

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7453248号
(P7453248)

(45)発行日 令和6年3月19日(2024.3.19)

(24)登録日 令和6年3月11日(2024.3.11)

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

請求項の数 20 (全39頁)

(21)出願番号	特願2021-556235(P2021-556235)	(73)特許権者	590000248
(86)(22)出願日	令和2年3月16日(2020.3.16)		コーニンクレッカ フィリップス エヌ
(65)公表番号	特表2022-525902(P2022-525902 A)		ヴェ
(43)公表日	令和4年5月20日(2022.5.20)		Koninklijke Philips
(86)国際出願番号	PCT/EP2020/057014		N.V.
(87)国際公開番号	WO2020/187807		オランダ国 5 6 5 6 アーヘー アイン
(87)国際公開日	令和2年9月24日(2020.9.24)		ドーフエン ハイテック キャンパス 5 2
審査請求日	令和5年3月15日(2023.3.15)		High Tech Campus 5 2 ,
(31)優先権主張番号	19163714.9		5 6 5 6 AG Eindhoven, N
(32)優先日	平成31年3月19日(2019.3.19)	(74)代理人	etherlands
(33)優先権主張国・地域又は機関	欧州特許庁(EP)		110001690
早期審査対象出願		(72)発明者	弁理士法人M&Sパートナーズ
			コッペンズ ジェロエン ジェラルドゥス
			ヘンリクス
			オランダ国 5 6 5 6 アーエー アイン
			最終頁に続く

(54)【発明の名称】 オーディオ装置およびその処理の方法

(57)【特許請求の範囲】

【請求項1】

受信機回路と、生成器回路と、レンダラ回路とを含む、オーディオ装置であって、

前記受信機回路は、信号を受信し、

前記信号は、オーディオシーンのためのデータを備え、

前記データは、入力オーディオソースデータと、音響オブジェクトデータとを備え、

前記入力オーディオソースデータは、少なくとも1つのオーディオソースのためであり、

前記音響オブジェクトデータは、少なくとも1つの音響オブジェクトのためであり、

前記音響オブジェクトデータは、音響カップリングデータおよび空間特性データを備え、

前記音響カップリングデータは、オーディオエネルギーの振動/機械的エネルギーへの変換

、および/または、振動/機械的エネルギーのオーディオエネルギーへの変換の特性を示し、

前記生成器回路は、オブジェクトオーディオソースのためのオブジェクトオーディオソ

ースデータを生成し、

前記オブジェクトオーディオソースデータは、前記音響オブジェクトによって前記オー

ディオシーンにおいて発せられるオーディオを表現し、

前記オーディオシーンにおいて発せられる前記オーディオは、前記少なくとも1つのオー

ディオソースからカップリングされ、

前記生成器回路は、前記音響カップリングデータ、前記空間特性データ、および前記入

力オーディオソースデータにตอบสนองして前記オブジェクトオーディオソースデータを生成し、

前記レンダラ回路は、前記オーディオシーンをレンダリングし、前記レンダリングするこ

10

20

とは、前記オブジェクトオーディオソースデータをレンダリングすることを含む、オーディオ装置。

【請求項 2】

前記生成器回路は、前記オブジェクトオーディオソースが前記空間特性データにตอบสนองして決定される空間的広がりを有するように前記オブジェクトオーディオソースを表現するために、前記オブジェクトオーディオソースデータを生成する、請求項 1 に記載のオーディオ装置。

【請求項 3】

前記生成器回路は、前記音響カップリングデータにตอบสนองして、前記オブジェクトオーディオソースデータのオーディオのためのオーディオレベルおよび周波数応答のうちの少なくとも 1 つを決定する、請求項 1 に記載のオーディオ装置。

10

【請求項 4】

前記音響カップリングデータは、前記音響オブジェクトのためのカップリング係数の第 1 のセットを備え、
前記カップリング係数のセットは、前記音響オブジェクトのためのカップリング伝達関数を示す、請求項 1 に記載のオーディオ装置。

【請求項 5】

前記カップリング伝達関数は、前記音響オブジェクトの振動から、前記音響オブジェクトから発せられるオーディオへの伝達関数である、請求項 4 に記載のオーディオ装置。

【請求項 6】

前記オブジェクトオーディオソースデータは、オーディオオブジェクトの振動および前記カップリング伝達関数から、前記音響オブジェクトによって前記オーディオシーンにおいて発せられるオーディオを表現する、請求項 5 に記載のオーディオ装置。

20

【請求項 7】

前記生成器回路は、前記音響オブジェクトの物理モデルにตอบสนองして前記振動を判定する、請求項 6 に記載のオーディオ装置。

【請求項 8】

前記生成器回路は、時間変化する力にตอบสนองして前記振動を判定し、前記時間変化する力は、前記音響オブジェクトに加えられる、請求項 6 に記載のオーディオ装置。

【請求項 9】

前記カップリング伝達関数は、前記音響オブジェクトに入射するオーディオから、前記音響オブジェクトの振動への伝達関数である、請求項 4 に記載のオーディオ装置。

30

【請求項 10】

前記生成器回路は、前記空間特性データによって示されるような前記音響オブジェクトの位置に対応する空間位置を有する前記オブジェクトオーディオソースを表現するために、前記オブジェクトオーディオソースデータを生成する、請求項 1 に記載のオーディオ装置。

【請求項 11】

前記生成器回路は、前記空間特性データによって示される前記少なくとも 1 つのオーディオソースの位置および前記音響オブジェクトの位置にตอบสนองして、前記オブジェクトオーディオソースデータを生成する、請求項 1 に記載のオーディオ装置。

40

【請求項 12】

前記オブジェクトオーディオソースデータは、前記空間特性データによって示されるリスニング位置および前記音響オブジェクトの位置にตอบสนองして生成される、請求項 1 に記載のオーディオ装置。

【請求項 13】

前記信号は、第 2 のオーディオソースのための入力オーディオソースデータを備え、
前記オブジェクトオーディオソースデータは、前記少なくとも 1 つのオーディオソースと前記第 2 のオーディオソースとからのオーディオの結合カップリングから、前記音響オブジェクトによって前記オーディオシーンにおいて発せられるオーディオを表現する、請

50

求項 1 に記載のオーディオ装置。

【請求項 1 4】

データ生成器回路と、信号生成器回路と、送信機回路とを備える、オーディオ装置であつて、

前記データ生成器回路は、オーディオシーンのためのデータを生成し、
前記データは、入力オーディオソースデータと、音響オブジェクトデータとを備え、
前記入力オーディオソースデータは、少なくとも1つのオーディオソースのためであり、
前記音響オブジェクトデータは、少なくとも1つの音響オブジェクトのためであり、
前記音響オブジェクトデータは、音響カップリングデータおよび空間特性データを備え、
前記音響カップリングデータは、オーディオエネルギーの振動/機械的エネルギーへの変換、
および/または、振動/機械的エネルギーのオーディオエネルギーへの変換の特性を示し、
前記データ生成器回路は、データ信号を生成し、
前記データ信号は、前記オーディオシーンのための前記データを備え、
前記データ生成器回路は、前記データ信号を送信する、オーディオ装置。

10

【請求項 1 5】

オーディオを処理する方法であつて、前記方法は、
信号を受信するステップと、オブジェクトオーディオソースのためのオブジェクトオーディオソースデータを生成するステップと、オーディオシーンをレンダリングするステップと

を有し、
前記信号は、オーディオシーンのためのデータを備え、
前記データは、入力オーディオソースデータと、音響オブジェクトデータとを備え、
前記入力オーディオソースデータは、少なくとも1つのオーディオソースのためであり、
前記音響オブジェクトデータは、前記オーディオシーンにおける少なくとも1つの音響オブジェクトのためであり、
前記音響オブジェクトデータは、音響カップリングデータおよび空間特性データを備え、
前記音響カップリングデータは、オーディオエネルギーの振動/機械的エネルギーへの変換、
および/または、振動/機械的エネルギーのオーディオエネルギーへの変換の特性を示し、
前記オブジェクトオーディオソースデータは、前記音響オブジェクトによって前記オーディオシーンにおいて発せられるオーディオを表現し、
前記オーディオシーンにおいて発せられる前記オーディオは、前記少なくとも1つのオーディオソースからカップリングされ、
生成器回路は、前記音響カップリングデータ、前記空間特性データ、および前記入力オーディオソースデータにตอบสนองして、前記オブジェクトオーディオソースデータを生成し、
前記レンダリングするステップは、前記オブジェクトオーディオソースデータをレンダリングするステップを含む、方法。

