

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7603055号
(P7603055)

(45)発行日 令和6年12月19日(2024.12.19)

(24)登録日 令和6年12月11日(2024.12.11)

(51)国際特許分類

F I

H O 4 R 3/02 (2006.01)

H O 4 R 3/02

請求項の数 24 (全84頁)

(21)出願番号	特願2022-506120(P2022-506120)	(73)特許権者	507236292
(86)(22)出願日	令和2年7月29日(2020.7.29)		ドルビー ラボラトリーズ ライセンシン
(65)公表番号	特表2022-542962(P2022-542962		グ コーポレーション
	A)		アメリカ合衆国 9 4 1 0 3 カリフォル
(43)公表日	令和4年10月7日(2022.10.7)		ニア州 サンフランシスコ マーケット
(86)国際出願番号	PCT/US2020/043958		ストリート 1 2 7 5
(87)国際公開番号	WO2021/021857	(73)特許権者	510185767
(87)国際公開日	令和3年2月4日(2021.2.4)		ドルビー・インターナショナル・アーベ
審査請求日	令和5年7月28日(2023.7.28)		ー
(31)優先権主張番号	P201930702		アイルランド ディー0 2 ヴィーケイ6
(32)優先日	令和1年7月30日(2019.7.30)		0 ダブリン グランド カナル ドックラ
(33)優先権主張国・地域又は機関			ンズ サー ジョン ロジャーソンズ キー
	スペイン(ES)		7 7 ブロック シー
(31)優先権主張番号	62/880,113	(74)代理人	100101683
(32)優先日	令和1年7月30日(2019.7.30)		弁理士 奥田 誠司
最終頁に続く		最終頁に続く	

(54)【発明の名称】 分散配置されたオーディオデバイスのための音響エコー除去制御

(57)【特許請求の範囲】

【請求項1】

オーディオセッション管理方法であって、

オーディオ環境内の複数のマイクロフォンの各々からの出力信号を受信するステップであって、前記複数のマイクロフォンの各々は前記オーディオ環境のマイクロフォン位置に存在し、前記出力信号は人の現在の発声に対応する信号を含む、ステップと、

前記出力信号に基づき、前記人に関するコンテキスト情報の1つ以上のアスペクトを決定するステップであって、前記コンテキスト情報は、前記人の推定される現在位置、および、前記人の1つ以上のマイクロフォン位置に対する推定される現在の近接度のうちの少なくとも1つを含む、ステップと、

前記人の前記推定される現在位置に最も近い前記マイクロフォン位置に最も近い最近位ラウドスピーカ内蔵オーディオデバイスを決定するステップと、

前記コンテキスト情報の前記1つ以上のアスペクトに少なくとも部分的に基づき、前記オーディオ環境の2つ以上のオーディオデバイスを選択するステップであって、前記2つ以上のオーディオデバイスの各々は少なくとも1つのラウドスピーカを含み、前記2つ以上のオーディオデバイスは、前記最近位ラウドスピーカ内蔵オーディオデバイスを含む、ステップと、

前記2つ以上のオーディオデバイス用のラウドスピーカフィード信号へとレンダリングされているオーディオデータに適用する、1種類以上のオーディオ処理変更を決定するステップであって、前記オーディオ処理変更は、前記人の前記推定される現在位置に最も近

い前記マイクロフォンにおける発話対エコー比を増大する効果を有し、前記エコーは、前記2つ以上のオーディオデバイスにより出力されたオーディオ音声の少なくとも一部を含み、前記最近位ラウドスピーカ内蔵オーディオデバイスに対する前記オーディオ処理変更の少なくとも1つは、前記2つ以上のオーディオデバイスのうちの他のオーディオデバイスに対するオーディオ処理変更とは異なり、前記1種類以上のオーディオ処理変更は、前記最近位ラウドスピーカ内蔵オーディオデバイスのラウドスピーカ再生レベルを低下させる、ステップと、

前記1種類以上のオーディオ処理変更を適用させるステップと、を含む方法。

【請求項2】

前記1種類以上のオーディオ処理変更は、スペクトル修正を含む、請求項1に記載の方法。

10

【請求項3】

前記1種類以上のオーディオ処理変更は、前記2つ以上のオーディオデバイスの前記ラウドスピーカのラウドスピーカ再生レベルを低下させる、請求項1または2に記載の方法。

【請求項4】

前記オーディオ環境の2つ以上のオーディオデバイスを選択するステップは、前記オーディオ環境のN個のラウドスピーカ内蔵オーディオデバイスを選択することを含み、Nは2よりも大きい整数である、請求項1から3のいずれか1つに記載の方法。

【請求項5】

前記オーディオ環境の前記2つ以上のオーディオデバイスを選択するステップは、マイクロフォン位置およびラウドスピーカ内蔵オーディオデバイス位置の少なくとも一方に対する、前記人の推定される現在位置に少なくとも部分的に基づく、請求項1から4のいずれか1つに記載の方法。

20

【請求項6】

前記1種類以上のオーディオ処理変更は、前記人の前記推定される現在位置から離れる方向に、オーディオ信号のレンダリングをワーピングするように、レンダリングプロセスを変更することを含む、請求項1から5のいずれか1つに記載の方法。

【請求項7】

前記スペクトル修正は、500Hzと3kHzとの間の周波数帯域においてオーディオデータのレベルを低下させることを含む、請求項2に記載の方法。

30

【請求項8】

前記1種類以上のオーディオ処理変更は、オーディオ再生信号の少なくとも1つの選択された周波数帯域に、少なくとも1つのギャップを挿入することを含む、請求項1から7のいずれか1つに記載の方法。

【請求項9】

前記1種類以上のオーディオ処理変更は、ダイナミックレンジ圧縮を含む、請求項1から8のいずれか1つに記載の方法。

【請求項10】

前記2つ以上のオーディオデバイスを選択するステップは、1つ以上のマイクロフォン位置についての信号対エコー比推定値に少なくとも部分的に基づく、請求項1から9のいずれか1つに記載の方法。

40

【請求項11】

前記2つ以上のオーディオデバイスを選択するステップは、前記信号対エコー比推定値が信号対エコー比閾値以下であるか否かを判定することに少なくとも部分的に基づく、請求項10に記載の方法。

【請求項12】

前記1種類以上のオーディオ処理変更を決定するステップは、前記信号対エコー比推定値に少なくとも部分的に基づくコスト関数の最適化に基づく、請求項10に記載の方法。

【請求項13】

前記コスト関数は、レンダリングパフォーマンスに少なくとも部分的に基づく、請求項

50

12に記載の方法。

【請求項14】

前記2つ以上のオーディオデバイスを選択するステップは、近接度推定値に少なくとも部分的に基づく、請求項1から13のいずれか1つに記載の方法。

【請求項15】

各マイクロフォンの前記出力信号から複数の現在の音響的特徴を決定するステップと、前記複数の現在の音響的特徴に分類器を適用するステップとをさらに含み、

前記分類器を適用するステップは、前記オーディオ環境内の複数のユーザゾーン内で前記人によってなされた複数の過去の発声から得られた、過去に決定された音響的特徴に対して訓練されたモデルを適用することを含み、

前記人に関するコンテキスト情報の1つ以上のアスペクトを決定するステップは、前記分類器からの出力に少なくとも部分的に基づき、前記人が現在位置しているユーザゾーンの推定値を決定することを含む、請求項1から14のいずれか1つに記載の方法。

【請求項16】

前記ユーザゾーンの推定値は、前記複数のマイクロフォンの幾何学的位置を参照せずに決定される、請求項15に記載の方法。

【請求項17】

前記現在の発声および前記過去の発声は、ウェイクワードの発声を含む、請求項15または16に記載の方法。

【請求項18】

前記コンテキスト情報の前記1つ以上のアスペクトに応じて少なくとも1つのマイクロフォンを選択するステップをさらに含む、請求項1から17のいずれか1つに記載の方法。

【請求項19】

前記1つ以上のマイクロフォンは、前記オーディオ環境の複数のオーディオデバイス内に設けられている、請求項1から18のいずれか1つに記載の方法。

【請求項20】

前記1つ以上のマイクロフォンは、前記オーディオ環境の1つのオーディオデバイス内に設けられている、請求項1から19のいずれか1つに記載の方法。

【請求項21】

前記1つ以上のマイクロフォン位置のうち少なくとも1つは、1つのオーディオデバイスの複数のマイクロフォンに対応している、請求項1から20のいずれか1つに記載の方法。

【請求項22】

請求項1から21のいずれか1つに記載の方法を実行するように構成された装置。

【請求項23】

請求項1から21のいずれか1つに記載の方法を実行するように構成されたシステム。

【請求項24】

1つ以上のデバイスを制御して、請求項1から21のいずれか1つに記載の方法を実行させる命令を含む、コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

〔関連出願への相互参照〕

本願は、2020年7月21日付け出願の米国仮特許出願第62/705,897号、2020年6月25日付け出願の米国仮特許出願第62/705,410号、2020年2月7日付け出願の米国仮特許出願第62/971,421号、2019年12月18日付け出願の米国仮特許出願第62/950,004号、2019年7月30日付け出願の米国仮特許出願第62/880,122号、2019年7月30日付け出願の米国仮特許出願第62/880,113号、2019年11月29日付け出願の欧州特許出願第19212391.7号および2019年7月30日付け出願のスペイン特許出願第P201

10

20

30

40

50

930702号に基づく優先権を主張するものであり、各出願の開示内容を全て本願に援用する。

【0002】

本願は、複数のオーディオデバイス（例えばスマートオーディオデバイス）のコーディネート（オーケストレーション）および実装を行い、かつ、オーディオデバイスによるオーディオ音声のレンダリングを制御する、システムおよび方法に関する。

【背景技術】

【0003】

オーディオデバイス（スマートオーディオデバイスを含むが、それらに限定されない）は、広く用いられており、多くの家庭において一般的な要素になりつつある。オーディオデバイスを制御する既存のシステムおよび方法は利益を提供するが、改良されたシステムおよび方法が望まれる。

【0004】

【表記と命名】

特許請求の範囲を含む本開示全体を通じて、「スピーカ」および「ラウドスピーカ」は、同義的に、単一のスピーカフィードによって駆動される任意の音響放射トランスデューサ（またはトランスデューサのセット）を示すように使用される。典型的なヘッドフォンセットは、2つのスピーカを含む。スピーカは、単一の共通のスピーカフィードまたは複数のスピーカフィードによって駆動されるような、複数のトランスデューサ（例えばウーファーとツイーター）を含むように実装され得る。いくつかの例において、スピーカフィード（単数または複数）は、異なるトランスデューサに接続された異なる回路ブランチにおいて異なる処理を受けてもよい。

【0005】

特許請求の範囲を含む本開示全体を通じて、信号またはデータに対して演算（例えば、信号またはデータに対するフィルタリング、スケーリング、変換、またはゲインの適用）を「行う」という表現は、信号またはデータに対して直接演算を行うこと、または信号またはデータの処理済みバージョン（例えば、演算の実行を受ける前に予備フィルタリングまたは前処理されたバージョンの信号）に対して演算を行うことの意味において広義で使用される。

【0006】

特許請求の範囲を含む本開示全体を通じて、「システム」という表現は、デバイス、システム、またはサブシステムの意味において広義で使用される。例えば、デコーダを実装するサブシステムは、デコーダシステムと呼ばれることがあり、そのようなサブシステムを含むシステム（例えば、複数の入力に応答してX個の出力信号を生成するシステムであって、入力のうちM個をサブシステムが生成し、他のX-M個の入力が外部ソースから受信される）は、デコーダシステムとも呼ばれ得る。

【0007】

特許請求の範囲を含む本開示全体を通じて、「プロセッサ」という用語は、データ（例えば、オーディオ、またはビデオもしくは他の画像データ）に対する演算を実行するためにプログラマブルであるかまたは他の方法で（例えば、ソフトウェアまたはファームウェアによって）構成可能なシステムまたはデバイスの意味において広義で使用される。プロセッサの例としては、フィールドプログラマブルゲートアレイ（または他の構成可能な集積回路またはチップセット）、オーディオまたは他のサウンドデータに対してパイプライン化処理を行うようにプログラムおよび／または他の方法で構成されたデジタルシグナルプロセッサ、プログラマブルな汎用プロセッサまたはコンピュータ、およびプログラマブルなマイクロプロセッサチップまたはチップセットなどが挙げられる。

【0008】

特許請求の範囲を含む本開示全体を通して、「接続する（couples）」または「接続された（coupled）」という用語は、直接的接続または間接的接続のいずれかを意味するように用いられる。したがって、第1のデバイスが第2のデバイスに接続される場合、この

10

20

30

40

50

接続は、直接的接続により実現されてもよいし、他のデバイスおよび接続を介した間接的接続により実現されてもよい。

【0009】

本明細書中において、「スマートデバイス」は、概して、ある程度対話的かつ／または自律的に動作可能な、ブルートゥース、ジグビー (Zigbee)、近距離無線通信、Wi-Fi、Li-Fi (light fidelity)、3G、4G、5Gなどの各種無線プロトコルを介して1つ以上の他のデバイス（またはネットワーク）と通信するように構成された電子デバイスである。スマートデバイスの、いくつかの代表的な例として、スマートフォン、スマートカー、スマートサーモスタット、スマートドアベル、スマートロック、スマート冷蔵庫、ファブレットおよびタブレット、スマートウォッチ、スマートバンド、スマートキーチェーン、ならびにスマートオーディオデバイスが挙げられる。「スマートデバイス」という用語は、人工知能などのユビキタスコンピューティングのいくつかの特性を示すデバイスを指してもよい。

10

【0010】

本明細書中において、「スマートオーディオデバイス」という表現は、単一目的オーディオデバイスまたは多目的オーディオデバイス（例えば、バーチャルアシスタント機能の少なくともいくつかの側面を実装するオーディオデバイス）のいずれかであるスマートデバイスを表すために用いられる。単一目的オーディオデバイスとは、少なくとも1つのマイクロフォンを含むかまたはそれに接続される（任意に少なくとも1つのスピーカおよび／またはカメラも含むかまたはそれに接続される）デバイス（例えば、スマートスピーカ、テレビ (TV) または携帯電話）であり、単一の目的を達成するために概してまたは主として設計されているデバイスである。例えば、TVは典型的には、プログラム素材からのオーディオを再生することができる（再生できると考えられている）が、ほとんどの場合、現代のTVは、何らかのオペレーティングシステムを実行し、その上で、テレビを見るためのアプリケーションを含む複数のアプリケーションがローカルに実行される。同様に、携帯電話機のオーディオ入力と出力は多くのことを行い得るが、これらは当該電話機上で実行されているアプリケーションによって提供されている。この意味で、スピーカ（単数または複数）およびマイクロフォン（単数または複数）を有する単一目的オーディオデバイスは、スピーカ（単数または複数）およびマイクロフォン（単数または複数）を直接使用するためのローカルアプリケーションおよび／またはサービスを実行するように構成されることが多い。ゾーンすなわちユーザー設定されたエリアにわたってオーディオの再生を実現するためにグループ化するように構成された、単一目的オーディオデバイスもある。

20

30

【0011】

多目的オーディオデバイスの1つの一般的な種類として、バーチャルアシスタント機能の少なくともいくつかの側面を実装するオーディオデバイスが挙げられるが、バーチャルアシスタント機能の他の側面は、多目的オーディオデバイスが通信を行うように構成された1つ以上のサーバなどの、1つ以上の他のデバイスにより実装されてもよい。本明細書において、そのような多目的オーディオデバイスを、「バーチャルアシスタント」と呼んでもよい。バーチャルアシスタントは、少なくとも1つのマイクロフォンを含むか、またはこのマイクロフォンに接続された（さらに、任意に、少なくとも1つのスピーカおよび／または少なくとも1つのカメラを含むか、もしくはそれ（それら）に接続された）デバイス（例えば、スマートスピーカまたは音声アシスタント内蔵デバイス）である。いくつかの例において、バーチャルアシスタントは、ある意味においてクラウドイネーブルなアプリケーション、もしくは、バーチャルアシスタント自体において完全には実装されないアプリケーションのために、（バーチャルアシスタントとは別個の）複数のデバイスを利用する能力を提供する。換言すると、バーチャルアシスタント機能の少なくともいくつかの側面、例えば音声認識機能は、インターネットなどのネットワークを介してバーチャルアシスタントが通信し得る1つ以上のサーバまたは他のデバイスにより、（少なくとも部分的に）実装されてもよい。複数のバーチャルアシスタントが、例えば離散的かつ条件

40

50

付きで定義された方法で、協働することがあってもよい。例えば、2つ以上のバーチャルアシスタントが、それらのうち1つ（すなわちウェイクワードを聞いたことを最も確信している1つ）が、そのウェイクワードに応答するという意味において、協働し得る。いくつかの実施態様において、コネクテッド状態にある複数のデバイスが、1つのメインアプリケーションによって管理される、一種の集合体を形成してもよい。その1つのメインアプリケーションは、バーチャルアシスタントであり得る（または、実装し得る）。

【0012】

本明細書中において、「ウェイクワード」とは、任意の音（例えば、人間によって発せられた単語、または他の何らかの音）の意味において広義で使用される。スマートオーディオデバイスは、（スマートオーディオデバイスに含まれるかまたは接続された少なくとも1つのマイクロフォン、または少なくとも1つの他のマイクロフォンを用いた）音の検出（「聞き取り（hearing）」）に応答して、目覚めるよう構成される。この文脈において「目覚める（awake）」とは、デバイスがサウンドコマンドを待つ（すなわち、耳を立てている）状態に入ることを表す。いくつかの事例において、本明細書において「ウェイクワード」と呼ばれ得るものは、複数のワード、例えばフレーズを含むことができる。

【0013】

本明細書中において、「ウェイクワード検出器」という表現は、リアルタイムのサウンド（例えば、発話）特徴と学習済みモデルとの間の整合性を連続的に探索するように構成されたデバイス（またはデバイスを構成するための命令を含むソフトウェア）を表す。典型的には、ウェイクワードイベントは、ウェイクワードが検出された確率が事前に定義された閾値を超えているとウェイクワード検出器によって判断されるたびに、トリガされる。例えば閾値は、他人受入率と本人拒否率との間の妥当な妥協点を与えるように調整された、所定の閾値であってもよい。ウェイクワードイベントの後、デバイスはコマンドに耳を立てる状態（「目覚めた（awakened）」状態または「アテンティブネス（attentiveness）」状態と呼ばれることがある）に入り、この状態において、受け取ったコマンドをより大規模でより計算集約的な認識器に渡し得る。

【0014】

本明細書において、「マイクロフォン位置」という表現は、1つ以上のマイクロフォンの位置を指す。いくつかの例において、1つのマイクロフォン位置は、1つのオーディオデバイス内に設けられた複数のマイクロフォンからなる1つのマイクロフォンアレイに対応していてもよい。例えば、マイクロフォン位置は、1つ以上のマイクロフォンを含むオーディオデバイス全体に対応する1つの位置であってもよい。いくつかのそのような例において、マイクロフォン位置は、1つのオーディオデバイスのマイクロフォンアレイの重心に対応する1つの位置であってもよい。しかし、いくつかの例において、マイクロフォン位置は、1つのマイクロフォンの位置であってもよい。いくつかのそのような例において、オーディオデバイスが有するマイクロフォンの数は、1つのみであってもよい。

【発明の概要】

【0015】

いくつかの開示された実施形態は、1つ以上のオーディオデバイスにおける全二重通信（full duplex）を成功させるための鍵となる基準を改善する、聴者または「ユーザ」エクスペリエンスの管理へのアプローチを提供する。この基準は、信号対エコー比（Signal to Echo ratio:SER）として知られており、また発話対エコー比（Speech to Echo Ratio）と呼ぶこともある。これは、1つ以上のマイクロフォンを介して環境（例えば部屋）からキャプチャされるボイス信号（または他の所望の信号）と、出力されたプログラムコンテンツ、対話型コンテンツ等から1つ以上のマイクロフォンを備えたオーディオデバイスにおいて提示されるエコーとの比として定義されてもよい。オーディオ環境の多数のオーディオデバイスは、ラウドスピーカおよびマイクロフォンの両方を内蔵している一方で、他の機能にも関わっているということが考えられる。しかし、このオーディオ環境の他のオーディオデバイスは、1つ以上のラウドスピーカを有するがマイクロフォンを備えていないかもしれないし、あるいは、1つ以上のマイクロフォンを有するがラウドスピーカを

10

20

30

40

50

備えていないかもしれない。いくつかの実施形態は、所与のユースケースまたはシナリオにおいて、ユーザの最も近くのラウドスピーカ（単数または複数）を使用することを故意に避けている（または、ユーザの最も近くのラウドスピーカを主として使用しない）。代替的にまたは追加的に、いくつかの実施形態は、オーディオ環境の1つ以上のマイクロフォンにおいてSERを増大させるために、オーディオ環境の1つ以上のラウドスピーカによりレンダリングされたオーディオデータに対して、1以上の他の種類のオーディオ処理変更を起こしてもよい。

【0016】

いくつかの実施形態は、コーディネート（オーケストレーション）されたオーディオデバイスを含むシステムを実装するように構成されている。いくつかの実施態様において、オーディオデバイスは、スマートオーディオデバイスを含み得る。いくつかのそのような実施態様によると、複数のスマートオーディオデバイスのうち2つ以上のスマートオーディオデバイスは、ウェイクワード検出器である（または、ウェイクワード検出器を実装するように構成されている）。したがって、そのような例において、複数のマイクロフォン（例えば、非同期マイクロフォン）が利用可能である。いくつかの例において、各マイクロフォンは、スマートオーディオデバイスのうち少なくとも1つに含まれていてもよいし、スマートオーディオデバイスのうち少なくとも1つと通信するように構成されてもよい。例えば、マイクロフォンの少なくともいくつかは、スマートオーディオデバイスのいずれにも含まれないが、スマートオーディオデバイスの少なくとも1つと通信するように（よって、マイクロフォンの出力が、スマートオーディオデバイスの少なくとも1つによりキャプチャ可能となるように）構成された、（例えば、家電機器内の）独立したマイクロフォンであってもよい。いくつかの実施形態において、各ウェイクワード検出器（またはウェイクワード検出器を含む各スマートオーディオデバイス）、もしくはシステムの別のサブシステム（例えば分類器）は、少なくともいくつかのマイクロフォン（例えば非同期マイクロフォン）から得られた複数の音響的特徴により駆動される分類器を適用することにより、人のゾーンを推定するように構成されている。いくつかの実施態様では、その目的は、人の正確な位置を推定することではなく、人の現在位置を含む離散ゾーン（discrete zone）のロバスト推定を形成することであってもよい。

【0017】

いくつかの実施態様において、人（本明細書中、「ユーザ」と呼ぶこともある）、スマートオーディオデバイスおよびマイクロフォンは、オーディオ環境（例えば、ユーザの住居、自動車または職場）内に存在する。このオーディオ環境内において、音声はユーザからマイクロフォンへと伝搬し得る。オーディオ環境は、複数の所定のゾーンを含んでもよい。いくつかの例によると、環境は、少なくとも次のゾーンを含んでもよい：調理エリア、ダイニングエリア、生活空間のオープンエリア、生活空間のテレビエリア（テレビソファを含む）等。システムの動作中、ユーザは任意の時刻において前記ゾーンのうちの1つに物理的に存在し（ユーザのゾーン）、ユーザのゾーンは時間により異なり得る、と仮定する。

【0018】

いくつかの例において、マイクロフォンは、非同期であってもよく（例えば、異なるサンプリングクロックを用いてデジタルサンプリングされてもよく）、かつ、ランダムに配置されてもよい（または、少なくとも所定の位置に配置されなくてもよい、対称に配置されなくてもよい、グリッド状に配置されなくてもよい、等）。いくつかの例において、ユーザのゾーンは、少なくとも部分的にウェイクワード検出器のうちの少なくとも1つから得られた複数の高レベルの特徴により駆動される、データ駆動型のアプローチを介して推定されてもよい。これらの特徴（例えば、ウェイクワード確信度および受信レベル）は、いくつかの例において、ほとんど帯域幅を使わないかもしれないし、また、非常に少ないネットワーク負荷で分類器を実装するデバイスへと（例えば非同期的に）送信されてもよい。

【発明が解決しようとする課題】

【0019】

いくつかの実施形態の側面は、スマートオーディオデバイスを実装すること、および／または、スマートオーディオデバイスをコーディネートさせることに関する。

【課題を解決するための手段】

【0020】

いくつかの開示された実施態様の側面は、1つ以上の開示された方法またはそのステップを実行するように構成された（例えば、プログラムされた）システムと、1つ以上の開示された方法またはそのステップを実行するためのコード（例えば1つ以上の開示された方法またはそのステップを実行するように実行可能なコード）を格納した、非一時的なデータの格納を実装する、有形の非一時的なコンピュータ読取可能媒体（例えば、ディスクまたは他の有形の記憶媒体）とを含む。例えば、いくつかの開示された実施形態は、1つ以上の開示された方法またはそのステップを含め、1つ以上の開示された方法またはそのステップを含むデータに対して各種演算のいずれかを実行するようにソフトウェアまたはファームウェアでプログラムされた、かつ／またはその他構成された、プログラム可能な汎用プロセッサ、デジタルシグナルプロセッサ、またはマイクロプロセッサであるか、もしくは、そのプロセッサを含み得る。そのような汎用プロセッサは、入力装置と、メモリと、アサートされたデータに回答して1つ以上の開示された方法（またはそのステップ）を実行するようにプログラムされた（かつ／またはその他構成された）処理用サブシステムとを含むコンピュータシステムであってもよいし、そのコンピュータシステムを含んでもよい。

【0021】

いくつかの実施態様において、制御システムは、1つ以上のオーディオセッション管理方法などの、本明細書中に開示された1つ以上の方法を実装するように構成されてもよい。いくつかのそのような方法は、オーディオ環境内の複数のマイクロフォンの各々からの出力信号を（例えば制御システムにより）受信するステップを含む。いくつかの例において、前記複数のマイクロフォンの各々は、前記オーディオ環境のマイクロフォン位置に存在する。いくつかの例において、前記出力信号は、人の現在の発声に対応する信号を含む。いくつかの例によると、前記出力信号は、雑音および／またはエコー等の、発話でないオーディオデータに対応する信号を含む。

【0022】

いくつかのそのような方法は、前記出力信号に基づき、前記人に関するコンテキスト情報の1つ以上のアスペクトを（例えば制御システムにより）決定するステップを含む。いくつかの例において、コンテキスト情報は、前記人の推定される現在位置および／または前記人の1つ以上のマイクロフォン位置に対する推定される現在の近接度を含む。いくつかのそのような方法は、前記コンテキスト情報の前記1つ以上のアスペクトに少なくとも部分的に基づき、前記オーディオ環境の2つ以上のオーディオデバイスを選択するステップを含む。いくつかの実施態様において、2つ以上のオーディオデバイスの各々は、少なくとも1つのラウドスピーカを含む。

【0023】

いくつかのそのような方法は、前記2つ以上のオーディオデバイス用のラウドスピーカフィード信号へとレンダリングされているオーディオデータに適用する、1種類以上のオーディオ処理変更を（例えば制御システムにより）決定するステップを含む。いくつかの例において、前記オーディオ処理変更は、1つ以上のマイクロフォンにおける発話対エコー比を増大する効果を有する。いくつかのそのような方法は、前記1種類以上のオーディオ処理変更を適用させるステップを含む。

【0024】

いくつかの実施態様によると、前記1種類以上のオーディオ処理変更は、前記2つ以上のオーディオデバイスの前記ラウドスピーカのラウドスピーカ再生レベルの低下を生じてもよい。いくつかの実施態様において、第1のオーディオデバイスに対する前記オーディオ処理変更の少なくとも1つは、第2のオーディオデバイスに対するオーディオ処理変更

とは異なってもよい。いくつかの例において、前記オーディオ環境の2つ以上のオーディオデバイスを（例えば制御システムにより）選択するステップは、前記オーディオ環境のN個のラウドスピーカ内蔵オーディオデバイスを選択することを含んでもよく、Nは2よりも大きい整数である。

【0025】

いくつかの実施態様において、前記オーディオ環境の前記2つ以上のオーディオデバイスを選択するステップは、マイクロフォン位置およびラウドスピーカ内蔵オーディオデバイス位置の少なくとも一方に対する、前記人の推定される現在位置に少なくとも部分的に基づいてもよい。いくつかのそのような実施態様によると、前記方法は、前記人の前記推定される現在位置に最も近い最近位ラウドスピーカ内蔵オーディオデバイス、または、前記人の前記推定される現在位置に最も近い前記マイクロフォン位置に最も近い最近位ラウドスピーカ内蔵オーディオデバイス、を決定するステップを含んでもよい。いくつかのそのような例において、前記2つ以上のオーディオデバイスは、前記最近位ラウドスピーカ内蔵オーディオデバイスを含んでもよい。

10

【0026】

いくつかの例において、前記1種類以上のオーディオ処理変更は、前記人の前記推定される現在位置から離れる方向に、オーディオ信号のレンダリングをワーピングするように、レンダリングプロセスを変更することを含む。いくつかの実施態様において、前記1種類以上のオーディオ処理変更は、スペクトル修正を含んでもよい。いくつかのそのような実施態様によると、前記スペクトル修正は、500 Hzと3 KHzとの間の周波数帯域においてオーディオデータのレベルを低下させることを含んでもよい。

20

【0027】

いくつかの実施態様において、前記1種類以上のオーディオ処理変更は、オーディオ再生信号の少なくとも1つの選択された周波数帯域に、少なくとも1つのギャップを挿入することを含んでもよい。いくつかの例において、前記1種類以上のオーディオ処理変更は、ダイナミックレンジ圧縮を含んでもよい。

【0028】

いくつかの実施態様によると、前記2つ以上のオーディオデバイスを選択するステップは、1つ以上のマイクロフォン位置についての信号対エコー比推定値に少なくとも部分的に基づいてもよい。例えば、前記2つ以上のオーディオデバイスを選択するステップは、前記信号対エコー比推定値が信号対エコー比閾値以下であるか否かを判定することに少なくとも部分的に基づいてもよい。いくつかの例において、前記1種類以上のオーディオ処理変更を決定するステップは、前記信号対エコー比推定値に少なくとも部分的に基づくコスト関数の最適化に基づいてもよい。例えば、前記コスト関数は、レンダリングパフォーマンスに少なくとも部分的に基づいてもよい。いくつかの実施態様において、前記2つ以上のオーディオデバイスを選択するステップは、近接度推定値に少なくとも部分的に基づいてもよい。

30

【0029】

いくつかの例において、前記方法は、各マイクロフォンの前記出力信号から複数の現在の音響的特徴を（例えば制御システムにより）決定するステップと、前記複数の現在の音響的特徴に分類器を適用するステップとを含んでもよい。いくつかの実施態様によると、前記分類器を適用するステップは、前記環境内の複数のユーザゾーン内で前記人によってなされた複数の過去の発声から得られた、過去に決定された音響的特徴に対して訓練されたモデルを適用することを含んでもよい。

40

【0030】

いくつかのそのような例において、前記人に関するコンテキスト情報の1つ以上のアスペクトを決定するステップは、前記分類器からの出力に少なくとも部分的に基づき、前記人が現在位置しているユーザゾーンの推定値を決定することを含んでもよい。いくつかの実施態様によると、前記ユーザゾーンの推定値は、前記複数のマイクロフォンの幾何学的位置を参照せずに決定されてもよい。いくつかの例において、前記現在の発声および

50

前記過去の発声は、ウェイクワードの発声であってもよいし、ウェイクワードの発声を含んでもよい。

【0031】

いくつかの実施態様によると、前記1つ以上のマイクロフォンは、前記オーディオ環境の複数のオーディオデバイス内に設けられてもよい。しかし、別の例において、前記1つ以上のマイクロフォンは、前記オーディオ環境の1つのオーディオデバイス内に設けられてもよい。いくつかの例において、前記1つ以上のマイクロフォン位置のうち少なくとも1つは、1つのオーディオデバイスの複数のマイクロフォンに対応していてもよい。いくつかの開示された方法は、前記コンテキスト情報の前記1つ以上のアスペクトに応じて少なくとも1つのマイクロフォンを選択するステップを含んでもよい。

10

【0032】

本開示の少なくともいくつかの側面は、オーディオセッション管理方法などの方法により実装されてもよい。本明細書中の別の場所で示すように、いくつかの例において、前記方法は、少なくとも部分的には、本明細書中に開示されているような制御方法により実装されてもよい。いくつかのそのような方法は、オーディオ環境内の複数のマイクロフォンの各々から出力信号を受信するステップを含む。いくつかの例において、前記複数のマイクロフォンの各々は、前記オーディオ環境のマイクロフォン位置に存在する。いくつかの例において、前記出力信号は、人の現在の発声に対応する信号を含む。いくつかの例によると、前記出力信号は、雑音および／またはエコー等の、発話でないオーディオデータに対応する信号を含む。

20

【0033】

いくつかのそのような方法は、前記出力信号に基づき、前記人に関するコンテキスト情報の1つ以上のアスペクトを決定するステップを含む。いくつかの例において、コンテキスト情報は、前記人の推定される現在位置および／または前記人の1つ以上のマイクロフォン位置に対する推定される現在の近接度を含む。いくつかのそのような方法は、前記コンテキスト情報の前記1つ以上のアスペクトに少なくとも部分的に基づき、前記オーディオ環境の2つ以上のオーディオデバイスを選択するステップを含む。いくつかの実施態様において、2つ以上のオーディオデバイスの各々は、少なくとも1つのラウドスピーカを含む。

【0034】

いくつかのそのような方法は、前記2つ以上のオーディオデバイス用のラウドスピーカフィード信号へとレンダリングされているオーディオデータに適用する、1種類以上のオーディオ処理変更を決定するステップを含む。いくつかの例において、前記オーディオ処理変更は、1つ以上のマイクロフォンにおける発話対エコー比を増大する効果を有する。いくつかのそのような方法は、前記1種類以上のオーディオ処理変更を適用させるステップを含む。

