、もしくは、画面に対するタッチ操作に基づいて、視点位置を中心に、仮想空間に配置された発音オブジェクトの仮想方位を回転させる。

40

【0137】

すると、指の向きを右から左あるいは左から右へ短時間に水平方向に変化させるジェスチャーや、画面153を構成するタッチスクリーンに対する右スライドや左スライドがされると、ユーザの視点位置の周りに並ぶ演奏者や仮想ディスプレイの動画が移動して、ユーザは、これらの様子を、順に見比べたり、その音声を聞き比べたりできるようになる。

【0138】

50

また、上記実施形態では、演奏者のアバターをつまんだり複数の指を画面153にタッチして回転させる操作により、アバターの位置や向きを編集可能としている。すなわち、情報処理装置101は、注目オブジェクトが特定されている間、撮影画像に含まれるユーザの手画像に基づくジェスチャー、もしくは、画面に対するタッチ操作に基づいて、注目オブジェクトの仮想空間における位置もしくは向きを変化させることができる。

【0139】

(まとめ)

以上説明した通り、本実施形態に係る情報処理装置は、カメラを有し、  
現実世界に固定された第1座標系における前記情報処理装置の第1向きを検知する検知部

10

前記カメラにより撮影されている撮影画像にユーザの顔画像が含まれていれば、前記撮影画像および前記顔画像から、前記情報処理装置に固定された第2座標系における前記ユーザの顔の第2向きを推定する推定部、

前記検知された第1向きと、前記推定された第2向きと、から、前記第1座標系における前記ユーザの顔の第3向きを算出する算出部、

前記算出された第3向きに応じた情報を出力する出力部  
を備えるように構成する。

【0140】

また、本実施形態に係る情報処理装置において、

前記情報処理装置は、前記ユーザが装着する音響機器に無線もしくは有線により接続され、

20

前記出力部は、前記情報を前記音響機器に出力するように構成することができる。

【0141】

また、本実施形態に係る情報処理装置において、

前記音響機器は、外音取り込み可能なヘッドフォン、イヤフォン、ネックスピーカー、骨伝導スピーカー、もしくは、補聴器である

ように構成することができる。

【0142】

また、本実施形態に係る情報処理装置において、前記出力部は、

30

仮想音源に対応付けられる仮想方位と、

前記算出された第3向きと、  
の角度差に応じた強度で前記仮想音源をミキシングした音声を前記情報として出力するように構成することができる。

【0143】

また、本実施形態に係る情報処理装置において、

前記仮想方位は、あらかじめ定められる

ように構成することができる。

【0144】

また、本実施形態に係る情報処理装置において、

40

前記検知部は、前記第1座標系における前記情報処理装置の第1位置をさらに検知し、

前記仮想音源に対応付けられる前記仮想方位は、

仮想音源が前記第1座標系において配置される仮想位置と、

前記検知された第1向きおよび第1位置と、

から、算出される

ように構成することができる。

【0145】

また、本実施形態に係る情報処理装置において、

前記情報処理装置は、前記検知された第1位置および第1向きに応じた映像情報を、前記カメラの撮影方向と同じ方向を表示方向とする画面に表示し、

50

前記顔画像の大きさに応じて、前記仮想音源の波形を補正するように構成することができる。

【0146】

また、本実施形態に係る情報処理装置において、前記撮影画像に前記ユーザの前記顔画像と、前記ユーザの手画像と、が含まれていれば、前記第2座標系における前記ユーザの前記顔と、前記ユーザの手と、の距離に応じて、前記仮想音源の波形を補正するように構成することができる。

【0147】

また、本実施形態に係る情報処理装置において、前記撮影画像に前記ユーザの手画像が含まれていれば、前記撮影画像の代表点と、前記手画像と、の距離に応じて、前記仮想音源の波形を補正するように構成することができる。