20

30

【請求項 1 6】

オーディオデータ信号を生成する方法であつて、前記方法は、
オーディオシーンのためのデータを生成するステップであつて、
前記データは、入力オーディオソースデータと、音響オブジェクトデータとを備え、
前記入力オーディオソースデータは、少なくとも1つのオーディオソースのためであり、
前記音響オブジェクトデータは、少なくとも1つの音響オブジェクトのためであり、
前記音響オブジェクトデータは、音響カップリングデータおよび空間特性データを備え、
前記音響カップリングデータは、オーディオエネルギーの振動/機械的エネルギーへの変換、
および/または、振動/機械的エネルギーのオーディオエネルギーへの変換の特性を示す、
データを生成するステップと、
前記オーディオデータ信号を生成するステップであつて、前記オーディオデータ信号は、
前記オーディオシーンのための前記データを備える、オーディオデータ信号を生成するステップと、

40

前記オーディオデータ信号を送信するステップとを有する、方法。

50

【請求項 17】

コンピュータプログラムがプロセッサにおいて実行された場合、請求項 15 に記載の方法を実行する、非一時的媒体に記憶されたコンピュータプログラム。

【請求項 18】

コンピュータプログラムがプロセッサにおいて実行された場合、請求項 16 に記載の方法を実行する、非一時的媒体に記憶されたコンピュータプログラム。

【請求項 19】

前記生成器回路は、前記音響カップリングデータにตอบสนองして、前記オブジェクトオーディオソースデータのオーディオのためのオーディオレベルおよび周波数応答のうち少なくとも 1 つを決定する、請求項 2 に記載のオーディオ装置。

10

【請求項 20】

前記生成器回路は、時間変化する力にตอบสนองして前記振動を判定し、前記時間変化する力は、前記音響オブジェクトに加えられる、請求項 7 に記載のオーディオ装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、オーディオ装置およびその処理の方法、特に、拡張 / 仮想現実アプリケーションのためのオーディオ処理に関するが、これに限定されない。

【背景技術】

【0002】

オーディオビジュアルコンテンツに基づく体感の多様性と範囲は、そのようなコンテンツを継続的に開発および導入する新しいサービスとその利用および消費の手法により、近年大幅に増大している。特に、多くの空間的でインタラクティブなサービス、アプリケーション、および体感が開発されており、ユーザに、より複雑な没入型の体感を提供できるようになっている。

20

【0003】

そのようなアプリケーションの例としては、仮想現実 (VR)、拡張現実 (AR)、および複合現実 (MR) アプリケーションがあり、これらは急速に主流になりつつあり、多くのソリューションが消費者市場に向けられている。多くの規格化団体によって、いくつかの規格化も進められている。そのような規格化活動は、たとえば、ストリーミング、ブロードキャスト、レンダリングなどを含む VR / AR / MR システムのさまざまな態様の規格化を積極的に進めている。

30

【0004】

VR アプリケーションは、異なる世界 / 環境 / シーンにいるユーザに対応するユーザ体感を提供する傾向がある一方、AR (複合現実 MR を含む) アプリケーションは、現在の環境にあるユーザに対応するユーザ体感を提供する傾向があるが、さらなる情報、または、仮想的なオブジェクトまたは情報が追加されている。したがって、VR アプリケーションは、完全に没入型の合成的に生成された世界 / シーンを提供する傾向があるが、AR アプリケーションは、ユーザが物理的に存在する実際のシーンにオーバーレイされる部分的な合成世界 / シーンを提供する傾向がある。しかしながら、これらの用語はしばしば置換可能に使用され、重複率が高い。以下では、仮想現実 / VR という用語は、仮想現実と拡張現実との両方を表すために使用される。

40

【0005】

例として、ますます人気が高まっているサービスは、ユーザがシステムと積極的かつ動的にインタラクトしてレンダリングのパラメータを変更し、ユーザの位置および向きにおける動きや変化に適応できるようにする手法で画像とオーディオとを提供することである。多くのアプリケーションにおいて非常に魅力的な特徴は、たとえば、視覚者が、提示されているシーン内を移動して「見回す」ことができるようにするなど、視覚者の有効な視覚位置および視覚方向を変える機能である。

【0006】

50

そのような特徴により、特に、仮想現実体感をユーザに提供できるようになる。これにより、ユーザは、仮想環境内を（比較的）自由に動き回り、自分の位置と、見ている場所とを動的に変えることができる。通常、そのような仮想現実アプリケーションは、シーンの3次元モデルに基づいており、モデルが、動的に評価され、特定の要求されたビューを提供する。このアプローチは、たとえば、1人で行うシューティングゲームのカテゴリなど、コンピュータやコンソール向けのゲームアプリケーションからよく知られている。

【0007】

また、特に仮想現実アプリケーションでは、提示される画像は、三次元画像であることが望ましい。実際、視覚者の没入感を最適化するために、通常、ユーザは、提示されたシーンを三次元シーンとして体感することが好ましい。実際、仮想現実体感では、ユーザが自分の位置、カメラの視点、および仮想世界に対する瞬間を選択できることが望ましいはずである。

10

【0008】

視覚的なレンダリングに加えて、ほとんどのVR/ARアプリケーションは、対応するオーディオ体感をさらに提供する。多くのアプリケーションでは、オーディオは、好ましくは、オーディオソースが、ビジュアルシーン内の対応するオブジェクトの位置に対応する位置から到着するように知覚される空間オーディオ体感を提供する。したがって、オーディオシーンとビデオシーンは、一貫性があり、両方とも完全な空間的な体感を提供するものとして認識されることが好ましい。

【0009】

たとえば、多くの没入型体感は、バイノーラルオーディオレンダリング技術を使用したヘッドフォン再生によって生成される仮想オーディオシーンによって提供される。多くのシナリオでは、そのようなヘッドフォン再生は、ユーザの頭の動きに反応してレンダリングできるように、ヘッドトラッキングに基づく場合があり、これにより、没入感が大幅に向上する。

20

【0010】

しかしながら、没入感が高く、パーソナライズされた自然な体感をユーザに提供するには、オーディオシーンのレンダリングが可能な限り現実的であることが重要であり、多くのVR体感など、オーディオビジュアル体感を組み合わせた場合、オーディオ体感がビジュアル体感と厳密に一致していること、すなわち、レンダリングされたオーディオシーンとビデオシーンとが厳密に一致していることが重要である。

30

【0011】

多くのアプリケーションでは、シーンを表現するオーディオビジュアルデータは、1つの、多くの場合は、中央のデバイスによって生成され、受信したオーディオビジュアルデータを処理して特定のユーザにローカルレンダリングを提供する個々のリモートエンティティに配信される。通常、これは、たとえば、シーン内のユーザの現在の動き、位置、および向き、または他のローカライズされたパラメータに適応し得る。したがって、多くのアプリケーションでは、受信したオーディオデータに基づいてオーディオシーンのローカルレンダリングと合成が実行される。

【0012】

そのようなアプリケーションをサポートするために、オーディオ表現及びデータのために多くのアプローチと、オーディオ規格化とが進められてきた。そのようなアプローチおよび規格では、さまざまなオーディオ構成要素が、空間情報とともに個々に表現され得る。オーディオシーンにおいて、オーディオ構成要素とソースを表現するためのさまざまなアプローチが、さまざまな規格およびアプローチにおいて使用され得る。

40

【0013】

たとえば、開発中のMPEG-Iパート4(Moving Picture Experts Group Immersive Audio Coding)規格では、オーディオソースが送信され、ユーザが6つの自由度(6DoF)のオーディオシーンをレンダリングするために使用される。これらの音ソースは、オブジェクト、チャンネル、および高次