30

【0035】

いくつかの実施態様によると、前記1種類以上のオーディオ処理変更は、前記2つ以上のオーディオデバイスの前記ラウドスピーカのラウドスピーカ再生レベルの低下を生じてもよい。いくつかの実施態様において、第1のオーディオデバイスに対する前記オーディオ処理変更の少なくとも1つは、第2のオーディオデバイスに対するオーディオ処理変更とは異なってもよい。いくつかの例において、前記オーディオ環境の2つ以上のオーディオデバイスを選択するステップは、前記オーディオ環境のN個のラウドスピーカ内蔵オーディオデバイスを選択することを含んでもよく、Nは2よりも大きい整数である。

40

【0036】

いくつかの実施態様において、前記オーディオ環境の前記2つ以上のオーディオデバイスを選択するステップは、マイクロフォン位置およびラウドスピーカ内蔵オーディオデバイス位置の少なくとも一方に対する、前記人の推定される現在位置に少なくとも部分的に基づいてもよい。いくつかのそのような実施態様によると、前記方法は、前記人の前記推定される現在位置に最も近い最近位ラウドスピーカ内蔵オーディオデバイス、または、前

50

記人の前記推定される現在位置に最も近い前記マイクロフォン位置に最も近い最近位ラウドスピーカ内蔵オーディオデバイス、を決定するステップを含んでもよい。いくつかのそのような例において、前記2つ以上のオーディオデバイスは、前記最近位ラウドスピーカ内蔵オーディオデバイスを含んでもよい。

【0037】

いくつかの例において、前記1種類以上のオーディオ処理変更は、前記人の前記推定される現在位置から離れる方向に、オーディオ信号のレンダリングをワーピングするように、レンダリングプロセスを変更することを含む。いくつかの実施態様において、前記1種類以上のオーディオ処理変更は、スペクトル修正を含んでもよい。いくつかのそのような実施態様によると、前記スペクトル修正は、500Hzと3KHzとの間の周波数帯域においてオーディオデータのレベルを低下させることを含んでもよい。

10

【0038】

いくつかの実施態様において、前記1種類以上のオーディオ処理変更は、オーディオ再生信号の少なくとも1つの選択された周波数帯域に、少なくとも1つのギャップを挿入することを含んでもよい。いくつかの例において、前記1種類以上のオーディオ処理変更は、ダイナミックレンジ圧縮を含んでもよい。

【0039】

いくつかの実施態様によると、前記2つ以上のオーディオデバイスを選択するステップは、1つ以上のマイクロフォン位置についての信号対エコー比推定値に少なくとも部分的に基づいてもよい。例えば、前記2つ以上のオーディオデバイスを選択するステップは、前記信号対エコー比推定値が信号対エコー比閾値以下であるか否かを判定することに少なくとも部分的に基づいてもよい。いくつかの例において、前記1種類以上のオーディオ処理変更を決定するステップは、前記信号対エコー比推定値に少なくとも部分的に基づくコスト関数の最適化に基づいてもよい。例えば、前記コスト関数は、レンダリングパフォーマンスに少なくとも部分的に基づいてもよい。いくつかの実施態様において、前記2つ以上のオーディオデバイスを選択するステップは、近接度推定値に少なくとも部分的に基づいてもよい。

20

【0040】

いくつかの例において、前記方法は、各マイクロフォンの前記出力信号から複数の現在の音響的特徴を決定するステップと、前記複数の現在の音響的特徴に分類器を適用するステップとを含んでもよい。いくつかの実施態様によると、前記分類器を適用するステップは、前記環境内の複数のユーザゾーン内で前記人によってなされた複数の過去の発声から得られた、過去に決定された音響的特徴に対して訓練されたモデルを適用することを含んでもよい。

30

【0041】

いくつかのそのような例において、前記人に関するコンテキスト情報の1つ以上のアスペクトを決定するステップは、前記分類器からの出力に少なくとも部分的に基づき、前記人が現在位置しているユーザゾーンの推定値を決定することを含んでもよい。いくつかの実施態様によると、前記ユーザゾーンの推定値は、前記複数のマイクロフォンの幾何学的位置を参照せずに決定されてもよい。いくつかの例において、前記現在の発声および前記過去の発声は、ウェイクワードの発声であってもよいし、ウェイクワードの発声を含んでもよい。

40

【0042】

いくつかの実施態様によると、前記1つ以上のマイクロフォンは、前記オーディオ環境の複数のオーディオデバイス内に設けられてもよい。しかし、別の例において、前記1つ以上のマイクロフォンは、前記オーディオ環境の1つのオーディオデバイス内に設けられてもよい。いくつかの例において、前記1つ以上のマイクロフォン位置のうち少なくとも1つは、1つのオーディオデバイスの複数のマイクロフォンに対応していてもよい。いくつかの開示された方法は、前記コンテキスト情報の前記1つ以上のアスペクトに応じて少なくとも1つのマイクロフォンを選択するステップを含んでもよい。

50

【0043】

本明細書に記載の演算、機能、および／または方法のいくつかまたは全ては、1つ以上の非一時的な媒体に格納された命令（例えばソフトウェア）に従い、1つ以上のデバイスにより実行されてもよい。そのような非一時的な媒体は、本明細書に記載されたようなメモリ装置を含んでもよい。これは、ランダムアクセスメモリ（RAM）装置、リードオンリーメモリ（ROM）装置を含むが、これらに限定されない。したがって、本開示に記載された主題の革新的ないくつかの側面は、ソフトウェアが格納された非一時的な媒体において実装され得る。

【0044】

例えば、ソフトウェアは、1つ以上のデバイスを制御して、オーディオ環境における複数のマイクロフォンの各々からの出力信号を受信するステップを含む方法を実行する命令を含んでもよい。いくつかの例において、前記複数のマイクロフォンの各々は、前記オーディオ環境のマイクロフォン位置に存在する。いくつかの例において、前記出力信号は、人の現在の発声に対応する信号を含む。いくつかの例によると、前記出力信号は、雑音および／またはエコー等の、発話でないオーディオデータに対応する信号を含む。

【0045】

いくつかのそのような方法は、前記出力信号に基づき、前記人に関するコンテキスト情報の1つ以上のアスペクトを決定するステップを含む。いくつかの例において、コンテキスト情報は、前記人の推定される現在位置および／または前記人の1つ以上のマイクロフォン位置に対する推定される現在の近接度を含む。いくつかのそのような方法は、前記コンテキスト情報の前記1つ以上のアスペクトに少なくとも部分的に基づき、前記オーディオ環境の2つ以上のオーディオデバイスを選択するステップを含む。いくつかの実施態様において、2つ以上のオーディオデバイスの各々は、少なくとも1つのラウドスピーカを含む。

【0046】

いくつかのそのような方法は、前記2つ以上のオーディオデバイス用のラウドスピーカフィード信号へとレンダリングされているオーディオデータに適用する、1種類以上のオーディオ処理変更を決定するステップを含む。いくつかの例において、前記オーディオ処理変更は、1つ以上のマイクロフォンにおける発話対エコー比を増大する効果を有する。いくつかのそのような方法は、前記1種類以上のオーディオ処理変更を適用させるステップを含む。

【0047】

いくつかの実施態様によると、前記1種類以上のオーディオ処理変更は、前記2つ以上のオーディオデバイスの前記ラウドスピーカのラウドスピーカ再生レベルの低下を生じてもよい。いくつかの実施態様において、第1のオーディオデバイスに対する前記オーディオ処理変更の少なくとも1つは、第2のオーディオデバイスに対するオーディオ処理変更とは異なってもよい。いくつかの例において、前記オーディオ環境の2つ以上のオーディオデバイスを選択するステップは、前記オーディオ環境のN個のラウドスピーカ内蔵オーディオデバイスを選択することを含んでもよく、Nは2よりも大きい整数である。

【0048】

いくつかの実施態様において、前記オーディオ環境の前記2つ以上のオーディオデバイスを選択するステップは、マイクロフォン位置およびラウドスピーカ内蔵オーディオデバイス位置の少なくとも一方に対する、前記人の推定される現在位置に少なくとも部分的に基づいてもよい。いくつかのそのような実施態様によると、前記方法は、前記人の前記推定される現在位置に最も近い最近位ラウドスピーカ内蔵オーディオデバイス、または、前記人の前記推定される現在位置に最も近い前記マイクロフォン位置に最も近い最近位ラウドスピーカ内蔵オーディオデバイス、を決定するステップを含んでもよい。いくつかのそのような例において、前記2つ以上のオーディオデバイスは、前記最近位ラウドスピーカ内蔵オーディオデバイスを含んでもよい。

【0049】

いくつかの例において、前記1種類以上のオーディオ処理変更は、前記人の前記推定される現在位置から離れる方向に、オーディオ信号のレンダリングをワーピングするように、レンダリングプロセスを変更することを含む。いくつかの実施態様において、前記1種類以上のオーディオ処理変更は、スペクトル修正を含んでもよい。いくつかのそのような実施態様によると、前記スペクトル修正は、500Hzと3kHzとの間の周波数帯域においてオーディオデータのレベルを低下させることを含んでもよい。

【0050】

いくつかの実施態様において、前記1種類以上のオーディオ処理変更は、オーディオ再生信号の少なくとも1つの選択された周波数帯域に、少なくとも1つのギャップを挿入することを含んでもよい。いくつかの例において、前記1種類以上のオーディオ処理変更は、ダイナミックレンジ圧縮を含んでもよい。

【0051】

いくつかの実施態様によると、前記2つ以上のオーディオデバイスを選択するステップは、1つ以上のマイクロフォン位置についての信号対エコー比推定値に少なくとも部分的に基づいてもよい。例えば、前記2つ以上のオーディオデバイスを選択するステップは、前記信号対エコー比推定値が信号対エコー比閾値以下であるか否かを決定することに少なくとも部分的に基づいてもよい。いくつかの例において、前記1種類以上のオーディオ処理変更を決定するステップは、前記信号対エコー比推定値に少なくとも部分的に基づくコスト関数の最適化に基づいてもよい。例えば、前記コスト関数は、レンダリングパフォーマンスに少なくとも部分的に基づいてもよい。いくつかの実施態様において、前記2つ以上のオーディオデバイスを選択するステップは、近接度推定値に少なくとも部分的に基づいてもよい。

【0052】

いくつかの例において、前記方法は、各マイクロフォンの前記出力信号から複数の現在の音響的特徴を決定するステップと、前記複数の現在の音響的特徴に分類器を適用するステップとを含んでもよい。いくつかの実施態様によると、前記分類器を適用するステップは、前記環境内の複数のユーザゾーン内で前記人によってなされた複数の過去の発声から得られた、過去に決定された音響的特徴に対して訓練されたモデルを適用することを含んでもよい。

【0053】

いくつかのそのような例において、前記人に関するコンテキスト情報の1つ以上のアスペクトを決定するステップは、前記分類器からの出力に少なくとも部分的に基づき、前記人が現在位置しているユーザゾーンの推定値を決定することを含んでもよい。いくつかの実施態様によると、前記ユーザゾーンの推定値は、前記複数のマイクロフォンの幾何学的位置を参照せずに決定されてもよい。いくつかの例において、前記現在の発声および前記過去の発声は、ウェイクワードの発声であってもよいし、ウェイクワードの発声を含んでもよい。

【0054】

いくつかの実施態様によると、前記1つ以上のマイクロフォンは、前記オーディオ環境の複数のオーディオデバイス内に設けられてもよい。しかし、別の例において、前記1つ以上のマイクロフォンは、前記オーディオ環境の1つのオーディオデバイス内に設けられてもよい。いくつかの例において、前記1つ以上のマイクロフォン位置のうち少なくとも1つは、1つのオーディオデバイスの複数のマイクロフォンに対応していてもよい。いくつかの開示された方法は、前記コンテキスト情報の前記1つ以上のアスペクトに応じて少なくとも1つのマイクロフォンを選択するステップを含んでもよい。

【0055】

本明細書に記載された主題の1つ以上の実施態様の詳細を、添付の図面および以下の説明に記載する。その他の特徴、側面、および利点は、以下の説明、図面、および特許請求の範囲から明らかになる。なお、以下の図形の相対的な寸法は、縮尺どおりに描かれていない場合もあることに留意されたい。

10

20

30

40

50

【図面の簡単な説明】

【0056】

【図1A】図1Aは、ある例によるオーディオ環境を示す。

【図1B】図1Bは、オーディオ環境の別の例を示す。

【図2A】図2Aは、本開示のさまざまな側面を実装することができる装置の構成要素の例を示すブロック図である。

【図2B】図2Bは、いくつかの実施態様によるオーディオセッション管理方法のブロックを含む流れ図である。

【図3A】図3Aは、別個のレンダリング制御および複数のデバイスに亘るリスニングまたはキャプチャロジックを実装するように構成されたシステムのブロック図である。

10

【図3B】図3Bは、別の開示された実施態様によるシステムのブロック図である。

【図3C】図3Cは、ある例によるエネルギーバランシングネットワーク (energy balancing network) を実装するように構成された実施形態のブロック図である。

【図4】図4は、オーディオ環境の1つ以上のマイクロフォンにおける発話対エコー比を増大し得るオーディオ処理の例を示すグラフである。

【図5】図5は、オーディオ環境の1つ以上のマイクロフォンにおける発話対エコー比を増大し得る別の種類のオーディオ処理を示すグラフである。

【図6】図6は、オーディオ環境の1つ以上のマイクロフォンにおける発話対エコー比を増大し得る別の種類のオーディオ処理を示す。

【図7】図7は、オーディオ環境の1つ以上のマイクロフォンにおける発話対エコー比を増大し得る別の種類のオーディオ処理を示すグラフである。

20

【図8】図8は、音量を下げようとするオーディオデバイスが、話している人に最も近いオーディオデバイスではないかもしれない例の図である。

【図9】図9は、非常に高いSERを有するデバイスがユーザの極めて近くにある状況を示す。

【図10】図10は、図2Aに示すような装置により実行され得る方法の一例の概略を示す流れ図である。

【図11】図11は、ゾーン分類器を実装するように構成された実施形態の一例の要素のブロック図である。

【図12】図12は、図2Aの装置200などの装置により実行され得る方法の一例の概略を示す流れ図である。

30

【図13】図13は、図2Aの装置200などの装置により実行され得る方法の別の例の概略を示す流れ図である。

【図14】図14は、図2Aの装置200などの装置により実行され得る方法の別の例の概略を示す流れ図である。

【図15】図15は、スピーカ起動電位の集合の一例およびオブジェクトレンダリング位置を示す図である。

【図16】図16は、スピーカ起動電位の集合の一例およびオブジェクトレンダリング位置を示す図である。

【図17】図17は、図2Aに示すような装置またはシステムにより実行され得る方法の一例の概略を示す流れ図である。

40

【図18】図18は、実施形態例における、スピーカ起動電位のグラフである。

【図19】図19は、実施形態例における、オブジェクトレンダリング位置のグラフである。

【図20】図20は、実施形態例における、スピーカ起動電位のグラフである。

【図21】図21は、実施形態例における、オブジェクトレンダリング位置のグラフである。

【図22】図22は、実施形態例における、スピーカ起動電位のグラフである。

【図23】図23は、実施形態例における、オブジェクトレンダリング位置のグラフである。

50

【発明を実施するための形態】

【0057】

現在、設計者らは、概して、オーディオデバイスを、娯楽、通信、および情報サービスの混合体であり得るオーディオ音声のための単一のインターフェイスポイントとして考えている。通知および音声制御にオーディオ音声を用いることは、視覚的または身体的な割込みを回避できるという利点を有する。拡大しつつあるデバイスのランドスケープは細分化され、より多くのシステムが我々の一対の耳を奪い合っている。

【0058】

対話型オーディオ音声の全ての形態において、全二重通信オーディオ性能を向上するという課題は、依然として困難なままである。ある室内において、この室内における伝送または情動的なキャプチャにとって関係のないオーディオ出力が存在する場合、このオーディオ出力をキャプチャされた信号から除去する（例えばエコー除去および／またはエコー抑制によって）ことが望ましい。いくつかの開示された実施形態は、1つ以上のデバイスにおける全二重通信を成功させるための鍵となる基準である、信号対エコー比（SER）を改善する、ユーザエクスペリエンスのアプローチおよび管理を提供する。

【0059】

このような実施形態は、ユーザの可聴範囲内に複数のオーディオデバイスが存在し、各オーディオデバイスが、所望の娯楽、通信、または情報サービスのためにユーザの位置で適切な音量のオーディオプログラム素材を提供することができるという状況において、有用であると考えられる。そのような実施形態の価値は、ユーザに対して同じくらい近くに3つ以上のオーディオデバイスが存在する場合に特に高いと考えられる。

【0060】

レンダリングアプリケーションがオーディオデバイスの主要な機能である場合があるので、可能な限り多くのオーディオ出力デバイスを用いたい場合がある。いくつかのオーディオデバイスがユーザのより近くにある場合、これらのオーディオデバイスは、音声を正確に配置する能力または特定のオーディオシグナリングおよびイメージングをユーザに届ける能力の点でより有利であり得る。しかし、これらのオーディオデバイスが1つ以上のマイクロフォンを含む場合、それらはユーザの声を拾うのに好ましいかもしれない。信号対エコー比の困難性とともに考えた場合、ユーザのより近くに存在するデバイスを単信（入力オンリー）モードで使用するか、または、ユーザがこのデバイスに近づくと、信号対エコー比が劇的に向上されることがわかる。

【0061】

開示されたさまざまな実施形態において、オーディオデバイスは、他の機能（例えば図1Aに示す機能）を提供しつつ、内蔵のスピーカおよびマイクロフォンの両方を有していてもよい。いくつかの開示された実施形態は、ある状況において、ユーザの最も近くにあるラウドスピーカ（単数または複数）をあえて主として用いないというコンセプトを実装する。

【0062】

アプリケーション（例えばクラウド型アプリケーション）間の、コネクテッドなオペレーティングシステムまたは仲介者排除（disintermediation）において、（オーディオ入力、出力、および／またはリアルタイムインタラクションを可能にする）多くの異なる種類のデバイスが含まれ得ると考える。そのようなデバイスの例として、ウェアラブルな家庭用オーディオデバイス、モバイルデバイス、自動モバイルコンピューティングデバイス、およびスマートスピーカが挙げられる。スマートスピーカは、クラウド型サービス用の、ネットワーク接続されたスピーカおよびマイクロフォンを含んでもよい。このようなデバイスの他の例は、スピーカおよび／またはマイクロフォンを内蔵しており、ライト、時計、テレビ、携帯情報端末、冷蔵庫およびゴミ箱を含む。いくつかの実施形態は、スマートホームハブなどのオーケストレーションデバイス（orchestrating device）またはオーディオセッション管理を行うように構成された他のデバイス（本明細書中、「オーディオセッションマネジャ」と呼ぶことがある）を介してオーディオ環境内の複数のオーディ

10

20

30

40

50

オーディオのオーケストレーションを行うための共通プラットフォームが存在する状況に特に関連している。いくつかのそのような実施態様は、デバイス固有ではない言語であって、オーケストレーションデバイスがソフトウェアアプリケーションにより特定された複数のユーザや場所の間でオーディオコンテンツのルーティング（routing）を行うような言語において、オーディオセッションマネージャとローカル実装されたソフトウェアアプリケーションとの間のコマンドを含み得る。いくつかの実施形態は、レンダリングを動的に管理する方法（例えば、最も近くに存在するデバイスから音声を遠ざけて空間イメージングを維持する制約を含む）、ならびに／もしくは、ゾーン内のユーザの位置を特定する方法、ならびに／もしくは、デバイス間およびユーザに対して相対的にデバイスのマッピングおよび位置特定を行う方法を実装する。

10

【0063】

典型的には、複数のスマートオーディオデバイスを含むシステムは、ユーザから（上で定義した）「ウェイクワード」をいつ聞いたかを示す必要があり、ユーザからのコマンドに注意を払っている（換言すると、ユーザからのコマンドがないかと耳をすましている）。

【0064】

図1Aは、ある例によるオーディオ環境を示す。いくつかの開示された実施形態は、任意の環境（例えば生活空間または職場）内に、例えば本明細書に開示されるように、音声を伝達しオーディオ音声をキャプチャすることができる多数のオーディオデバイスが存在するというシナリオにおいて特に有用であり得る。図1Aのシステムは、さまざまな開示された実施形態に応じて構成されてもよい。

20

【0065】

図1Aは、オーディオインタラクションのための1組のスマートオーディオデバイス（デバイス1.1）と、オーディオ出力のためのスピーカ（1.3）と、制御可能なライト（1.2）とを含む、オーディオ環境（生活空間）の図である。他の開示された実施態様と同様、図1Aにおける要素の種類、個数および配置は例示に過ぎない。他の実施態様は、より多くの要素、より少ない要素、および／または異なる要素を提供し得る。いくつかの例において、マイクロフォン1.5のうち1つ以上のマイクロフォンは、デバイス1.1、ライト1.2、スピーカ1.3のうちの1つの一部であってもよいし、デバイス1.1、ライト1.2、スピーカ1.3のうちの1つと関連付けられてもよい。代替的にまたは追加的に、マイクロフォン1.5のうち1つ以上のマイクロフォンは、当該環境の別の部分（例えば、壁、天井、家具、家電、または当該環境の別のデバイス）に取り付けられてもよい。ある例において、デバイス1.1の各々は、少なくとも1つのマイクロフォン1.5を含む（かつ／または少なくとも1つのマイクロフォン1.5に接続されている）。図1Aには図示しないが、いくつかのオーディオ環境は、1つ以上のカメラを含んでもよい。いくつかの開示された実施態様によると、オーディオ環境の1つ以上のデバイス（例えば、デバイス1.1のうちの1つ以上のデバイスなどの、オーディオセッション管理のために構成されたデバイス、オーディオセッションマネージャを実装するデバイス、スマートホームハブ等）は、ウェイクワードやコマンドなどを発したユーザ（1.4）が何処にいるか（例えば、生活空間のどのゾーンにいるか）を推定することができてもよい。図1Aに示すシステムの1つ以上のデバイス（例えば、デバイス1.1）は、さまざまな開示された実施形態を実装するように構成されてもよい。さまざまな方法を用いて、図3のデバイスから情報を集約的に得て、ウェイクワードを発したユーザの位置推定値を提供することができる。いくつかの開示された方法によると、図1Aのマイクロフォン1.5から情報が集約的に得られ、その情報は、ウェイクワードを発したユーザの位置推定値を提供するように構成された分類器を実装するデバイス（例えばオーディオセッション管理のために構成されたデバイス）に提供される。

30

40

【0066】

生活空間（例えば図1Aの生活空間）内に、1組の通常のアクティビティゾーンが存在し、このゾーンにおいて、人は、タスクまたはアクティビティを実行するか、またはしきいを越える。これらのエリア（本明細書において「ユーザゾーン」と呼ぶ）は、いくつか

50

の例において、幾何学的位置 (geometric location) の座標または他の指標を特定することなく、ユーザにより定義されてもよい。いくつかの例のよると、人の「コンテキスト」は、ユーザが現在居るユーザゾーンまたはそのユーザゾーンの推定値を含んでもよいし、もしくは、それに一致してもよい。図 1 A において、ユーザゾーンは以下のゾーンを含む：

1. キッチンシンクおよび調理エリア (生活空間の左上領域内) ；
2. 冷蔵庫のドア (シンクおよび調理エリアの右側) ；
3. ダイニングエリア (生活空間の左下領域内) ；
4. 生活空間のオープンエリア (シンクおよび調理エリアならびにダイニングエリアの右側) ；
5. テレビソファ (オープンエリアの右側) ；
6. テレビ自身；
7. テーブル；
8. ドアエリアまたは通路 (生活空間の右上領域内) 。

他のオーディオ環境は、より多くのユーザゾーン、より少ないユーザゾーン、ならびに／もしくは、1つ以上の寝室ゾーン、ガレージゾーン、パティオまたはデッキゾーンなど他の種類のユーザゾーンを含んでもよい。

【0067】

いくつかの実施形態によると、音 (例えば、ウェイクワードまたは注意喚起する他の信号) が起こった (または生じた) 場所を推定する (例えば、音が起こった場所の不確実な推定値 (uncertain estimate) を決定する) システムは、その推定値に一定の信頼をおいていてもよい (または複数の仮説を有してもよい)。例えば、人がオーディオ環境の複数のユーザゾーンの境界の近くにたまたま居る場合、この人の位置の不確実な推定値は、この人がこれら複数のゾーンの各ゾーン内に居るという一定の確信を含んでもよい。音声インターフェイスの従来のいくつかの実施態様において、音声アシスタントの声は一度に一カ所のみから発せられ、1つの位置 (例えば、図 1 A において、8 カ所のスピーカ位置 (1. 1 および 1. 3) のうちの 1 つ) に対して 1 つの選択肢とすることが要求される。しかし、単純な仮定のロールプレイに基づく、(そのような従来の実施態様において) アシスタントの声の発生源の選択された位置 (つまり、音声アシスタントに含まれるかまたは接続されたスピーカの位置) が、注意を表出するための自然な返答のフォーカスポイントである尤度が低いものであり得ることが明らかである。

【0068】

図 1 B は、オーディオ環境の別の例を示す。図 1 B に示す別のオーディオ環境は、直接的な発話 102 を発するユーザ 101 と、1組のスマートオーディオデバイス 103 および 105、オーディオ出力のためのスピーカ、およびマイクロフォンを含むシステムとを含む。このシステムは、いくつかの開示された実施態様に従って構成されてもよい。ユーザ 101 (本明細書中、「話者」と呼ぶこともある) により発せられた発話は、システムの 1つ以上の要素により、ウェイクワードであると認識され得る。

【0069】

より具体的には、図 1 B のシステムの要素は以下を含む。

- 102：直接的で局所的な声 (ユーザ 101 によって生成される) 。
- 103：音声アシスタントデバイス (1つ以上のラウドスピーカに接続される)。デバイス 103 は、デバイス 105 よりもユーザ 101 の近くに配置されている。デバイス 103 を「近位」デバイスと呼び、デバイス 105 を「遠位」デバイスと呼ぶこともある。
- 104：近位デバイス 103 内の (またはそれに接続された) 複数のマイクロフォン。
- 105：(1つ以上のラウドスピーカに接続された) 音声アシスタントデバイス。
- 106：遠位デバイス 105 内の (またはそれに接続された) 複数のマイクロフォン。
- 107：家電機器 (例えば、ランプ) 。
- 108：家電機器 107 内の (またはそれに接続された) 複数のマイクロフォン。いくつかの例において、マイクロフォン 108 の各々は、分類器を実装するように構成され

10

20

30

40

50

たデバイス（いくつかの例において、デバイス103または105の少なくとも一方）と通信するように構成されてもよい。いくつかの実施態様において、分類器を実装するように構成されたデバイスはまた、CHASMまたはスマートホームハブを実装するように構成されたデバイスなどの、オーディオセッション管理のために構成されたデバイスであってもよい。

【0070】

図1Bのシステムはまた、少なくとも1つの分類器（例えば、後で説明する図11の分類器1107）を含んでもよい。例えば、デバイス103（またはデバイス105）は、分類器を含んでもよい。代替的にまたは追加的に、分類器は、デバイス103および／またはデバイス105と通信するよう構成され得る別のデバイスにより実装されてもよい。いくつかの例において、分類器は、別のローカルデバイス（例えば、環境109内のデバイス）により実装されてもよい。それに対して、別の例では、分類器は、環境109の外部に配置されたりモートデバイス（例えば、サーバ）により実装されてもよい。

【0071】

いくつかの実施態様によると、少なくとも2つのデバイス（例えば、図1Aのデバイス1.1、図1Bのデバイス103および105など）は、（例えば、オーディオセッション管理のために構成されたデバイスなどの、オーケストレーションデバイスの制御下において）任意の様態で協働して、それらデバイス間でオーディオ音声の共同制御され得るように音声を配送する。例えば、2つのデバイス103および105は、個別にまたは共同で音声を再生してもよい。ある単純なケースにおいて、デバイス103および105は、共同の対として動作して、各々がオーディオ音声の一部分をレンダリングする（例えば、一般性を失わずに、ステレオ信号の一方が実質的にL側をレンダリングし、他方が実質的にR側をレンダリングする）。

【0072】

家電機器107（または別のデバイス）は、ユーザ101の最も近くに存在しかつラウドスピーカを有していない1つのマイクロフォン108を含んでもよい。その場合、この特定のオーディオ環境およびこのユーザ101の特定の位置に対して、デバイス105および／または家電機器107のスピーカ（単数または複数）により再生されるオーディオ音声に対するオーディオ処理を変更することによっても改善し得ない、好ましい信号対エコー比または発話対エコー比（SER）が既に存在する状況を考える。いくつかの実施形態において、そのようなマイクロフォンは存在しない。

【0073】

いくつかの開示された実施形態は、検出可能で有意なSERパフォーマンスインパクトを提供する。いくつかの実施態様は、ゾーン位置特定および／または動的可変レンダリングの側面を実装することなく、そのような利点を提供する。しかし、いくつかの実施形態は、デバイスから離れる方向にサウンドオブジェクト（またはオーディオオブジェクト）を斥ける（repelling）または「ワーピング（warping）」するようにレンダリングすることを含む、オーディオ処理変更を実装する。特定のオーディオデバイスや特定の位置などから離れる方向にオーディオオブジェクトをワーピングする理由は、いくつかの例において、人間の発話をキャプチャするために使用される特定のマイクロフォンにおける信号対エコー比を改善するためである。そのようなワーピングは、1つ、2つ、または3つ以上の近くに存在するオーディオデバイスの再生レベルを下げることを含んでもよいが、このことに限定されない。いくつかの場合において、SERを改善するためのオーディオ処理の変更は、オーディオ処理の変更が実装される（例えば再生レベルを下げる）1つ、2つ、またはそれ以上の近くに存在するオーディオデバイスが、ユーザの最も近くに存在するオーディオデバイス、ユーザの発話をキャプチャするために使用される特定のマイクロフォンの最も近くに存在するオーディオデバイス、および／または、注目している音声の最も近くに存在するオーディオデバイスとなるように、ゾーン検出技術により通知されてもよい。

【0074】

いくつかの実施形態の側面は、コンテキストと、決定 (decision) と、オーディオ処理変更 (本明細書中、「レンダリング変更」と呼ぶ) とを含む。いくつかの例において、これらの側面は以下のとおりである。

コンテキスト (CONTEXT) (位置および／または時間など)。いくつかの例において、位置および時間はコンテキストの一部であり、各々が様々な方法で提供または決定され得る。

決定 (DECISION) (閾値または変更の連続的な調整を含み得る)。この構成要素は、単純であっても複雑であってもよく、それは特定の実施形態に応じて決まる。いくつかの実施形態において、決定は、例えばフィードバックに応じて、連続的に為されてもよい。いくつかの例において、決定は、例えば後で説明するような好循環のフィードバックの安定性などの、システムの安定性を生み得る。

10

レンダリング (RENDER) (オーディオ処理変更の本質)。本明細書中「レンダリング (rendering)」と表記するが、オーディオ処理変更は、レンダリング変更を含んでも含まなくてもよく、それは特定の実施態様に依拠して決まる。いくつかの実施態様において、かろうじて知覚可能なオーディオ処理変更の実施態様から、厳密かつ明白なオーディオ処理変更のレンダリングの実施態様までを含む、オーディオ処理変更に複数のオプションが存在する。

【0075】

いくつかの例において、「コンテキスト」は、位置および意図に関する情報を含んでもよい。例えば、コンテキスト情報は、少なくとも、ユーザの現在位置に一致するユーザゾーンの推定値などの、ユーザの位置についての大まかな知識を含んでもよい。コンテキスト情報は、オーディオオブジェクト位置 (例えば、ユーザによるウェイクワードの発話に一致するオーディオオブジェクト位置) に一致してもよい。いくつかの例において、コンテキスト情報は、オブジェクトまたは個人が音を鳴らしたタイミングおよび尤度に関する情報を含んでもよい。コンテキストの例として、以下のものが挙げられるが、それらに限定されない。

20

A. 尤もらしい位置が何処なのかを知ること。これは、以下の事項に基づく。

i) 弱検出または低確率検出 (例えば、もしかすると注目している音声であり得るが、作用するのに十分に明瞭であることもないこともあり得る、音声の検出)。

ii) 特定の起動 (例えば、発声されかつ明瞭に検出されたウェイクワード)。

30

iii) 習慣およびパターン (例えば、テレビの近くにあるソファなどの所与の位置が、ソファに座ってテレビでビデオ素材を視聴して関連するオーディオ音声を聞いている1人以上の人に関連付けられるというような、パターン認識に基づく。)

iv) および／または他の様式 (modality) に基づく他の形態の近接度検知 (例えば、オーディオ環境の家具内または家具上に設けられた) 1つ以上の赤外線センサ、カメラ、容量型センサ、高周波 (RF) センサ、熱センサ、圧力センサ、ウェアラブルビーコン等の統合。

B. 例えば改善された検出度で人が聞きたい音声の尤度を知るまたは推定すること。これは、以下の事項のいくつかまたは全てを含み得る。

i) ウェイクワード検出など、任意のオーディオ音声の検出に基づくイベント。

40

ii) 既知のアクティビティまたは一連のイベント (例えば、ビデオコンテンツの表示の一時休止、スクリプト化された自動音声認識 (ASR) 型の対話型コンテンツにおけるインタラクションのための空間、もしくは、アクティビティの変更および／または (テレビ会議の1人以上の参加者による一時休止などの) 全二重通信通信アクティビティの対話ダイナミクスの変更) に基づくイベントまたはコンテキスト。

iii) 追加的な他の様式の感覚入力 (sensory input)

iv) 任意の様態で頻繁にリスニングを改良するという選択肢 — 向上された準備状態または改良されたリスニング。

【0076】

A (尤もらしい位置が何処なのかを知ること) と B (例えば改善された検出度でユーザ

50

が聞きたい音声の尤度を知るまたは推定すること)との重要な差を説明すると、Aは、まだ聞くべきものがあるか否かを必ずしも知る必要なく、特定の位置情報または知識を含むが、それに対してBは、正確に何処で聞くかを必ずしも知る必要なく、特定のタイミングまたはイベント情報により注目する。当然、AおよびBのいくつかのアスペクトにおいて重複する部分があり得る。例えば、ウェイクワードの弱検出または完全検出が、位置およびタイミングの両方に関する情報を有する。