10

【0148】

また、本実施形態に係る情報処理装置において、前記撮影画像に前記顔画像が含まれていなければ、前記強度を既定値とするように構成することができる。

【0149】

また、本実施形態に係る情報処理装置において、前記仮想音源は、仮想空間に配置された発音オブジェクトに対応付けられ、前記情報処理装置は、前記発音オブジェクトが配置された前記検知された第1位置および第1向きに応じた視点位置および視線方向により観察した前記仮想空間の様子を、前記カメラの撮影方向と同じ方向を表示方向とする画面に表示し、

20

特定条件が満たされると、前記仮想音源に対応付けられる仮想方位と、前記算出された第3向きと、の角度差が最小の発音オブジェクトを、前記ユーザによる注目オブジェクトとして特定し、

前記出力部は、前記算出された第3向きに応じた情報を出力するのにかえて、前記特定された注目オブジェクトに対応付けられる仮想音源を、他の仮想音源に優先して出力し、

前記特定された注目オブジェクトを他の発音オブジェクトより強調して前記画面に表示し、

30

解除条件が満たされると、前記注目オブジェクトとしての特定を解除するように構成することができる。

【0150】

また、本実施形態に係る情報処理装置において、前記発音オブジェクトは、音声とともに再生される動画であり、前記注目オブジェクトが特定されている間、前記情報処理装置は、前記注目オブジェクトを前記画面内の所定位置に所定拡大率で表示し、前記出力部は、前記注目オブジェクトに対応付けられる仮想音源を所定の増幅率で、他の仮想音源をミュートして、ミキシングした音声を出し、

40

前記注目オブジェクトとしての特定が解除されると、前記情報処理装置は、前記注目オブジェクトとしての特定が解除された発音オブジェクトの仮想方位が前記算出された第3向きに一致するように、前記視点位置を中心に、前記仮想空間に配置された発音オブジェクトの仮想方位を回転させる

ように構成することができる。

【0151】

また、本実施形態に係る情報処理装置において、前記情報処理装置は、前記撮影画像に含まれる前記ユーザの手画像に基づくジェスチャー、もしくは、前記画面に対するタッチ操作に基づいて、前記視点位置を中心に、前記仮想空間に配置された発音オブジェクトの仮想方位を回転させる

50

ように構成することができる。

【0152】

また、本実施形態に係る情報処理装置において、

前記発音オブジェクトは、音声を発するアバターであり、

前記注目オブジェクトが特定されている間、前記情報処理装置は、前記撮影画像に含まれる前記ユーザの手画像に基づくジェスチャー、もしくは、前記画面に対するタッチ操作に基づいて、前記注目オブジェクトの前記仮想空間における位置もしくは向きを変化させるように構成することができる。

【0153】

本実施形態に係る情報処理方法は、カメラを有する情報処理装置が、

現実世界に固定された第1座標系における前記情報処理装置の第1向きを検知し、

前記カメラにより撮影されている撮影画像にユーザの顔画像が含まれていれば、前記撮影画像および前記顔画像から、前記情報処理装置に固定された第2座標系における前記ユーザの顔の第2向きを推定し、

前記検知された第1向きと、前記推定された第2向きと、から、前記第1座標系における前記ユーザの顔の第3向きを算出し、

前記算出された第3向きに応じた情報を出力する

ように構成する。

【0154】

本実施形態に係るプログラムは、カメラを有するコンピュータを、

現実世界に固定された第1座標系における前記情報処理装置の第1向きを検知する検知部

、前記カメラにより撮影されている撮影画像にユーザの顔画像が含まれていれば、前記撮影画像および前記顔画像から、前記情報処理装置に固定された第2座標系における前記ユーザの顔の第2向きを推定する推定部、

前記検知された第1向きと、前記推定された第2向きと、から、前記第1座標系における前記ユーザの顔の第3向きを算出する算出部、

前記算出された第3向きに応じた情報を出力する出力部

として機能させるように構成する。

【0155】

当該プログラムは、非一時的なコンピュータ読取可能な情報記録媒体に記録して配布、販売することができる。また、コンピュータ通信網等の一時的な伝送媒体を介して配布、販売することができる。