50

アンビソニックス (H O A) になる。

【 0 0 1 4 】

オーディオオブジェクトは、音ソースの表現であるため、通常は音ソースの位置に関連付けられる。対照的に、オーディオチャンネルはスピーカ信号の表現である。通常、2つ以上のオーディオチャンネルは互いに関連付けられ、固定位置においてレンダリングされる。それらは通常、1つまたは複数の音ソースをとともに表現し、振幅パニングが2つ以上のオーディオチャンネルの位置の間に知覚される局在化を引き起こすように、音ソースは2つ以上のオーディオチャンネルによって表現される。H O A の場合、オーディオ表現はマイクロフォンの原理に基づいており、各 H O A 信号は、特定の指向特性を有する (仮想的な) マイクロフォンを表現する。指向特性は、球面調和関数に基づいているため、H O A 表現のさまざまな信号を組み合わせ、特定の方向から受信した音波に対応するオーディオ信号を得ることができる。したがって、オーディオオブジェクトとオーディオチャンネルがオーディオ放射を表現する場合、H O A は、空間の特定のポイントにおいて受信されるオーディオを表現する。

10

【 0 0 1 5 】

オーディオオブジェクトの場合、広がり概念を使用して、音ソースの (3 D) サイズを示す。通常、オブジェクトは、ポイントソースとしてレンダリングされる単一の信号として表現され、たとえば、それを単一の H R T F ペアと畳み込み、バイノーラル処理を実行する。広がりパラメータ / 特徴により、レンダリングは、ユーザに知覚される幅および高さを導入できる。

20

【 0 0 1 6 】

H O A は、ほとんどのビットストリームにおける音響音および拡散音を含む可能性があるが、ドライ音ソース成分を含む場合もある。また、チャンネルはドライソースであるか、あるいは音響を含むこともできる (ドライ音ソース / 信号は、処理されていないものを指す場合がある。すなわち、時間間隔は、生の未処理の元の音に対応する場合があり、たとえば、https://en.wikipedia.org/wiki/Audio_signal_processing を参照) 。オブジェクトソースと一部のチャンネルの場合、M P E G - I デコーダは、ユーザの位置に応じて音響をシミュレートするための音響モデルを含む必要がある。

30

【 0 0 1 7 】

しかしながら、そのようなアプローチは、多くのアプリケーションのために、多くのシナリオにおいて、効率的な性能を提供する可能性があるが、すべての状況およびシナリオにおいて最適であるとは限らない。一部のアプリケーションでは、最適ではない品質を体感する場合があります、オーディオシーンが、完全に自然にまたは現実的ではないと認識される場合があります。状況によっては、オーディオは、提供されたビデオシーンに完全に対応していない場合や、対応する自然環境で知覚されるものに完全に対応していない場合があります。

【 発明の概要 】

【 発明が解決しようとする課題 】

【 0 0 1 8 】

したがって、オーディオ処理、特に仮想 / 拡張 / 複合現実体感 / アプリケーション、アプリケーションのために改善されたアプローチが有利である。特に、操作性の改善、柔軟性の向上、複雑さの軽減、実施の容易化、オーディオ体感の向上、オーディオおよびビジュアルシーンのより一貫した知覚、カスタマイズ性の向上、パーソナライズ化の向上を可能にするアプローチと、改善された仮想現実体感と、改善されたオーディオ品質と、および / または、改善された性能および / または操作が有利となるであろう。

40

【 0 0 1 9 】

したがって、本発明は、好ましくは、上記の欠点の1つまたは複数、単独または任意の組合せで軽減、低減、または排除することを目指す。

【 課題を解決するための手段 】

【 0 0 2 0 】

50

本発明の態様によれば、オーディオシーンのためのデータを備える信号を受信するための受信機であって、データは、オーディオシーンにおける少なくとも第1のオーディオソースのための入力オーディオソースデータと、少なくとも1つの音響オブジェクトのための音響オブジェクトデータとを備え、音響オブジェクトデータは、音響オブジェクトのための音響カップリングデータおよび空間特性データを備える、受信機と、第1のオーディオソースからのオーディオのカップリングから音響オブジェクトによってオーディオシーンにおいて発せられるオーディオを表現するオブジェクトオーディオソースのためのオブジェクトオーディオソースデータを生成するための生成器であって、音響カップリングデータ、空間特性データ、および入力オーディオソースデータに応答してオブジェクトオーディオソースデータを生成するように構成される、生成器と、オーディオシーンをレンダリングするためのレンダラとを備え、レンダリングすることは、オブジェクトオーディオソースデータをレンダリングすることを含む、オーディオ装置が提供される。

10

【0021】

本発明は、多くの実施形態において、改善されたオーディオレンダリングを提供することができ、特に、オーディオシーンの、改善された表現およびレンダリングを提供することができる。多くのシナリオでは、シーンのより現実的な表現が実現され、改善された、および/または、より自然なユーザ体感が実現され得る。多くの実施形態において、効果的な処理が達成され得る。このアプローチは、他の多くのオーディオ処理およびレンダリングアプローチと互換性があり、ユーザに提示されるオーディオに追加の補完的な寄与を提供し得る。

20

【0022】

カップリングは、オーディオエネルギーの、振動/機械的エネルギーへの変換、および/または、振動/機械的エネルギーのオーディオエネルギーへの変換を反映し得、音響オブジェクトのための音響カップリングデータは、音響オブジェクトの、オーディオエネルギーの振動/機械的エネルギーへの変換、および/または、振動/機械的エネルギーのオーディオエネルギーへの変換の特性を示す場合がある。音響カップリングデータは、音響オブジェクトの材料および/または構成および/または組成を示すデータを含む、音響オブジェクトの機械的特性を示すデータを含み得る。

【0023】

入力オーディオソースデータは、特に、第1のオーディオソースのオーディオ信号データおよび空間特性データを備え得る。オブジェクトオーディオソースデータは、オブジェクトオーディオソースのためのオーディオ信号データおよび空間特性データを備え得る。オーディオソースのための空間データは、オーディオソースの位置および/または広がりを示し得る。

30

【0024】

生成器は、オーディオ構成要素の空間特性データによって示される空間拡張および/または位置に対応する空間拡張および/または位置を有するオーディオソースに対応するオブジェクトオーディオソースデータを生成するように構成される。生成器は、特に第1のオーディオソースの入力オーディオ信号を、音響カップリングデータに依存する伝達関数とフィルタリングすることによってなど、音響カップリングデータ(および入力オーディオ信号データ)から決定されるレベル/周波数応答に対応するオブジェクトオーディオソースデータを生成するように構成される。

40

【0025】

本発明の任意選択の特徴によれば、生成器は、空間特性データに依りて決定される空間的広がりを有するオブジェクトオーディオソースを表現するために、オブジェクトオーディオソースデータを生成するように構成される。

【0026】

これにより、オーディオシーンのレンダリングが改善され、オーディオシーンのより現実的な知覚になることがよくある。特に、生成器は、空間特性データによって示されるように、音響オブジェクトの広がりと同じ空間的広がりを有するオブジェクトオーディオソ

50

ースを表現するために、オブジェクトオーディオソースデータを生成し得る。

【0027】

本発明の任意選択の特徴によれば、生成器は、音響カップリングデータに応答して、オブジェクトオーディオソースデータのオーディオのためのオーディオレベルおよび周波数応答のうちの少なくとも1つを決定するように構成される。

【0028】

これにより、オーディオシーンのレンダリングが改善され、オーディオシーンのより現実的な知覚になることがよくある。

【0029】

本発明の任意選択の特徴によれば、音響カップリングデータは、音響オブジェクトのカップリング係数の第1のセットを備え、カップリング係数のセットは、音響オブジェクトのためのカップリング伝達関数を示す。

10

【0030】

これは、カップリング効果の特に効率的で有利な表現を提供し得る。カップリング係数のセットは、カップリング伝達関数に対応するフィルタを記載し得る。いくつかの実施形態では、カップリング係数のセットは、単一の係数のみを備え得る。

【0031】

本発明の任意選択の特徴によれば、カップリング伝達関数は、音響オブジェクトの振動から、音響オブジェクトから発せられるオーディオへの伝達関数である。

【0032】

これは、多くの実施形態において、カップリング効果の特に有利で柔軟な表現を提供し得る。これは、高度な柔軟性を提供し、さまざまな効果の表現およびレンダリングを可能とし得る。

20

【0033】

本発明の任意選択の特徴によれば、生成器は、オーディオオブジェクトの振動およびカップリング伝達関数から、音響オブジェクトによってオーディオシーンにおいて発せられるオーディオを表現するオブジェクトオーディオソースデータを生成するように構成される。