【0077】

いくつかのユースケースにとって、重要なことは、「コンテキスト」が、聞きたいと思う位置（例えば、人および／または最も近くに存在するマイクロフォンの位置）およびタイミングの両方に関する情報を含むということである。このコンテキストの情報は、1つ以上の関連付けられた決定、および、1つ以上の起こり得るオーディオ処理変更（例えば1つ以上の起こり得るレンダリング変更）を駆動する。したがって、さまざまな実施形態は、コンテキストを形成するために使用され得る各種情報に基づき、多くの可能性を考慮に入れる。

【0078】

次に、「決定」アスペクトについて説明する。このアスペクトは、例えば、関連付けられたオーディオ処理が変更される1つ、2つ、または3つ以上の出力デバイスを決定することを含んでもよい。このような決定を形成する1つの単純な方法は、以下のようなものである。

【0079】

コンテキストからの情報（例えば、位置および／またはイベント（もしくは、ある意味において、その位置に関して重要な何かがあるということに対する信頼））が与えられると、いくつかの例において、オーディオセッションマネージャは、その位置からオーディオ環境内のいくつかまたは全てのオーディオデバイスまでの距離を決定または推定し得る。いくつかの実施形態において、オーディオセッションマネージャはまた、オーディオ環境のいくつかまたは全てのオーディオデバイスについて、各ラウドスピーカ（または1組のラウドスピーカ）に対する1組の起動電位を生成し得る。いくつかのそのような例によると、1組の起動電位は、 $[f_1, f_2, \dots, f_n]$ と決定され得、一般性を失わずに、 $[0, \dots, 1]$ の範囲内であり得る。別の例において、決定の結果は、各デバイスの目標の発話対エコー比改善値 $[s_1, s_2, \dots, s_n]$ を、「レンダリング」アスペクトに記述（説明）してもよい。さらなる例において、起動電位および発話対エコー比改善値の両方は、「決定」アスペクトにより生成され得る。

【0080】

いくつかの実施形態において、起動電位は、「レンダリング」アスペクトが、所望のマイクロフォン位置でSERが改善されることを確実にする度合いを伝達する。いくつかのそのような例において、 f_n の最大値は、レンダリングされたオーディオ音声を積極的に音量低減（ダッキング）するまたはワーピングすることを示し、もしくは、値 s_n が与えられた場合、オーディオ音声は、 s_n の発話対エコー比を達成するように、限定かつ音量低減（ダッキング）されることを示す。0.5により近い f_n の中間値は、いくつかの実施形態において、中程度のレンダリング変更のみが要求され、これらの位置に向けてオーディオ源をワーピングすることが適切であることを示し得る。さらに、いくつかの実施形態において、 f_n の低い値は、減衰するのに重要でないとみなされ得る。いくつかのそのような実施形態において、閾値レベル以下の f_n 値はアサートされなくてもよい。いくつかの例によると、閾値レベル以下の f_n 値は、オーディオコンテンツのレンダリングをワーピングして向ける先の位置に一致してもよい。いくつかの例において、閾値レベル以下の f_n 値に一致するラウドスピーカは、後で説明するプロセスに応じて再生レベルを上昇されてもよい。

【0081】

いくつかの実施形態によると、上記方法（または後で説明する別の方法の1つ）は、全ての選択されたオーディオデバイスに対する、例えば、オーディオ環境の各デバイス、オ

10

20

30

40

50

オーディオ環境の1つ以上のデバイス、オーディオ環境の2つ以上のデバイス、オーディオ環境の3つ以上のデバイスなどに対する、選択されたオーディオ処理変更の各々について、制御パラメータを生成するために使用されてもよい。オーディオ処理変更の選択は、特定の実施態様に依じて異なってもよい。例えば、決定は、

- － オーディオ処理を変更しようとする2つ以上のラウドスピーカからなる1組のラウドスピーカと、

- － この1組の2つ以上のラウドスピーカに対するオーディオ処理を変更する範囲と、

を決定することを含んでもよい。変更の範囲は、いくつかの例において、設計または決定された範囲の文脈において決定されてもよく、これは、当該1組のラウドスピーカに含まれる1つ以上のラウドスピーカの能力に少なくとも部分的に基づき得る。いくつかの例において、各ラウドスピーカの能力は、周波数特性、再生レベル限界値、および／または1つ以上のラウドスピーカダイナミクス処理アルゴリズムのパラメータを含み得る。

【0082】

例えば、1つの設計上の選択事項として、特定の状況における最良の選択肢がラウドスピーカの音量を下げることであり得る。いくつかのそのような例において、オーディオ処理変更の最大範囲および／または最小範囲が決定されてもよい。例えば、任意のラウドスピーカの音量を下げる範囲が、例えば15 dB、20 dB、25 dB等の特定の閾値に限定される。いくつかのそのような実施態様において、決定は、1つ、2つ、または3つ以上のラウドスピーカを選択するヒューリスティクスまたはロジックに基づいてもよく、かつ、注目しているアクティビティの確信度やラウドスピーカ位置などに基づいてもよい。決定は、1つ、2つ、または3つ以上のラウドスピーカにより再生されるオーディオ音声を、最小値および最大値の範囲（例えば0 dB～20 dB）内の任意の量だけ音量低減（ダッキング）することであってもよい。決定方法（またはシステム要素）は、いくつかの例において、各ラウドスピーカ内蔵オーディオデバイスにつき1組の起動電位を生成してもよい。

【0083】

ある単純な例において、決定プロセスは、1つを除いて全てのオーディオデバイスがレンダリング起動変更の値0を有すると決定することや、その1つのオーディオデバイスがレンダリング起動変更の値1を有すると決定することのような単純なことであってもよい。オーディオ処理変更（例えば、音量低減（ダッキング））の設計およびオーディオ処理変更の範囲（例えば、時定数）は、いくつかの例において、決定ロジックから独立していてもよい。このアプローチにより、単純かつ効果的な設計が得られる。

【0084】

しかし、別の実施態様は、2つ以上のラウドスピーカ内蔵オーディオデバイスを選択することと、2つ以上のラウドスピーカ内蔵オーディオデバイスのうち、少なくとも2つ、少なくとも3つ（および、いくつかの例において、全て）のオーディオデバイスに対するオーディオ処理を変更することを含んでもよい。いくつかのそのような例において、第1のオーディオデバイスに対するオーディオ処理変更のうちの少なくとも1つ（例えば再生レベルの低減）は、第2のオーディオデバイスに対するオーディオ処理変更とは異なってもよい。これらオーディオ処理変更の間の差は、いくつかの例において、各オーディオデバイスの位置に対する人の推定される現在位置またはマイクロフォン位置に少なくとも部分的に基づいてもよい。いくつかのそのような実施態様によると、オーディオ処理変更は、レンダリングプロセスを変更することの一部として、異なるラウドスピーカ位置において異なるスピーカ起動電位を適用して、注目している人の推定される現在位置から離れる方向に、オーディオ信号のレンダリングをワーピングすることを含んでもよい。これらオーディオ処理変更の間の差は、いくつかの例において、ラウドスピーカの能力に少なくとも部分的に基づいてもよい。例えば、オーディオ処理変更が、バスレンジにおけるオーディオ音声のレベルを低下させることを含む場合、そのような変更は、バスレンジにおいて大音量再生が可能な1つ以上のラウドスピーカを含むオーディオデバイスに対して、より積極的に適用されてもよい。

10

20

30

40

50

【0085】

次に、オーディオ処理変更アспект（本明細書中、「レンダリング変更」アспектと呼ぶこともある）に関するさらなる詳細を説明する。本開示は、このアспектを、「最近位スピーカの音量低下（turn nearest down）」（例えば、1つ、2つ、または3つ以上の最近位スピーカにより再生されるオーディオコンテンツがレンダリングされる音量を低減すること）と呼ぶことがあるが、（本明細書の他の箇所に示すように、）より一般的には、多くの実施態様において実装され得るのは、所望のオーディオ発信源（例えばウェイクワードを話した人）をキャプチャまたは検知する能力に関して、信号対エコー比の全体的な推定、測定および／または基準を改善することに向けられた、オーディオ処理に対する1つ以上の変更である。いくつかの場合において、オーディオ処理変更（例えば、レンダリングされたオーディオコンテンツの音量を「低下」すること）は、結果の量のいくつかの連続的なパラメータによって調整される（または調整され得る）。例えば、ラウドスピーカの音量を下げるという文脈において、いくつかの実施態様は、調節可能な（例えば連続的に調節可能な）量の減衰（dB）を適用することができてよい。いくつかのそのような例において、調節可能な量の減衰は、ぎりぎり感知し得る変更のための第1の範囲（例えば、0～3 dB）と、SERに特に効果的な改善をもたらすが、聴者が絶対に感知し得る第2の範囲（例えば、0～20 dB）とを有してもよい。

10

【0086】

上記スキーマ（コンテキスト（CONTEXT）、決定（DECISION）、ならびにレンダリング（RENDER）またはレンダリング変更（RENDERING CHANGE））を実装するいくつかの実施形態において、（例えば、ユーザもしくは別の個人またはシステム要素の「最近位」に位置するラウドスピーカまたはデバイスについて）「最近位」の特定の確固たる境界（hard boundary）が存在しなくてもよく、一般性を失わずに、レンダリング変更は、以下のAおよびBのうちの1つ以上を変化させる（例えば、連続的に変化させる）ことを含み得るかまたは含んでいる。

20

【0087】

A. 1つ以上のオーディオデバイスからのオーディオ出力を低減するように、出力を変更するモード。ここで、オーディオ出力の変更は、以下のi)～v)のうちの1つ以上を含み得る。

i) オーディオデバイス出力の全体のレベルを低下すること（1つ以上のラウドスピーカの音量を下げる、1つ以上のラウドスピーカを消音すること）。

30

ii) 例えば、我々が検出したいオーディオ音声のスペクトルとは異なる出力を生成するように設計された略線形のイコライズ（EQ）フィルタを用いて、1つ以上のラウドスピーカの出力のスペクトルを成形すること。いくつかの例において、人間の声を検出するために出力スペクトルが成形されている場合、フィルタは、約500 Hz～3 kHzの範囲内の周波数を下げ得る（例えば、この周波数範囲の各端部において±5%または±10%）、もしくは、低周波帯および高周波帯を強調して中間帯域（例えば、約500 Hz～3 kHzの範囲）にスペースが残るようにラウドネス（loudness）を成形し得る。

iii) 出力の上限またはピークを変更して、ピークレベルの低下、ならびに／もしくは、オーディオ検出のために達成されたSERを生成するシステム全体（例えば、タイムドメインダイナミックレンジ圧縮器またはマルチバンド周波数依存型圧縮器）の一部である任意のエコー除去のパフォーマンスを追加的に低下させ得る歪成分の低減のいずれかを行うこと。このようなオーディオ信号修正は、オーディオ信号の振幅を効果的に低減し得、かつ、ラウドスピーカのエクスカッションを制限することに寄与し得る。

40

iv) エネルギーを低減する傾向がある方法でオーディオ音声を空間的にステアリングすること、もしくは、例えば本明細書中に記載した「ワーピング」例におけるように、1つ以上のラウドスピーカの出力を、システム（例えばオーディオ処理マネージャ）がより高いSERを実現している1つ以上のマイクロフォンに接続すること。

v) 後で説明するギャップ挿入の例と同様に、一時的なタイムスライシングまたはタイムアジャストメントを用いて、オーディオ音声の細断片（glimpses）を得るのに十

50

分な、スプース時間周波数のより低い出力の「ギャップ」または期間を生成すること。

v i) 上記方法の何らかの組み合わせにて、オーディオ音声を変更すること。

【0088】

B. エネルギーの保存、ならびに／もしくは、特定のリスニング位置または幅広い1組のリスニング位置における連続性の生成。例えば、以下のi) およびii) のうちの1つ以上を含む。

i) いくつかの例において、1つのラウドスピーカから除去されたエネルギーは、別のラウドスピーカに追加のエネルギーを提供することにより補償され得る。いくつかの例において、全体のラウドネスは同じままであるか、または、実質的に同じままである。このことは、必須の特徴ではないが、コンテンツを失うことなく、「最近位の」デバイスまたは最近位の1組のデバイスのオーディオ処理に、より厳格な変更を施すことを可能にする効果的な手段であり得る。しかし、連続性および／またはエネルギーの保存は、複雑なオーディオ出力およびオーディオシーンを扱う場合に特に関係し得る。

ii) 起動の時定数。特に、オーディオ処理の変更は、通常状態に戻される（例えば1000～10000ms）よりも少し早く（例えば100～200ms）適用され得、その結果、オーディオ処理の変更は、感知可能である場合、意図的に見えるが、次に変更状態から通常状態に戻ることは、任意の実際のイベントまたは（ユーザの視点から見た）変更に関係があるように見えないかもしれない。いくつかの例において、ほとんど感知できないほど遅いかもしれない。

【0089】

ここで、コンテキストおよび決定をどのように定式化しかつ決定するかのさらなる例を説明する。

【0090】

実施形態A

（コンテキスト（CONTEXT））一例として、コンテキスト情報は、以下のように数学的に定式化され得る。

$H(a, b)$ 、デバイスaとデバイスbとの間のおおよその物理的距離（メートル）：

【数1】

$$H(a, b) \neq 0 \in \mathbb{R}$$

$$\forall a, b \in D$$

（ここで、Dはシステムにおける全てのデバイスの組を示し、Sは各デバイスにおける推定SERである）は、次のように表現され得る。

【数2】

$$S(a) \in \mathbb{R} \quad \forall a \in D$$

【0091】

HおよびSを決定する：

Hは、デバイスの物理的位置の特性（property）であり、以下の（1）および（2）によって決定または推定され得る。

（1）ユーザによる直接的な指示。例えば、スマートフォンまたはタブレット装置を用いて、環境の平面図または同様の図式表示上に、デバイスのおおよその位置をマーキングまたは指示する。このようなデジタルインターフェイスは、スマートホームデバイスの

構成、グループ化、名前、目的およびIDの管理において、既にありふれたものである。例えば、このような直接的な指示は、アマゾンアレクサ（Amazon Alexa）スマートフォンアプリケーション、Sonos S2コントローラアプリケーション、または同様のアプリケーションを介して提供されてもよい。

(2) 例えば、J. Yang and Y. Chen, "Indoor Localization Using Improved RSS-Based Trilateration Methods," GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference, Honolulu, HI, 2009, pp. 1-6, doi: 10.1109/GLOCOM.2009.5425237 および／またはMardeni, R. & Othman, Shaifull & Nizam, (2010) "Node Positioning in ZigBee Network Using Trilateration Method Based on the Received Signal Strength Indicator (RSSI)" 46 (本明細書中、これら両方の文献を参考として援用する)に開示されているように、ブルートゥース、Wi-Fi、ジグビー等の一般的な無線通信技術の測定された信号強度(受信信号強度表示またはRSSIと呼ぶこともある)を用いて基礎的三辺測定の問題を解き、デバイス間の物理的距離の推定値を生成する。

10

【0092】

$S(a)$ は、デバイス a における発話対エコー比の推定値である。定義によれば、発話対エコー比 (dB) は、次の式により与えられる。

【数3】

$$S(a) = \hat{\mu}_{Sp} - \hat{\mu}_E$$

20

【0093】

上記式において、

$\hat{\mu}_{Sp}$

は発話エネルギーの推定値 (dB) であり、

$\hat{\mu}_E$

30

はエコー除去後の残留エコーエネルギーの推定値 (dB) である。これらの量を推定するためのさまざまな方法を、本明細書中に開示している。例えば、以下のものが挙げられる。

【0094】

(1) 発話エネルギーおよび残留エコーエネルギーは、デバイスのマイクロフォンとスピーカとの間の音響的接続およびオンボードのエコー除去回路のパフォーマンスを考慮して、特定のデバイスについて実行されたオフラインの測定プロセスにより推定されてもよい。いくつかのそのような例において、平均発話エネルギーレベル「Avg Speech」は、公称距離におけるデバイスにより測定された人間の発話の平均レベルによって決定されてもよい。例えば、マイクロフォン内蔵デバイスから1メートル離れて立っている少数の人々からの発話は、生成中に当該デバイスにより記録されてもよく、エネルギーを平均して「Avg Speech」を生成してもよい。いくつかのそのような例によると、平均残留エコーエネルギーレベル「Avg Echo」は、生成中に当該デバイスからの音楽コンテンツを再生し、オンボードのエコー除去回路を実行してエコー残留信号を生成することにより推定されてもよい。音楽コンテンツの小さいサンプルについてのエコー残留信号のエネルギーを平均することを用いて、Avg Echoを推定してもよい。デバイスがオーディオ音声を再生していない場合、Avg Echoは名目上の低い値(−96.0 dBなど)に設定されてもよい。いくつかのそのような実施態様において、発話エネルギーおよび残留エコーエネルギーは、下のように表現され得る。

40

【数4】

50

$$\widehat{\mu}_{Sp} = AvgSpeech_{dB}$$

$$\widehat{\mu}_E = AvgEcho_{dB}$$

【0095】

(2) いくつかの例によると、平均発話エネルギーは、音声区間検出器 (VAD) により決定されたユーザの発声に対応するマイクロフォン信号のエネルギーを取得することにより決定されてもよい。いくつかのそのような例において、平均残留エコーエネルギーは、VADが発話を示していない場合のマイクロフォン信号のエネルギーによって推定されてもよい。xが、デバイスaのマイクロフォンの、あるサンプリングレートにおけるパルス符号変調 (PCM) サンプルであり、Vが、音声区間に対応するサンプルに対して値1.0、その他の場合には値0.0をとるVADフラグである場合、発話エネルギーおよび残留エコーエネルギーは、次のように表現され得る。

【数5】

$$\widehat{\mu}_{Sp} = 10 * \log_{10} \left(\sum_n V[n] x[n]^2 \right)$$

$$\widehat{\mu}_E = 10 * \log_{10} \left(\sum_n (1 - V[n]) x[n]^2 \right)$$

【0096】

(3) 前述の方法に対してさらに、いくつかの実施態様において、マイクロフォン内のエネルギーは、確率変数として扱われ、VAD決定に基づき別個にモデル化されてもよい。発話エネルギーおよびエコーエネルギーそれぞれの統計的モデルSpおよびEは、任意の数の統計的モデリング技術を用いて推定され得る。S(a)を近似するための発話およびエコーの両方についての平均値 (dB) は、それぞれSpおよびEから導出され得る。このことを達成する一般的な方法は、統計的信号処理の分野に存在する。例えば、

- ・ エネルギーのガウス分布を仮定し、バイアスされた2次統計値

$\hat{\mu}$

および

$\hat{\sigma}$

を算出する。

- ・ 離散的ビンからなるエネルギー値のヒストグラムを作成して、マルチモードであり得る分布を得る。ここで、混合モデル (例えばガウス混合モデル) について期待値最大化 (EM) パラメータ推定のステップを適用した後、混合モデルにおける下位分布 (sub-distributions) のいずれかに属する最も大きい平均値

$\hat{\mu}$ を用い得る。

10

20

30

40

50

【0097】

(決定 (DECISION))

本明細書の別の箇所で説明するように、さまざまな開示された実施態様において、決定アスペクトは、いずれのデバイスが、レンダリング修正などのオーディオ処理修正を受信したかを決定し、さらにいくつかの実施形態においては、いずれのデバイスが、所望のSER改善のためにどの程度の修正が要求されているかについての表示を受信したかを決定する。いくつかのそのような実施形態は、例えば集合Dに含まれる全てのデバイスにおけるSの最大値を見つけることによって決定される、最良の初期SER値を有するデバイスにおけるSERを改善するように構成されてもよい。他の実施形態は、履歴的使用パターンに基づいて決定される、ユーザにより定期的に話しかけられるデバイスにおけるSERを日和見的 (opportunisticly) に改善するように構成されてもよい。他の実施形態は、複数のマイクロフォン位置におけるSERの改善を試みるように構成されてもよい。例えば、以下の説明のために複数のデバイスを選択する。

10

【0098】

1つ以上のマイクロフォン位置が決定されると、いくつかのそのような実施態様において、所望のSER改善値 (SERI) が次のように決定され得る。

【数6】

$$SERI = S(m) - TargetSER [dB]$$

20

【0099】

上記式において、mは改善されるデバイス／マイクロフォン位置を示し、TargetSERは閾値であり、これは、使用中のアプリケーションにより設定され得る。例えば、ウェイクワード検出アルゴリズムは、大語彙音声認識器よりも低い動作SERを許容し得る。TargetSERの典型的な値は、-6dB～12dBのオーダーであり得る。前に言及したように、いくつかの実施形態において、S(m)が既知でないかまたは容易に推定されない場合、典型的な反響の室内またはセッティングにおいて記録された発話およびエコーのオフライン測定値に基づく任意のプリセット値で十分であり得る。いくつかの実施形態は、0～1の範囲の f_n を特定することにより、オーディオ処理 (例えばレンダリング) を修正しようとするデバイスを決定してもよい。他の実施形態は、オーディオ処理 (例えばレンダリング) を修正すべき程度を、発話対エコー比改善値 (デシベル) s_n の単位で特定することを含んでもよい。ここで、 s_n は、次式のように算出され得る。

30

【数7】

$$s_n = SERI * f_n$$

【0100】

いくつかの実施形態は、例えば次式のように、デバイスのジオメトリから直接 f_n を算出してもよい。

40

【数8】

$$f_n = 1 - \left(\frac{H(m, n)}{\max_{i \in D} H(m, i)} \right)^2$$

50

【0101】

上記式において、 m は、オーディオ処理（例えばレンダリング）の最も大きい修正のために選択されるデバイスのインデックスである。他の実施態様は、デバイスジオメトリ上の関数を緩和または平滑化するという他の選択肢を含んでもよい。

【0102】

実施形態B（ユーザゾーンの参照）

いくつかの実施形態において、本開示のコンテキストアспектおよび決定アспектは、1つ以上のユーザゾーンが存在する文脈において生成され得る。本明細書において後で詳細に説明するように、1組の音響的特徴

$$W(j) = [w_1^T(j) \dots w_N^T(j)]^T$$

10

を用いて、事後確率

$$p(C_k|W(j))$$

（ C_k はある1組のゾーンラベルであり、

$$k = \{1 \dots K\}$$

20

であり、環境内に K 個の異なるユーザゾーンが存在する）を推定してもよい。各オーディオデバイスを各ユーザゾーンに関連付けることは、本明細書中で説明する訓練プロセスの一部としてユーザ自身により、あるいは、例えばアレクサスマートフォンアプリまたはSonos S2コントローラスマートフォンアプリなどのアプリケーションを介して、実現されてもよい。例えば、いくつかの実施態様は、 j 番目のデバイスをゾーンラベル C_k を有するユーザゾーンに関連付けることを、

$$z(C_k, n) \in [0, 1]$$

と表現してもよい。いくつかの実施形態において、

30

$$z(C_k, n)$$

および事後確率

$$p(C_k|W(j))$$

の両方が、考慮されたコンテキスト情報であってもよい。いくつかの実施形態は、代わりに、音響的特徴 $W(j)$ 自体を、コンテキストの一部と考えてもよい。他の実施形態において、これらの量（

40

$$z(C_k, n)$$

、事後確率

$$p(C_k|W(j))$$

、および音響的特徴 $W(j)$ 自体)のうちの2つ以上、ならびに／もしくは、これらの量の組み合わせが、コンテキスト情報の一部であってもよい。

【0103】

50

さまざまな実施形態の決定アスペクトは、デバイスの選択において1つ以上のユーザゾーンに関係付けられた量を用いてもよい。zおよびpの両方が利用可能である場合、次式のような決定例が為され得る。

【数9】

$$f_n = \max_k (z(C_k, n)p(C_k|W(j)))^\delta \quad k = \{1 \dots K\}$$

そのような実施形態によると、ユーザが居る尤度が最も高いユーザゾーンとの関連度が最も高いデバイスは、適用されるオーディオ処理（例えばレンダリング）の変更が最も大きい。いくつかの例において、 δ は、 $[0.5, 4.0]$ の範囲内の正の数であってもよい。いくつかのそのような例によると、 δ を用いてレンダリング変更の範囲を空間的に制御してもよい。そのような実施態様において、 δ が0.5であると選択される場合、より多くのデバイスがより大きいレンダリング変更を受信するのに対して、値が4.0である場合、レンダリング変更は、最尤ユーザゾーンの最も近くに存在するデバイスのみに限定される。

【0104】

本発明者らはまた、音響的特徴 $W(j)$ が決定アスペクトにおいて直接的に使用されるような、別の種類の実施形態を考える。例えば、発声jに関連付けられたウェイクワード確信度スコアを $w_n(j)$ とする場合、デバイスの選択は、次式に従って為され得る。

【数10】

$$f_n = \left(\frac{w_n(j)}{\max_k w_k(j)} \right)^\delta \quad k = \{1 \dots K\}$$

上記式において、 δ は前述の例と同様に解釈され、特定のウェイクワードシステムについて起こり得るウェイクワード確信度の典型的な分布を補償するという有用性をさらに有する。大部分のデバイスが高いウェイクワード確信度を報告する傾向がある場合、レンダリング変更アプリケーションの空間特異性（spatial specificity）を増大するように、 δ に例えば3.0等の比較的高い数値が選択され得る。ユーザがデバイスからさらに離れるとウェイクワード確信度が急に下降する傾向がある場合、レンダリング変更アプリケーション内により多くのデバイスを含むように、 δ に1.0または0.5等の比較的低い数値が選択され得る。いくつかの別の実施態様において、デバイスのマイクロフォンにおける発話レベルの推定値などの音響的特徴および／またはユーザの発声のDR比（direct to reverb ratio）について上に示したものと同様の式を、ウェイクワード確信度の代わりに使用し得ることが、本明細書から理解され得る。

【0105】

図2Aは、本開示のさまざまな側面を実施できる装置またはシステムの構成要素の例を示すブロック図である。本明細書の他の図面と同様、図2Aに示す要素の種類および個数は例示に過ぎない。他の実施態様は、より多くの、より少ない、かつ／または異なる種類および個数の要素を含んでもよい。いくつかの例によると、装置200は、本明細書に開示する方法の少なくともいくつかを実行するように構成されたデバイスであってもよいし、そのようなデバイスを含んでもよい。いくつかの実施態様において、装置200は、スマートスピーカ、ラップトップコンピュータ、携帯電話、タブレットデバイス、スマートホームハブ、または本明細書に開示する方法の少なくともいくつかを実行するように構成された別のデバイスであってもよいし、その別のデバイスを含んでもよい。いくつかの実施態様において、装置200は、オーディオセッションマネージャを実装するように構成さ

れてもよい。いくつかのそのような実施態様において、装置 200 は、サーバであってもよいし、サーバを含んでもよい。

【0106】

この例において、装置 200 は、インターフェイスシステム 205 および制御システム 210 を含む。インターフェイスシステム 205 は、いくつかの実施態様において、ソフトウェアアプリケーションを実行している（または、実行するように構成された）1つ以上のデバイスと通信するように構成されてもよい。そのようなソフトウェアアプリケーションは、「アプリケーション」または単に「アプリ」と呼ぶこともある。インターフェイスシステム 205 は、いくつかの実施態様において、制御情報およびアプリケーションに関する関連データを交換するように構成されてもよい。インターフェイスシステム 205 は、いくつかの実施態様において、オーディオ環境の1つ以上の他のデバイスと通信するように構成されてもよい。オーディオ環境は、いくつかの例において、家庭オーディオ環境であってもよい。インターフェイスシステム 205 は、いくつかの実施態様において、オーディオ環境のオーディオデバイスとの間で、制御情報および関連データを交換するように構成されてもよい。制御情報および関連データは、いくつかの例において、1つ以上のアプリケーションであって、装置 200 がそのアプリケーションと通信するように構成された、1つ以上のアプリケーションに関係する。

10

【0107】

インターフェイスシステム 205 は、いくつかの実施態様において、オーディオデータを受信するように構成されてもよい。オーディオデータは、オーディオ環境の少なくともいくつかのスピーカにより再生されることが予定されているオーディオ信号を含んでもよい。オーディオデータは、1つ以上のオーディオ信号および関連する空間データを含んでもよい。空間データは、例えば、チャンネルデータおよび／または空間メタデータを含んでもよい。インターフェイスシステム 205 は、オーディオ環境の1組のラウドスピーカのうちの少なくともいくつかのラウドスピーカに、レンダリングされたオーディオ信号を提供するように構成されてもよい。インターフェイスシステム 205 は、いくつかの実施態様において、環境内の1つ以上のマイクロフォンからの入力を受信するように構成されてもよい。

20

【0108】

インターフェイスシステム 205 は、1つ以上のネットワークインターフェイスおよび／または（1つ以上のユニバーサルシリアルバス（USB）インターフェイスなどの）1つ以上の外部デバイスインターフェイスを含んでもよい。いくつかの実施態様によると、インターフェイスシステム 205 は、1つ以上の無線インターフェイスを含んでもよい。インターフェイスシステム 205 は、1つ以上のマイクロフォン、1つ以上のスピーカ、表示システム、タッチセンサシステム、および／またはジェスチャセンサシステムなどの、ユーザインターフェイスを実装する1つ以上のデバイスを含んでもよい。いくつかの例において、インターフェイスシステム 205 は、制御システム 210 とメモリシステム（図 2A に示す任意のメモリシステム 215 など）との間に、1つ以上のインターフェイスを含んでもよい。しかし、いくつかの例において、制御システム 210 はメモリシステムを含んでもよい。

30

40

【0109】

制御システム 210 は、例えば、シングルチップまたはマルチチップの汎用プロセッサ、デジタルシグナルプロセッサ（DSP）、特定用途向け集積回路（ASIC）、フィールドプログラマブルゲートアレイ（FPGA）または他のプログラマブルロジックデバイス、ディスクリートゲートまたはトランジスタロジック、ならびに／もしくは、ディスクリートハードウェアコンポーネントを含んでもよい。

【0110】

いくつかの実施態様において、制御システム 210 は、複数のデバイス内に設けられてもよい。例えば、制御システム 210 の一部が、本明細書で図示する環境の1つに存在するデバイスに設けられ、制御システム 210 の別の部分が、サーバやモバイルデバイス（

50

例えばスマートフォンまたはタブレットコンピュータ) など、当該環境の外部に存在するデバイスに設けられてもよい。他の例において、制御システム 210 の一部が、本明細書で図示する環境の 1 つに存在するデバイスに設けられ、制御システム 210 の別の部分が、当該環境の 1 つ以上の他のデバイスに設けられてもよい。例えば、制御システムの機能は、この環境の複数のスマートオーディオデバイスに分配されてもよいし、オーケストレーションデバイス (本明細書中、「オーディオセッションマネジャ」または「スマートホームハブ」と呼ぶものなど) と、この環境の 1 つ以上の他のデバイスとによって共有されてもよい。インターフェイスシステム 205 はまた、いくつかのそのような例において、複数のデバイスに設けられてもよい。

【0111】

いくつかの実施態様において、制御システム 210 は、本明細書に開示した方法を少なくとも部分的に実行するように構成されてもよい。いくつかの例によると、制御システム 210 は、オーディオセッション管理方法を実装するように構成されてもよい。このオーディオセッション管理方法は、いくつかの例において、オーディオ環境の 2 つ以上のオーディオデバイス用のラウドスピーカフィード信号へとレンダリングされているオーディオデータに適用する、1 種類以上のオーディオ処理変更を決定することを含んでもよい。いくつかの実施態様によると、オーディオ処理変更は、オーディオ環境内の 1 つ以上のマイクロフォンにおける発話対エコー比を増大する効果を有してもよい。

【0112】

本明細書に開示した方法のいくつかまたは全ては、1 つ以上の非一時的な媒体に格納された命令 (例えばソフトウェア) に応じて、1 つ以上のデバイスにより実行されてもよい。そのような非一時的な媒体は、本明細書に記載されたようなメモリデバイス (ランダムアクセスメモリ (RAM) デバイス、リードオンリーメモリ (ROM) などを含むが、これらに限定されない) を含んでもよい。1 つ以上の非一時的な媒体は、例えば、図 2 A に示す任意のメモリシステム 215 および／または制御システム 210 内に設けられてもよい。したがって、本開示に記載された主題のさまざまな革新的側面が、ソフトウェアを格納している 1 つ以上の非一時的な媒体において実装され得る。ソフトウェアは、例えば、少なくとも 1 つのデバイスを制御して、オーディオセッション管理方法を実装する命令を含んでもよい。ソフトウェアは、いくつかの例において、オーディオ環境の 1 つ以上のオーディオデバイスを制御して、オーディオデータの獲得、処理および／または提供を行う命令を含んでもよい。ソフトウェアは、いくつかの例において、オーディオ環境の 2 つ以上のオーディオデバイス用のラウドスピーカフィード信号へとレンダリングされているオーディオデータに適用する、1 種類以上のオーディオ処理変更を決定することを含んでもよい。いくつかの実施態様によると、オーディオ処理変更は、オーディオ環境内の 1 つ以上のマイクロフォンにおける発話対エコー比を増大する効果を有してもよい。ソフトウェアは、例えば、図 2 A の制御システム 210 などの制御システムの 1 つ以上の構成要素により実行可能であってもよい。

【0113】

いくつかの例において、装置 200 は、図 2 A に示す任意のマイクロフォンシステム 220 を含んでもよい。任意のマイクロフォンシステム 220 は、1 つ以上のマイクロフォンを含んでもよい。いくつかの実施態様において、前記マイクロフォンのうち 1 つ以上のマイクロフォンは、スピーカシステムのスピーカやスマートオーディオデバイス等の、別のデバイスの一部であってもよいし、別のデバイスに関連付けられてもよい。いくつかの例において、装置 200 は、マイクロフォンシステム 220 を含まなくてもよい。しかし、いくつかのそのような実施態様において、装置 200 は、それでもなお、オーディオ環境内の 1 つ以上のマイクロフォンのためのマイクロフォンデータを、インターフェイスシステム 210 を介して受信するように構成されてもよい。