【0156】

本実施形態に係るコンピュータ読取可能な非一時的な情報記録媒体は、上記のプログラムが記録されるように構成する。

【0157】

本発明は、本発明の広義の精神と範囲を逸脱することなく、様々な実施の形態及び変形が可能とされるものである。また、上述した実施の形態は、この発明を説明するためのものであり、本発明の範囲を限定するものではない。すなわち、本発明の範囲は、実施の形態ではなく、特許請求の範囲によって示される。そして、特許請求の範囲内及びそれと同等の発明の意義の範囲内で施される様々な変形が、この発明の範囲内とみなされる。

本願においては、日本国に対して令和3年(2021年)4月20日(火)に出願した特許出願特願2021-070745を基礎とする優先権を主張するものとし、指定国の法令が許す限り、当該基礎出願の内容を本願に取り込むものとする。

【産業上の利用可能性】

【0158】

本発明によれば、現実世界におけるユーザの顔の向きを推定し、これに応じた情報を出力するための情報処理装置、情報処理方法、プログラム、ならびに、情報記録媒体を提供することができる。

10

20

30

40

50

## 【符号の説明】

## 【0159】

- 101 情報処理装置
- 111 検知部
- 112 推定部
- 113 算出部
- 114 出力部
- 151 カメラ
- 152 音響機器
- 153 画面
- 301 特定部
- 302 解除部