【0034】

これは、多くの実施形態において、カップリング効果の特に有利で柔軟な表現を提供し得る。

30

【0035】

本発明の任意選択の特徴によれば、生成器は、音響オブジェクトの物理モデルに응答して振動を判定するように構成される。

【0036】

これにより、レンダリングされるオーディオシーンが改善され、多くの場合、より現実的になる。

【0037】

本発明の任意選択の特徴によれば、生成器は、音響オブジェクトに加えられる時間変化する力に응答して振動を判定するように構成される。

40

【0038】

これにより、レンダリングされるオーディオシーンが改善され、多くの場合、より現実的になる。多くの実施形態では、たとえば、機械的な動きの結果として生成されるオーディオなど、追加の効果およびオーディオをレンダリングできる場合がある。

【0039】

本発明の任意選択の特徴によれば、カップリング伝達関数は、音響オブジェクトに入射するオーディオから、音響オブジェクトの振動へのものである。

【0040】

これは、多くの実施形態において、カップリング効果の特に有利で柔軟な表現を提供し得る。

50

【0041】

本発明の任意選択の特徴によれば、生成器は、空間特性データによって示されるような音響オブジェクトの位置に対応する空間位置を有するオブジェクトオーディオソースを表現するために、オブジェクトオーディオソースデータを生成するように構成される。

【0042】

これにより、オーディオシーンのレンダリングが改善され、オーディオシーンのより現実的な知覚になることがよくある。特に、生成器は、空間特性データによって示されるように、音響オブジェクトの位置と同じ空間位置を有するオブジェクトオーディオソースを表現するために、オブジェクトオーディオソースデータを生成し得る。

【0043】

本発明の任意選択の特徴によれば、生成器は、空間特性データによって示される第1のオーディオソースの位置および音響オブジェクトの位置に応答して、オブジェクトオーディオソースデータを生成するように構成される。

【0044】

これは、多くの実施形態において、オーディオシーンの改善されたレンダリングを提供し得る。

【0045】

本発明の任意選択の特徴によれば、生成器は、空間特性データによって示されるリスニング位置および音響オブジェクトの位置に応答して、オブジェクトオーディオソースデータを生成するように構成される。

【0046】

これは、多くの実施形態において、オーディオシーンの改善されたレンダリングを提供し得る。

【0047】

本発明の任意選択の特徴によれば、信号は、第2のオーディオソースのための入力オーディオソースデータを備え、生成器は、第1のオーディオソースおよび第2のオーディオソースからのオーディオの結合カップリングから、音響オブジェクトによってオーディオシーンにおいて発せられるオーディオを表現するオブジェクトオーディオソースデータを生成するように構成される。

【0048】

これは、多くの実施形態において、オーディオシーンの改善されたレンダリングを提供し得る。

【0049】

多くの実施形態では、第1のオーディオソースおよび第2のオーディオソースからのオーディオの結合カップリングから音響オブジェクトによってオーディオシーンにおいて発せられるオーディオを表現するオブジェクトオーディオソースデータの生成は、第1のオーディオソースのための音響カップリングデータ、空間特性データ、および入力オーディオソースデータに回答して生成されるオーディオデータと、第2のオーディオソースのための音響カップリングデータ、空間特性データ、およびオーディオ信号データに回答して生成されるオーディオデータとの重ね合わせによるものであり得る。

【0050】

本発明の態様によれば、オーディオシーンのためのデータを生成するためのデータ生成器であって、データは、オーディオシーンにおける少なくとも第1のオーディオソースのための入力オーディオソースデータと、少なくとも1つの音響オブジェクトのための音響オブジェクトデータとを備え、音響オブジェクトデータは、音響オブジェクトのための音響カップリングデータおよび空間特性データを備える、データ生成器と、オーディオシーンのためのデータを備えるデータ信号を生成するための信号生成器と、データ信号を送信するための送信機とを備える、オーディオ装置が提供される。

【0051】

本発明の態様によれば、オーディオを処理する方法が提供され、この方法は、オーディ

10

20

30

40

50

オーディオシーンのためのデータを備える信号を受信することであって、データは、オーディオシーンにおける少なくとも第1のオーディオソースのための入力オーディオソースデータと、少なくとも1つの音響オブジェクトのための音響オブジェクトデータとを備え、音響オブジェクトデータは、音響オブジェクトのための音響カップリングデータおよび空間特性データを備える、受信することと、第1のオーディオソースからのオーディオのカップリングから音響オブジェクトによってオーディオシーンにおいて発せられるオーディオを表現するオブジェクトオーディオソースのためのオブジェクトオーディオソースデータを生成することであって、生成器は、音響カップリングデータ、空間特性データ、および入力オーディオソースデータにตอบสนองしてオブジェクトオーディオソースデータを生成するように構成される、生成することと、オーディオシーンをレンダリングすることとを有し、レンダリングすることは、オブジェクトオーディオソースデータをレンダリングすることを含む。

10

【0052】

本発明の態様によれば、生成およびオーディオデータ信号の方法が提供され、この方法は、オーディオシーンのためのデータを生成することと、データは、オーディオシーンにおける少なくとも第1のオーディオソースのための入力オーディオソースデータと、少なくとも1つの音響オブジェクトのための音響オブジェクトデータとを備え、音響オブジェクトデータは、音響オブジェクトのための音響カップリングデータおよび空間特性データを備える、生成することと、オーディオシーンのためのデータを備えるオーディオデータ信号を生成することと、オーディオデータ信号を送信することとを有する。

20

【0053】

本発明のこれらおよび他の態様、特徴および利点は、以下に記載される実施形態から明らかであり、参照して解明されるであろう。

【0054】

本発明の実施形態は、例としてのみ、図面を参照して記載される。

【図面の簡単な説明】**【0055】**

【図1】図1は、クライアントサーバベースの仮想現実システムの例を示す図である。

【図2】図2は、本発明のいくつかの実施形態によるオーディオ装置の要素の例を示す図である。

30

【図3】図3は、本発明のいくつかの実施形態によるオーディオ装置の要素の例を示す図である。

【図4】図4は、オーディオシーンにおけるオーディオ伝播の例を示す図である。

【図5】図5は、音響オブジェクトのオーディオ効果の例を示す図である。

【図6】図6は、本発明のいくつかの実施形態によるオーディオ装置の要素の例を示す図である。

【図7】図7は、音響オブジェクトのオーディオ効果間関係を示す図である。

【図8】図8は、本発明のいくつかの実施形態による、音響オブジェクトの空間的広がりをレンダリングする例を示す図である。

【発明を実施するための形態】

40

【0056】

ユーザが仮想世界または拡張世界を移動可能とする（拡張現実および複合現実を含む）仮想現実体感はますます人気が高まっており、そのような要求を満たすためのサービスが開発されている。そのような多くのアプローチでは、視覚的およびオーディオデータは、ユーザ（または視覚者）の現在の姿勢を反映するように動的に生成され得る。

【0057】

この分野では、配置と姿勢という用語は、位置および/または方向/向きの一般的な用語として使用される（たとえば、[https://en.wikipedia.org/wiki/Pose_\(computer_vision\)](https://en.wikipedia.org/wiki/Pose_(computer_vision))を参照）。たとえば、物理的なオブジェクト、カメラ、ヘッド、またはビューからなる位置と方向/向きの組合せは、

50

姿勢または配置と呼ばれる場合がある。したがって、配置または姿勢の指標は、最大6つの値/構成要素/自由度を備え得、各値/構成要素は、典型的には、対応する物理的オブジェクトの位置/場所または向き/方向の個々の特性を記載する。もちろん、多くの状況で、配置または姿勢は、たとえば、1つまたは複数の構成要素が固定または無関係であると思なされる場合、より少ない構成要素で表現され得る(たとえば、すべての物理的オブジェクトが同じ高さで水平な向きを有していると思なされる場合、4つの構成要素が、物理的オブジェクトの姿勢の完全な表現を提供する場合がある)。以下では、姿勢という用語は、(可能な最大自由度に対応する)1から6の値で表現できる位置および/または向きを指すために使用される。