【0114】

いくつかの実施態様によると、装置 200 は、図 2 A に示す任意のラウドスピーカシステム 225 を含んでもよい。任意のスピーカシステム 225 は、1 つ以上のラウドスピー

10

20

30

40

50

力を含んでもよい。本明細書中、ラウドスピーカを「スピーカ」と呼ぶこともある。いくつかの例において、任意のラウドスピーカシステム 225 の少なくともいくつかのラウドスピーカは、任意の位置に配置されてもよい。例えば、任意のラウドスピーカシステム 225 の少なくともいくつかのスピーカは、Dolby5.1、Dolby5.1.2、Dolby7.1、Dolby7.1.4、Dolby9.1、Hamasaki22.2などの、任意の標準的な規定のスピーカレイアウトに対応しない位置に配置されてもよい。いくつかのそのような例において、任意のラウドスピーカシステム 225 の少なくともいくつかのラウドスピーカは、空間にとって都合のいい位置（例えばラウドスピーカを収容する空間がある位置）に配置されてよく、任意の標準的な規定のスピーカレイアウトに配置しなくてもよい。いくつかの例において、装置 200 は、任意のラウドスピーカシステム 225 を含まなくてもよい。

10

【0115】

いくつかの実施態様において、装置 200 は、図 2 A に示す任意のセンサシステム 230 を含んでもよい。任意のセンサシステム 230 は、1つ以上のカメラ、タッチセンサ、ジェスチャセンサ、人感センサ（motion detectors）などを含んでもよい。いくつかの実施態様によると、任意のセンサシステム 230 は、1つ以上のカメラを含んでもよい。いくつかの実施態様において、カメラは、自立型カメラであってもよい。いくつかの例において、任意のセンサシステム 230 の1つ以上のカメラは、スマートオーディオデバイス内に設けられてもよい。ここで、スマートオーディオデバイスは、単用途オーディオデバイスまたはバーチャルアシスタントであってもよい。いくつかのそのような例において、任意のセンサシステム 230 の1つ以上のカメラは、テレビ、携帯電話、またはスマートスピーカ内に設けられてもよい。いくつかの例において、装置 200 は、センサシステム 230 を含まなくてもよい。しかし、いくつかのそのような実施態様において、装置 200 は、それでもなお、オーディオ環境内の1つ以上のセンサのためのセンサデータを、インターフェイスシステム 210 を介して受信するように構成されてもよい。

20

【0116】

いくつかの実施態様において、装置 200 は、図 2 A に示す任意の表示システム 235 を含んでもよい。任意の表示システム 235 は、1つ以上の発光ダイオード（LED）ディスプレイなど、1つ以上の表示装置を含んでもよい。いくつかの例において、任意の表示システム 235 は、1つ以上の有機発光ダイオード（OLED）ディスプレイを含んでもよい。装置 200 が表示システム 235 を含むいくつかの例において、センサシステム 230 は、表示システム 235 の1つ以上の表示装置に最も近いタッチセンサシステムおよび／またはジェスチャセンサシステムを含んでもよい。いくつかのそのような実施態様によると、制御システム 210 は、表示システム 235 を制御して、1つ以上のグラフィカルユーザインターフェイス（GUI）を提示するように構成されてもよい。

30

【0117】

いくつかの例によると、装置 200 は、スマートオーディオデバイスであってもよいし、スマートオーディオデバイスを含んでもよい。いくつかのそのような実施態様において、装置 200 は、ウェイクワード検出器であってもよいし、ウェイクワード検出器を（少なくとも部分的に）実装してもよい。例えば、装置 200 は、バーチャルアシスタントであってもよいし、バーチャルアシスタントを（少なくとも部分的に）実装してもよい。

40

【0118】

図 2 B は、いくつかの実施態様によるオーディオセッション管理方法のブロックを含む流れ図である。方法 250 のブロックは、本明細書に記載の他の方法と同様、必ずしも図示した順番で実行される必要はない。いくつかの実施態様において、方法 250 のブロックのうち、1つ以上のブロックを同時に実行してもよい。さらに、方法 250 のいくつかの実施態様は、図示したブロックおよび／または説明したブロックよりも多くの個数のブロックまたは少ない個数のブロックを含んでもよい。方法 250 のブロックは、図 2 A を参照して上で説明した制御システム 210 などの制御システムまたは開示した他の制御システム例のうちの1つであり得る（または、その制御システムを含み得る）1つ以上のデバイスにより実行されてもよい。いくつかの実施態様によると、方法 250 のブロックは

50

、少なくとも部分的に、本明細書においてオーディオセッションマネージャと呼ぶものを実装しているデバイスにより実行されてもよい。

【0119】

この例によると、ブロック255は、オーディオ環境内の複数のマイクロフォンの各々から出力信号を受信することを含む。この例において、複数のマイクロフォンの各々はオーディオ環境のマイクロフォン位置に設けられており、出力信号は人の現在の発声に対応する信号を含む。いくつかの例において、現在の発声は、ウェイクワードの発声であってもよい。しかし、出力信号は、人が話していない時間に対応する信号を含み得る。そのような信号は、例えば、エコーや雑音などのベースラインレベルを設定するために使用され得る。

10

【0120】

この例において、ブロック260は、出力信号に基づき、人に関するコンテキスト情報の1つ以上のアスペクトを決定することを含む。この実施態様において、コンテキスト情報は、この人の推定される現在位置、および／または、1つ以上のマイクロフォン位置に対するこの人の推定される現在の近接度を含む。上述のとおり、本明細書で使用する「マイクロフォン位置」という表現は、1つ以上のマイクロフォンの位置を示す。いくつかの例において、1つのマイクロフォン位置は、1つのオーディオデバイス内の複数のマイクロフォンからなる1つのマイクロフォンアレイに対応していてもよい。例えば、マイクロフォン位置は、1つ以上のマイクロフォンを含むオーディオデバイス全体に対応する1つの位置であってもよい。いくつかのそのような例において、マイクロフォン位置は、1つのオーディオデバイスのマイクロフォンアレイの重心に対応する1つの位置であってもよい。しかし、いくつかの例において、マイクロフォン位置は、1つのマイクロフォンの位置であってもよい。いくつかのそのような例において、オーディオデバイスは、1つのマイクロフォンだけを有していてもよい。

20

【0121】

いくつかの例において、コンテキスト情報を決定することは、人が現在位置しているユーザゾーンの推定値を生成することを含んでもよい。いくつかのそのような例は、各マイクロフォンの出力信号から複数の現在の音響的特徴を決定することと、これら複数の現在の音響的特徴に分類器を適用することとを含んでもよい。分類器を適用することは、例えば、環境内の複数のユーザゾーン内でその人によってなされた複数の過去の発声から得られた過去に決定された音響的特徴に対して訓練されたモデルを適用することを含んでもよい。いくつかのそのような例において、その人に関するコンテキスト情報の1つ以上のアスペクトを決定することは、分類器からの出力に少なくとも部分的に基づき、その人が現在位置しているユーザゾーンの推定値を決定することを含んでもよい。いくつかのそのような例において、ユーザゾーンの推定値は、複数のマイクロフォンの幾何学的位置を参照することなく決定されてもよい。いくつかの例によると、現在の発声および過去の発声は、ウェイクワードの発声であってもよいし、ウェイクワードの発声を含んでもよい。

30

【0122】

この実施態様によると、ブロック265は、コンテキスト情報の1つ以上のアスペクトに少なくとも部分的に基づいて、オーディオ環境の2つ以上のオーディオデバイスを選択することを含み、2つ以上のオーディオデバイスの各々は、少なくとも1つのラウドスピーカを含んでいる。いくつかの例において、オーディオ環境の2つ以上のオーディオデバイスを選択することは、オーディオ環境のN個のラウドスピーカ内蔵オーディオデバイス（Nは2より大きい整数）を選択することを含んでもよい。いくつかの例において、オーディオ環境の2つ以上のオーディオデバイスを選択すること、または、オーディオ環境のN個のラウドスピーカ内蔵オーディオデバイスを選択することは、オーディオ環境の全てのラウドスピーカ内蔵オーディオデバイスを選択することを含んでもよい。

40

【0123】

いくつかの例において、オーディオ環境の2つ以上のオーディオデバイスを選択することは、マイクロフォン位置および／またはラウドスピーカ内蔵オーディオデバイス位置に

50

対する人の推定される現在位置に少なくとも部分的に基づいてもよい。いくつかのそのような例は、この人の推定される現在位置に最も近い最近位ラウドスピーカ内蔵オーディオデバイスを決定すること、または、この人の推定される現在位置に最も近いマイクロフォン位置に最も近い最近位ラウドスピーカ内蔵オーディオデバイスを決定することを含んでもよい。いくつかのそのような例において、2つ以上のオーディオデバイスは、最近位ラウドスピーカ内蔵オーディオデバイスを含んでもよい。

【0124】

いくつかの実施態様によると、2つ以上のオーディオデバイスを選択することは、信号対エコー比推定値が信号対エコー比閾値以下であるか否かを決定することに少なくとも部分的に基づいてもよい。

【0125】

この例によると、ブロック270は、2つ以上のオーディオデバイス用のラウドスピーカフィールド信号へとレンダリングされているオーディオデータに適用する、1種類以上のオーディオ処理変更を決定することを含む。この実施態様において、オーディオ処理変更は、1つ以上のマイクロフォンにおける発話対エコー比を増大する効果を有する。いくつかの例において、1つ以上のマイクロフォンは、オーディオ環境の複数のオーディオデバイス内に設けられてもよい。しかし、いくつかの実施態様によると、1つ以上のマイクロフォンは、オーディオ環境の1つのオーディオデバイス内に設けられてもよい。いくつかの例において、オーディオ処理変更は、2つ以上のオーディオデバイスのラウドスピーカのラウドスピーカ再生レベルの低下を生じてもよい。

【0126】

いくつかの例によると、第1のオーディオデバイスに対するオーディオ処理変更の少なくとも1つは、第2のオーディオデバイスに対するオーディオ処理変更とは異ってもよい。例えば、オーディオ処理変更（単数または複数）は、第1のオーディオデバイスの第1のラウドスピーカのラウドスピーカ再生レベルに第1の低下を生じ、かつ、第2のオーディオデバイスの第2のラウドスピーカのラウドスピーカ再生レベルに第2の低下を生じてもよい。いくつかのそのような例において、ラウドスピーカ再生レベルの低下は、人の推定される現在位置に対する（または人の推定される現在位置に最も近いマイクロフォン位置に対する）近接度がより高いオーディオデバイスに対して、相対的により大きくなってもよい。

【0127】

しかし、本発明者らは、いくつかの例において為し得る多くの種類のオーディオ処理変更を考える。いくつかの実施態様によると、1種類以上のオーディオ処理変更は、レンダリングプロセスを変更して、人の推定される現在位置から離れる方向（または人の推定される現在位置に最も近いマイクロフォン位置から離れる方向に）に、オーディオ信号のレンダリングをワーピングすることを含んでもよい。

【0128】

いくつかの実施態様において、1種類以上のオーディオ処理変更は、スペクトル修正を含んでもよい。例えば、スペクトル修正は、500Hzと3kHzとの間の周波数帯域においてオーディオデータのレベルを低下させることを含んでもよい。他の例において、スペクトル修正は、より高い最大周波数および／またはより低い最小周波数を有する周波数帯域においてオーディオデータのレベルを低下させることを含んでもよい。いくつかの実施態様によると、1種類以上のオーディオ処理変更は、オーディオ再生信号の少なくとも1つの選択された周波数帯域に、少なくとも1つのギャップを挿入することを含んでもよい。

【0129】

いくつかの実施態様において、1種類以上のオーディオ処理変更を決定することは、信号対エコー比推定値に少なくとも部分的に基づくコスト関数の最適化に基づいてもよい。いくつかの例において、コスト関数は、レンダリングパフォーマンスに少なくとも部分的に基づいてもよい。

10

20

30

40

50

【0130】

この例によると、ブロック275は、1種類以上のオーディオ処理変更を適用させることを含んでもよい。いくつかの例において、ブロック275は、オーディオ環境内でオーディオ処理を制御している1つ以上のデバイスにより、1種類以上のオーディオ処理変更を適用することを含んでもよい。別の例において、ブロック275は、（例えば、オーディオセッションマネージャからのコマンドまたは制御信号を介して）オーディオ環境の1つ以上の他のデバイスにより、1種類以上のオーディオ処理変更を適用させることを含んでもよい。

【0131】

方法250のいくつかの実施態様は、コンテキスト情報の1つ以上のアспектに依拠して少なくとも1つのマイクロフォンを選択することを含んでもよい。いくつかのそのような実施態様において、方法250は、1つ以上のマイクロフォン位置に対する人の推定される現在の近接度に応じて、少なくとも1つのマイクロフォンを選択することを含んでもよい。方法250のいくつかの実施態様は、ユーザゾーンの推定値に応じて少なくとも1つのマイクロフォンを選択することを含んでもよい。いくつかのそのような実施態様によると、方法250は、選択されたマイクロフォン（単数または複数）から受信したマイクロフォン信号に応じて、バーチャルアシスタント機能を少なくとも部分的に実装することを含んでもよい。いくつかのそのような実施態様において、方法250は、選択されたマイクロフォン（単数または複数）から受信したマイクロフォン信号に少なくとも部分的に基づくテレビ会議機能を提供することを含んでもよい。

【0132】

いくつかの実施形態は、レンダリングおよびマッピングを実装するように構成され、かつ、オーディオ処理を変更する（例えば、最も近くのラウドスピーカのうちの1つ、2つ、またはそれ以上のスピーカの音量を下げる）ために、ソフトウェアまたは他のロジック報告（manifestations of logic）を用いるように（例えば、ロジックを実装するシステム要素を含むように）構成された、（2つ以上のデバイス、例えば、スマートオーディオデバイスを含む）システムを提供する。ロジックは、オーディオセッションマネージャを実装するように構成されたデバイスなどの、管理者（supervisor）を実装してもよい。管理者は、いくつかの例において、レンダリングのために構成されたシステム要素とは別個に実行され得る。

【0133】

図3Aは、別個のレンダリング制御および複数のデバイスに亘るリスニングまたはキャプチャロジックを実装するように構成されたシステムのブロック図である。他の開示された図面と同様、図3A、図3Bおよび図3Cに示す要素の個数、種類および配置は例示に過ぎない。他の実施態様は、より多くの要素、より少ない要素、および／または異なる種類の要素を含んでもよい。例えば、他の実施態様は、4つ以上のオーディオデバイス、異なる種類のオーディオデバイスなどを含んでもよい。

【0134】

図3A、図3Bおよび図3Cに示すモジュールは、本開示において図示および説明する他のモジュールと同様、特定の例に応じて、ハードウェア、ソフトウェア、ファームウェア等を介して実装されてもよい。いくつかの実施態様において、開示されたモジュール（いくつかの例において、「要素」と呼ぶこともある）のうちの1つ以上のモジュールは、図2Aを参照して説明した制御システム210などの制御システムを介して実装されてもよい。いくつかのそのような例において、開示されたモジュールのうちの1つ以上のモジュールは、1つ以上のそのような制御システムにより実行されるソフトウェアに従って実装されてもよい。

【0135】

図3Aの要素は、以下のものを含む。

オーディオデバイス302、303および304（いくつかの例において、スマートオーディオデバイスであってもよい）。この例によると、オーディオデバイス302、3

10

20

30

40

50

03および304の各々は、少なくとも1つのラウドスピーカおよび少なくとも1つのマイクロフォンを含む。

【0136】

— 要素300は、オーディオデバイス302、303および304のうちの1つ以上のオーディオデバイスに亘って再生される、オーディオデータを含む、ある形態のコンテンツを示す。コンテンツ300は、線形のコンテンツであっても対話型のコンテンツであってもよく、これは特定の実施態様に依拠して決まる。

【0137】

— モジュール301は、レンダリングロジックに依拠してレンダリングを行うことを含む（但し、このことに限定されない）、オーディオ処理のために構成されている。例えば、いくつかの実施形態において、モジュール301は、単に、3つのオーディオデバイス302、303および304の全てに対して均等に、コンテンツ300のオーディオ（例えば、モノラルまたはステレオ）を複製するように構成されてもよい。いくつかの別の実施形態において、オーディオデバイス302、303および304のうちの1つ以上のオーディオデバイスは、レンダリング機能を含む（但し、このことに限定されない）、オーディオ処理機能を実装するように構成されてもよい。

【0138】

— 要素305は、オーディオデバイス302、303および304に分配される信号を示す。いくつかの例において、信号305は、スピーカフィールド信号であってもよいし、スピーカフィールド信号を含んでもよい。上述のとおり、いくつかの実施形態において、モジュール301の機能は、オーディオデバイス302、303および304のうちの1つ以上のオーディオデバイスを介して実装されてもよく、この場合、信号305は、オーディオデバイス302、303および304のうちの1つ以上のオーディオデバイスに限定されてもよい。しかし、図3Aでは、それらを1組のスピーカフィールド信号として示す。なぜなら、いくつかの実施形態（例えば、後で図4を参照して説明する実施形態）は、信号305の単純な最終的な遮断（final interception）または後処理を実装するからである。

【0139】

— 要素306は、オーディオデバイス302、303および304のマイクロフォンによりキャプチャされた生のマイクロフォン信号を示す。

【0140】

— モジュール307は、マイクロフォン信号処理ロジックと、いくつかの例において、マイクロフォン信号キャプチャロジックとを実装するように構成されている。この例において、オーディオデバイス302、303および304の各々が1つ以上のマイクロフォンを有するので、キャプチャされた生信号306は、モジュール307により処理される。いくつかの実施形態において、ここでのとおり、モジュール307は、エコー除去および／またはエコー検出機能を実装するように構成されてもよい。

【0141】

— 要素308は、モジュール301～モジュール307により供給される、ローカルなエコー参照信号および／またはグローバルなエコー参照信号を示す。この例によると、モジュール307は、ローカルなエコー参照信号および／またはグローバルなエコー参照信号308に依拠して、エコー除去および／またはエコー検出機能を実装するように構成される。いくつかの実施形態において、マイクロフォンキャプチャ処理および／または生マイクロフォン信号の処理は、オーディオデバイス302、303および304の各々におけるローカルなエコー除去および／または検出ロジックとともに、分配されてもよい。キャプチャおよびキャプチャ処理の特定の実施形態が、レンダリングに施す任意の変更がSER全体ならびにキャプチャ処理およびロジックの効力に与えるインパクトを算出しかつ理解するという発想にとって、重要なわけではない。

【0142】

— モジュール309は、（例えば、所望のオーディオ音声を、特定の単一の位置また

10

20

30

40

50

は幅広い位置から発せられたものとして知覚させることを目的として、) キャプチャされたオーディオ音声の混合または組み合わせの全体を実装するシステム要素である。いくつかの実施態様において、モジュール307はまた、要素309の混合機能を提供してもよい。

【0143】

— モジュール310は、発声されたものについて、もしくは、オーディオ環境内で注目しているアクティビティが為されたか否かについて、何らかの決定を行うために、検出されたオーディオ音声の処理を行う最終アスペクトを実装するシステム要素である。モジュール310は、例えば、オーディオ環境内で人々が行っている行為に関するコンテキスト、オーディオ環境内の全体の雑音レベルの程度などについて、自動音声認識 (ASR) 機能や、背景雑音レベルおよび／または背景雑音の種類を検知する機能を提供してもよい。いくつかの実施態様において、モジュール310の機能のうち、いくつかの機能または全ての機能は、オーディオデバイス302、303および304が設けられたオーディオ環境の外側で (例えば、クラウド型サービスプロバイダの1つ以上のデバイス (例えば1つ以上のサーバ) において) 実装されてもよい。

【0144】

図3Bは、別の開示された実施態様によるシステムのブロック図である。この例において、図3Bに示すシステムは、図3Aのシステムの要素を含み、いくつかの開示された実施形態に応じた機能を含むように、図3Aのシステムを拡張する。図3Bのシステムは、動作中の分散型オーディオシステムに適用される、コンテキスト (CONTEXT)、決定 (DECISION) およびレンダリングアクション (RENDERING ACTION) のアスペクトを実装する要素を含む。いくつかの例によると、コンテキスト (CONTEXT)、決定 (DECISION) およびレンダリングアクション (RENDERING ACTION) のアスペクトを実装する要素へのフィードバックは、アクティビティ (例えば、検出された発話) が存在する場合の確信度の上昇、もしくは、アクティビティの感覚 (the sense of activity) を自信をもって減少し (アクティビティの低尤度)、オーディオ処理を初期状態に戻す能力の、いずれかを生じさせてもよい。

【0145】

図3Bの要素は、以下のものを含む。

— モジュール351は、コンテキスト (CONTEXT) のステップを示す (そして実装する) システム要素であって、例えば、オーディオ音声をより良く検出すること (例えば1つ以上のマイクロフォンにおいて発話対エコー比を増大すること) が望まれ得る位置の表示と、我々が聞きたいという尤度または感覚 (例えば、ウェイクワードまたはコマンドなどの発話が1つ以上のマイクロフォンによりキャプチャされることの尤度) とを獲得する。この例において、モジュール351および353は、制御システム (この例では、図2Aの制御システム210) を介して実装される。いくつかの実施態様において、ブロック301および307もまた、制御システム (いくつかの例において、制御システム210であり得る) により実装されてもよい。いくつかの実施態様によると、ブロック356、357および358もまた、制御システム (いくつかの例において、制御システム210であり得る) により実装されてもよい。

【0146】

— 要素352は、モジュール351へのフィードバック経路を示す。この例において、フィードバック352は、モジュール310により提供される。いくつかの実施形態において、フィードバック352は、コンテキストを決定することに関連し得るマイクロフォン信号のキャプチャから得られる、オーディオ処理 (ASRのためのオーディオ処理など) の結果に対応してもよい。例えば、ウェイクワードの弱検出または早期検出もしくは発話アクティビティの低検出の感覚を用いて、リスニングの改善 (例えば、1つ以上のマイクロフォンにおける発話対エコー比の上昇) を必要とするコンテキストの確信度または感覚を上昇し始めてもよい。

【0147】

10

20

30

40

50

－ モジュール 3 5 3 はシステム要素であって、このシステム要素において（またはこのシステム要素によって）、いずれのオーディオデバイスのオーディオ処理を変更するか、および、オーディオ処理を変更する量に関して決定が為される。モジュール 3 5 3 は、特定の実施態様に依拠して、オーディオデバイスの種類および／または能力（例えば、ラウドスピーカ能力、エコー抑制能力等）やオーディオデバイスの尤もらしい向き等の、特定のオーディオデバイス情報を用いても、用いなくてもよい。いくつかの例において後で説明するように、モジュール 3 5 3 の決定実行プロセスは、スマートスピーカまたは他のラウドスピーカと比較して、ヘッドホンデバイスについて、大きく異なってもよい。

【0148】

－ 要素 3 5 4 は、制御経路 3 5 5（信号経路 3 5 5 と呼ぶこともある）を介して個々のレンダリングブロックへと出力される、モジュール 3 5 3 の出力である。この例において、モジュール 3 5 3 の出力は、図 3 B において f_n 値として示される、1 組の制御関数である。この 1 組の制御関数は、この信号経路 3 5 5 が当該オーディオ環境に限定されるように、（例えば、無線伝送を介して）伝達されてもよい。この例において、制御関数は、モジュール 3 5 6、3 5 7 および 3 5 8 に提供される。

【0149】

－ モジュール 3 5 6、3 5 7 および 3 5 8 は、オーディオ処理（出力されたレンダリング（いくつかの実施形態のレンダリング（RENDER）アспект）を含み得るが、これに限定されない）を変更するように構成されたシステム要素である。この例において、モジュール 3 5 6、3 5 7 および 3 5 8 は、出力 3 5 4 の制御関数（この例では f_n 値）により起動が制御される。いくつかの実施態様において、モジュール 3 5 6、3 5 7 および 3 5 8 の機能は、ブロック 3 0 1 を介して実装されてもよい。

【0150】

図 3 B の実施形態および他の実施態様において、好循環のフィードバックが起こり得る。（いくつかの例において、自動音声認識（ASR）を実装し得る）要素 3 1 0 の出力 3 5 2 が、いくつかの例に依拠して発話を検出した場合、それが弱検出であっても（例えば、低確信度であっても）、コンテキスト（CONTEXT）要素 3 5 1 は、オーディオ環境内のどのマイクロフォン（単数または複数）が音声をキャプチャしたか（例えば、どのマイクロフォン（単数または複数）がエコー以外のエネルギーを最も多く有したか）に基づき、位置を推定してもよい。いくつかのそのような例によると、決定（DECISION）ブロック 3 5 3 は、オーディオ環境の 1 つ、2 つ、または 3 つ以上のラウドスピーカを選択してもよく、かつ、レンダリングの変更に関する小さい値（例えば、 $f_n = 0.25$ ）をアクティベート（activate）してもよい。全体で 20 dB の音量低減（ダッキング）をした場合、この値は、選択されたデバイス（単数または複数）において、平均的な人間の聴力で感知し得る約 5 dB の音量減少を生じる。時定数および／またはイベント検出と組み合わせた場合に、そしてオーディオ環境の他のラウドスピーカが同様のコンテンツを再生している場合、レベルの低減はより感知されにくくなり得る。ある例において、音量を下げるのは、オーディオデバイス 3 0 3（話している人 3 1 1 の最も近くのオーディオデバイス）であり得る。他の例において、オーディオデバイス 3 0 2 および 3 0 3 両方の音量を、いくつかの例においては（例えば人 3 1 1 に対する推定近接度に応じて）異なる量だけ、下げてもよい。他の例において、オーディオデバイス 3 0 2、3 0 3 および 3 0 4 全ての音量を、いくつかの例においては異なる量だけ、下げてもよい。オーディオデバイス 3 0 2、3 0 3 および 3 0 4 のうちの 1 つ以上のオーディオデバイスによる再生のレベルを下げた結果、発話対エコー比は、人 3 1 1 の近くの 1 つ以上のマイクロフォン（例えば、オーディオデバイス 3 0 3 の 1 つ以上のマイクロフォン）において増大され得る。従って、人 3 1 1 が話し続けた（例えば、ウェイクワードを繰り返し続けた、または、コマンドを発し続けた）場合、システムは人 3 1 1 の声をより良く「聞く」ことができる。いくつかのそのような実施態様において、次の期間のあいだ（例えば次の数秒間）に、いくつかの例においては連続的な様態で、システム（例えば、少なくとも部分的にブロック 3 5 1 および 3 5 3 を介して実装されるオーディオセッションマネジャ）が、例えば f_2

10

20

30

40

50

= 1 を選択することにより、人 3 1 1 の近くの 1 つ以上のラウドスピーカの音量をオフにすることへと素早く切り替わってもよい。

【0151】

図 3 C は、ある例によるエネルギーバランシングネットワークを実装するように構成された実施形態のブロック図である。図 3 C は、図 3 B のシステムの要素を含み、かつ、エネルギーの補償を実装する要素（例えば、要素 3 7 1）を含むように（例えば、「他のデバイスの音量を少し上げる」ように）図 3 B のシステムを拡張した、システムのブロック図である。

【0152】

いくつかの例において、図 3 C のシステム（または図 3 C のようなシステム）のオーディオセッション管理（オーディオセッションマネジャ）のために構成されたデバイスは、1 つ以上のマイクロフォンにおいて発話対エコー比を増大するために適用される、オーディオ処理（例えば、1 つ以上の選択されたラウドスピーカ（例えば、 $f_n > 0$ である制御信号を受信するオーディオデバイスのラウドスピーカ）のレベルの低減）の結果失われた、聴者（3 1 1）におけるバンデッドエネルギー（banded energy）を評価してもよい。その後、オーディオセッションマネジャは、SER 型のオーディオ処理変更を補償するために、レベルの上昇および／または他の形態のエネルギーバランシングを、オーディオ環境の他のスピーカに適用してもよい。

【0153】

少し関係があるコンテンツをレンダリングしており、かつ、関連しているかまたは類似のスペクトルを有するオーディオ成分が、オーディオ環境の複数のラウドスピーカにより再生されている場合（単純な例はモノラル再生）によくあることだが、エネルギーバランシングをそれほど行う必要がないかもしれない。例えば、オーディオ環境内に、1 ～ 2 の割合の範囲の距離（1 が最も近い）を空けて 3 つのラウドスピーカが存在する場合、それらのラウドスピーカにより同一のコンテンツが再生されている場合は、最も近くのラウドスピーカの音量を 6 dB 下げても、2 ～ 3 dB のインパクトにしかない。最も近くのラウドスピーカをオフにしても、聴者の位置における音声にとって、全体で 3 ～ 4 dB のインパクトにしかない。

【0154】

（例えば、ギャップの挿入または空間的ステアリングの）より複雑な状況において、いくつかの例では、エネルギーの保存（energy preservation）および知覚の連続性（perceptual continuity）の形態は、より多元的なエネルギーバランスであり得る。

【0155】

図 3 C において、コンテキスト（CONTEXT）を実装する要素（単数または複数）は、いくつかの例において、ウェイクワード弱検出のオーディオレベル（近接度の相互性（reciprocity of proximity））であってもよい。換言すると、コンテキスト（CONTEXT）を決定することの一例は、検出されたエコーを介して検出された何らかのウェイクワード発声のレベルに基づいてもよい。このような方法は、発話対エコー比を決定するステップを実際に含んでもよいし、含まなくてもよく、このことは特定の実施態様に依拠して決まる。しかし、いくつかの例において、複数のマイクロフォン位置の各々において検出されたウェイクワードの発声のレベルを検出および評価するだけで、コンテキスト（CONTEXT）の十分なレベルが提供され得る。

【0156】

（例えば、図 3 C のシステムにおいて）コンテキスト（CONTEXT）を実装するシステム要素により実装される方法のいくつかの例は、以下の事項を含むが、それらに限定されない。

【0157】

— ウェイクワードの一部を検出すると、マイクロフォン内蔵オーディオデバイスへの近接度が、ウェイクワード確信度から推断され得る。ウェイクワード発声のタイミングもまた、ウェイクワード確信度から推断され得る。

10

20

30

40

50

【0158】

— 生のマイクロフォン信号に適用されたエコー除去およびエコー抑制に加えて、いくつかのオーディオアクティビティが検出される。いくつかの実施態様は、1組のエネルギーレベルおよび分類を用いて、そのオーディオアクティビティがボイスアクティビティであることがどれくらい尤もらしいかを決定する（ボイスアクティビティ検出）。このプロセスは、ボイスアクティビティの確信度または尤度を決定してもよい。声の位置は、インタラクションの類似の状況にとって最良のマイクロフォンの可能性に基づき得る。例えば、オーディオセッションマネージャを実装しているデバイスは、ユーザがよく居る位置の近くには存在しない壁設置デバイスよりも、ユーザがよく居る位置かまたはその位置の近くにあるテーブル上のデバイスなど、1つのマイクロフォン内蔵オーディオデバイスがユーザの最も近くに存在することを、予め知っているかもしれない。

10

【0159】

（例えば、図3Cのシステムにおいて）決定（DECISION）を実装するシステム要素の実施形態例は、ボイスアクティビティに関する確信度値を決定すること、および、いずれのデバイスが最も近くのマイクロフォン内蔵オーディオデバイスであるかを決定することを行うように構成された要素である。

【0160】

図3Cのシステム（および他の実施形態）において、任意の位置におけるSERを増大するために適用するオーディオ処理変更（単数または複数）の量は、距離およびボイスアクティビティに関する確信度の関数であってもよい。

20

【0161】

（例えば、図3Cのシステムにおいて）レンダリング（RENDERING）を実装する方法のいくつかの例は、以下の事項を含む。

dBのみを下げる、および／または、

スピーチバンドイコライゼーション（EQ）（例えば、後で図4を参照して説明する）および／または、

レンダリングの変更の時変調（Time modulation）（図5を参照して説明する）、および／または、

一時的なタイムスライシングまたはタイムアジャストメントを用いて、注目しているオーディオ音声の細断片（glimpses）を得るのに十分なスパース時間周波数のより低い出力の「ギャップ」または期間を生成（例えば、オーディオコンテンツに挿入）すること。後で、図9を参照していくつかの例を説明する。

30

【0162】

図4は、オーディオ環境の1つ以上のマイクロフォンにおける発話対エコー比を増大し得るオーディオ処理の例を示すグラフである。図4のグラフは、スペクトル修正の例を提供する。図4において、スペクトル修正は、発話に対応していることがわかっている周波数（これらの例において、およそ200Hz～10KHzの範囲内の周波数（例えば、この範囲の高い側および／または低い側の周波数の5%または10%以内））のレベルを低減することを含んでいる。他の例は、異なる周波数帯域（例えば、およそ500Hzと3KHzとの間（例えば、この範囲の高い側および／または低い側の周波数の5%または10%以内））内の周波数のレベルを低減することを含んでもよい。いくつかの実施態様において、この範囲の外側の周波数は、スペクトル修正により生じるラウドネスの低下を少なくとも部分的に補償するために、より高いレベルで再生されてもよい。

40

【0163】

図4の要素は、以下のものを含む。

601：フラットEQを示す曲線。

602：示された周波数範囲の部分的減衰を示す曲線。このような部分的減衰は、比較的感知されにくい、それにもかかわらず、音声検出に有用なインパクトを与え得る。

603：示された周波数範囲の顕著に大きい減衰を示す曲線。曲線603が示すようなスペクトル修正は、発話の聞き取りに大きいインパクトを与え得る。いくつかの例にお

50

いて、曲線 6 0 3 が示すような積極的なスペクトル修正は、全ての周波数のレベルを大幅に低減する選択肢を提供し得る。

【0 1 6 4】

いくつかの例において、オーディオセッションマネージャは、曲線 6 0 1、6 0 2 および 6 0 3 が示すシーケンスなどの、時変 (time-varying) スペクトル修正に一致するオーディオ処理変更を生じ得る。