10

20

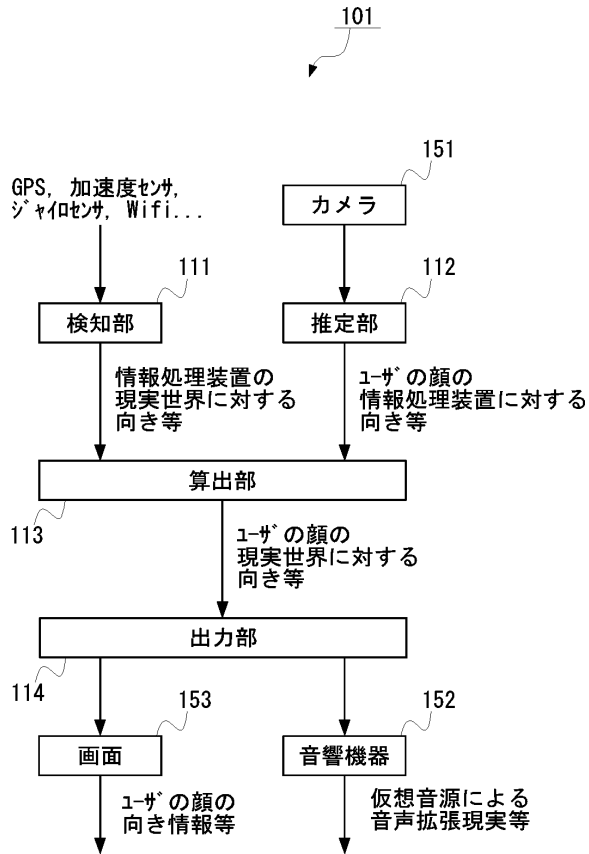
30

40

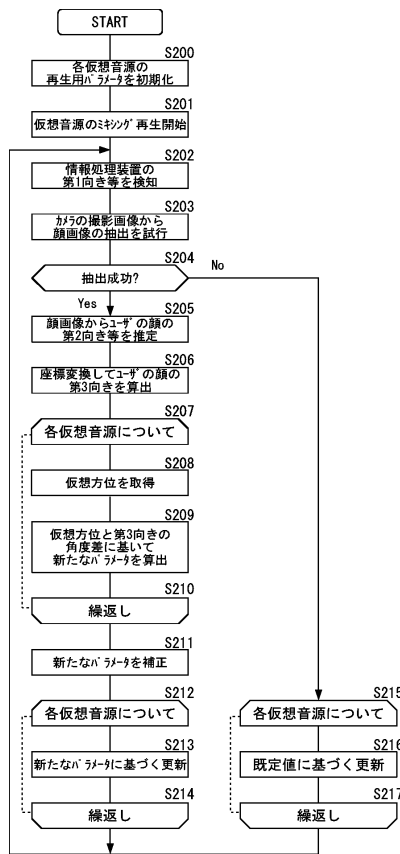
50

【図面】

【図 1】



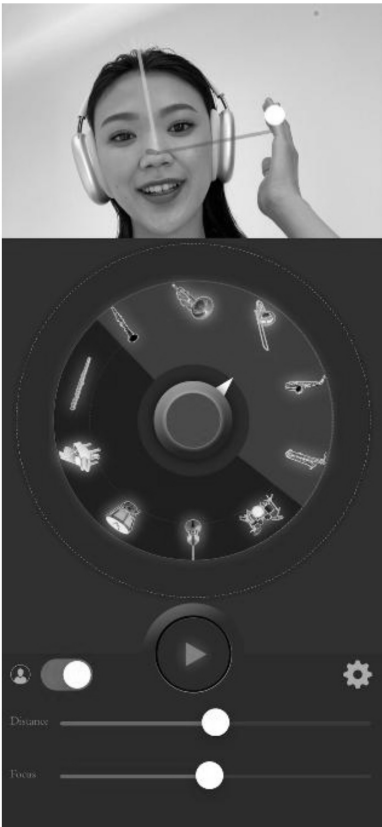
【図 2】



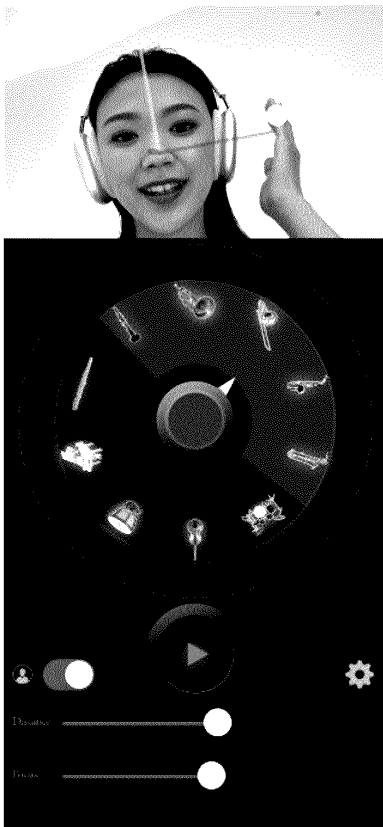
10

20

【図 3】



【図 4】

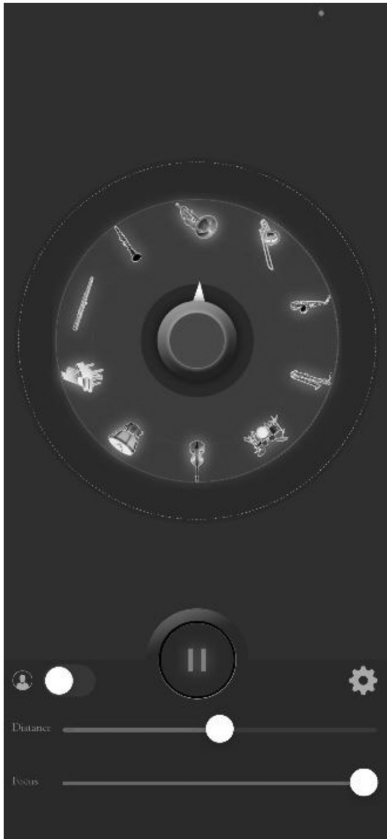


30

40

50

【図 5】



【図 6】



10

20

【図 7】



【図 8】