【0058】

多くのVRアプリケーションは、最大の自由度、すなわち、位置と向きの各々に3つの自由度、合計6つの自由度を有する姿勢に基づいている。したがって、姿勢は、6つの自由度を表現する6つの値のセットまたはベクトルによって表現され得、したがって、姿勢ベクトルは、3次元位置および/または3次元方向の指標を提供することができる。しかしながら、他の実施形態では、姿勢は、より少ない値によって表現され得ることが理解されよう。

【0059】

視覚者に最大の自由度を提供することに基づくシステムまたはエンティティは、通常、6つの自由度(6DoF)を有すると呼ばれる。多くのシステムおよびエンティティは、向きまたは位置のみを提供し、これらは通常、3つの自由度(3DoF)を有するものとして知られている。

【0060】

6つの自由度で、ユーザは、3次元デカルト座標系の3次元すべてにおいて移動できる。または、リスナに関しては、左右、前後、および上下方向である。残りの3つの自由度は、デカルト座標系の3軸(たとえば、航空機の操縦で知られているヨー、ピッチ、ロール、たとえばhttps://en.wikipedia.org/wiki/Aircraft_principal_axesを参照)の方向に沿った回転である。

【0061】

通常、仮想現実アプリケーションは、左目および右目のために別々のビュー画像の形式で3次元出力を生成する。次に、これらは、VRヘッドセットの、通常は個々の左目および右目ディスプレイなどの適切な手段によってユーザに提示され得る。他の実施形態では、1つまたは複数のビュー画像は、たとえば、裸眼立体視ディスプレイ上に提示されるか、または実際にいくつかの実施形態では、(たとえば、従来の二次元ディスプレイを使用して)単一の二次元画像のみが生成され得る。

【0062】

同様に、所与の視覚者/ユーザ/リスナの姿勢のために、シーンのオーディオ表現を提供することができる。オーディオシーンは通常、オーディオソースが目的の位置から発生していると認識される空間的な体感を提供するためにレンダリングされる。多くの実施形態では、オーディオシーンは、シーン空間内を移動する可能性のあるオーディオソースによって動的に変化し得る。また、ユーザの姿勢が変化すると、通常、ユーザの姿勢に対するオーディオソースの相対位置が変化する。したがって、オーディオソースの空間的な知覚は、ユーザに対する新しい位置を反映するように変化する必要がある。したがって、オーディオレンダリングは、ユーザの姿勢に応じて適合され得る。

【0063】

視覚者またはユーザの姿勢入力、さまざまなアプリケーションにおいてさまざまな手法で決定され得る。多くの実施形態では、ユーザの物理的な動きを直接追跡することができる。たとえば、ユーザ領域を見渡すカメラは、ユーザの頭を検出して追跡する(または目さえも(アイトラッキングする))ことができる。多くの実施形態では、ユーザは、外部および/または内部手段によって追跡できるVRヘッドセットを装着し得る。たとえば、ヘッドセットは、ヘッドセット、したがって頭の動きおよび回転に関する情報を提供す

10

20

30

40

50

る加速度計およびジャイロスコープを備え得る。いくつかの例では、VRヘッドセットは、信号を送信するか、または外部センサがVRヘッドセットの位置を決定することを可能にする（たとえば、視覚的な）識別子を備え得る。

【0064】

いくつかのシステムでは、視覚者の姿勢は、たとえば、ユーザがジョイスティックまたは同様の手動入力を手動で制御することによって、手動の手段によって提供され得る。たとえば、ユーザは、一方の手で第1のアナログジョイスティックを制御することによって、仮想シーン内で仮想的な視覚者を手動で動かし、他方の手で第2のアナログジョイスティックを手動で動かすことによって、仮想的な視覚者が見ている方向を手動で制御することができる。

10

【0065】

いくつかのアプリケーションでは、手動と自動とのアプローチの組合せを使用して、入力視覚者姿勢を生成できる。たとえば、ヘッドセットは頭の向きを追跡し、シーン内の視覚者の動き/位置は、ジョイスティックを使用してユーザによって制御され得る。

【0066】

多くのシステムでは、シーンを記載するデータが、中央のソースから個々のユーザのローカルデバイスに送信される。たとえば、いくつかの異なるビュー位置からのシーンのビューを表現する視覚データは、中央サーバからローカルクライアントに送信され得る。次に、この視覚データに基づいて、ローカルデバイスにおけるローカルレンダラが、視覚者の現在位置の特定のビューを合成し得る。したがって、（現在の視覚者の姿勢とは無関係に）いくつかのキャプチャまたはアンカ位置からのシーン表現を、シーンの表現としてローカルデバイスに送信することができ、ローカルデバイスは、これを処理して、現在の視覚者の姿勢のビューを動的に生成することができる。

20

【0067】

同様に、リモートデバイス/サーバは、仮想オーディオ環境を反映するオーディオシーンを生成し得る。これは、多くの実施形態において、仮想オーディオ環境における異なるオーディオソースの相対位置に対応するオーディオ要素を生成することによって行われ得、これらは、対応する位置において知覚されるようにレンダリングされる。

【0068】

たとえば、リモートデバイスは、オーディオシーンを表現するオーディオデータを生成し、オーディオ構成要素/オブジェクト/信号、またはオーディオシーンにおける異なるオーディオソースに対応する他のオーディオ要素を、（たとえば、物理的オブジェクトを移動するために動的に変化し得る）これらの位置を示す位置情報とともに送信し得る。オーディオ要素は、特定の位置に関連付けられた要素を含み得るが、より分散されたまたは拡散されたオーディオソースのための要素を含み得る。たとえば、一般的な（ローカライズされていない）背景音、周囲音、拡散残響などを表現するオーディオ要素が提供され得る。

30

【0069】

オーディオデータは、多くの実施形態において、シーンの音響特性を記述するメタデータも含み得る。このメタデータは、音響オクルージョンまたは反射を引き起こす（オーディオ以外の）要素の情報を含み得る。

40

【0070】

次に、ローカルVRデバイスは、たとえば、オーディオ構成要素のオーディオソースの相対位置を反映する適切なバイノーラル処理を適用することによって、オーディオ要素を適切にレンダリングし得る。音響メタデータを使用して、ユーザに到達するオーディオに対するシーンのさらなる効果をモデル化することができる。これらは、オーディオ要素のオーディオ信号に直接影響を与える場合もあれば、残響などの追加のオーディオ要素を導入する場合もある。

【0071】

したがって、VRサービスのオーディオ側の場合、中央サーバは、いくつかの実施形態

50

では、オーディオシーンを表現するオーディオデータを生成することができ、特に、ローカルクライアント/デバイスによってレンダリングできるいくつかのオーディオ要素によってこのオーディオシーンを表現し得る。オーディオ要素は、特に、オーディオオブジェクト、オーディオチャンネル、高次アンビソニックスのうちの1つまたは複数を含み得る。

【0072】

図1は、中央サーバ101が、たとえば、インターネットのようなネットワーク105を介して、いくつかのリモートクライアント103と連絡を取るそのようなVRシステムの例を示している。中央サーバ101は、潜在的に多数のリモートクライアント103を同時にサポートするように構成され得る。

【0073】

そのようなアプローチは、多くのシナリオにおいて、たとえば、異なるデバイスの複雑さとリソース要求、通信要件などの間の改善されたトレードオフを提供し得る。

【0074】

図2は、以後、オーディオレンダラ200とも呼ばれ、多くのアプリケーションおよびシナリオにおいて、改善されたオーディオレンダリングを提供し得るオーディオ装置の要素を示している。特に、オーディオ装置は、多くのVRアプリケーションに対して改善されたレンダリングを提供することができ、オーディオ装置は、図1のリモートクライアント103に対してオーディオ処理およびレンダリングを実行するように特に構成され得る。

【0075】

図3は、以後、オーディオ信号生成器300とも呼ばれ、多くのアプリケーションおよびシナリオにおいて、改善されたオーディオデータ信号を生成し得るオーディオ装置の要素を示している。特に、オーディオ装置は、多くのVRアプリケーションのための改善されたオーディオビジュアルデータストリームの一部として、改善されたオーディオデータ信号を提供することができ、オーディオ信号生成器300は、図1のVRサーバ101のためのオーディオ処理およびレンダリングを実行するように特に構成され得る。