【0 1 6 5】

いくつかの例によると、1 つ以上のスペクトル修正が、オフィス、寝室、眠っている乳児などの位置から離れる方向に、再生されたオーディオ音声を「ワーピングする」ことを生じるレンダリング変更の文脈など、他のオーディオ処理変更の文脈において使用されてもよい。そのようなワーピングに関連付けて使用されるスペクトル修正 (単数または複数) は、例えば、バス周波数範囲 (例えば、2 0 ~ 2 5 0 H z) 内のレベルを低減してもよい。

【0 1 6 6】

図 5 は、オーディオ環境の 1 つ以上のマイクロフォンにおける発話対エコー比を増大し得る別の種類のオーディオ処理を示すグラフである。この例において、縦軸は 0 ~ 1 の範囲の「f」値を示し、横軸は時間 (秒) を示す。図 5 は、レンダリング効果の起動の時間に対する、(曲線 7 0 1 で示す) 軌線の図である。いくつかの例において、モジュール 3 5 6、3 5 7 または 3 5 8 のうちの 1 つ以上は、図 5 に示す種類のオーディオ処理を実装してもよい。この例によると、(曲線 7 0 1 が示す) 時定数の非対称性は、システムが、短時間 (例えば 1 0 0 m s ~ 1 秒) で制御された値 (f _ n) に調整するが、かなり時間をかけて (例えば 1 0 秒以上) 値 f _ n (値 7 0 3) からゼロへと緩和することを示す。いくつかの例において、2 秒と N 秒との間の間隔は、複数秒 (例えば 4 ~ 1 0 秒の範囲内) であってもよい。

【0 1 6 7】

また、図 5 には、この例では最大値が f _ n に等しい階段状の第 2 の起動曲線 7 0 2 を示す。この実施態様によると、上昇する段差は、コンテンツ自体のレベルの急激な変化 (例えば、有声開始 (voice onset) または音節レート (syllable rate)) に一致する。

【0 1 6 8】

上述のとおり、いくつかの実施態様において、一時的なタイムスライシングまたは周波数調節により、注目しているオーディオ音声の細断片 (glimpses) を得るのに十分なスパス時間周波数出力の「ギャップ」または期間を、(例えばオーディオコンテンツにギャップを挿入することにより) 生成し得る (例えば、オーディオコンテンツおよびその認識の、「ギャップがある状態 (gappiness)」の範囲を拡大または減少する)。

【0 1 6 9】

図 6 は、オーディオ環境の 1 つ以上のマイクロフォンにおける発話対エコー比を増大し得る別の種類のオーディオ処理を示す。図 6 は、1 つの例に従って強制ギャップ (forced gaps) が挿入された、修正されたオーディオ再生信号のスペクトログラムの一例である。より具体的には、図 6 のスペクトログラムを生成するために、再生信号の周波数帯域に強制ギャップ G 1、G 2 および G 3 を挿入して、修正されたオーディオ再生信号を生成した。図 6 に示すスペクトログラムにおいて、横軸に沿った位置は時間を示し、縦軸に沿った位置は、任意の時点における修正されたオーディオ再生信号のコンテンツの周波数を示す。

【0 1 7 0】

各小領域 (各小領域は、縦および横の座標を有する点を中心とする) におけるドットの密度は、対応する周波数および時点における、修正されたオーディオ再生信号のコンテンツのエネルギーを示す (より密度の高い領域は、より大きいエネルギーを有するコンテンツを示し、より密度の低い領域は、より小さいエネルギーを有するコンテンツを示す)。したがって、ギャップ G 1 は、ギャップ G 2 または G 3 が存在する時間 (または期間) よりも早い時間 (つまり期間内) に存在し、かつ、ギャップ G 1 は、ギャップ G 2 または G 3 が挿入された周波数帯域よりも高い周波数帯域に挿入されている。

10

20

30

40

50

【0171】

再生信号に強制ギャップを導入することは、（例えば、ユーザおよびユーザの環境をより良く聞かため）デバイスがコンテンツのプレイバックストリームを一時停止する単信方式のデバイス動作とは区別される。いくつかの開示された実施形態に応じて再生信号に強制ギャップを導入することは、再生中に導入されたギャップの結果として生じるアーティファクトが知覚される可能性を大幅に低減（または除去）するように、好適には、強制ギャップがユーザに与えるインパクトがゼロまたは最小となるが、再生環境におけるマイクロフォンの出力信号が強制ギャップを示す（例えば、ギャップを利用してパーベシブナリスニング方法を実装し得る）ように、最適化され得る。いくつかの開示された実施形態に応じて導入された強制ギャップを用いることにより、パーベシブナリスニングシステムは、音響エコーキャンセラを用いなくても、非再生音声（例えば、再生環境におけるバックグラウンドアクティビティおよび／またはバックグラウンドノイズを示す音声）をモニターし得る。

10

【0172】

いくつかの例によると、単一のチャンネルからの時間スペクトル出力に複数のギャップを挿入してもよい。これらギャップは、「ギャップを通して聞く」という、疎らな感覚の進歩したリスニング能力を生じ得る。

【0173】

図7は、オーディオ環境の1つ以上のマイクロフォンにおける発話対エコー比を増大し得る別の種類のオーディオ処理を示すグラフである。この実施形態において、オーディオ処理変更は、ダイナミックレンジ圧縮を含む。

20

【0174】

この例は、ダイナミックレンジを限定する2つの極値の間の遷移を含む。曲線801が示すあるケースでは、オーディオセッションマネージャは、ダイナミックレンジ制御を適用させない。一方、曲線802が示す他のケースでは、オーディオセッションマネージャは、比較的積極的なリミッターを適用させる。曲線802に対応するリミッターは、オーディオ出力のピークを、10 dB以上低減し得る。いくつかの例によると、圧縮比は、たった3:1である。いくつかの実施形態において、曲線802（または別のダイナミックレンジ圧縮曲線）は、デバイスのピーク出力から-20 dB（または約-20 dB）の位置（例えば、+/-1 dB以内、+/-2 dB以内、+/-3 dB以内など）に屈曲部（knee）を含み得る。

30

【0175】

次に、レンダリング（RENDERING）（例えば、図3Bまたは図3Cのシステムにおける、1つ以上のマイクロフォンにおける発話対エコー比を増大する効果を有するオーディオ処理変更）を実装するシステム要素の実施形態の別の例を説明する。本実施形態において、エネルギーバランシングが実行される。上述のとおり、ある単純な例において、オーディオセッションマネージャは、オーディオ環境の1つ以上のマイクロフォンにおいてSERを増大するための他のオーディオ処理変更の結果失われた、聴者の位置またはゾーンにおけるオーディオ音声のバンドドエネルギー（banded energy）を評価してもよい。その後、オーディオセッションマネージャは、この聴者の位置またはゾーンにおいて失われたエネルギーを埋め合わせるブーストを、他のスピーカに追加してもよい。

40

【0176】

少し関係があるコンテンツをレンダリングしており、かつ、関連しているかまたは類似のスペクトルを有する成分が複数のデバイスに存在する場合（単純な例はモノラル再生）によくあることだが、あまりに何もする必要はないかもしれない。例えば、1~2の割合の範囲の距離（1が最も近い）を空けて3つのラウドスピーカが存在する場合、（それらのラウドスピーカにより同一のコンテンツが再生されている場合、）最も近くのラウドスピーカの音量を6 dB下げても、2~3 dBのインパクトにしかない。また、最も近くのラウドスピーカをオフにしても、聴者の位置における音声にとって、おそらく全体で3~4 dBのインパクトにしかない。

50

【0177】

次に追加の実施形態の側面を説明する。

【0178】

1. 「最も近い (NEAREST)」の定義における2次因子

以下の2つの例が示すように、「近接度」または「最も近い」の尺度は、距離の単純な尺度でなくてもよく、推定される発話対エコー比を含むスカラーランキング (scalar ranking) であってもよい。オーディオ環境の複数のオーディオデバイスが同一でない場合、各ラウドスピーカ内蔵オーディオデバイスは、そのラウドスピーカ (単数または複数) から自分自身のマイクロフォン (単数または複数) への接続が異なってもよく、発話対エコー比におけるエコーレベルに大きな影響を与える。また、これらのオーディオデバイスは、マイクロフォンの配置が異なっており、リスニングのために (例えば、特定の方向からの音声を検出するため、または、オーディオ環境の特定の位置における音声またはその位置からの音声を検出するために)、相対的により適していたり、より適していなかったりする。したがって、いくつかの実施態様において、計算 (決定 (DECISION)) は、近接度およびヒアリングの相互性 (reciprocity of hearing) よりも、因子として考慮され得る。

10

【0179】

図8は、音量を下げようとするオーディオデバイスが、話している人に最も近いオーディオデバイスではないかもしれない例の図である。この例において、オーディオデバイス802は、オーディオデバイス805よりも、話している人100に近い。いくつかの例によると、図8に示すような状況において、オーディオセッションマネージャは、異なるベースラインSERおよびオーディオデバイス特性を考慮し、オーディオプレゼンテーション (audio presentation) への出力の低減のインパクトに対する、人101の発話をより良くキャプチャできるように出力を下げることのベネフィットの、最良のコスト／ベネフィット比でデバイス (単数または複数) の音量を下げてよい。

20

【0180】

図8は、「最も近い (nearest)」のより機能的な尺度に複雑さと有用性が存在し得る例を示す。この例では、人101が、音 (発話102) をたてており、オーディオセッションマネージャはこの音をキャプチャするように構成されている。また、2つのオーディオデバイス802および805が設けられており、これら両方が、ラウドスピーカ (806および804) ならびにマイクロフォン (803および807) を有している。マイクロフォン803が、人101により近いオーディオデバイス802のラウドスピーカ804の非常に近くにあることを考慮すると、このデバイスのラウドスピーカの音量を下げて、適したSERを生じ得ないかもしれない。この例において、オーディオデバイス805のマイクロフォン807は、ビーム形成を行う (概して、より好ましいSERを生じる) ように構成されており、したがって、オーディオデバイス805のラウドスピーカの音量を下げることは、オーディオデバイス802のラウドスピーカの音量を下げるよりも、インパクトが小さいかもしれない。いくつかのそのような例において、最適な決定 (DECISION) は、ラウドスピーカ806の音量を下げることであり得る。

30

【0181】

図9を参照して別の例を説明する。ここでは、2つのデバイス (一方が1対のヘッドホンであり、他方がスマートスピーカ) において生じ得るベースラインSERの最も大きい差について考える。

40

【0182】

図9は、非常に高いSERを有するデバイスがユーザの極めて近くにある状況を示す。図9において、ユーザ101は、ヘッドホン902を装着し、音声102を話して (発して) おり、この音声102は、ヘッドホン902のマイクロフォン903およびスマートスピーカデバイス904のマイクロフォンの両方によりキャプチャされる。この場合、スマートスピーカデバイス904はまた、ヘッドホンに合う任意の音声を生成してもよい (例えば、イマージブサウンド (没入感の高いサウンド) のための近位／遠位レンダリング

50

）。確かに、ヘッドホン 902 は、ユーザ 101 に対して最も近い出力デバイスであるが、ヘッドホンから最も近いマイクロフォン 903 へのエコーの経路はほぼ存在せず、したがって、このデバイスの S E R は非常に高く、ヘッドホンが聴者に対して略全てのレンダリング効果を与えるので、デバイスの音量を下げた場合に非常に大きなインパクトを与える。この場合、スマートスピーカ 904 の音量を下げることは、部分的なものに過ぎずかつレンダリング全体についての変更に逆らうものではあり（近くに居る他の聴者がその音声を聞いている）、実際のアクションは決定されていないかもしれないが、より有益であり得る。というのも、スピーカの音量を下げることはまたその他にオーディオ処理のパラメータを変更することが、オーディオ環境内に提供されるオーディオ音声をより良いものへと変えるように、ユーザピックアップの S E R を改善し得るからである。ある意味において、ヘッドホンにおける固有のデバイス S E R のおかげで既に十分機能的である。

10

【0183】

複数のスピーカおよび分散配置されたマイクロフォンを有しかつ所与のサイズを超える大きさのデバイスに関して、いくつかの条件下において、多数のスピーカおよび多数のマイクロフォンを有する単一のオーディオデバイスを、たまたま強固に接続された別個のデバイスからなる 1 つのコンステレーションと考えることができる。この場合、音量を下げる決定を、個々のスピーカに適用してもよい。したがって、いくつかの実施態様において、オーディオセッションマネージャは、このタイプのオーディオデバイスを、独立したマイクロフォンおよびラウドスピーカからなる 1 つの集合体であると考え得る。それに対して、他の例において、オーディオセッションマネージャは、このタイプのオーディオデバイスを、複合的なスピーカとマイクロフォンのアレイを有する 1 つのデバイスと考え得る。また、単一のデバイスに設けられたスピーカを個別のデバイスとして扱うことと、複数のラウドスピーカを有する単一のオーディオデバイスにおいて、レンダリングの 1 つのアプローチが空間的ステアリングであるという考えとの間に二重性が存在する（このことは、必然的に、単一のオーディオデバイスに設けられたラウドスピーカの出力に差分変化を与える）ということが理解され得る。

20

【0184】

移動する聴者の近くのオーディオデバイスからの空間的イメージングのセンシティブさを、最も近くのオーディオデバイス（単数または複数）が避けることの 2 次的効果に関し、多くの場合、移動する聴者の近くにラウドスピーカが存在したとしても、最も近いラウドスピーカ（単数または複数）から特定のオーディオオブジェクトまたはレンダリング素材を再生することに意味がないことがあり得る。これは、直接的オーディオ音声経路のラウドネスが直接的に変化するという事実に関連する（ $1/r^2$ （ r は音声伝搬する距離））。そして、ラウドスピーカが任意の聴者に近づく（ $r \rightarrow 0$ ）、全体の混合音声に対する、このラウドスピーカにより再生されている音声のレベルの安定性が劣化する。

30

【0185】

いくつかのそのような例において、（例えば）以下のような実施形態を実装することが有利であり得る。

— コンテキスト (CONTEXT) は、誰かがテレビで視ている番組についてオーディオ音声を聞くことができることが常に有用であると想定される、一般的なリスニングエリア（例えば、テレビの近くのソファ）である。

40

— 決定 (DECISION) : （例えば、ソファの近くの）一般的なリスニングエリアにおけるコーヒーテーブル上に置かれたスピーカを有するデバイスについて、 $f_{n=1}$ と設定する。

— レンダリング (RENDERING) : デバイスをオフにし、何処か別の場所でエネルギーがレンダリングされる。

【0186】

このオーディオ処理変更のインパクトは、ソファに座っている人にとってのより良いリスニングである。コーヒーテーブルがソファの一方の端にある場合、この方法は、聴者の

50

このオーディオデバイスに対する近接度の、センシティブさを排除し得る。いくつかの例において、このオーディオデバイスは、例えばサラウンドチャンネルにとって、理想的な位置に存在する一方、ソファを横切ってこのスピーカまで20dBのレベル差があり得るという事実は、聴者／話者の正確な位置が分からないのであれば、この最も近いデバイスの音量を下げるかまたはオフにするのがよいということを意味する。

【0187】

図10は、図2Aに示すような装置により実行され得る方法の一例の概略を示す流れ図である。方法1000のブロックは、本明細書に開示する他の方法と同様、必ずしも図示した順序どおり実行する必要はない。さらに、このような方法は、図示および／または説明したブロックより多い個数または少ない個数のブロックを含んでもよい。この実施態様において、方法1000は、環境内におけるユーザの位置を推定することを含む。

10

【0188】

この例において、ブロック1005は、環境内の複数のマイクロフォンの各々から出力信号を受信することを含む。この例において、複数のマイクロフォンの各々は、環境のマイクロフォン位置に設けられている。この例によると、出力信号は、ユーザの現在の発声に対応している。いくつかの例において、現在の発声は、ウェイクワードの発声であってもよいし、ウェイクワードの発声を含んでもよい。ブロック1005は、例えば、インターフェイスシステム（図2Aのインターフェイスシステム205など）を介して環境内の複数のマイクロフォンの各々から出力信号を受信する制御システム（図2Aの制御システム120など）を含んでもよい。

20

【0189】

いくつかの例において、環境内のマイクロフォンのうちの少なくともいくつかのマイクロフォンは、1つ以上の他のマイクロフォンにより提供される出力信号に対して非同期の出力信号を提供してもよい。例えば、複数のマイクロフォンのうちの第1のマイクロフォンが、第1のサンプリングクロックに基づきオーディオデータをサンプリングし、複数のマイクロフォンのうちの第2のマイクロフォンが、第2のサンプリングクロックに基づきオーディオデータをサンプリングしてもよい。いくつかの例において、環境内のマイクロフォンの少なくとも1つが、スマートオーディオデバイスに含まれてもよいし、スマートオーディオデバイスと通信するように構成されてもよい。

【0190】

この例によると、ブロック1010は、各マイクロフォンの出力信号から、複数の現在の音響的特徴を決定することを含む。この例において、「現在の音響的特徴」は、ブロック1005の「現在の発声」から得られた音響的特徴である。いくつかの実施態様において、ブロック1010は、1つ以上の他のデバイスから、複数の現在の音響的特徴を受信することを含んでもよい。例えば、ブロック1010は、1つ以上の他のデバイスにより実装される1つ以上のウェイクワード検出器から、複数の現在の音響的特徴のうちの少なくともいくつかを受信することを含んでもよい。代替的にまたは追加的に、いくつかの実施態様において、ブロック1010は、出力信号から複数の現在の音響的特徴を決定することを含んでもよい。

30

【0191】

音響的特徴が単一のデバイスにより決定されるか複数のデバイスにより決定されるかに関わらず、音響的特徴は非同期的に決定され得る。音響的特徴が複数のデバイスにより決定される場合、音響的特徴を決定するプロセスをコーディネートさせるようにデバイスが構成されていなければ、音響的特徴は概して非同期的に決定され得る。音響的特徴が単一のデバイスにより決定される場合、いくつかの実施態様において、単一のデバイスは各マイクロフォンの出力信号を異なる時間に受信し得るので、音響的特徴は、それでもなお、非同期的に決定され得る。いくつかの例において、環境内のマイクロフォンのうち少なくともいくつかのマイクロフォンが、1つ以上の他のマイクロフォンにより提供される出力信号に対して非同期である出力信号を提供し得るので、音響的特徴は非同期的に決定され得る。

40

50

【0192】

いくつかの例において、音響的特徴は、ウェイクワード確信度指数 (wakeword confidence metric)、ウェイクワード長さ指数 (wakeword duration metric)、および／または、少なくとも1つの受信レベル指数 (received level metric) を含んでもよい。受信レベル指数は、マイクロフォンにより検出された音声の受信レベルを示し、かつ、マイクロフォンの出力信号のレベルに対応し得る。

【0193】

代替的にまたは追加的に、音響的特徴は、以下のうちの1つ以上を含み得る。

- ・ 音響モデルに対する1-best (ビタビ) 配列に沿った、各ウェイクワード状態についての平均状態エントロピー (純度)。
- ・ ウェイクワード検出器の音響モデルに対するCTC-損失 (コネクショニスト時間分類損失)。
- ・ ウェイクワード検出器は、ウェイクワード確信度に加えて、話者のマイクロフォンからの距離の推定値および／またはRT60推定値を提供するように訓練されてもよい。距離推定値および／またはRT60推定値は、音響的特徴であってもよい。
- ・ マイクロフォンにおける広帯域受信レベル／パワーの代わりに、もしくはマイクロフォンにおける広帯域受信レベル／パワーに加えて、音響的特徴は、多数のログ／メル／バーク間隔の周波数帯域における受信レベルであってもよい。周波数帯域は、特定の実施態様に応じて変動し得る (例えば、2つの周波数帯域、5つの周波数帯域、20個の周波数帯域、50個の周波数帯域、1オクターブの周波数帯域、または1／3オクターブの周波数帯域)。
- ・ 過去のある時点におけるスペクトル情報のケプストラム表現。これは、バンドパワーの対数をDCT (離散コサイン変換) することにより算出される。
- ・ 人間の発話について重み付けされた周波数帯域におけるバンドパワー。例えば、音響的特徴は、特定の周波数帯域 (例えば400 Hz～1.5 kHz) のみに基づいてもよい。この例において、より高い周波数およびより低い周波数は無視してもよい。
- ・ 帯域あたりまたはピンあたりの、ボイスアクティビティ検出器確信度。
- ・ 音響的特徴は、長期雑音推定値に少なくとも部分的に基づいて、信号対雑音比が不十分なマイクロフォンを無視してもよい。
- ・ 「ピーキネス (peakiness)」という発話の尺度としての尖度 (Kurtosis)。尖度は、長いリバーブテイル (残響音) によるスミアリング (smearing) の指標であり得る。
- ・ 推定ウェイクワード開始時間。開始および期間は、フレーム内で等しい、または、全てのマイクロフォンの間で等しいことが期待される。外れ値は、信頼できない推定値の手がかりたり得る。このことは、必ずしもサンプルに対してではなく、例えば数十ミリ秒のフレームに対して、あるレベルのシンクロニー (synchrony) を想定している。

【0194】

この例によると、ブロック1015は、複数の現在の音響的特徴に分類器を適用することを含む。いくつかのそのような例において、分類器を適用することは、環境内の複数のユーザゾーン内でユーザによってなされた複数の過去の発声から得られた、過去に決定された音響的特徴に対して訓練されたモデルを適用することを含んでもよい。本明細書中、

【0195】

いくつかの例において、ユーザゾーンは、シンクエリア、調理エリア、冷蔵庫エリア、ダイニングエリア、ソファエリア、テレビエリア、寝室エリア、および／または、出入り口エリアを含んでもよい。いくつかの例によると、これらユーザゾーンのうち1つ以上のゾーンは、所定のユーザゾーンであってもよい。いくつかのそのような例において、1つ以上の所定のユーザゾーンは、訓練プロセスの間、ユーザにより選択可能である。

【0196】

いくつかの実施態様において、分類器を適用することは、過去の発声に対して訓練した混合ガウスモデルを適用することを含んでもよい。いくつかのそのような実施態様による

と、分類器を適用することは、過去の発声の、正規化されたウェイクワード確信度、正規化された平均受信レベル、または、最大受信レベルのうちの1つ以上に対して訓練された混合ガウスモデルを適用することを含んでもよい。しかし、別の実施態様において、分類器を適用することは、本明細書に開示する他のモデルのうちの1つなど、異なるモデルに基づいてもよい。いくつかの例において、このモデルは、ユーザゾーンのラベルが付いた訓練データを用いて訓練されてもよい。しかし、いくつかの例において、分類器を適用することは、ユーザゾーンのラベルが付いていない、無ラベル訓練データを用いて訓練されたモデルを適用することを含む。

【0197】

いくつかの例において、過去の発声は、ウェイクワードの発声であってもよく、ウェイクワードの発声を含んでもよい。いくつかのそのような例によると、過去の発声および現在の発声は、同じウェイクワードの発声であってもよい。

【0198】

この例において、ブロック1020は、分類器からの出力に少なくとも部分的に基づいて、ユーザが現在位置しているユーザゾーンの推定値を決定することを含む。いくつかのそのような例において、この推定値は、複数のマイクロフォンの幾何学的位置を参照せずに決定されてもよい。例えば、この推定値は、個々のマイクロフォンの座標を参照せずに決定されてもよい。いくつかの例において、この推定値は、ユーザの幾何学的位置を推定することなく決定されてもよい。

【0199】

方法1000のいくつかの実施態様は、推定されるユーザゾーンに応じて、少なくとも1つのスピーカを選択することを含んでもよい。いくつかのそのような実施態様は、少なくとも1つの選択されたスピーカを制御して、推定されるユーザゾーンに音を提供することを含んでもよい。代替的にまたは追加的に、方法1000のいくつかの実施態様は、推定されるユーザゾーンに応じて、少なくとも1つのマイクロフォンを選択することを含んでもよい。いくつかのそのような実施態様は、少なくとも1つの選択されたマイクロフォンにより出力された信号を、スマートオーディオデバイスに提供することを含んでもよい。

【0200】

図11は、ゾーン分類器を実装するように構成された実施形態の一例の要素のブロック図である。この例によると、システム1100は、環境（例えば図1Aまたは図1Bに示すような環境）の少なくとも一部において分散配置された複数のラウドスピーカ1104を含む。この例において、システム1100は、マルチチャネルラウドスピーカレンダラー1101を含む。この実施態様によると、マルチチャネルラウドスピーカレンダラー1101の出力は、ラウドスピーカ駆動信号（スピーカ1104を駆動するスピーカフィード信号）およびエコー参照信号の両方として働く。この実施態様において、エコー参照信号は、複数のラウドスピーカ参照チャンネル1102を介して、エコー管理サブシステム1103に提供される。ここで、エコー参照信号は、レンダラー1101から出力されたスピーカフィード信号のうちの少なくともいくつかを含んでいる。

【0201】

この実施態様において、システム1100は、複数のエコー管理サブシステム1103を含む。この例によると、エコー管理サブシステム1103は、1つ以上のエコー抑制プロセスおよび／または1つ以上のエコー除去プロセスを実装するように構成されている。この例において、エコー管理サブシステム1103の各々は、ウェイクワード検出器1106のうちの1つに、対応するエコー管理出力1103Aを提供する。エコー管理出力1103Aは、エコー管理サブシステム1103のうちの関連する1つのエコー管理サブシステムへの入力と比較して、減衰されたエコーを有する。

【0202】

この実施態様によると、システム1100は、環境（例えば図1Aまたは図1Bに示す環境）の少なくとも一部において分散配置されたN個のマイクロフォン1105（Nは整数）を含む。これらのマイクロフォンは、アレイマイクロフォンおよび／またはスポット

10

20

30

40

50

マイクロフォンを含み得る。例えば、環境内に配置された1つ以上のスマートオーディオデバイスは、マイクロフォンのアレイを含んでもよい。この例において、マイクロフォン1105の出力は、エコー管理サブシステム1103への入力として提供される。この実施態様によると、エコー管理サブシステム1103の各々は、個々のマイクロフォン1105またはマイクロフォン1105の個々の群または部分集合の出力をキャプチャする。
【0203】

この例において、システム1100は、複数のウェイクワード検出器1106を含む。この例によると、ウェイクワード検出器1106の各々は、エコー管理サブシステム1103の1つからのオーディオ出力を受信し、複数の音響的特徴1106Aを出力する。各エコー管理サブシステム1103から出力される音響的特徴1106Aは、ウェイクワード確信度、ウェイクワードの長さ、および受信レベルの尺度を含み得る（但し、これらに限定されない）。3つの音響的特徴1106Aを示す3つの矢印を、各エコー管理サブシステム1103から出力されているように図示しているが、別の実施態様において、より多くの数またはより少ない数の音響的特徴1106Aが出力されてもよい。さらに、これら3つの矢印は略垂直の線に沿って分類器1107に当たっているが、このことは、分類器1107が、全てのウェイクワード検出器1106から同時に音響的特徴1106Aを必ず受信することを示すものではない。本明細書の別の箇所で示すように、音響的特徴1106Aは、いくつかの例において、非同期的に決定され得る、かつ／または、非同期的に分類器に提供され得る。

【0204】

この実施態様によると、システム1100は、ゾーン分類器1107（分類器1107と呼ぶこともある）を含む。この例において、分類器は、環境内の複数のマイクロフォン1105（例えば全てのマイクロフォン1105）について、複数の特徴1106Aを、複数のウェイクワード検出器1106から受信する。この例によると、ゾーン分類器1107の出力1108は、ユーザが現在位置しているユーザゾーンの推定値に対応している。いくつかのそのような例によると、出力1108は、1つ以上の事後確率に対応してもよい。ユーザが現在位置しているユーザゾーンの推定値は、ベイズ統計学に基づく最大事後確率であってもよいし、その最大事後確率に対応してもよい。

【0205】

次に、分類器の実施態様例を説明する。この分類器は、いくつかの例において、図11のゾーン分類器1107に一致してもよい。 $x_i(n)$ を、離散時間 n における i 番目（ $i = \{1 \dots N\}$ ）のマイクロフォン信号とする（つまり、マイクロフォン信号 $x_i(n)$ は、 N 個のマイクロフォン1105の出力である）。エコー管理サブシステム1103において N 個の信号 $x_i(n)$ の処理を行うことにより、それぞれ離散時間 n において、「クリーン（clean）な」マイクロフォン信号 $e_i(n)$ が生成される（ $i = \{1 \dots N\}$ ）。この例において、図11において1103Aで示すクリーンな信号 $e_i(n)$ は、ウェイクワード検出器1106に供給される。ここで、各ウェイクワード検出器1106は、図11において1106Aで示す特徴のベクトル $w_i(j)$ を生成する（ $j = \{1 \dots J\}$ は、 j 番目のウェイクワード発声に対応するインデックスである）。この例において、分類器1107は、特徴の総集合

$$W(j) = [w_1^T(j) \dots w_N^T(j)]^T$$

を入力とする。

【0206】

いくつかの実施態様によると、1組のゾーンラベル C_k （ $k = \{1 \dots K\}$ ）は、環境内における多数（ K 個）の異なるユーザゾーンに対応していてもよい。例えば、ユーザゾーンは、ソファゾーン、キッチンゾーン、リーディングチェアゾーンなどを含んでもよい。いくつかの例は、キッチンまたは他の部屋の中に、複数のゾーンを定義してもよい。例えば、キッチンエリアは、シンクゾーン、調理ゾーン、冷蔵庫ゾーン、およびダイニングゾ

ーンを含んでもよい。同様に、リビングルームエリアは、ソファゾーン、テレビゾーン、リーディングチェアゾーン、1つ以上の出入り口ゾーンなどを含んでもよい。これらのゾーンのゾーンラベルは、例えば訓練期間のあいだ、ユーザにより選択可能であってもよい。
【0207】

いくつかの実施態様において、分類器1107は、例えばベイズ分類器を用いることにより、特徴の組 $W(j)$ の事後確率

$$p(C_k|W(j))$$

を推定する。確率

$$p(C_k|W(j))$$

は、(j番目の発声およびk番目のゾーンについて、ゾーン C_k の各々および発声の各々について)ユーザがゾーン C_k の各々に存在する確率を示し、分類器1107の出力1108の一例である。

【0208】

いくつかの例によると、訓練データは、ゾーン(例えばソファゾーン)を選択または定義することをユーザに促すことにより、(例えば各ユーザゾーンについて)集められてもよい。訓練プロセスは、選択または定義されたゾーンの近傍において、訓練発声(ウェイクワードの発声など)を行うことをユーザに促すことを含んでもよい。ソファゾーンの例において、訓練プロセスは、ソファの中央および両端において、訓練発声を行うことをユーザに促すことを含んでもよい。訓練プロセスは、ユーザゾーン内の各位置において、訓練発声を複数回繰り返すことをユーザに促すことを含んでもよい。その後、別のユーザゾーンに移動して、全ての指定されたユーザゾーンがカバーされるまで訓練発声を続けるように、ユーザに促してもよい。

【0209】

図12は、図2Aの装置200などの装置により実行され得る方法の一例の概略を示す流れ図である。方法1200のブロックは、本明細書に開示する他の方法と同様、必ずしも図示した順序どおり実行する必要はない。さらに、このような方法は、図示および／または説明したブロックより多い個数または少ない個数のブロックを含んでもよい。この実施態様において、方法1200は、環境内におけるユーザの位置を推定するための分類器を訓練することを含む。

【0210】

この例において、ブロック1205は、環境の第1のユーザゾーン内の複数の位置の各々において、少なくとも1回の訓練発声を行うことをユーザに促すことを含む。訓練発声は、いくつかの例において、ウェイクワード発声の1つ以上の事例であってもよい。いくつかの実施態様によると、第1のユーザゾーンは、ユーザにより選択および／または定義された任意のユーザゾーンであってもよい。いくつかの例において、制御システムは、対応するゾーンラベル(例えば、前に説明したゾーンラベル C_k のうちの1つの対応する事例)を生成してもよく、かつ、ゾーンラベルを、第1のユーザゾーンについて得られた訓練データに関連付けてもよい。

【0211】

自動化された促進システムを用いて、これらの訓練データを収集してもよい。上述のとおり、装置200のインターフェイスシステム205は、1つ以上のマイクロフォン、1つ以上のスピーカ、表示システム、タッチセンサシステムおよび／またはジェスチャセンサシステムなど、ユーザインターフェイスを実装する1つ以上のデバイスを含んでもよい。例えば、装置200は、訓練プロセスの間、ユーザに対して、以下の促進メッセージを、表示システムのスクリーン上に表示してもよいし、1つ以上のスピーカを介して通知してもよい。

- ・ 「ソファに移動」
- ・ 「顔を左右に振りつつウェイクワードを10回言う」
- ・ 「ソファとリーディングチェアとの間の中間位置に移動し、ウェイクワードを10回言う」
- ・ 「料理するときのようにキッチンに立ち、ウェイクワードを10回言う」

【0212】

この例において、ブロック1210は、環境内の複数のマイクロフォンの各々から、第1の出力信号を受信することを含む。いくつかの例において、ブロック1210は、環境内のアクティブなマイクロフォンの全てから、第1の出力信号を受信することを含んでもよい。それに対して、他の例では、ブロック1210は、環境内のアクティブなマイクロフォンの全てを含む部分集合から、第1の出力信号を受信することを含んでもよい。いくつかの例において、環境内のマイクロフォンのうちの少なくともいくつかのマイクロフォンは、1つ以上の他のマイクロフォンにより提供される出力信号に対して非同期の出力信号を提供してもよい。例えば、複数のマイクロフォンのうちの第1のマイクロフォンが、第1のサンプリングクロックに基づきオーディオデータをサンプリングし、複数のマイクロフォンのうちの第2のマイクロフォンが、第2のサンプリングクロックに基づきオーディオデータをサンプリングしてもよい。