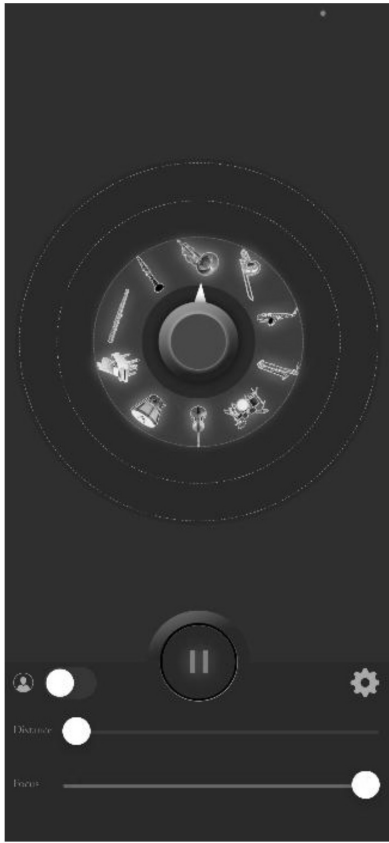


30

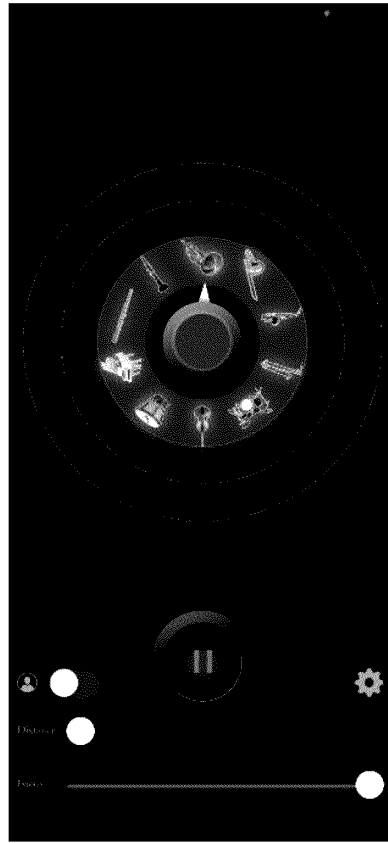
40

50

【図 9】



【図 10】



10

20

【図 11】



【図 12】

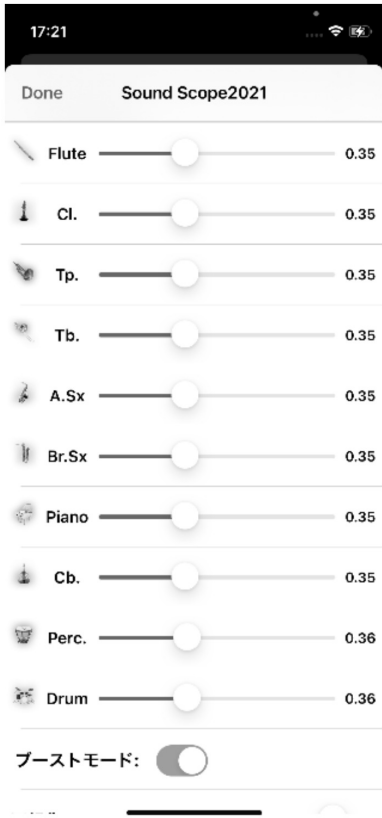


30

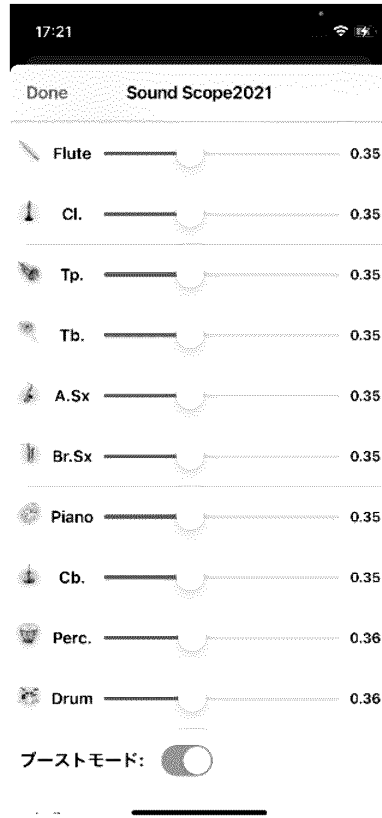
40

50

【 図 1 3 】



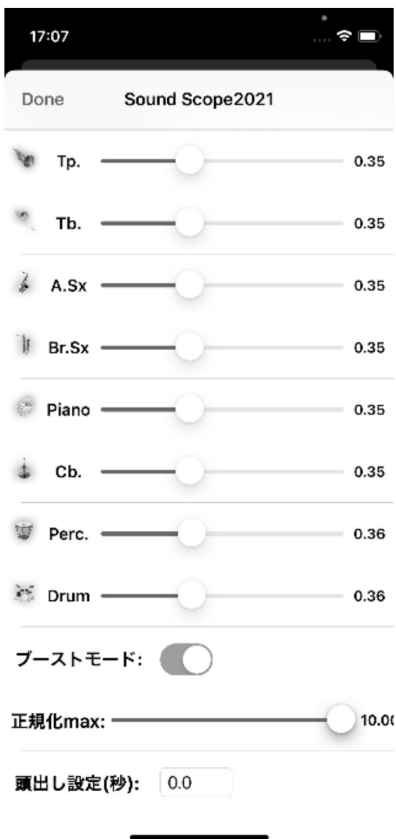
【 図 1 4 】



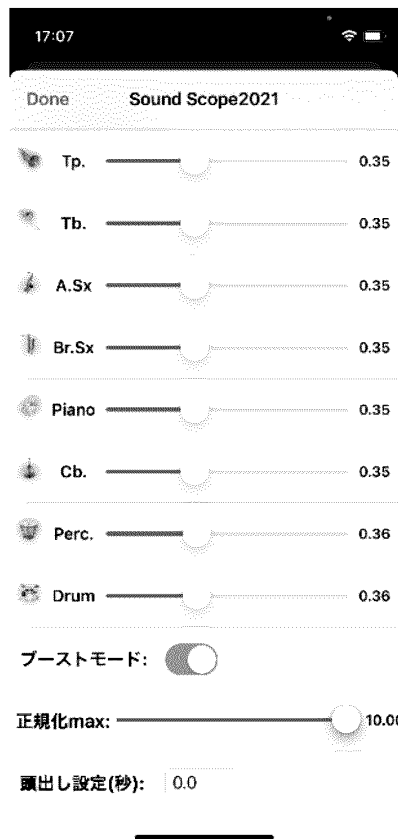
10