【0076】

オーディオ信号生成器300は、オーディオシーンの表現を提供するデータを生成するように構成されたオーディオデータ生成器301を備える。データは、シーン内のいくつかのオーディオソースのデータを含み得、データは、ソースによって生成されたオーディオ(信号)を記載するオーディオデータ、ならびに典型的にはシーンの空間特性の記載を提供する空間情報を備える。そのような空間データは、典型的には、シーン内のオーディオソースの位置を示すデータ、および/または、オーディオソースの空間分布/広がり記載を含み得る。空間分布は、たとえば、データソースが拡散して空間的に分散しているか否か(たとえば、周囲ノイズまたは背景ノイズ)、または明確に画定されたポイントソースであるか否かを反映し得る。所与のオーディオソースのための空間データは、たとえば、オーディオソースのサイズを示し得る。オーディオシーンにおける所与のオーディオソースのオーディオソースデータは、オーディオソースのためのオーディオ信号データおよび空間特性データを特に備え得る。

【0077】

オーディオソースによって生成されたオーディオは、通常、エンコードされたオーディオ信号として提供される。オーディオソースは通常、オーディオオブジェクト、チャンネル、またはHOAとして表現され得る。

【0078】

オーディオソースは、特に、ポイントソースとしてレンダリングされることを意図された、関連付けられた位置メタデータを備えた単一のオーディオ信号であり得る。それに加えて、空間的広がり(サイズ)がメタデータの一部である場合があり、音ソースを、特定のサイズでレンダリングする必要があることを示す。(周波数に依存する)指向性パターンなどのさらなるメタデータは、音ソース定義の一部であり得る。

【0079】

この例では、オーディオシーンを特徴付けるデータは、オーディオソースに限定されず

10

20

30

40

50

、1つまたは複数の音響オブジェクトの記述も含むことができる。音響オブジェクトは、音響環境に影響を与える可能性のあるシーン内のオブジェクトであり得、特に、少なくとも1つのオーディオソースから、オーディオシーンにおける少なくとも1つのポイントにおいて受信されるオーディオに影響を及ぼし得るオブジェクトであり得る。したがって、音響オブジェクトは、オーディオソースからのオーディオの伝播に影響を与える可能性のあるシーンにおける任意のオブジェクトであり得る。音響オブジェクトは、通常、パッシブ音響オブジェクトである場合がある（後で記載されるように、通常、入射オーディオまたは機械的衝撃にตอบสนองしてオーディオを生成する場合がある）。音響オブジェクトは、空気とは異なる音響特性を有するオブジェクトであり得る。

【0080】

したがって、オーディオシーンのために生成されるデータは、オーディオを生成するオーディオソースのみならず、オーディオに影響を与える可能性のある音響オブジェクトを含む。場合によっては、オブジェクトオーディオソースは、特定の音響オブジェクトに関連付けられ得る。

【0081】

図3を参照して示すように、オーディオデータ生成器301は、オーディオデータ生成器301によって生成されたオーディオシーンのためのデータを備えるデータ信号を生成するように構成された信号生成器303に結合される。典型的には、信号生成器303は、オーディオシーンを特徴付けるデータと、ビジュアルシーンを特徴付ける対応するビデオデータとの両方を含むオーディオビジュアルデータストリームを生成する。したがって、データストリームは、オーディオビジュアルシーンを特徴付ける。信号生成器303は、任意の適切な符号化アルゴリズムおよびフォーマットを使用して、ビデオおよびオーディオデータを符号化するように構成され得る。

【0082】

生成された符号化されたオーディオビジュアルデータストリームは、任意の適切な通信チャネルを介してデータ信号を送信するように構成された送信機305に供給され得る。図1～図3の特定の例では、送信機305は、ネットワーク105を介してクライアントデバイス101にデータ信号を送信することを可能にする適切なネットワークインターフェースを備え得る。

【0083】

図2を参照して示すように、オーディオレンダラ200は、オーディオ信号生成器300からデータ信号を受信する受信機201を備える。したがって、受信機201は、データが、1つ、典型的にはより多くのオーディオソースのための入力オーディオソースデータ、ならびに1つ、典型的にはより多くの音響オブジェクトのための音響オブジェクトデータを含むオーディオシーンを表現するデータを受信する。

【0084】

入力オーディオソースデータは、生成されたオーディオの時間および/または周波数特性を具体的に記載することができ、（時間とともに変化する）単一のオーディオ信号を特に表現し得る。通常、データは、オーディオソースの空間データも含む。受信データにおいて表現される所与の入力オーディオソースのための入力オーディオソースデータは、所与の入力オーディオソースのためのオーディオ信号データおよび空間特性データを特に備え得る。

【0085】

受信機201は、オーディオデータの特徴付けるデータが供給されるレンダラ203に結合される。レンダラ203は、たとえば、個々のオーディオソースを個々にレンダリングし、結果として得られる信号を組み合わせることによって、オーディオシーンをレンダリングし得る。レンダリングは、オーディオソースのために受信された空間情報と、当業者に知られているようにオーディオが生成されるリスニング位置とに基づき、空間処理を含み得る。たとえば、ヘッドフォン出力の場合、バイノーラル処理を使用して、適切な空間位置からリスナに到達するように知覚されるオーディオソースを備えたヘッドフォン信

10

20

30

40

50

号を生成することができる。スピーカ出力の場合、スピーカレンダリングアルゴリズム（たとえば、V B A P）を適用して、オーディオ要素からスピーカ信号を生成することができる。

【0086】

レンダラ203はさらに、音響オブジェクトの特性を反映するようにオーディオソースを処理するように構成され得る。たとえば、オーディオソースのための信号レベルは、オーディオソースと、リスニング位置との間の音響オブジェクトに起因する減衰に対応するように低減され得る。

【0087】

レンダラ203は、特に、オーディオ信号をスピーカ（ヘッドフォンを含む）フィード / 駆動信号に変換する処理要素であり得る。たとえば、ヘッドフォンを装着している場合、この信号は、H R T FまたはB R I Rを使用してレンダリングされるが、スピーカ消費のために、音ソースとリスナの（仮想的）位置に対するスピーカの相対位置に応じて、信号がスピーカにマッピングされる。

10

【0088】

レンダラ203は、所与のアプリケーションのスピーカに直接供給される信号を生成することに限定されないが、生成された信号は、おそらく、さらに処理されることが理解されるであろう。たとえば、信号は、ラウドネスレベリング、ダイナミックレンジ圧縮（D R C）、（真の）ピーク制限、または他の処理ステップの対象となり得る。さらに、ラウドスピーカフィードへのレンダリングの後に、バイノーラルレンダリングを介したヘッドフォンスピーカフィードへのレンダリングが続く場合がある。

20

【0089】

オーディオレンダラ200は、シーンに追加のオーディオソースを導入するように構成された生成器205をさらに備える。生成器205は、少なくとも第1のオーディオソースからのオーディオのカップリングから、音響オブジェクトによってオーディオシーンにおいて発せられるオーディオを表現するオブジェクトオーディオソースのためのオブジェクトオーディオソースデータを特に生成し得る。

【0090】

システムでは、データ信号は、少なくとも1つの音響オブジェクトのための音響オブジェクトデータを含み、音響オブジェクトデータは、特に、第1の音響オブジェクトのための音響カップリングデータと空間特性データとの両方を含む。音響オブジェクトは、ボックスベース、球ベース、またはポリゴンベースの空間記述を提供する空間メタデータによって表現され得る。それはたとえば、リスナに面しているときの要素の有効な幅と高さの寸法を示す2つのサイズ値で表現され得る。

30

【0091】

そのようなデータに基づいて、生成器205は、以降、オブジェクトオーディオソースと呼ばれ、他のオーディオ（すなわち、入力オーディオソースからのオーディオ）を用いて、このカップリングによって音響オブジェクトから生成された音を反映する、新しいオーディオソースを生成する。この追加のオーディオソースは、シーンのレンダリングに含まれ得る。特に、オブジェクトオーディオソースデータは、受信信号においてオーディオソースがどのように表現されるかに対応するオーディオ信号データおよび空間データによって、追加のオブジェクトオーディオソースを表現し得る。特に、オブジェクトオーディオソースは、生成されたオーディオを記載するオーディオ信号データ、および生成されたオーディオの位置および / または広がりを記載する空間データによって表現され得る。