【0213】

この例において、複数のマイクロフォンの各々は、環境のマイクロフォン位置に設けられている。この例において、第1の出力信号は、第1のユーザゾーンから受信した、検出された訓練発声の事例に対応している。ブロック1205は、環境の第1のユーザゾーン内の複数の位置の各々において、少なくとも1回の訓練発声を行うことをユーザに促すことを含むので、この例では、「第1の出力信号」という用語は、第1のユーザゾーンについての訓練発声に対応する全ての出力信号からなる集合を指す。他の例において、「第1の出力信号」という用語は、第1のユーザゾーンについての訓練発声に対応する全ての出力信号からなる部分集合を指してもよい。

【0214】

この例によると、ブロック1215は、第1の出力信号の各々から、1つ以上の第1の音響的特徴を決定することを含む。いくつかの例において、第1の音響的特徴は、ウェイクワード確信度指数および／または受信レベル指数を含んでもよい。例えば、第1の音響的特徴は、正規化されたウェイクワード確信度指数、正規化された平均受信レベルの指標、および／または、最大受信レベルの指標を含んでもよい。

【0215】

上述のとおり、ブロック1205は、環境の第1のユーザゾーン内の複数の位置の各々において、少なくとも1回の訓練発声を行うことをユーザに促すことを含むので、この例では、「第1の出力信号」という用語は、第1のユーザゾーンについての訓練発声に対応する全ての出力信号からなる集合を指す。したがって、この例では、「第1の音響的特徴」という用語は、第1のユーザゾーンについての訓練発声に対応する全ての出力信号からなる集合から得られた音響的特徴からなる集合を指す。したがって、この例において、第1の音響的特徴の集合は、第1の出力信号の集合と少なくとも同程度の大きさである。例えば、出力信号の各々から2つの音響的特徴が決定される場合、第1の音響的特徴の集合は、第1の出力信号の集合の2倍の大きさになる。

【0216】

この例において、ブロック1220は、分類器モデルを訓練して、第1のユーザゾーンと第1の音響的特徴との間に相関関係を形成することを含む。分類器モデルは、例えば、本明細書に開示した分類器モデルのうちのいずれかであってもよい。この実施態様によると、分類器モデルは、複数のマイクロフォンの幾何学的位置を参照せずに訓練される。換言すると、この例において、訓練プロセスの間、複数のマイクロフォンの幾何学的位置に関するデータ（例えばマイクロフォン座標データ）は、分類器モデルに提供されない。

【0217】

図13は、図2Aの装置200などの装置により実行され得る方法の別の例の概略を示す流れ図である。方法1300のブロックは、本明細書に開示する他の方法と同様、必ずしも図示した順序どおり実行する必要はない。例えば、いくつかの実施態様において、ブロック1325の音響的特徴決定プロセスの少なくとも一部を、ブロック1315またはブロック1320よりも先に実行してもよい。さらに、このような方法は、図示および／または説明したブロックより多い個数または少ない個数のブロックを含んでもよい。この実施態様において、方法1300は、環境内におけるユーザの位置を推定するための分類器を訓練することを含む。方法1300は、方法1200を環境の複数のユーザゾーンに展開する例を提供する。

【0218】

この例において、ブロック1305は、環境のユーザゾーン内のある位置において、少なくとも1回の訓練発声を行うことをユーザに促すことを含む。いくつかの例において、ブロック1305は、ブロック1305がユーザゾーン内の単一の位置に関係していることを除き、図12のブロック1205を参照して上で説明した状態で実行されてもよい。訓練発声は、いくつかの例において、ウェイクワード発声の1つ以上の事例であってもよい。いくつかの実施態様によると、ユーザゾーンは、ユーザにより選択および／または定義された任意のユーザゾーンであってもよい。いくつかの例において、制御システムは、対応するゾーンラベル（例えば、前に説明したゾーンラベル C_k のうちの1つの対応する事例）を生成してもよく、かつ、ゾーンラベルを、ユーザゾーンについて得られた訓練データに関連付けてもよい。

【0219】

この例によると、ブロック1310は、図12のブロック1210を参照して上で説明したのと実質的に同じように実行される。しかし、この例において、ブロック1310のプロセスは、必ずしも訓練データが獲得された第1のユーザゾーンではなく、任意のユーザゾーンに対して一般化されている。したがって、ブロック1310から受信した出力信号は、「環境内の複数のマイクロフォンの各々からの出力信号であって、複数のマイクロフォンの各々は、環境のマイクロフォン位置に設けられており、出力信号は、ユーザゾーンから受信した検出された訓練発声の事例に対応している。」この例において、「出力信号」という用語は、ユーザゾーンのある位置における1回以上の訓練発声に対応する全ての出力信号からなる集合を指す。他の例において、「出力信号」という用語は、ユーザゾーンのある位置における1回以上の訓練発声に対応する全ての出力信号からなる部分集合を指す。

【0220】

この例によると、ブロック1315は、現在のユーザゾーンについて十分な訓練データが獲得できたか否かを判定することを含む。いくつかのそのような例において、ブロック1315は、閾値回数の訓練発声に対応する出力信号が、現在のユーザゾーンについて得られたか否かを判定することを含んでもよい。代替的にまたは追加的に、ブロック1315は、現在のユーザゾーン内の閾値箇所数の位置における訓練発声に対応する出力信号が得られたか否かを判定することを含んでもよい。得られていないと判定された場合、この例では、方法1300はブロック1305に戻り、同じユーザゾーン内のある位置において、少なくとも1回追加の発声を行うことをユーザに促す。

【0221】

しかし、ブロック1315において、現在のユーザゾーンについて十分な訓練データが得られたと判定された場合、この例では、プロセスはブロック1320へと続く。この例によると、ブロック1320は、追加のユーザゾーンについて訓練データを得るか否かを判定する。いくつかの例によると、ブロック1320は、ユーザが過去に特定した各ユーザゾーンについて、訓練データが得られたか否かを判定することを含んでもよい。他の例において、ブロック1320は、最小個数のユーザゾーンについて、訓練データが得られたか否かを判定することを含んでもよい。最小個数は、ユーザにより選択されていてもよい。他の例において、最小個数は、環境毎に推奨された最小個数や、環境内の部屋毎に推

10

20

30

40

50

奨された最小個数などであってもよい。

【0222】

ブロック1320において、追加のユーザゾーンについて訓練データを獲得すべきだと判定された場合、この例では、プロセスはブロック1322へと続く。ブロック1322は、環境の別のユーザゾーンに移動することをユーザに促すことを含む。いくつかの例において、次のユーザゾーンは、ユーザにより選択可能であってもよい。この例によると、プロセスは、ブロック1322の促進ステップの後に、ブロック1305へと続く。いくつかのそのような例において、ブロック1322の促進ステップの後に、ユーザが新たなユーザゾーンに到達したことを確認するようユーザに促してもよい。いくつかのそのような例によると、ブロック1305の促進ステップの前に、ユーザが新たなユーザゾーンに到達したことを確認するようユーザに要求してもよい。

10

【0223】

ブロック1320において、追加のユーザゾーンについて訓練データを獲得すべきではないと判定された場合、この例では、プロセスはブロック1325へと続く。この例において、方法1300は、K個のユーザゾーンについて訓練データを得ることを含む。この実施態様において、ブロック1325は、訓練データが得られた1番目～K番目のユーザゾーンの各々に対応する1番目～H番目の出力信号から、1番目～G番目の音響的特徴を決定することを含む。この例において、「第1の出力信号」という用語は、第1のユーザゾーンについての訓練発声に対応する全ての出力信号からなる集合を指す。また、「H番目の出力信号」という用語は、K番目のユーザゾーンについての訓練発声に対応する全ての出力信号からなる集合を指す。同様に、「第1の出力信号」という用語は、第1の出力信号から決定された音響的特徴の集合を指し、「G番目の音響的特徴」という用語は、H番目の出力信号から決定された音響的特徴の集合を指す。

20

【0224】

これらの例によると、ブロック1330は、分類器モデルを訓練して、1番目～K番目のユーザゾーンと1番目～K番目の音響的特徴との間にそれぞれ相関関係を形成することを含む。分類器モデルは、例えば、本明細書に開示した分類器モデルのうちのいずれかであってもよい。

【0225】

上記の例において、ユーザゾーンは、（例えば、前に説明したゾーンラベル C_k のうちの1つの対応する事例にしたがって）ラベリングされる。しかし、モデルは、ラベリングされたユーザゾーンに応じて訓練されてもよいし、ラベリングされていないユーザゾーンに応じて訓練されてもよく、このことは、特定の実施態様に依拠して決まる。ラベリングされている場合、各訓練発声は、例えば次式のように、ユーザゾーンに対応するラベルと対にされてもよい。

30

【数11】

$$p(C_k | \mathbf{W}(j)) = \begin{cases} 1 & k = k' \\ 0 & \text{otherwise} \end{cases},$$

40

【0226】

分類器モデルを訓練することは、ラベリングされた訓練データに最も適合するものを決定することを含んでもよい。一般性を失わずに、分類器モデルにとって適切な分類アプローチは、以下のものを含み得る。

- ・ ベイズ分類器であって、例えば、クラスあたりの（per-class）分布が、多変量正規分布、完全共分散（full-covariance）混合ガウスモデル、または対角共分散（diagonal-covariance）混合ガウスモデルにより記述（説明）される、ベイズ分類器、
- ・ ベクトル量子化、
- ・ 最近傍（k-平均）、

50

- ・ 1つの出力が各クラスに対応している、SoftMax出力レイヤを有するニューラルネットワーク、
- ・ サポートベクターマシン（SVM）、ならびに／もしくは、
- ・ 勾配ブースティングマシン（GBM）などの、ブースティング技術。

【0227】

ラベリングされない場合を実装する一例において、データは自動的にK個のクラス（Kは未知であり得る）に分割されてもよい。ラベリングされていない自動分割は、例えば、古典的なクラスタリング技術（例えば、k-平均アルゴリズムまたは混合ガウスモデリング）を用いることにより実行され得る。

【0228】

ロバスト性を向上させるために、分類器モデルの訓練に正則化が適用されてもよく、かつ、新たな発声が行われると、時間の経過に伴いモデルパラメータが更新されてもよい。

【0229】

次に、実施形態のさらなる側面を説明する。

【0230】

音響的特徴の集合の一例（例えば、図11の音響的特徴1106A）は、ウェイクワード確信度の尤度、最も確信度の高いウェイクワードの推定される長さに対する平均受信レベル、および、最も確信度の高いウェイクワードの推定される長さに対する最大受信レベルを含んでもよい。特徴は、各ウェイクワード発声について、その最大値に対して正規化されてもよい。訓練データはラベリングされてもよく、完全共分散混合ガウスモデル（GMM）を訓練して、訓練ラベルの期待値を最大化してもよい。推定されるゾーンは、事後確率を最大化するクラスであってもよい。

【0231】

いくつかの実施形態の上記説明は、促進された収集プロセスの間に収集された訓練データの集合から、音響ゾーンモデルを学習することについて議論した。そのモデルにおいて、訓練時間（または設定モード）ならびにランタイム（またはレギュラーモード）は、マイクロフォンシステムが配置され得る2つの異なるモードと考え得る。このスキームへの展開は、オンライン学習であり、ここでは、音響ゾーンモデルの一部または全てが、オンラインで（例えば、ランタイムに、またはレギュラーモードで）学習または適合される。換言すると、「ランタイム」プロセスにおいて分類器を適用して、（例えば、図10の方法1000に従って）ユーザが現在位置しているユーザゾーンの推定値を生成した後であっても、いくつかの実施態様において、分類器を訓練するプロセスは継続してもよい。

【0232】

図14は、図2Aの装置200などの装置により実行され得る方法の別の例の概略を示す流れ図である。方法1400のブロックは、本明細書に開示する他の方法と同様、必ずしも図示した順序どおり実行する必要はない。さらに、このような方法は、図示および／または説明したブロックより多い個数または少ない個数のブロックを含んでもよい。この実施態様において、方法1400は、環境内におけるユーザの位置を推定する「ランタイム」プロセスの間、分類器の継続的訓練を含む。方法1400は、本明細書において「オンライン学習モード」と呼ぶものの一例である。

【0233】

この例において、方法1400のブロック1405は、方法1000のブロック1005～1020に対応する。ここで、ブロック1405は、分類器からの出力に少なくとも部分的に基づき、ユーザが現在位置しているユーザゾーンの推定値を提供することを含む。この実施態様によると、ブロック1410は、ブロック1405の推定値に関する暗黙的または明示的なフィードバックを得ることを含む。ブロック1415において、分類器は、ブロック1405において受信されたフィードバックに従って更新される。ブロック1415は、例えば、1つ以上の強化学習方法を含んでもよい。ブロック1415からブロック1405へと延びる点線の矢印が示唆するとおり、いくつかの実施態様において、方法1400は、ブロック1405に戻ることを含んでもよい。例えば、方法1400は

10

20

30

40

50

、更新されたモデルを適用することに基づき、将来のある時点においてユーザが位置するユーザゾーンの将来の推定値を提供することを含んでもよい。

【0234】

フィードバックを得るための明示的な技術は、以下のことを含んでもよい。

- ・ 音声ユーザインターフェイス（UI）を用いて、予測が正しいかどうか否かをユーザに尋ねる。例えば、以下を示す音声をユーザに提供してもよい。：『あなたはソファに座っていると思います。「正しい」または「間違い」で教えてください。』

- ・ いつでもボイスUIを用いて間違っただ予測を訂正し得ることをユーザに知らせる。（例えば、以下を示す音声をユーザに提供してもよい。：『私に話しかければ、あなたが何処にいるか予測できます。もし私の予測が間違いであれば、「アマンダ、私はソファには座っていない。リーディングチェアに座っている。」のように教えてください。』）

- ・ いつでもボイスUIを用いて正しい予測に対して報い得ることをユーザに知らせる。（例えば、以下を示す音声をユーザに提供してもよい。：『私に話しかければ、あなたが何処にいるか予測できます。もし私の予測が正しければ、「アマンダ、正解だ。私はソファに座っている。」のように教えてください。そうすれば私の予測はさらに改善されます。』）

- ・ フィードバックを与えるためにユーザが操作できる、物理ボタンまたは他のUI要素を含む（例えば、物理デバイス上もしくはスマートフォンアプリ内の、サムズアップ（いいね）ボタンおよび／またはサムズダウン（低評価）ボタン）。

【0235】

ユーザが居るユーザゾーンを予測することの目的は、例えば、ウェイクワードに続くコマンドをより良く認識するために、ユーザの音響ゾーンからより効果的に音をピックアップすることを試みる、マイクロフォン選択スキームまたは適応型ビームフォーミングスキームに情報を与えることであり得る。そのようなシナリオにおいて、ゾーン予測の品質に関するフィードバックを得る暗黙的な技術は、以下の事項を含み得る。

- ・ ウェイクワードに続くコマンドの誤認識をもたらす予測に罰則を課する。誤認識を示し得るプロキシとしては、ユーザが、例えば「アマンダ、ストップ!」のような取消コマンド様のものを発することにより、コマンドに対する音声アシスタントの返答を中断することを含み得る；

- ・ 音声認識器がコマンドを正しく認識したことの低確信度をもたらす予測を罰する。多くの自動音声認識システムが、結果とともに確信度レベルを返す能力を有し、この目的のために使用することができる；

- ・ 第2パスウェイクワード検出器が高確信度でウェイクワードを遡及的に検出することの失敗をもたらす予測を罰する；および／または、

- ・ ウェイクワードの高確信度の認識および／またはユーザのコマンドの正しい認識をもたらす予測を強化する。

【0236】

以下に説明するのは、第2パスウェイクワード検出器が高い確信度でウェイクワードを遡及的に検出することを失敗する例である。環境内のマイクロフォンからの現在の発声に対応する出力信号を得た後で、かつ、（例えば、マイクロフォンと通信するために構成された複数の第1パスウェイクワード検出器を介して）出力信号に基づいて音響的特徴を決定した後に、音響的特徴が分類器に提供されると仮定する。換言すると、音響的特徴は、検出されたウェイクワード発声に対応するとみなす。さらに、現在の発声をした人はゾーン3（この例では、リーディングチェアに対応する）に居る可能性が最も高いと、分類器が決定すると仮定する。例えば、ボイスコマンド認識のためにクラウド型バーチャルアシスタントに送信されるために、ゾーン3内の人の声を聞くのに最適であるということが既知である、特定のマイクロフォンまたは学習した複数のマイクロフォンの組み合わせが存在してもよい。

【0237】

さらに、音声認識のためにどのマイクロフォン（単数または複数）を使用するかを決定

10

20

30

40

50

した後で、かつ、人の発話が実際にバーチャルアシスタントサービスに送信される前に、コマンド認識のためにあなたが送信しようとしている、ゾーン3について選択されたマイクロフォン（単数または複数）により検出された発話に対応するマイクロフォン信号に対して、第2パスウェイワード検出器が演算を行うと仮定する。ウェイクワードが実際に発せられたことに関して、第2パスウェイワード検出器が複数の第1パスウェイワード検出器と一致しない場合、それはおそらく、分類器がゾーンを間違えて予測したからである。したがって、分類器は、罰せられなければならない。

【0238】

1つ以上のウェイクワードが話された後の、ゾーンマッピングモデルの事後更新（事後更新）のための技術は、以下のものを含んでもよい。

- ・ 混合ガウスモデル（GMM）または最近傍モデルの最大事後（MAP）適合、ならびに／もしくは、

- ・ 例えば、ニューラルネットワークの強化学習であって、例えば、適切な「ワン-ホット（one-hot）」（正確な予測の場合）または「ワン-コールド（one-cold）」（不正確な予測の場合）の正解ラベル（ground truth label）をSoftMax出力に関連付け、オンラインバックプロパゲーションを適用して、新たなネットワークの重みを決定することにより実現される、強化学習。

【0239】

この文脈におけるMAP適合のいくつかの例は、ウェイクワードが話される度に、GMM内の平均を調節することを含んでもよい。このようにして、平均は、後続のウェイクワードが話された場合に観測される音響的特徴により近づく。代替的にまたは追加的に、そのような例は、ウェイクワードが話される度に、GMM内の分散／共分散または混合重み情報（mixture weight information）を調節することを含んでもよい。

【0240】

例えば、MAP適合スキームは、次の通りであってもよい。

$$\mu_{i, \text{new}} = \mu_{i, \text{old}} * \alpha + x * (1 - \alpha)$$

【0241】

上記式において、 $\mu_{i, \text{old}}$ は混合体内の*i*番目のガウスの平均値を示し、 α はどの程度積極的にMAP適合が起こるべきかを制御するパラメータを示し（ α は、[0.9, 0.999]の範囲内であり得る）、 x は新たなウェイクワード発声の特徴ベクトルを示す。インデックス「*i*」は、ウェイクワード時刻において話者の位置を含むことの、最も高い事前確率を返す混合要素に対応する。

【0242】

あるいは、混合要素の各々は、例えば次式のように、ウェイクワードを含むことの事前確率に応じて調節されてもよい。

$$M_{i, \text{new}} = \mu_{i, \text{old}} * \beta_i * x * (1 - \beta_i)$$

【0243】

上記式において、 $\beta_i = \alpha * (1 - P(i))$ であり、ここで $P(i)$ は、観測値 x が混合要素*i*に帰せられることの事前確率を示す。

【0244】

強化学習の一例において、3つのユーザゾーンが設けられてもよい。ある特定のウェイクワードについて、モデルは、3つのユーザゾーンについて確率が[0.2, 0.1, 0.7]であると予測すると仮定する。第2の情報源（例えば第2パスウェイワード検出器）が、第3のゾーンが正しかったことを確認した場合、正解ラベルは[0, 0, 1]（ワン-ホット）であり得る。ゾーンマッピングモデルの事後更新は、ニューラルネットワークを介してエラーをバックプロパゲーションすることを含んでもよく、これは、同じ入力が再度示された場合にニューラルネットワークがゾーン3をより強く予測するであろうことを事実上意味する。逆に、第2の情報源が、ゾーン3は不正確な予測であったことを示す場合、一例において、正解ラベルは[0.5, 0.5, 0.0]であり得る。ニューラルネットワークを介してエラーをバックプロパゲーションすることにより、将来同じ入

10

20

30

40

50

力が示された場合に、モデルがゾーン3を予測する可能性が低くなる。

【0245】

フレキシブルレンダリングにより、空間オーディオ音声を、任意の個数の任意に配置されたスピーカ上でレンダリングすることが可能になる。家庭内のスマートオーディオデバイス（例えばスマートスピーカ）を含む（但しそれに限定されない）、オーディオデバイスの広範な配置を鑑みて、消費者製品が、オーディオ音声のフレキシブルレンダリングを実行し、そのようにレンダリングされたオーディオ音声を再生することを可能にする、フレキシブルなレンダリング技術を実現する必要がある。

【0246】

フレキシブルレンダリングを実装する複数の技術が開発されている。それら技術は、レンダリングの問題を、コスト関数最小化の1つとして扱う。ここで、コスト関数は2つの項からなる：第1項はレンダラーが実現しようとする所望の空間印象をモデリングし、第2項はスピーカを起動することにコストを割り当てる。今のところ、この第2項は、スパースな解を生成することに集中している。ここで、レンダリングされているオーディオ音声の所望の空間位置の近傍にあるスピーカのみが起動される。

【0247】

消費者環境における空間的オーディオ音声の再生は、典型的には、例えば、5.1サラウンドおよび7.1サラウンドなど、規定された位置に配置された規定数のラウドスピーカに関連付けられている。これらの場合において、コンテンツは、関連付けられたラウドスピーカに対して特定のオーサリングされており、かつ、各ラウドスピーカに対して1つ設けられた離散チャンネルとしてエンコードされている（例えば、DolbyDigitalまたはDolbyDigitalPlusなど）。より最近では、このコンテンツと特定のラウドスピーカ位置との関連を断ち切った、没入感の高いオブジェクト型の空間オーディオフォーマットが導入されている（DolbyAtmos）。その代わり、コンテンツは、個々のオーディオオブジェクトの集合として記述（説明）されてもよい。ここで、各オーディオオブジェクトは、3次元空間内の当該オーディオオブジェクトの所望の知覚位置を記述（説明）している、おそらく時変のメタデータを有している。再生時において、コンテンツは、再生システムにおけるラウドスピーカの個数および位置に適合したレンダラーにより、ラウドスピーカフィールド信号に変換される。しかし、そのようなレンダラーの多くは、依然として、1組のラウドスピーカの位置を、1組の規定されたレイアウト（例えば、DolbyAtmosでの、3.1.2、5.1.2、7.1.4、9.1.6など）のうちの1つとして制約している。

【0248】

そのような制約されたレンダリングを越えんとして、任意の位置に配置された任意の個数のラウドスピーカ上で、オブジェクト型のオーディオ音声柔軟にレンダリングされることを可能にする方法が開発されている。これらの方法は、リスニング空間内のラウドスピーカの個数および物理的な位置をレンダラーが知っていることを要求する。そのようなシステムを平均的な消費者にとって実用的なものとするため、ラウドスピーカの位置を特定する自動化された方法が望ましい。そのような方法の1つは、おそらくはラウドスピーカと共に配置された多数のマイクロフォンを使用することをあてにしている。ラウドスピーカを介してオーディオ信号を再生し、マイクロフォンで録音することにより、各ラウドスピーカとマイクロフォンとの間の距離が推定される。これらの距離から、ラウドスピーカおよびマイクロフォン両方の位置が導き出される。

【0249】

消費者空間内にオブジェクト型の空間オーディオ音声導入されると同時に、アマゾンエコー製品群などの、いわゆる「スマートスピーカ」の急速な選定が行われる。これらのデバイスが非常に人気なのは、無線接続性および統合型音声インターフェイス（例えば、アマゾンのアレクサ）により得られる、そのシンプルさおよび便利さに帰せられるものであり得る。しかし、これらのデバイスの音波性能（sonic capabilities）は、概して、特に空間オーディオ音声に対して、限定されている。たいていの場合において、これらのデバイスは、モノラル再生またはステレオ再生に制約されている。しかし、上記のフレキ

シブルレンダリング技術および自動位置特定技術を、複数のオーケストレーションされたスマートスピーカと組み合わせることにより、消費者による設定が極めてシンプルなままでありつつ、非常に洗練された空間再生能力を有するシステムをもたらす得る。消費者は、無線接続性のおかげでスピーカ配線を用いる必要なく、どこでも都合のいい場所に、望む個数のスピーカを配置することができ、かつ、内蔵のマイクロフォンを用いて、関連付けられたフレキシブルレンダラーのために、スピーカの位置を自動的に特定することができる。

【0250】

従来のフレキシブルレンダリングアルゴリズムは、特定の所望の知覚された空間的印象を、可能な限り厳密に達成するように設計されている。オーケストレーションされたスマートスピーカからなるシステムにおいて、時折、この空間的印象を維持することは、最も重要でなくてもよいし、所望の目的でなくてもよい。例えば、誰かが同時に統合型音声アシスタントに話しかけようとする、検出された発話を含むマイクロフォン信号の信号対雑音比および／または信号対エコー比（S E R）を増大するように、所定のマイクロフォンの近くのスピーカにおいて相対的な再生レベルを低減するような状態で、空間レンダリングをしばらくの間変化させることが望ましいかもしれない。本明細書に記載したいくつかの実施形態は、例えば1つ以上の追加の目的を達成するため、そのような空間レンダリングに対する動的改変を可能にするように、既存のフレキシブルレンダリング方法の改変例として実装されてもよい。

【0251】

既存のフレキシブルレンダリング技術は、重心振幅パンニング（Center of Mass Amplitude Panning (CMAP)）およびフレキシブルバーチャリゼーション（Flexible Virtualization (FV)）を含む。概観的に説明すると、これら技術の両方は、2つ以上のスピーカからなる1組のスピーカ上で再生するために、1つ以上のオーディオ信号（各オーディオ信号は、関連付けられた所望の知覚された空間位置を有する）からなる1組のオーディオ信号をレンダリングする。ここで、この1組のスピーカの相対的な起動は、スピーカ上で再生される前記オーディオ信号の知覚された空間位置のモデル、ならびに、スピーカの位置に対するオーディオ信号の所望の知覚された空間位置の近接度の関数である。このモデルは、目的の空間位置の近くの聴者がオーディオ信号を聞くこと、ならびに、どのスピーカを使用してこの空間印象を達成するかを近接度の項が制御することを確実にする。とりわけ、近接度の項は、オーディオ信号の所望の知覚された空間位置の近くのスピーカを起動することを好む。CMAPおよびFVの両方について、この関数関係は、次式のように、2つの項（1項は空間アスペクトを示し、1項は近接度を示す）の和として記述されたコスト関数から好適に導き出される。

【数12】

$$C(\mathbf{g}) = C_{\text{spatial}}(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) + C_{\text{proximity}}(\mathbf{g}, \vec{o}, \{\vec{s}_i\})$$

(1)

ここで、集合

$\{\vec{s}_i\}$

はM個のラウドスピーカからなる1組のラウドスピーカの位置を示し、

$\vec{o}$

はオーディオ信号の所望の知覚された空間位置を示し、 $\mathbf{g}$ はスピーカ起動電位のM次元ベ

クトルを示す。C M A Pについて、このベクトルにおける各起動電位は、スピーカあたりのゲインを示す。一方、F Vについて、各起動電位はフィルタを示す（後者において、 g は、特定の周波数における複素数のベクトルと同一視され得、かつ、異なる g が複数の周波数に亘って算出されて、フィルタを形成する）。起動電位の最適なベクトルは、次式のように、複数の起動電位に亘るコスト関数を最小化することにより見出される。

【数 1 3】

$$\mathbf{g}_{opt} = \min_{\mathbf{g}} C(\mathbf{g}, \vec{o}, \{\vec{s}_i\})$$

(2 a)

10

【 0 2 5 2 】

コスト関数の所与の定義では、

$\mathbf{g}_{opt}$

の成分の間の相対的なレベルは適切であるが、上記最小化の結果得られた最適な起動電位の絶対レベルを制御することは難しい。この問題を取り扱うために、その後、

$\mathbf{g}_{opt}$

20

の正規化を行って、起動電位の絶対レベルを制御してもよい。例えば、単位長を有するようにベクトルを正規化するのが望ましいかもしれない。これは、一般的に使用されるコンスタントパワーパンニングルールに従って行われる。

【数 1 4】

$$\bar{\mathbf{g}}_{opt} = \frac{\mathbf{g}_{opt}}{\|\mathbf{g}_{opt}\|}$$

(2 b)

30

【 0 2 5 3 】

フレキシブルレンダリングアルゴリズムの正確な振る舞いは、コスト関数の2つの項 $C_{spatial}$ および $C_{proximity}$ の特定の構成により規定される。C M A P の場合、 $C_{spatial}$ は、1組のラウドスピーカから再生されるオーディオ信号の知覚された空間位置を、関連付けられた起動ゲイン g_i （ベクトル $\mathbf{g}$ の要素）で重み付けされたラウドスピーカの位置の重心に配置するモデルから導出される。

【数 1 5】

$$\vec{o} = \frac{\sum_{i=1}^M g_i \vec{s}_i}{\sum_{i=1}^M g_i}$$

(3)

40

式 (3) は、その後、所望のオーディオ位置と起動されたラウドスピーカにより生成された位置との二乗誤差を示す空間コストへと変形される。

【数 1 6】

50

$$C_{spatial}(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) = \|(\sum_{i=1}^M g_i) \vec{o} - \sum_{i=1}^M g_i \vec{s}_i\|^2 = \|\sum_{i=1}^M g_i (\vec{o} - \vec{s}_i)\|^2 \quad (4)$$

F V の場合、コスト関数の空間項は、異なるように定義される。その目的は、聴者の左耳および右耳におけるオーディオオブジェクト位置

$\vec{o}$

に対応するバイノーラルレスポンス $\mathbf{b}$ を生成することである。コンセプト上、 $\mathbf{b}$ はフィルタの 2×1 ベクトル（片耳に 1 つのフィルタ）であるが、より簡便には、特定の周波数における複素数の 2×1 ベクトルとして扱う。この特定の周波数における表現で進めると、所望のバイノーラルレスポンスは、オブジェクト位置で索引付けされた 1 組の H R T F から取得し得る。

【数 1 7】

$$\mathbf{b} = HRTF\{\vec{o}\}$$

(5)

それと同時に、ラウドスピーカにより聴者の耳において生成された 2×1 バイノーラルレスポンス $\mathbf{e}$ は、 $2 \times M$ 音響伝達マトリクス $\mathbf{H}$ に複素スピーカ起動電位値（complex speaker activation value）の $M \times 1$ ベクトル $\mathbf{g}$ を掛け合わせたものとしてモデル化される。

【数 1 8】

$$\mathbf{e} = \mathbf{H}\mathbf{g}$$

(6)

音響伝達マトリクス $\mathbf{H}$ は、聴者の位置に対する、ラウドスピーカの位置の集合

$\{\vec{s}_i\}$

に基づきモデル化される。最後に、コスト関数の空間成分は、所望のバイノーラルレスポンス（式（5））とラウドスピーカにより生成されたバイノーラルレスポンス（式（6））との二乗誤差として定義される。

【数 1 9】

$$C_{spatial}(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) = (\mathbf{b} - \mathbf{H}\mathbf{g})^*(\mathbf{b} - \mathbf{H}\mathbf{g})$$

(7)

簡便には、式（4）および式（7）に定義した C M A P および F V についてのコスト関数の空間項は、両方とも、スピーカ起動電位 $\mathbf{g}$ の関数としての二次行列に変形され得る。

【数 2 0】

$$C_{spatial}(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) = \mathbf{g}^* \mathbf{A} \mathbf{g} + \mathbf{B} \mathbf{g} + \mathbf{C}$$

(8)

ここで、AはM x Mの正方行列であり、Bは1 x Mベクトルであり、Cはスカラーである。行列Aは階数が2であり、したがって、M>2である場合、空間誤差項がゼロに等しいスピーカ起動電位gが無限に存在する。コスト関数の第2項C_{proximity}を導入することにより、この不確実性が除かれ、結果として、他の可能な解と比較して、知覚に関して有益な特性を有する特定の解が得られる。CMAFおよびFVの両方について、C_{proximity}は、所望のオーディオ信号位置

10

 $\vec{o}$

から離れた位置

 $\vec{s}_i$

に存在するスピーカの起動が、所望の位置の近くに配置されたスピーカの起動よりも大きい罰則を受けるように、構成される。この構成により、スパース（疎ら）であるスピーカ起動電位の最適な集合が生成され（ここで、所望のオーディオ信号の位置に近接したスピーカのみが有意に起動される）、実際に、結果として、1組のスピーカの周囲における聴者の移動に対して知覚上よりロバストな、オーディオ信号の空間再生が得られる。

20

【0254】

この目的のために、コスト関数の第2項C_{proximity}は、スピーカ起動電位の絶対値の二乗の距離加重合計として定義されてもよい。これは、次式のように、行列の形でコンパクトに表現される。

【数21】

$$C_{proximity}(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) = \mathbf{g}^* \mathbf{D} \mathbf{g}$$

30

(9a)

ここで、Dは、所望のオーディオ位置と各スピーカとの間の距離罰則の対角行列である。

【数22】

$$\mathbf{D} = \begin{bmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_M \end{bmatrix}, \quad d_i = distance(\vec{o}, \vec{s}_i)$$

40

(9b)

距離罰則関数は、多くの形式をとり得るが、次式は有用なパラメータ化である。

$$distance(\vec{o}, \vec{s}_i) = \alpha d_0^2 \left(\frac{\|\vec{o} - \vec{s}_i\|}{d_0} \right)^\beta$$

50

(9c)

ここで、

$$\|\vec{o} - \vec{s}_i\|$$

は所望のオーディオ位置とスピーカ位置との間のユークリッド距離であり、 α および β は調節可能なパラメータである。パラメータ α は罰則の全体的な強度を示す。 d_0 は距離罰則の空間的範囲に対応し（およそ距離 d_0 以上離れた位置のラウドスピーカが罰せられる）、 β は距離 d_0 における罰則の開始の急激度（abruptness）を表す。