20

【 図 1 5 】



【 図 1 6 】



30

40

50

【図 17】



【図 18】



10

20

【図 19】



【図 20】

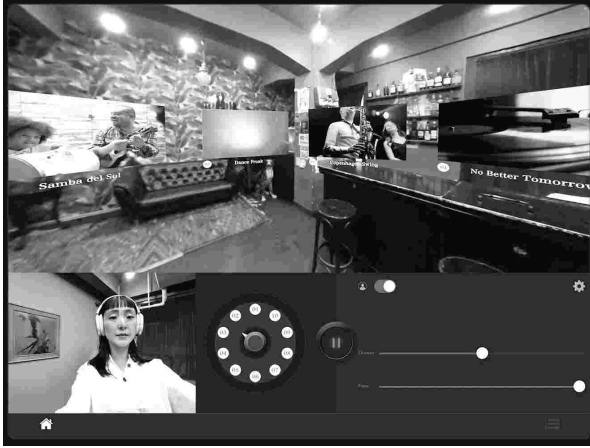


30

40

50

【 2 1 】

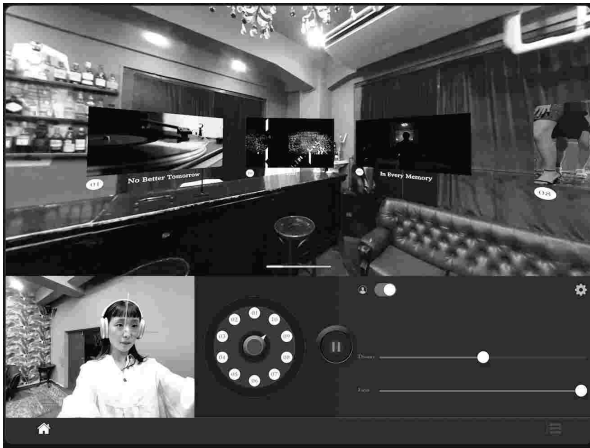


【 2 2 】

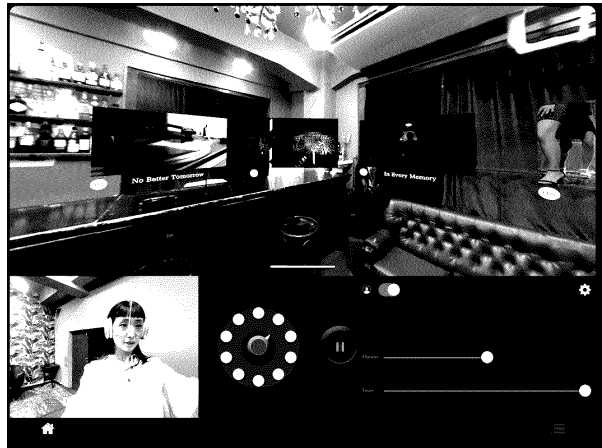


10

【 2 3 】



【 2 4 】



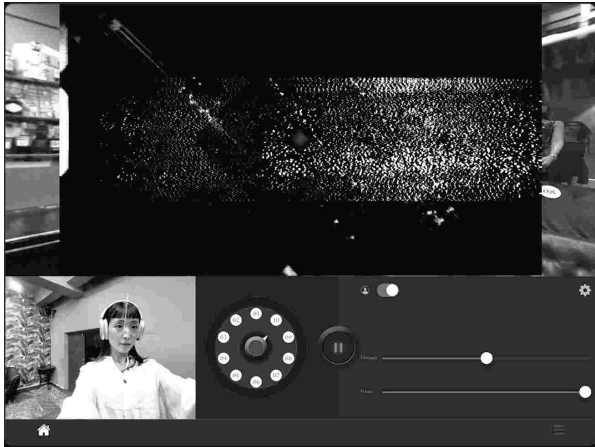
20

30