40

【0092】

音響モデルは、環境内の音の進行をモデル化するために、特に、オブジェクトが環境内のオーディオにどのように影響するかをモデル化するために使用できる。通常、モデル化された音波は、さまざまな固体材料と相互作用する。典型的な音響モデルには、到来する音波の反射、吸収、および（オブジェクトを通過する）透過を含む。レンダラ203は、オーディオソースをレンダリングし、それらが環境によってどのように影響を受けるかを

50

評価するとき、そのような効果を考慮し得る。

【0093】

特にオクルージョンの場合、モデルは、オクルージョンシーンオブジェクトがどこにあるか、およびそれらの音響特性が何であるかを考慮し得る。したがって、受信されたデータは、音響オブジェクトを記載するデータを特に含むことができ、通常、ほとんどの場合、視覚的にも表現されるオブジェクトの（簡略化された）形状を記載することができる。

【0094】

オクルージョンは、絶対的なものと見なすことができ、オクルージョンオブジェクトは、音を通過させないが、多くのオブジェクトでは、オクルージョンは完全ではなく、一部のオーディオは通過できる。これは、図4および図5に例示されるように、ユーザが依然として直接経路の一部を受け取ることができるように、通常、材料を通過するエネルギーの（周波数に依存する）一部を示す透過係数を提供することによって表現することができる。

10

【0095】

ほとんどの場合、ユーザはまた、反射と残響によって音ソースのエネルギーの一部を受け取る。これは、オクルージョン要素が、空気中のすべての（比較的短い）経路を遮断する場合には当てはまらない可能性がある。

【0096】

材料が異なれば、オクルージョン特性も異なる。壁はカーテンよりも強く塞ぐため、透過係数はカーテンよりも壁の方が低くなる。また、周波数特性が異なる場合があり、通常、レンガの壁は、非常に低い周波数しか通過させない。記載されている2つの極端なもの間の特性を有する異なるオクルージョン材料が存在する可能性がある。

20

【0097】

したがって、レンダラ203は、音響オブジェクトを介して、または反射および/または残響を介して、異なる経路を反映するオーディオソースを生成しようとするレンダリングを実行することができる。

【0098】

図6は、異なる効果を別々にレンダリングすることによって、入力オーディオソース信号がどのようにレンダリングされ得るかを示す要素の例を示す。この例は、たとえば、オーディオオブジェクトなどのドライ音ソース信号に適用することができる。

【0099】

特に、第1の経路601は、環境内の残響を反映するようにオーディオソースを処理することができる。たとえば、オーディオソース信号は、適切な残響フィルタまたはモデルを使用してフィルタリングし得る。第2の経路603は、たとえば、個々の反射のタイミングおよびレベルに対応する係数を有するフィルタで、オーディオソース信号をフィルタリングすることによって、早期の反射を反映するようにオーディオソースを処理し得る。第3の経路605は、たとえば、直接伝搬経路に依存するオーディオソース信号に減衰を適用することによって、直接経路を表現するためにオーディオソース信号を処理する。タイミング、減衰、および周波数応答が、経路ごとに異なる可能性があるだけでなく、空間効果も異なる可能性があることが理解されよう。特に、残響の場合、オーディオは、空間的な明確さのない拡散と見なされ得、直接経路は、オーディオソースの方向から到着すると見なされ得、反射は、オーディオ環境の特有の特性に応じて、異なる方向から到着すると見なされ得、通常、直接経路よりも空間的に画定されていないと見なされ得る。直接信号は、オーディオソース位置から直接到着するようにレンダリングことができ、空間的な広がりには比較的小さい可能性がある。

30

40

【0100】

音響オブジェクトによる直接経路減衰は、多くの場合、オーディオソースからの信号を減衰させる透過係数を適用することによってモデル化できる。しかしながら、発明者は、多くのシナリオにおいて、これは、オーディオソースの完全なまたは理想的な表現、およびこれが音響オブジェクトによってどのように影響されるかを提供しない可能性があることに気付いた。

50

【 0 1 0 1 】

たとえば、このアプローチは、壁ではなく、カーテンの現実的なモデルである可能性がある。これは、壁の強力なローパス透過挙動に起因する可能性があるが、薄い石膏ボードの壁や木製のドアなど、これらの両極端の間にある他の材料の場合、高周波成分が多くなり、一般に減衰が少なくなる。これらのケースが透過係数でモデル化されている場合、局在化が可能となるが、実際にはそうではない。

【 0 1 0 2 】

記載されたシステムでは、オーディオソースによって生成されたオーディオに対する音響オブジェクトの影響は、音響オブジェクトのための音響カップリングデータおよび空間特性データを備える音響オブジェクトデータによってさらに表現される。これは、この経路が、シーンに新しいオーディオソースを導入することによって、オーディオソース生成器 2 0 5 を含む第 4 の経路 6 0 7 において処理され得る。

10

【 0 1 0 3 】

カップリングは、オブジェクトのオーディオと振動 / (発振) 運動がどのように相互作用し、互いに変換されるかを表現する音響効果である。実施形態において、入射するオーディオは、オブジェクトの振動を引き起こす可能性があり、それによって、オーディオエネルギー (の一部) が、オブジェクトの機械的エネルギーに変換される可能性がある。また、オブジェクトの振動 / (発振) 運動は、オーディオを生成し、それによって、振動 / 機械的エネルギーをオーディオエネルギーに変換する場合がある。実際、カップリング効果は、多くの実施形態において、音響オブジェクトへの入射オーディオが、音響オブジェクトを振動で振動させ、次にオーディオを生成するオーディオ効果を提供し得る。

20

【 0 1 0 4 】

音ソースは、エネルギーを放射し、ほとんどの環境やシーンでは、そのエネルギーの一部は、リスナの耳に到達する前に、1 つまたは複数の音響オブジェクトに衝突する。音響オブジェクトは、通常、材料特性が空気とは異なる空間領域である可能性がある (そして、通常、実世界シーンにおける物理的オブジェクト、または仮想シーンのための物理的オブジェクトを表現する仮想オブジェクトに対応する場合がある) 。これらの要素の境界上で、音響インピーダンスは、上記のように、およびたとえば、図 5 に例示されるように、いくつかの音響効果を生じさせる。そのような音響効果は、通常、反射、透過、回折 (散乱) 、および吸収であり得る。

30

【 0 1 0 5 】

たとえば、多孔質または可撓性の材料は、エネルギーを吸収および反射する可能性があるが、エネルギーの一部を、その背後の空気に通過させる。そうしなければ、カーテンの後ろで起こっていることを聞くことはできない。

【 0 1 0 6 】

吸収される音 / エネルギーは、さまざまな変換効果によって引き起こされ得る。一部のオーディオは放散され、材料内で効率的に熱エネルギーに変換され得る。しかしながら、オーディオエネルギーの一部は、材料の動きに変換される可能性があり、したがって、対応するオブジェクト / 表面の振動に変換される可能性がある。しかしながら、熱エネルギーへの変換の第 1 の効果は、対応するオーディオエネルギーがオーディオ領域 / 環境から除去されることを実際に意味するが、第 2 の効果は、さらなるオーディオ効果を引き起こす可能性がある。

40

【 0 1 0 7 】

特に、音響オブジェクトの振動は、空気が音響オブジェクトによって移動され、したがって、音響オブジェクトの遠側 / 閉塞側を含む音の生成をもたらす可能性がある。したがって、オーディオを振動運動に変換した後、この振動運動を音に変換することができる。さらに、たとえば、空間的広がりおよび周波数範囲を含むこの音の特徴は、音響オブジェクトの特性に実質的に依存する可能性があり、直線透過から生じる音とは実質的に異なる可能性がある。たとえば、隣接する部屋間の音の伝播では、壁、ドア、または他のオブジェクトの全体が振動している可能性があり、比較的小さな音ソースではない広い領域が音

50

を生成しているため、局在化がより困難になる。

【0108】

図7は、起こり得る異なる効果間の関係の例を示す。オーディオエネルギーと振動/機械的エネルギーとの変換効果は、カップリングとして知られており、建物のさまざまな部分の間の遮音性を向上させるために、主に建物の建設設計において考慮される。所与の音響オブジェクトに対して異なる音響効果が生じる度合いは、音響オブジェクトの構造および材料または材料の組合せを含む音響オブジェクトの特定の特性に依存する。