【0255】

式(8)および式(9a)で定義されたコスト関数の2つの項を組み合わせることにより、コスト関数全体が得られる。

【数24】

$$C(\mathbf{g}) = \mathbf{g}^* \mathbf{A} \mathbf{g} + \mathbf{B} \mathbf{g} + \mathbf{C} + \mathbf{g}^* \mathbf{D} \mathbf{g} = \mathbf{g}^* (\mathbf{A} + \mathbf{D}) \mathbf{g} + \mathbf{B} \mathbf{g} + \mathbf{C}$$

(10)

【0256】

このコスト関数の、 $\mathbf{g}$ に対する導関数をゼロに設定し、 $\mathbf{g}$ について解くことにより、スピーカ起動電位の最適解が得られる。

【数25】

$$\mathbf{g}_{opt} = \frac{1}{2} (\mathbf{A} + \mathbf{D})^{-1} \mathbf{B}$$

(11)

【0257】

概して、式(11)における最適解は、負の値のスピーカ起動電位を生じ得る。フレキシブルレンダラーのCMAP構成にとって、そのような負の起動電位は望ましくないかもしれないが、したがって、式(11)は、全ての起動電位が正の値であると限定して、最小化されてもよい。

【0258】

図15および図16は、スピーカ起動電位の集合の一例およびオブジェクトレンダリング位置を示す図である。これらの例において、スピーカ起動電位およびオブジェクトレンダリング位置は、4度、64度、165度、-87度、および-4度のスピーカ位置に対応している。図15は、スピーカ起動電位1505a、1510a、1515a、1520aおよび1525aを示す。これらは、上記特定のスピーカ位置について、式(11)に対する最適解を含んでいる。図16は、個々のスピーカ位置を、ドット1605、1610、1615、1620および1625としてプロットしている。これらは、それぞれ、スピーカ起動電位1505a、1510a、1515a、1520aおよび1525aに対応している。図16はまた、多数の考え得るオブジェクト角度（ドット1630a）に対する、理想のオブジェクト位置（換言すると、オーディオオブジェクトがレンダリングされる位置）、ならびに、それらオブジェクトについての、対応する実際のレンダリング位置（点線1640aにより理想のオブジェクト位置に連結されたドット1635a）を示している。

【0259】

ある種の実施形態は、複数のコーディネート（オーケストレーション）されたスマート

10

20

30

40

50

オーディオデバイスのうちの少なくとも1つ（例えば、全てまたはいくつか）による再生のために、オーディオ音声をレンダリングする方法を含む。例えば、ユーザの自宅に設けられた（システムに含まれる）1組のスマートオーディオデバイスは、さまざまな同時のユースケースに対処するようにオーケストレーションされてもよい。このようなケースは、全てまたはいくつかのスマートオーディオデバイスによる（つまり、全てまたはいくつかのスマートオーディオデバイスのスピーカ（単数または複数）による）再生のために、（実施形態に従って）オーディオ音声のフレキシブルレンダリングを行うことを含む。レンダリングに対する動的修正を要求する、システムとのインタラクションが多数想定されている。そのような修正は、空間忠実性（spatial fidelity）に焦点をあててもよいが、必ずしもその限りではない。

10

【0260】

いくつかの実施形態は、1組のスマートオーディオデバイスのうちの少なくとも1つ（例えば、全てまたはいくつか）による再生のために（もしくは、別の1組のスピーカのうちの少なくとも1つ（例えば、全てまたはいくつか）による再生のために）、オーディオ音声をレンダリングする方法である。レンダリングは、コスト関数の最小化を含んでもよい。ここで、コスト関数は、少なくとも1つの動的なスピーカ起動電位の項を含む。そのような動的なスピーカ起動電位の項の例として、以下に列挙するものが挙げられる（但し、それらには限定されない）：

- ・ 1人以上の聴者に対するスピーカの近接度；
- ・ 引力または斥力（attracting or repelling force）に対するスピーカの近接度；
- ・ いくつかの位置（例えば、聴者の位置、または乳児部屋）に対するスピーカの可聴度；
- ・ スピーカ能力（例えば、周波数特性および歪み）；
- ・ スピーカの、他のスピーカとの同期性；
- ・ ウェイクワード性能；ならびに、
- ・ エコー除去性能。

20

【0261】

動的なスピーカ起動電位の項（単数または複数）は、さまざまな振る舞いのうちの少なくとも1つを可能にしてもよい。これらの振る舞いは、特定のスマートオーディオデバイスから離れる方向に、オーディオ音声の空間的プレゼンテーション（spatial presentation）をワーピングすることにより、そのマイクロフォンが話者の発声をより良く聞くことができるようにすること、もしくは、二次的オーディオストリームが、スマートオーディオデバイスのスピーカ（単数または複数）からより良く聞こえるようにすることを含む。

30

【0262】

いくつかの実施形態は、コーディネート（オーケストレーション）された複数のスマートオーディオデバイスのスピーカ（単数または複数）による再生のためのレンダリングを実装する。他の実施形態は、別の1組のスピーカのスピーカ（単数または複数）による再生のためのレンダリングを実装する。

【0263】

（いくつかの実施形態に応じて実装される）フレキシブルレンダリング方法を、1組のワイヤレススマートスピーカ（または他のスマートオーディオデバイス）と組み合わせることにより、極めて能力が高くかつ使い易い空間オーディオレンダリングシステムを得ることができる。そのようなシステムとのインタラクションを想定する際に、システムの使用中に起こり得る他の目的のために最適化するために、空間レンダリングに動的な修正を加えることが望ましいことが明白になる。この目的を達成するために、ある種の実施形態は、（スピーカ起動電位が既に開示した空間項および近接度項の関数である）既存のフレキシブルレンダリングアルゴリズムを、レンダリングされているオーディオ信号、1組のスピーカ、および／または他の外部入力1つ以上の特性に基づき、1つ以上の追加の動的に設定可能な関数で増大する。いくつかの実施形態によると、式（1）に与えられた既存のフレキシブルレンダリングのコスト関数は、次式に応じて、これら1つ以上の追加の

40

50

依存性で増大される。

【数 2 6】

$$C(\mathbf{g}) = C_{\text{spatial}}(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) + C_{\text{proximity}}(\mathbf{g}, \vec{o}, \{\vec{s}_i\}) + \sum_j C_j \left(\mathbf{g}, \{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j \right) \tag{12}$$

【 0 2 6 4】

10

式 (1 2) において、項

$$C_j \left(\mathbf{g}, \{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j \right)$$

は、追加のコスト項を示す。ここで、

$$\{\hat{o}\}$$

はレンダリングされている（例えば、オブジェクト型オーディオプログラムの）オーディオ信号の 1 つ以上の特性からなる集合を示し、

20

$$\{\hat{s}_i\}$$

はオーディオ音声がレンダリングされているスピーカの 1 つ以上の特性からなる集合を示し、

$$\{\hat{e}\}$$

は 1 つ以上の追加の外部入力を示す。各項

30

$$C_j \left(\mathbf{g}, \{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j \right)$$

は、集合

$$\{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j$$

により包括的に表される、オーディオ信号、スピーカ、および／または外部入力の 1 つ以上の特性の組み合わせに関して、起動電位 g の関数としてコストを返す。なお、集合

40

$$\{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j$$

は、

$$\{\hat{o}\}$$

、

$$\{\hat{s}_i\}$$

、または

50

$\{\hat{e}\}$

のうち、最低でも1つの要素を含むことを理解されたい。

【0265】

以下に、

 $\{\hat{o}\}$

の例を挙げる（但し、それらに限定されない）。

- ・ オーディオ信号の、所望の知覚された空間位置；
- ・ オーディオ信号のレベル（おそらく時変）；および／または
- ・ オーディオ信号のスペクトル（おそらく時変）。

以下に、

 $\{\hat{s}_i\}$

の例を挙げる（但し、それらに限定されない）。

- ・ リスニング空間内のラウドスピーカの位置；
- ・ ラウドスピーカの周波数特性；
- ・ ラウドスピーカの再生レベルの上下限；
- ・ スピーカ内のダイナミクス処理アルゴリズムのパラメータ（リミッターゲイン（limiter gains）など）；
- ・ 各スピーカから他のスピーカへの音響伝達の測定値または推定値；
- ・ スピーカにおけるエコー除去性能の尺度；および／または、
- ・ スピーカ間における相対的な同期。

以下に、

 $\{\hat{e}\}$

の例を挙げる（但し、それらに限定されない）。

- ・ 再生空間内の1人以上の聴者または話者の位置；
- ・ 各ラウドスピーカからリスニング位置への音響伝達の測定値または推定値；
- ・ 話者から1組のラウドスピーカへの音響伝達の測定値または推定値；
- ・ 再生空間内の何らかの他のランドマークの位置；および／または、
- ・ 再生空間内における各スピーカから何らかの他のランドマークへの音響伝達の測定値または推定値。

【0266】

式（12）に規定する新たなコスト関数で、先に式（2a）および式（2b）で特定されたgについての最小化および考え得る事後正規化を介して、最適な1組の起動電位が見出され得る。

【0267】

図17は、図2Aに示すような装置またはシステムにより実行され得る方法の一例の概略を示す流れ図である。方法1700のブロックは、本明細書に開示する他の方法と同様、必ずしも図示した順序どおり実行する必要はない。さらに、このような方法は、図示および／または説明したブロックより多い個数または少ない個数のブロックを含んでもよい。方法1700のブロックは、図2Aに示す制御システム210などの制御システムであり得る（または制御システムを含み得る）、1つ以上のデバイスにより実行されてもよい。

【0268】

この実施態様において、ブロック1705は、制御システムにより、インターフェイス

10

20

30

40

50

システムを介して、オーディオデータを受信することを含む。この例において、オーディオデータは、1つ以上のオーディオ信号および関連付けられた空間データを含む。この実施態様によると、空間データは、オーディオ信号に対応する、目標の知覚された空間位置を示す。いくつかの例において、目標の知覚された空間位置は、例えば、DolbyAtmos位置メタデータなどの位置メタデータにより示されるように、明示的であってもよい。他の例において、目標の知覚された空間位置は、暗黙的であってもよい。例えば、目標の知覚された空間位置は、Dolby5.1、Dolby7.1または他のチャンネル型オーディオフォーマットに基づくチャンネルに関連付けられた仮定の位置であってもよい。いくつかの例において、ブロック1705は、インターフェイスシステムを介してオーディオデータを受信する制御システムのレンダリングモジュールを含む。

10

【0269】

この例によると、ブロック1710は、制御システムにより、環境の1組のラウドスピーカを介した再生のためのオーディオデータをレンダリングして、レンダリングされたオーディオ信号を生成することを含む。この例において、オーディオデータに含まれる1つ以上のオーディオ信号の各々をレンダリングすることは、コスト関数を最適化することにより、環境内の1組のラウドスピーカの相対的な起動を決定することを含む。この例によると、コストは、環境内の1組のラウドスピーカ上で再生される場合のオーディオ信号の知覚された空間位置のモデルの関数である。この例において、コストはまた、1組のラウドスピーカのうちの各ラウドスピーカの位置に対する、オーディオ信号の目標の知覚された空間位置の近接度の尺度の関数でもある。この実施態様において、コストはまた、1つ以上の追加の動的に設定可能な関数の関数でもある。この例において、動的に設定可能な関数は、以下の事項のうち1つ以上の事項に基づく。1人以上の聴者に対するラウドスピーカの近接度；引力位置に対するラウドスピーカの近接度。ここで、引力は、引力位置により近接した、相対的により高いラウドスピーカ起動電位を好む要素である。；斥力位置に対するラウドスピーカの近接度。ここで、斥力は、斥力位置により近接した、相対的により低いラウドスピーカ起動電位を好む要素である。；環境内の他のラウドスピーカに対する、各ラウドスピーカの能力；ラウドスピーカの、他のラウドスピーカとの同期；ウェイクワード性能；または、エコー除去性能。

20

【0270】

この例において、ブロック1715は、レンダリングされたオーディオ信号を、環境の1組のラウドスピーカのうちの少なくともいくつかのラウドスピーカに、インターフェイスシステムを介して提供することを含む。

30

【0271】

いくつかの例によると、知覚された空間位置のモデルは、聴者の左右の耳において、オーディオオブジェクト位置に対応するバイノーラルレスポンスを生成してもよい。代替的にまたは追加的に、知覚された空間位置のモデルは、1組のラウドスピーカから再生されるオーディオ信号の知覚された空間位置を、ラウドスピーカの関連付けられた起動ゲインで重み付けされた、1組のラウドスピーカの位置の重心に配置してもよい。

【0272】

いくつかの例において、1つ以上の追加の動的に設定可能な関数は、1つ以上のオーディオ信号のレベルに少なくとも部分的に基づいてもよい。いくつかの例において、1つ以上の追加の動的に設定可能な関数は、1つ以上のオーディオ信号のスペクトルに少なくとも部分的に基づいてもよい。

40

【0273】

方法1700のいくつかの例は、ラウドスピーカレイアウト情報を受信することを含む。いくつかの例において、1つ以上の追加の動的に設定可能な関数は、環境内のラウドスピーカの各々の位置に少なくとも部分的に基づいてもよい。

【0274】

方法1700のいくつかの例は、ラウドスピーカ仕様情報を受信することを含む。いくつかの例において、1つ以上の追加の動的に設定可能な関数は、各ラウドスピーカの能力

50

に少なくとも部分的に基づいてもよい。ここで、各ラウドスピーカの能力は、周波数特性、再生レベルの上下限值、または1つ以上のラウドスピーカダイナミクス処理アルゴリズムのうちの1つ以上を含み得る。

【0275】

いくつかの例によると、1つ以上の追加の動的に設定可能な関数は、各ラウドスピーカから他のラウドスピーカへの音響伝達の測定値または推定値に少なくとも部分的に基づいてもよい。代替的にまたは追加的に、1つ以上の追加の動的に設定可能な関数は、環境内の1人以上の人々のうちの聴者または話者の位置に少なくとも部分的に基づいてもよい。代替的にまたは追加的に、1つ以上の追加の動的に設定可能な関数は、各ラウドスピーカから聴者または話者の位置への音響伝達の測定値または推定値に少なくとも部分的に基づいてもよい。音響伝達の推定値は、例えば、各ラウドスピーカと聴者または話者の位置との間に存在し得る壁、家具、または他の物体に少なくとも部分的に基づいてもよい。

10

【0276】

代替的にまたは追加的に、1つ以上の追加の動的に設定可能な関数は、環境内の1つ以上のラウドスピーカではない物体またはランドマークのオブジェクト位置に少なくとも部分的に基づいてもよい。いくつかのそのような実施態様において、1つ以上の追加の動的に設定可能な関数は、各ラウドスピーカからオブジェクト位置またはランドマーク位置への音響伝達の測定値または推定値に少なくとも部分的に基づいてもよい。

【0277】

1つ以上の適切に定義された追加のコスト項を用いてフレキシブルレンダリングを実装することにより、数多の新しくかつ有用な振る舞いが達成されてもよい。以下に列挙する全ての振る舞いの例は、望ましくないと考えられる一定の条件下において、所与のラウドスピーカを罰することに関してキャストिंग (cast) される。最終結果として、これらのラウドスピーカは、1組のオーディオ信号の空間レンダリングにおいて、比較的起動され難い (activated less)。これらの場合の多くにおいて、空間レンダリングに任意の修正を加えることは別に、単に望ましくないラウドスピーカの音量を下げることを考え得るが、そのような戦略は、オーディオコンテンツの全体のバランスを著しく損ねるかもしれない。例えば、ミキシングした音声に含まれる或る成分が、全く聞こえなくなるかもしれない。一方、開示した実施態様では、これらの罰則付与をレンダリングのコア最適化に統合することにより、レンダリングを適合させ、残りの罰則の小さいスピーカを用いて考え得る最良の空間レンダリングを実行することを可能にする。これは、はるかに上質で、適合性があり、かつ効果的な解決方法である。

20

【0278】

ユースケースの例は以下のものを含むが、それらに限定されるものではない。

- ・ リスニングエリア周辺に、よりバランスのとれた空間的プレゼンテーションを提供する。

○ 空間オーディオ音声は、目的のリスニングエリアから概ね同じ距離にあるラウドスピーカに亘って、最良の状態で提示されることがわかっている。ラウドスピーカからリスニングエリアまでの距離の平均と比較して、かなり近くにあるラウドスピーカまたはかなり離れたラウドスピーカが罰せられて、その起動が低減されるように、コストが構成されてもよい。

40

【0279】

- ・ 聴者または話者から離れる方向もしくは聴者または話者に近づく方向に、オーディオ音声を移動する。

○ システムのユーザが、システムの（またはシステムに関連付けられた）スマート音声アシスタントに話しかけようとする場合、話者により近いラウドスピーカを罰するコストを生成することが有益かもしれない。このようにして、これらのラウドスピーカは比較的起動され難く (activated less)、それらの関連付けられたマイクロフォンは、話者の発声をより良く聞くことが可能になる。

○ リスニング空間内の他者に対する再生レベルを最小化する、1人の聴者にとっ

50

てより親密な体験を提供するために、聴者の位置から離れた位置にあるスピーカを重く罰することにより、聴者に最も近いスピーカのみが最も有意に起動されるようにしてもよい。

【0280】

・ ランドマーク、ゾーン、またはエリアから離れる方向もしくはランドマーク、ゾーン、またはエリアに近づく方向に、オーディオ音声を移動する。

○ リスニング空間の近傍の所与の位置は、要注意であると考え得る（乳児部屋、乳児のベッド、オフィス、読書エリア、学習エリアなど）。そのような場合、この位置、ゾーン、またはエリアに近いスピーカの使用を罰するように、コストが構成されてもよい。

○ あるいは、上記と同じ場合（または同様の場合）について、特に（取り付けられたまたは関連付けられたマイクロフォンを有する）スピーカのうちの1つが乳児部屋内部に設けられている場合、スピーカのシステムは、各スピーカから乳児部屋内への音響伝達の測定値を生成済みであり得る。この場合、乳児部屋に対するスピーカの物理的近接度を用いるのではなく、測定された乳児部屋内への音響伝達が高いスピーカを使用することを罰するように、コストを構成してもよい。

【0281】

・ スピーカの能力の最適使用

○ 異なるラウドスピーカの能力は、大きく異なり得る。例えば、ある人気のスマートスピーカは、限定的な低周波数能力の、1.6インチフルレンジドライバを1つだけ備えている。一方、別のスマートスピーカは、はるかに能力の高い3インチウーファを備えている。これらの能力は、概して、スピーカの周波数特性に反映されるので、スピーカに関連付けられた周波数特性の集合を、コスト項に利用し得る。特定の周波数において、周波数特性の点で他のスピーカと比べて能力の劣るスピーカは、罰則が与えられ、したがって、このスピーカが起動される程度はより低くなり得る。いくつかの実施態様において、このような周波数特性値は、スマートラウドスピーカと共に格納され、その後、フレキシブルレンダリングを最適化する責任を負った演算ユニットに伝達されてもよい。

【0282】

○ 多くのスピーカが複数のドライバを備えており、各ドライバは、異なる周波数帯域を再生する責任を負っている。例えば、ある任意のスマートスピーカは、低周波数用のウーファと高周波数用のツイーターを備えた2way設計である。典型的には、そのようなスピーカは、フルレンジの再生オーディオ信号を複数の適切な周波数帯域に分割して、それぞれのドライバに送信するクロスオーバー回路を備えている。あるいは、そのようなスピーカは、個々のドライバの能力に関する情報（周波数特性など）だけでなく、個々のドライバへの再生アクセスを、フレキシブルレンダラーに提供する。直前に説明したようなコスト項を適用することにより、いくつかの例において、フレキシブルレンダラーは、2つのドライバの間のクロスオーバーを、異なる周波数におけるそれらドライバの相対的な能力に基づき、自動的に形成してもよい。

【0283】

○ 上述の周波数特性の使用例は、スピーカの固有の能力に焦点を当てているが、リスニング環境に配置されたスピーカの能力を正確に反映しなくてもよい。所与の場合において、目的のリスニング位置で測定されたスピーカの周波数特性は、何らかの較正処置を介して利用可能であってもよい。スピーカの使用をさらに良く最適化すべく、そのような測定値を、予備計算された周波数特性の代わりに用いてもよい。例えば、あるスピーカは、ある特定の周波数において固有の能力が高いかもしれないが、その配置（例えば、壁や家具の裏側）のせいで、目的のリスニング位置における周波数特性が大きく限定され得る。この周波数特性をキャプチャしかつ適切なコスト項に与えられる測定値により、そのようなスピーカの有意な起動を防止することができる。

【0284】

○ 周波数特性は、ラウドスピーカの再生能力の1つのアスペクトにすぎない。多くの小型ラウドスピーカは、再生レベルが上昇するにつれ、まず歪み始めて、その後、特に低い周波数について、エクスカージョン限界に到達する。そのような歪みを低減するた

10

20

30

40

50

めに、多くのラウドスピーカが、再生レベルを、周波数に亘って可変であり得る、いくつかの限界閾値よりも低いレベルに抑制する、ダイナミクス処理を実装する。スピーカがこれら閾値の近傍または閾値上にある一方で、フレキシブルレンダリングに参加している他のスピーカがそうでない場合、この限界に達しているスピーカにおいて信号レベルを低減し、このエネルギーを他の比較的負担の少ないスピーカに振り向けることは、理に適っている。そのような振る舞いは、関連付けられたコスト項を適切に設定することにより、いくつかの実施形態に従って、自動的に達成され得る。そのようなコスト項は、以下の事項のうちの1つ以上を含んでもよい。

【0285】

・ ラウドスピーカの限界閾値に関して、全体の再生音量を監視する。例えば、音量レベルがその限界閾値に近いラウドスピーカに、より大きい罰則を与えてもよい。

10

【0286】

・ ラウドスピーカ限界閾値（おそらく周波数に亘って可変）に対する関係において、動的な信号レベル（おそらく周波数に亘って可変）を監視する。例えば、監視された信号レベルがその限界閾値により近いラウドスピーカは、より大きい罰則を与えられてもよい。

【0287】

・ 制限ゲイン（limiting gains）など、ラウドスピーカのダイナミクス処理のパラメータを直接監視する。いくつかのそのような例において、これらパラメータがより大きい制限を示すラウドスピーカは、より大きい罰則を与えられてもよい。

【0288】

・ 増幅器によりラウドスピーカへと給送される、実際の瞬間的な電圧、電流、および電力を監視して、ラウドスピーカが、線形範囲において動作しているか否かを判定する。例えば、より低い線形性で動作しているラウドスピーカが、より大きい罰則を与えられ得る。

20

【0289】

○ 一体型のマイクロフォンおよび対話型音声アシスタントを備えたスマートスピーカは、典型的には、ある種のエコー除去技術を用いて、録音マイクロフォンにより採録された、スピーカから再生されているオーディオ信号のレベルを低減する。この低減が大きくなるほど、スピーカは、空間内の話者の発声を聞き取りかつ理解できる可能性が高まる。エコーキャンセラの残余が一貫して高い場合、このことは、スピーカが、エコー経路の予測が難しい非線形領域内へと駆動されていることを示しているかもしれない。そのような場合、このスピーカから離れる方向に信号エネルギーを振り向けることは、理に適っているかもしれない、したがって、エコー除去性能を考慮したコスト項が有益かもしれない。そのようなコスト項は、その関連付けられたエコーキャンセラの性能が低いスピーカに、高いコストを割り当て得る。

30

【0290】

○ 複数のラウドスピーカ上で空間オーディオ音声をレンダリングする場合に、予測可能なイメージングを達成するために、概して、1組のラウドスピーカ上での再生を、時間全体に亘って適度に同期させることが要求される。有線接続されたラウドスピーカにとっては当然であるが、多数の無線ラウドスピーカで、同期をとるのは難しく、最終結果が変動し得る。そのような場合、各ラウドスピーカは、目的のスピーカとの同期の相対的な程度を報告することが可能かもしれない、この同期の程度は、同期コスト項に与えられてもよい。いくつかのそのような例において、同期の程度がより低いラウドスピーカは、より大きい罰則が与えられてもよく、したがって、レンダリングから排除されてもよい。さらに、ある種のオーディオ信号（例えば、拡散または無指向性の再生を意図した、ミキシングしたオーディオ音声の成分）には、厳密な同期が要求されなくてもよい。いくつかの実施態様において、成分はメタデータでそのようにタグが付されてもよく、同期コスト項は、罰則が低減されるように修正されてもよい。

40

【0291】

次に、実施形態のさらなる例を説明する。式（9 a）および式（9 b）に定義された近

50

接度コストと同様、新たなコスト関数の項

$$C_j(\mathbf{g}, \{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j)$$

の各々を、スピーカ起動電位の絶対値の二乗の加重合計として表現することが、都合がよいかもしれない。例えば、次式のとおりである。

【数 2 7】

$$C_j(\mathbf{g}, \{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j) = \mathbf{g}^* \mathbf{W}_j (\{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j) \mathbf{g}$$

10

(1 3 a)

ここで、 $\mathbf{W}_j$ は、項 j について起動するスピーカ i に関連付けられたコストを説明する重み

$$w_{ij} = w_{ij}(\{\{\hat{o}\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j)$$

の対角行列である。

【数 2 8】

$$\mathbf{W}_j = \begin{bmatrix} w_{1j} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{Mj} \end{bmatrix}$$

20

(1 3 b)

【0 2 9 2】

式 (1 3 a) および式 (1 3 b) を、式 (1 0) に与えた CMAP および FV のコスト関数を二次行列に変換したものと組み合わせることにより、式 (1 2) に与えた (いくつかの実施形態の) 一般的な拡大されたコスト関数の、潜在的に有益な実施態様が得られる。

30

【数 2 9】

$$C(\mathbf{g}) = \mathbf{g}^* \mathbf{A} \mathbf{g} + \mathbf{B} \mathbf{g} + \mathbf{C} + \mathbf{g}^* \mathbf{D} \mathbf{g} + \sum_j \mathbf{g}^* \mathbf{W}_j \mathbf{g} = \mathbf{g}^* (\mathbf{A} + \mathbf{D} + \sum_j \mathbf{W}_j) \mathbf{g} + \mathbf{B} \mathbf{g} + \mathbf{C}$$

(1 4)

【0 2 9 3】

新たなコスト関数の項がこのように定義されると、コスト関数全体は二次行列のままであり、起動電位の最適な集合 $\mathbf{g}_{opt}$ は、式 (1 4) の微分を介して、次式のように見出し得る。

40

【数 3 0】

$$\mathbf{g}_{opt} = \frac{1}{2} (\mathbf{A} + \mathbf{D} + \sum_j \mathbf{W}_j)^{-1} \mathbf{B}$$

(1 5)

【0 2 9 4】

重みの項 w_{ij} の各々を、ラウドスピーカの各々についての、所与の連続の罰則値

50

$$p_{ij} = p_{ij}(\{\{\delta\}, \{\hat{s}_i\}, \{\hat{e}\}\}_j)$$

の関数として考えることが有益である。ある実施形態例において、この罰則値は、（レンダリングしようとする）オブジェクトから考慮されたラウドスピーカまでの距離である。別の実施形態例において、この罰則値は、所与のラウドスピーカがいくつかの周波数を再生できないことを表す。この罰則値に基づき、重みの項 w_{ij} を、次のようにパラメータ化できる。

【数 3 1】

$$w_{ij} = \alpha_j f_j \left(\frac{p_{ij}}{\tau_j} \right)$$

(16)

ここで、 α_j は（重み項の全体の強度を考慮に入れた）前因子（pre-factor）である。 τ_j は罰則閾値である（この罰則閾値の周辺においてまたはこの罰則閾値を超えると、重みの項が有意になる）。 $f_j(x)$ は単調増加関数である。例えば、

$$f_j(x) = x^{\beta_j}$$

の場合、重みの項は次式の形を有する。

【数 3 2】

$$w_{ij} = \alpha_j \left(\frac{p_{ij}}{\tau_j} \right)^{\beta_j}$$

(17)

ここで、 α_j 、 β_j 、 τ_j は、それぞれが罰則の全体的な強度、罰則の開始の急激度、および罰則の範囲を示す、調節可能なパラメータである。これらの調節可能な値を設定する際には、コスト項 C_j の別の追加のコスト項に対する相対的な効果が、 $C_{spatial}$ および $C_{proximity}$ と同様、所望の結果を達成するために適切なものとなるように、留意する必要がある。例えば、経験則から言うと、あるものが、他のものを明確に支配するために、特定の罰則を望む場合、その強度 α_j を、次に最も大きい罰則強度の約 10 倍大きく設定することが適切であり得る。

【0295】

全てのラウドスピーカが罰せられた場合、後処理において全ての重み項から最小の罰則を減じて、これらスピーカのうち少なくとも 1 つが罰せられないようにするのが、多くの場合において好都合である。

【数 3 3】

$$w_{ij} \rightarrow w'_{ij} = w_{ij} - \min_i(w_{ij})$$

(18)

【0296】

上述のとおり、本明細書に記載の新たなコスト関数項（および、他の実施形態に従って使用される、同様の新たなコスト関数項）を用いて実現され得る、考え得るユースケースが多数存在する。次に、より具体的な詳細を、3つの例（聴者または話者に向かう方向にオーディオ音声を移動する例、聴者または話者から離れる方向にオーディオ音声を移動する例、ならびに、ランドマークから離れる方向にオーディオ音声を移動する例）により説明する。

【0297】

第1の例では、本明細書において「引力」と呼ぶものを用いて、ある位置に向かってオーディオ音声を引っ張る。ここで、ある位置とは、いくつかの例において、聴者または話者の位置、ランドマークの位置、家具の位置などであり得る。この位置を、本明細書中、

10

$\vec{l}_j$

から i 番目のスピーカまでの距離により与えられ、閾値 τ_j は、全てのスピーカについてのこれら距離のうちの最大値により与えられる。

【数34】

20

$$p_{ij} = \|\vec{l}_j - \vec{s}_i\|$$

(19a)

【数35】

$$\tau_j = \max_i \|\vec{l}_j - \vec{s}_i\|$$

30

(19b)

【0298】

聴者または話者に向かう方向にオーディオ音声を「引っ張る」ことのユースケースを説明するために、具体的には、 $\alpha_j = 20$ 、 $\beta_j = 3$ 、

$\vec{l}_j$

を180度の聴者／話者の位置（プロットの底部中央）に対応するベクトルに設定する。 α_j 、 β_j および

40

$\vec{l}_j$

のこれらの値は、例示に過ぎない。いくつかの実施態様において、 α_j は1～100の範囲内であってもよく、 β_j は1～25の範囲内であってもよい。図18は、実施形態例における、スピーカ起動電位のグラフである。この例において、図18は、スピーカ起動電位1505b、1510b、1515b、1520bおよび1525bを示す。これらは、 w_{ij} で表される引力を加えた、図15および図16と同じスピーカ位置についてのコスト関数に対する最適解を含んでいる。図19は、実施形態例における、オブジェクトレ

50

ンダリング位置のグラフである。この例において、図 19 は、多数の考え得るオブジェクト角度についての、対応する理想的なオブジェクト位置 1630b と、それらオブジェクトについての、対応する実際のレンダリング位置 1635b とを示す。実際のレンダリング位置 1635b は、点線 1640b で、理想的なオブジェクト位置 1630b に連結されている。実際のレンダリング位置 1635b の、固定された位置

$\vec{l}_j$

へと向かう斜めの向きは、コスト関数への最適解に対する、アトラクター重み付けのインパクトを示している。

【0299】

第2および第3の例では、「斥力」を用いて、ある位置から離れる方向にオーディオ音声を「押す」。ここで、ある位置とは、人の位置（例えば、聴者の位置、話者の位置など）であってもよいし、ランドマークの位置、家具の位置など、別の位置であってもよい。いくつかの例において、斥力を用いて、リスニング環境の或るエリアまたはゾーン（オフィスエリア、読書エリア、ベッドまたは寝室エリア（例えば、乳児用のベッドまたは寝室）など）から離れる方向にオーディオ音声を押してもよい。いくつかのそのような例によると、或る特定の位置を、ゾーンまたはエリアの代表として用いてもよい。例えば、乳児用ベッドを代表する位置は、乳児の頭の推定位置や、乳児に対応する推定される音声源の位置などであってもよい。この位置を、本明細書中、「斥力位置」または「斥位置」と呼ぶ場合もある。本明細書で用いる「斥力」は、斥力位置により近くなるほど、相対的により低いラウドスピーカ起動電位を好む要素である。この例によると、式（19）の引力と同様に、固定された斥位置

$\vec{l}_j$

に対して、 p_{ij} および τ_j を次のように定義する。

【数36】

$$p_{ij} = \max_i \|\vec{l}_j - \vec{s}_i\| - \|\vec{l}_j - \vec{s}_i\|$$

(19c)

【数37】

$$\tau_j = \max_i \|\vec{l}_j - \vec{s}_i\|$$

(19d)

【0300】

聴者または話者から離れる方向にオーディオ音声を押すことのユースケースを説明するために、具体的には、 $\alpha_j = 5$ 、 $\beta_j = 2$ 、

$\vec{l}_j$

を180度の聴者／話者の位置（プロットの底部中央）に対応するベクトルに設定する。

α_j 、 β_j および

$\vec{l}_j$

これらの値は、例示に過ぎない。上述のとおり、いくつかの例において、 α_j は 1 ～ 100 の範囲内であってもよく、 β_j は 1 ～ 25 の範囲内であってもよい。図 20 は、実施形態例における、スピーカ起動電位のグラフである。この例によると、図 20 は、スピーカ起動電位 1505 c、1510 c、1515 c、1520 c および 1525 c を示す。これらは、 w_{ij} で表される斥力を加えた、前の図面と同じスピーカ位置についてのコスト関数に対する最適解を含んでいる。図 21 は、実施形態例における、オブジェクトレンダリング位置のグラフである。この例において、図 21 は、多数の考え得るオブジェクト角度についての、理想的なオブジェクト位置 1630 c と、それらオブジェクトについての、対応する実際のレンダリング位置 1635 c とを示す。実際のレンダリング位置 1635 c は、点線 1640 c で、理想的なオブジェクト位置 1630 c に連結されている。実際のレンダリング位置 1635 c の、固定された位置