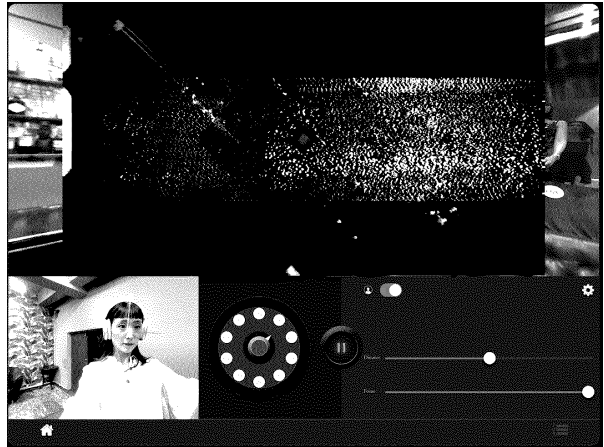
40

50

【図 25】

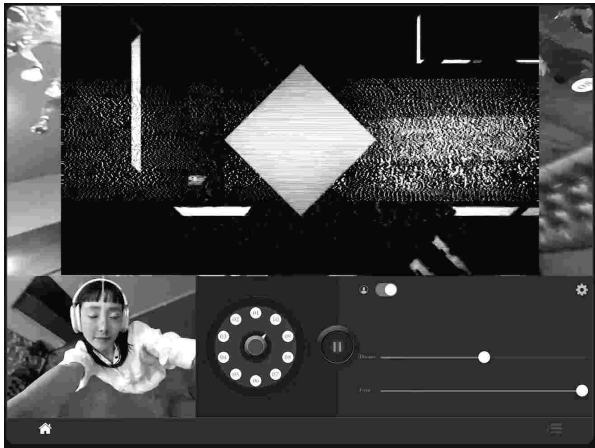


【図 26】

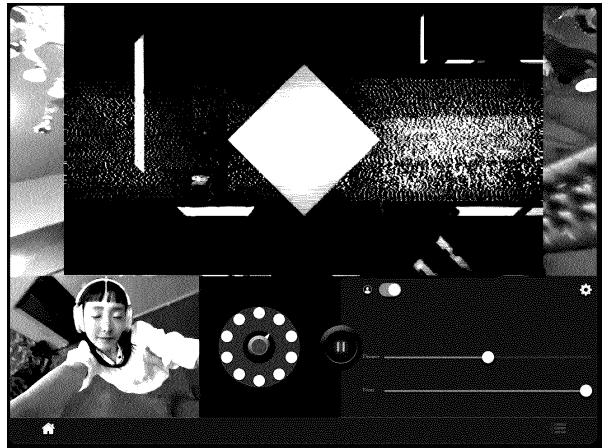


10

【図 27】



【図 28】



20

30

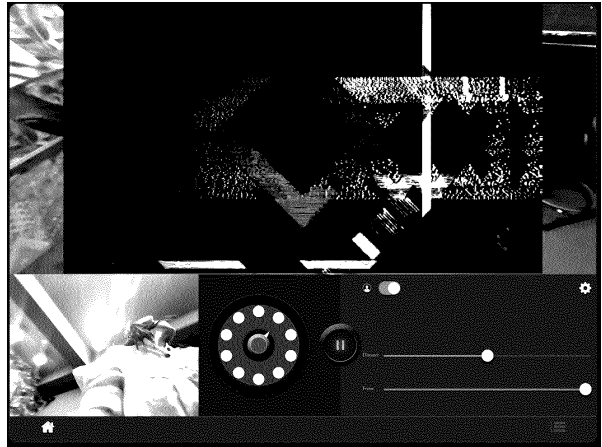
40

50

【図 29】

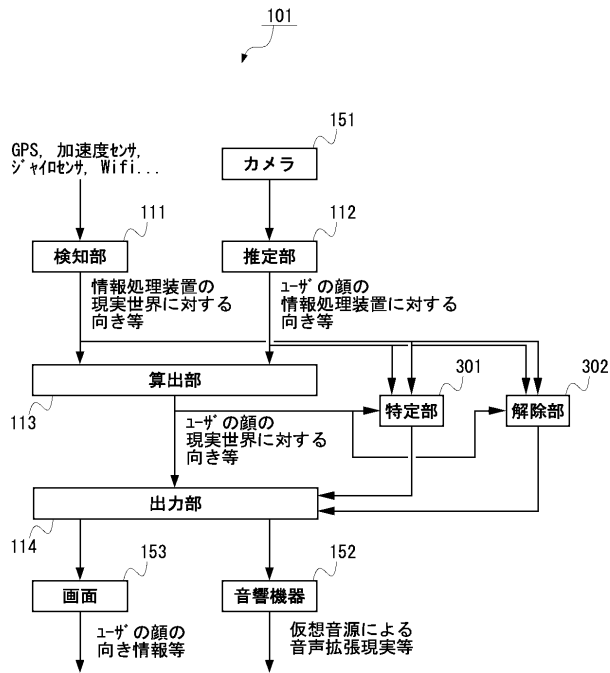


【図 30】



10

【図 31】



20

30

40

50

## フロントページの続き

- (56)参考文献 特開2008-92193(JP,A)  
特開2017-92732(JP,A)  
国際公開第2019/026597(WO,A1)  
特開2019-126033(JP,A)  
国際公開第2020/184021(WO,A1)  
米国特許出願公開第2019/0335288(US,A1)  
塚田学他,SDM3602:音楽イベントのための自由視聴点映像音声のインタラクティブ再生  
,マルチメディア,分散,協調とモバイル(DICOMO2017)シンポジウム論文集,一般社  
団法人情報処理学会,2017年06月,Vol.2017, No.1, pp.1460-1467
- (58)調査した分野 (Int.Cl.,DB名)  
H04S 1/00-7/00  
H04R 3/00