 $\vec{l}_j$

から離れる斜めの向きは、コスト関数への最適解に対する、リペラー (repeller) 重み付けのインパクトを示している。

【0301】

第3のユースケースの例は、就寝中の乳児の部屋へと続くドアなどの、音響的に要注意であるランドマークから離れる方向にオーディオ音声を「押す」ことである。1つ前の例と同様に、

 $\vec{l}_j$

を180度のドア位置（プロットの底部中央）に対応するベクトルに設定する。より強い斥力を達成し、かつ、音場を、主要なリスニング空間の前方部分内へと全体的に歪めるために、 $\alpha_j = 20$ 、 $\beta_j = 5$ と設定する。図 22 は、実施形態例における、スピーカ起動電位のグラフである。再び、この例において、図 22 は、スピーカ起動電位 1505 d、1510 d、1515 d、1520 d および 1525 d を示す。これらは、より強い斥力を加えた、同じ集合のスピーカ位置に対する最適解を含んでいる。図 23 は、実施形態例における、オブジェクトレンダリング位置のグラフである。再び、この例において、図 23 は、多数の考え得るオブジェクト角度についての、理想的なオブジェクト位置 1630 d と、それらオブジェクトについての、対応する実際のレンダリング位置 1635 d とを示す。実際のレンダリング位置 1635 d は、点線 1640 d で、理想的なオブジェクト位置 1630 d に連結されている。実際のレンダリング位置 1635 d の斜めの向きは、コスト関数への最適解に対する、より強いリペラー重み付けのインパクトを示している。

【0302】

図 2B の方法 250 のさらなる例において、ユースケースは、オーディオ環境における2つ以上のオーディオデバイスの選択（ブロック 265）と、「斥ける」力をオーディオ音声に適用すること（ブロック 275）とに依拠している。前の例によると、2つ以上のオーディオデバイスの選択は、いくつかの例において、値 f_{un} （オーディオ処理変更が起こる程度を制御する無単位のパラメータ）の形をとり得る。多くの組み合わせが可能である。1つの単純な例において、斥力に対応する重みは、

$$w_{ij} = \alpha_j f_i^{\beta_j}$$

として直接的に選択されてもよく、「決定」アспектにより選択されたデバイスを罰する。

【0303】

重みを決定する前述の例に対してさらに、いくつかの実施態様において、重みは次式のように決定されてもよい。

【数38】

$$w_{ij} = \alpha_j \left(f_i + \left(\frac{p_{ij}}{\tau_j} \right) \right)^{\beta_j}$$

10

【0304】

上記式において、 α_j 、 β_j 、 τ_j は、式(17)を参照して既に説明したように、それぞれが罰則の全体的な強度、罰則の開始の急激度、および罰則の範囲を示す、調節可能なパラメータである。したがって、上記式は、複数の罰則項の組み合わせとして理解されてもよく、これは、複数の同時のユースケースから起こるものである。例えば、オーディオ音声は、先行する例において説明した項 p_{ij} および項 τ_j を用いて、要注意であるランドマークから「押し離され」、また一方では、決定アスペクトにより決定された項 f_i を用いてSERを改善するのが望ましいマイクロフォン位置からも「押し離され」る。

【0305】

前の例はまた、発話対エコー比改善値(デシベル)で直接的に表現された s_{-n} を導入している。いくつかの実施形態は、部分的に、 s_{-n} の値(dB)に基づき、 α の値および β の値(それぞれ、罰則の強度および罰則の開始の急激度)を選択することを含んでもよく、 w_{ij} について上で示した式は、 α_j および β_j の代わりに、それぞれ、 α_{ij} および β_{ij} を用い得る。例えば、 $s_{-i} = -20$ dBの値は、 i 番目のスピーカを起動する高いコストに対応し得る。いくつかのそのような例において、 α_{ij} は、コスト関数の他の項 $C_{spatial}$ および $C_{proximity}$ の典型的な値よりも何倍も高い値に設定してもよい。例えば、 α の新たな値は、

20

$$\alpha'_{ij} = \alpha_j * 10^{-\frac{s_i}{20}}$$

30

により決定され得る。これは、 $s_{-i} = -20$ dBの値について、結果として、コスト関数における通常の値よりも10倍大きい値の α_{ij} になり得る。 β_{ij} を、 $0.5 < \beta_{ij} < 1.0$ の範囲内に設定されるように修正することは、いくつかの例において、 s_{-i} の大きい負の値に基づき、適切な修正であり得、 i 番目のスピーカの周辺のかかなり大きい領域から離れる方向にオーディオ音声を「押す」。例えば、 s_{-i} の値は、次式に応じて、 β_{ij} にマッピングされ得る。

【数39】

$$\beta_{ij} = \max \left(0.5, \min \left(1.0, 1.5 + \frac{s_i}{30.0} \right) \right)$$

40

この例において、 $s_{-i} = -20.0$ dBについて、 β_{ij} は0.8333であり得る。

【0306】

実施形態例の複数の側面は、以下の列挙実施形態例(EEE)を含む。

EEE1. ユーザからのボイスコマンドを検出するために信号対エコー比を改善する方法(またはシステム)であって、

a. 出力オーディオプログラム素材を生成するために、複数のデバイスが使用中である。

b. これらデバイスについて、聴者からの距離または順序付けられた関係の既知の集

50

合がある。

c. システムは、ユーザからの距離が最も短いデバイスの音量を選択的に下げる。

【0307】

EEE 2. EEE 1の方法またはシステムであって、信号の検出は、任意の雑音発生オブジェクトからの信号検出、または、1組のデバイスまでの距離の関係が既知であるオーディオ監視の所望の地点からの信号検出を含む、方法またはシステム。

【0308】

EEE 3. デバイスの順序付けを行うEEE 1またはEEE 2の方法またはシステムであって、距離と、名目上の音源距離 (nominal source distance) についてのデバイスの信号対エコー比とを考慮することを含む。

【0309】

EEE 4. EEE 1~EEE 3のいずれかの方法またはシステムであって、順序付けは、ユーザに対するデバイスの一般化された近接度およびこれのおおよその相互関係 (reciprocity) を考慮して、最も効果的な信号対エコー比改善値を推定し、この意味でデバイスの順序付けを行う。いくつかの開示された実施態様の側面は、1つ以上の開示された方法を実行するように構成された (例えば、プログラムされた) システムまたはデバイスと、1つ以上の開示された方法またはそのステップを実装するためのコードを格納した、有形のコンピュータ読取可能媒体 (例えばディスク) とを含む。例えば、システムは、1つ以上の開示された方法またはそのステップを含め、1つ以上の開示された方法またはそのステップを含むデータに対して各種演算のいずれかを実行するようにソフトウェアまたはファームウェアでプログラムされた、かつ／またはその他構成された、プログラム可能な汎用プロセッサ、デジタルシグナルプロセッサ、またはマイクロプロセッサであるか、もしくは、そのプロセッサを含み得る。そのような汎用プロセッサは、入力装置と、メモリと、アサートされたデータに応答して1つ以上の開示された方法 (またはそのステップ) を実行するようにプログラムされた (かつ／またはその他構成された) 処理用サブシステムとを含むコンピュータシステムであってもよいし、そのコンピュータシステムを含んでもよい。

【0310】

いくつかの開示された実施形態は、1つ以上の開示された方法の実行を含む、オーディオ信号 (単数または複数) に対して要求された処理を実行するように設定された (例えば、プログラムされたかその他設定された)、設定可能な (例えばプログラム可能な) デジタルシグナルプロセッサ (DSP) として実装される。あるいは、いくつかの実施形態 (またはその要素) は、1つ以上の開示された方法またはそのステップを含め、1つ以上の開示された方法またはそのステップを含む各種演算のいずれかを実行するようにソフトウェアまたはファームウェアでプログラムされた、かつ／またはその他構成された、汎用プロセッサ (例えば、入力装置およびメモリを含み得る、パーソナルコンピュータ (PC) もしくは他のコンピュータシステムまたはマイクロプロセッサ) として実装されてもよい。あるいは、いくつかの開示された実施形態の要素は、1つ以上の開示された方法またはそのステップを実行するように設定された (例えば、プログラムされた) 汎用プロセッサまたはDSPとして実装され、システムはまた、他の要素 (例えば、1つ以上のラウドスピーカおよび／または1つ以上のマイクロフォン) も含む。1つ以上の開示された方法またはそのステップを実行するように設定された汎用プロセッサは、典型的には、入力装置 (例えば、マウスおよび／またはキーボード) と、メモリと、表示装置とに接続され得る。

【0311】

いくつかの開示された実施態様の別の側面は、1つ以上の開示された方法の任意の実施形態またはそのステップを実行するためのコード (例えば、その実施形態を実行するように実行可能なコーダー) を格納した、コンピュータ読取可能媒体 (例えば、ディスクまたは他の有形の記憶媒体) である。

【0312】

本明細書中、具体的な実施形態および応用を説明したが、本明細書および特許請求の範

10

20

30

40

50

図に記載した内容の範囲から逸脱することなく、本明細書で説明した実施形態および応用に多くの改変を為し得ることは、当業者には明らかである。所与の実施態様を図示しかつ説明したが、本開示は、説明しかつ図示した具体的な実施形態もしくは説明した具体的な方法に限定されないことを理解されたい。

【図面】

【図 1 A】

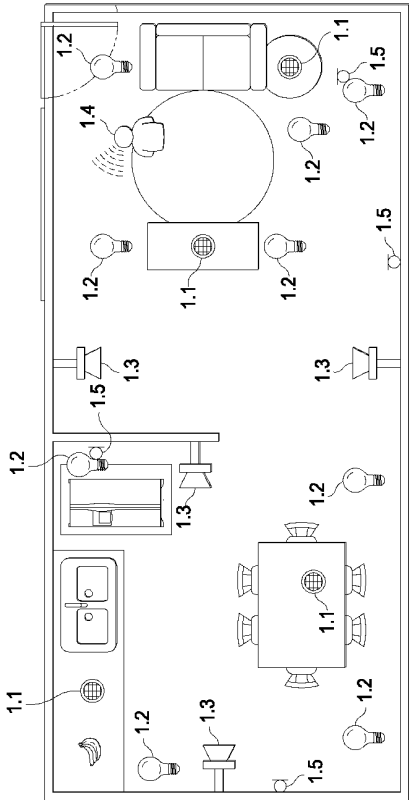


Figure 1A

【図 1 B】

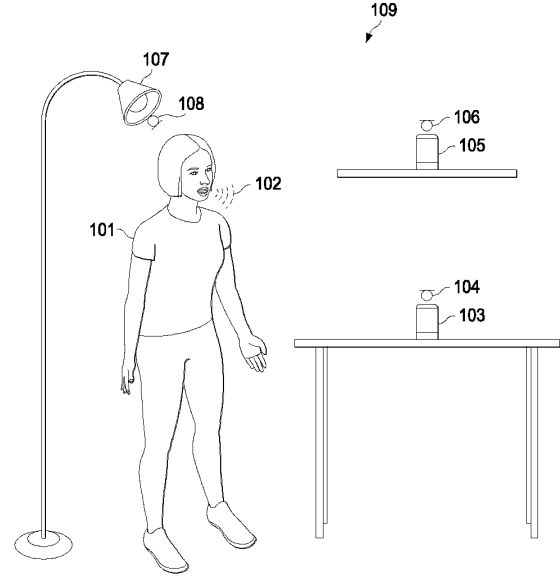


Figure 1B

10

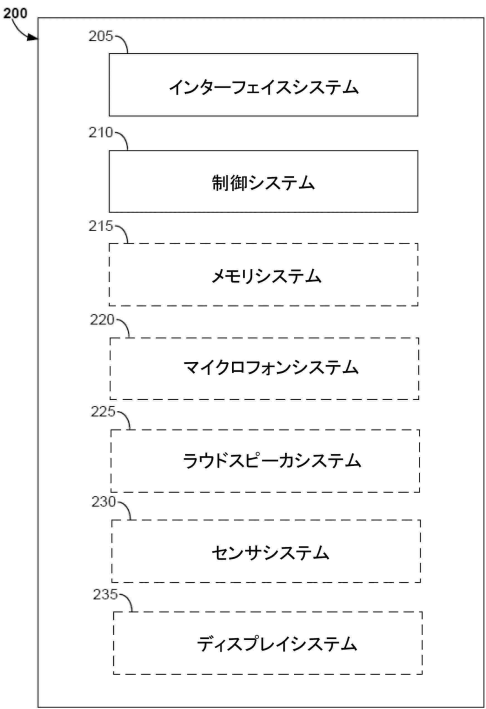
20

30

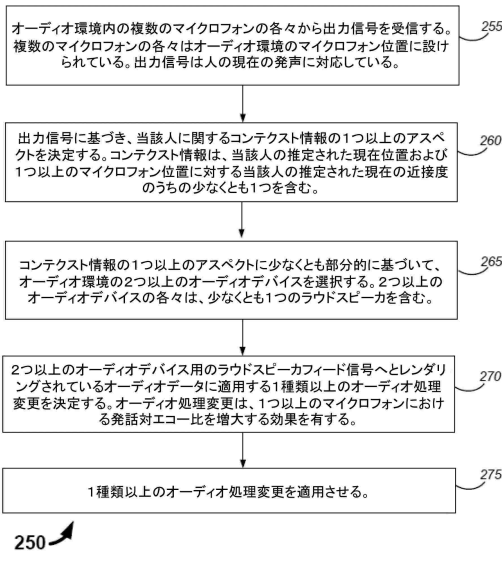
40

50

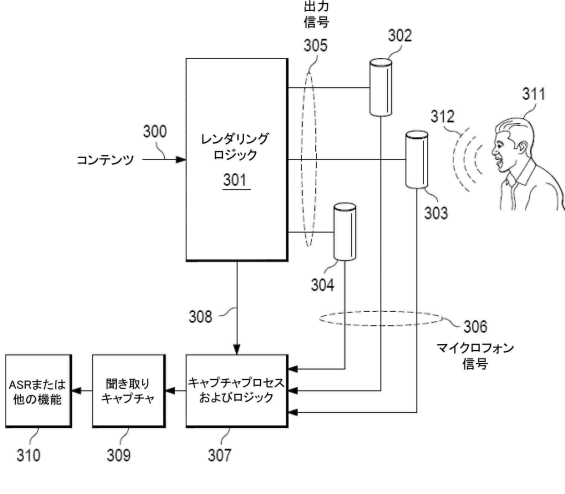
【図 2 A】



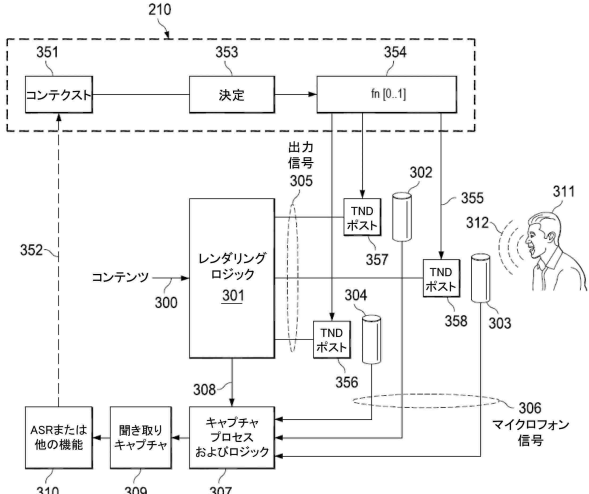
【図 2 B】



【図 3 A】



【図 3 B】



10

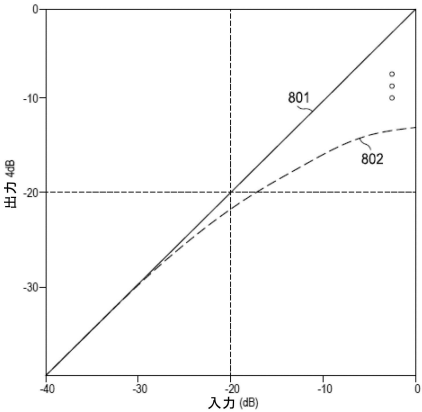
20

30

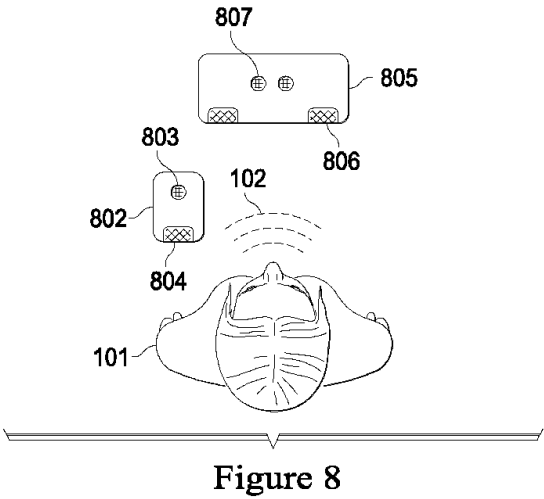
40

50

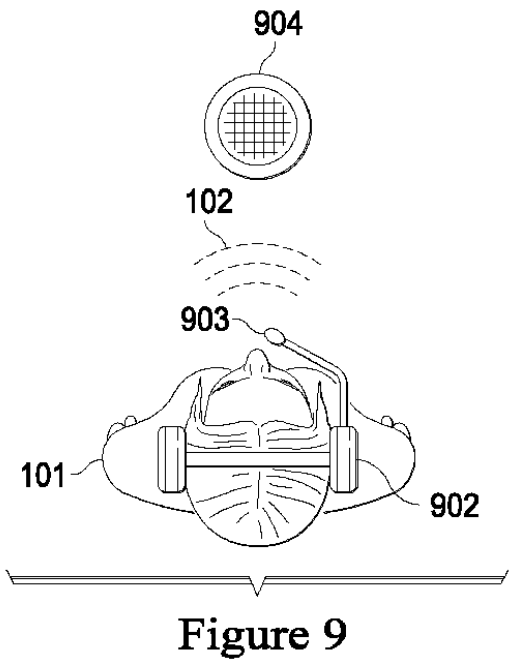
【図 7】



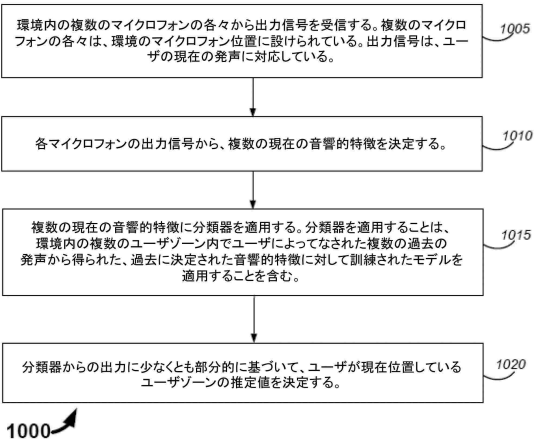
【図 8】



【図 9】



【図 10】



10

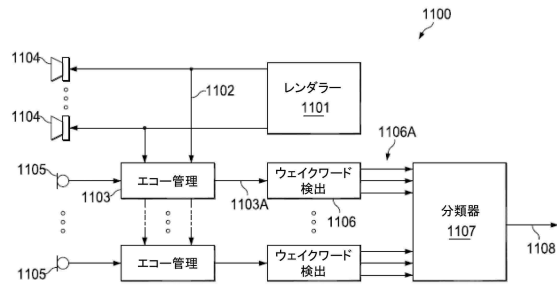
20

30

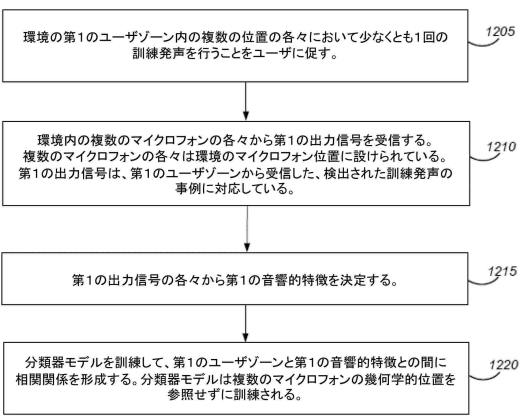
40

50

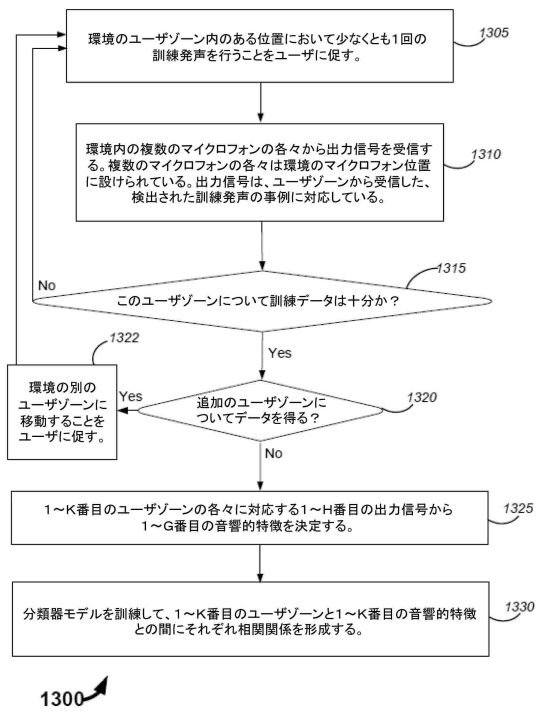
【図 1 1】



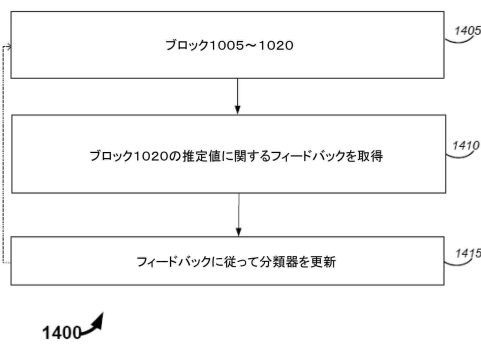
【図 1 2】



【図 1 3】



【図 1 4】



10

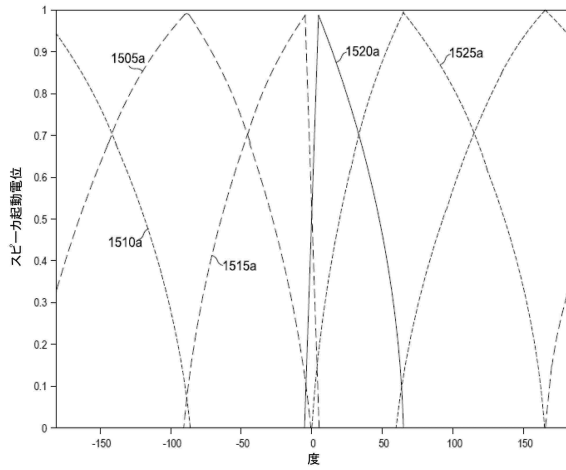
20

30

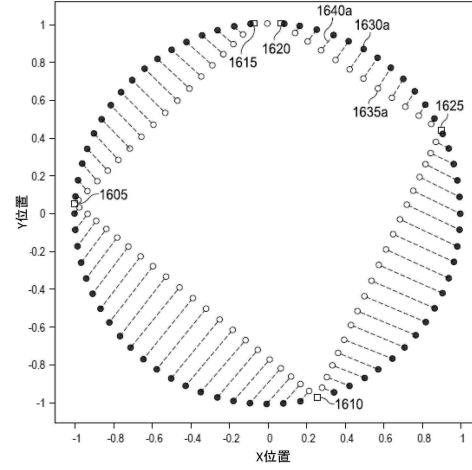
40

50

【図 15】



【図 16】



【図 17】

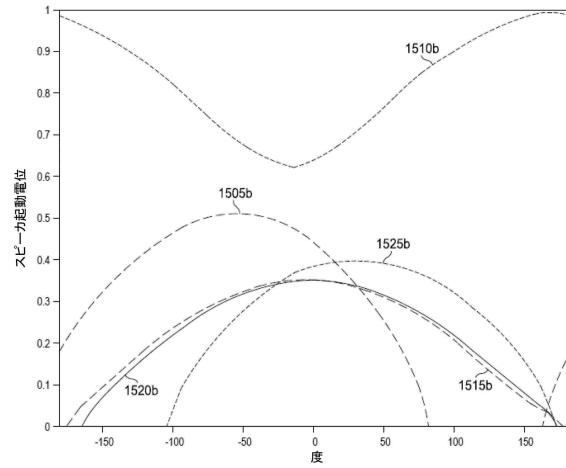
制御システムにより、インターフェイスシステムを介して、オーディオデータを受信する。オーディオデータは、1つ以上のオーディオ信号および関連付けられた空間データを含む。空間データは、オーディオ信号に対応する、目標の知覚された空間位置を示す。

制御システムにより、環境の1組のラウドスピーカを介した再生のためのオーディオデータをレンダリングして、レンダリングされたオーディオ信号を生成する。オーディオデータに含まれる1つ以上のオーディオ信号の各々をレンダリングすることは、コストを最適化することにより、環境内の1組のラウドスピーカの相対的な起動を決定することを含む。このコストは、環境内の1組のラウドスピーカ上で再生される場合の再生されたオーディオ信号の知覚された空間位置のモデル、1組のラウドスピーカのうちの各ラウドスピーカの位置に対するオーディオ信号の目標の知覚された空間位置の近接度の尺度、および、1つ以上の追加の動的に設定可能な関数、の関数である。1つ以上の追加の動的に設定可能な関数は、以下のうちの1つ以上に基づく。1人以上の聴者に対するラウドスピーカの近接度；引力位置に対するラウドスピーカの近接度。ここで、引力は、引力位置により近接した、相対的に高いラウドスピーカ起動電位を好む要素である。；斥力位置に対するラウドスピーカの近接度。ここで、斥力は、斥力位置により近接した、相対的に低いラウドスピーカ起動電位を好む要素である。；環境内の他のラウドスピーカに対する、各ラウドスピーカの能力；ラウドスピーカの、他のラウドスピーカとの同期；ウェイクワード性能；または、エコー除去性能。

レンダリングされたオーディオ信号を、環境の1組のラウドスピーカのうちの少なくともいくつかのラウドスピーカに、インターフェイスシステムを介して提供する。

1700

【図 18】



10

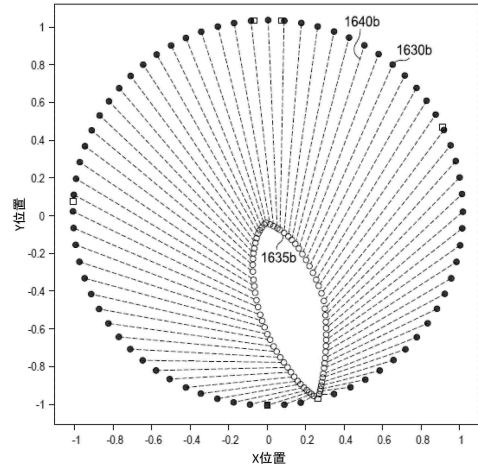
20

30

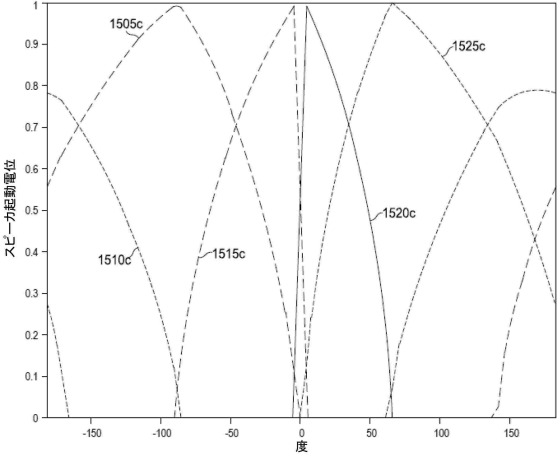
40

50

【図 1 9】

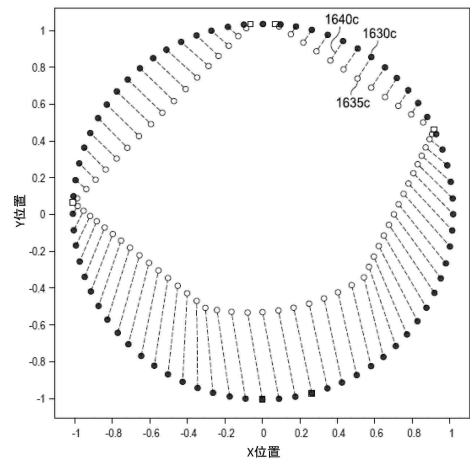


【図 2 0】

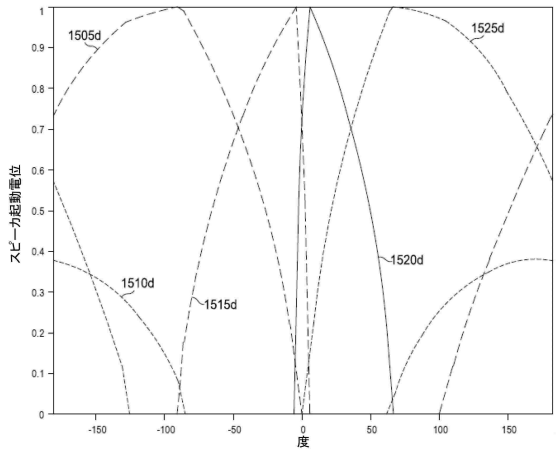


10

【図 2 1】



【図 2 2】



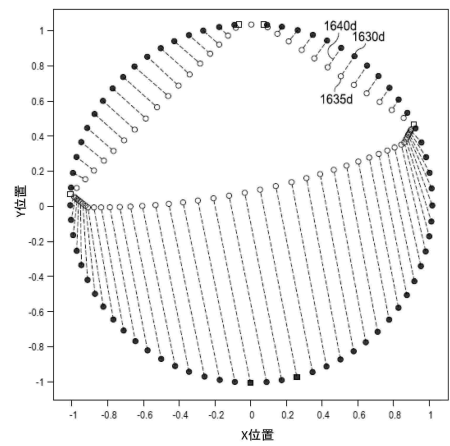
20

30

40

50

【図 2 3】



10

20

30

40

50

フロントページの続き

(33)優先権主張国・地域又は機関

米国(US)

(31)優先権主張番号 62/880,122

(32)優先日 令和1年7月30日(2019.7.30)

(33)優先権主張国・地域又は機関

米国(US)

(31)優先権主張番号 19212391.7

(32)優先日 令和1年11月29日(2019.11.29)

(33)優先権主張国・地域又は機関

欧州特許庁(EP)

(31)優先権主張番号 62/950,004

(32)優先日 令和1年12月18日(2019.12.18)

(33)優先権主張国・地域又は機関

米国(US)

(31)優先権主張番号 62/971,421

(32)優先日 令和2年2月7日(2020.2.7)

(33)優先権主張国・地域又は機関

米国(US)

(31)優先権主張番号 62/705,410

(32)優先日 令和2年6月25日(2020.6.25)

(33)優先権主張国・地域又は機関

米国(US)

(31)優先権主張番号 62/705,897

(32)優先日 令和2年7月21日(2020.7.21)

(33)優先権主張国・地域又は機関

米国(US)

(74)代理人 100155000

弁理士 喜多 修市

(74)代理人 100188813

弁理士 川喜田 徹

(74)代理人 100202197

弁理士 村瀬 成康

(72)発明者 ディキンズ, グレン エヌ.

オーストラリア 2 2 2 6 ニューサウスウェールズ州, コモ, ダブリーズ ロード 1 6

(72)発明者 ハインズ, クリストファー グラハム

アメリカ合衆国 9 4 1 0 3 カリフォルニア州 サンフランシスコ マーケット ストリート 1 2 7 5, ドルビー ラボラトリーズ インコーポレイテッド内

(72)発明者 グナワン, デイビッド

アメリカ合衆国 9 4 1 0 3 カリフォルニア州 サンフランシスコ マーケット ストリート 1 2 7 5, ドルビー ラボラトリーズ インコーポレイテッド内

(72)発明者 カートライト, リチャード ジェイ.

アメリカ合衆国 9 4 1 0 3 カリフォルニア州 サンフランシスコ マーケット ストリート 1 2 7 5, ドルビー ラボラトリーズ インコーポレイテッド内

(72)発明者 シーフェルド, アラン ジェイ.

アメリカ合衆国 9 4 1 0 3 カリフォルニア州 サンフランシスコ マーケット ストリート 1 2 7 5, ドルビー ラボラトリーズ インコーポレイテッド内

(72)発明者 アルテアガ, ダニエル

アメリカ合衆国 9 4 1 0 3 カリフォルニア州 サンフランシスコ マーケット ストリート 1 2 7 5, ドルビー ラボラトリーズ インコーポレイテッド内

-
- (72)発明者 トーマス，マーク アール．ピー．
 アメリカ合衆国 9 4 1 0 3 カリフォルニア州 サンフランシスコ マーケット ストリート 1 2
 7 5，ドルビー ラボラトリーズ インコーポレイテッド内
- (72)発明者 ランド，ジョシュア ビー．
 アメリカ合衆国 9 4 1 0 3 カリフォルニア州 サンフランシスコ マーケット ストリート 1 2
 7 5，ドルビー ラボラトリーズ インコーポレイテッド内
- 審査官 ▲徳▼田 賢二
- (56)参考文献 特開 2 0 0 7 - 1 4 2 5 9 5 (J P , A)
 特開 2 0 1 1 - 2 2 8 8 9 9 (J P , A)
 国際公開第 2 0 0 9 / 0 0 4 9 8 3 (W O , A 1)
- (58)調査した分野 (Int.Cl., D B 名)
 H 0 4 R 3 / 0 2