【0109】

図1から図3のシステムでは、データ信号は、音響オブジェクトのための音響カップリングデータおよび空間特性データを含む1つまたは複数の音響オブジェクトのための音響オブジェクトデータを含むように生成される。音響カップリングデータは、オーディオエネルギーの振動エネルギーへの変換、および/または、振動エネルギーのオーディオエネルギーへの変換の特性の指標を備える。多くの状況において、音響カップリングデータは、代替的または追加的に、音響オブジェクトの振動性能、特に振動効果が音響オブジェクト内でどのように伝播し、分散されるのかについての指標を含み得る。

【0110】

さらに、データは、オーディオレンダラ200によって、オーディオシーンにおいて新しいオーディオソースを生成するために使用される。したがって、受信データによって記載される入力オーディオソースに加えて、オーディオレンダラ200のオーディオソース生成器205は、少なくとも第1の入力オーディオソースからのオーディオのカップリングからの音響オブジェクトによって、オーディオシーンで発せられるオーディオを表現するオブジェクトオーディオソースのためのオブジェクトオーディオソースデータを生成するように構成される。オブジェクトオーディオソースデータは、特に、生成されるオーディオを記載するオーディオ信号データ、および発せられたオーディオの空間特性を示す空間データを含み得る。特に、空間データは、生成された音の位置および広がり(空間分布)を表現し得る。

【0111】

オーディオソース生成器205は、受信された音響カップリングデータ、音響オブジェクトについて受信された空間特性データ、ならびに特定のオーディオソース、および少なくとも第1のオーディオソースのオーディオ信号データに基づいて、オブジェクトオーディオソースデータを生成する。

【0112】

複雑さの低い例として、音響カップリングデータは、カップリングを介して生成されるオーディオ信号エネルギーの量を示すカップリング係数を備え得るので、したがって、カップリング係数は、第1のオーディオソースからのオーディオエネルギーを、振動エネルギーに変換する場合、および、この振動エネルギーを、オーディオエネルギーに変換する場合の変換損失を示し得る。したがって、たとえば、単にカップリング係数によって第1のオーディオソース信号をスケールアップするだけで、このエネルギーに対応するエネルギー/レベルで新しいソースを生成することができる。このオーディオソースのための空間データは、音響オブジェクトの位置および/または空間的広がりに基づいて決定される新しいソースの位置および/または広がり/または空間的広がり/分布は、直線透過から生じるものとは異なり、音響オブジェクトの空間特性(および、通常は機械的特性)に依存する。特に、多くの実施形態では、新しいオーディオソースは、音響オブジェクトの位置に対応する位置を有し、音響オブジェクトのサイズに対応する空間的広がり/分布を有するように生成され得る。たとえば、音響オブジェクトが壁に対応する場合、新しいソースは、壁全体、すなわち、生成された新しいオブジェクトのオーディオソースの広がりから発生するように生成され得る。

【0113】

生成されたオブジェクトオーディオソースは、他のオーディオソースと同じ手法でレンダリングされ得る。オブジェクトオーディオソースデータは、決定された特性、特に、決

10

20

30

40

50

定された位置および広がりで見られるようにオブジェクトオーディオソースをレンダリングできるレンダラ 203 に供給され得る。オーディオソースを、知覚された位置および広がりを有するようにレンダリングするための多くの異なるアルゴリズムが知られており、任意の適切なそのようなアプローチおよびアルゴリズムが使用され得ることが理解されよう。

【0114】

特に、オーディオソースサイズを作成するにはさまざまな手法がある。これは通常、音ソース信号のバージョンを、たとえば、信号のエネルギーの4分の1である、リスニング位置からの音響オブジェクトの知覚空間領域の四隅におけるように、複数の位置でレンダリングすることによって実現される。いくつかの実施形態は、リスナによって知覚されるように、音響オブジェクトの空間領域の四隅よりも多い（または少ない）位置、および/または異なる位置に、オブジェクト信号のバージョンをレンダリングすることができる。

10

【0115】

具体例として、図8は、壁または障壁が、リスニング位置からどのように見えるかを示している。そのような壁から得られるオーディオは、単一のポイントからではなく、壁の表面全体から発生していると見なすことができる。そのような知覚は、たとえば、図8におけるXによって示されるように、中央および各隅に、サブオーディオソースを生成することによって生成され得る。これらのサブオーディオソースのおのおのは、壁に対応する音響オブジェクトに対して決定されたオーディオを発するようにレンダリングされ得る。特定の例では、各サブオーディオソースのレベルは、たとえば、オーディオソースに対して決定された合計レベルの5分の1であり得る。さらに、いくつかの空間的な拡張または拡散効果（たとえば、いくつかの追加の残響または非相関フィルタ（通常は、ランダムな位相応答を有するオールパスフィルタ））が、レンダリングされたオーディオに適用され得る。したがって、組み合わせられた効果は、音響オブジェクト/壁 801 全体のサイズを有する単一のオーディオソースの知覚に至り得る。

20

【0116】

より高度な実施形態では、音ソース信号のエネルギーは、音ソース信号のバージョン全体に不均一に分散され、受信された音ソースが音響オブジェクトに最も近い場所に依りて、音響オブジェクトの特定の領域に重点を置く。

【0117】

したがって、オーディオレンダラ 200 は、音響オブジェクトのサイズおよび位置にしたがって、空間的広がりを有する1つまたは複数の新しい音ソースを導入し得る。新しい音ソースは、別の音ソースによって生成されたオーディオエネルギーが、音響オブジェクトの振動に変わることによって、音を生成する音響オブジェクトを表現する。

30

【0118】

多くの実施形態では、複数のオーディオソースからのオーディオが、考慮され得る。実際、2つのオーディオソースからの音響オブジェクトからのオーディオ入射が考慮され得る。たとえば、2つの異なるオーディオソースから音響オブジェクトに入射するオーディオは、たとえば、重ね合わせ、またはレイトレーシング、または他の任意のアルゴリズムによって、最初に決定され得る。組み合わせられたオーディオは、単一のソースからのオーディオであるかのように処理され得る。すなわち、たとえば、振動に変換され、振動挙動がモデル化され、結果として生じるオーディオがそこから生成され得る。

40

【0119】

他の実施形態では、個々のオーディオソースを個々に処理することができ、結果として得られる出力オーディオを組み合わせることができる。したがって、いくつかの実施形態では、オーディオソース生成器 205 は、第1のオーディオソースと第2のオーディオソースとからのオーディオの結合されたカップリングから、音響オブジェクトによってオーディオシーンにおいて発せられるオーディオを表現するオブジェクトオーディオソースデータを生成し得る。

【0120】

50

オーディオソース生成器 205 は、特に、第 1 のオーディオソースのための音響カップリングデータ、空間特性データ、およびオーディオ信号データにตอบสนองして生成されたオーディオ（信号）データと、第 2 のオーディオソースのための音響カップリングデータ、空間特性データ、およびオーディオ信号データにตอบสนองして生成されたオーディオ（信号）データとの重ね合わせによってこれを行う。

【0121】

多くの実施形態では、そのような組み合わせられたオブジェクトオーディオソースデータは、オーディオソース信号の 1 つまたは複数のフィルタリングされた重ね合わせとして生成されるオーディオソース信号データによって生成され得、フィルタリングは、少なくとも、典型的には、音響オブジェクトの周波数依存のカップリング係数によって決定される。

10

【0122】

記載されたシステムでは、音響オブジェクトに関連付けられたカップリングメタデータが、オーディオシーンを表現するデータ信号に含まれる。多くの実施形態では、音響カップリングデータは、振動エネルギーへのおよび/または振動エネルギーからのカップリングのレベルおよび/または周波数応答を表現し得る。多くの実施形態では、カップリングメタデータは、カップリングのための伝達関数の形態を採り得る。伝達関数は、たとえば、FIR または IIR フィルタのフィルタ伝達関数を表す 1 つまたは複数の係数など、1 つまたは複数のカップリング係数によって表現され得る。最も単純な実施形態では、伝達関数は、周波数に依存しないスケールングであり得、音響カップリングデータは、たとえば、入射オーディオエネルギーから、カップリングプロセスによって生成されたオーディオエネルギーへの信号エネルギーの減衰を示す単一のカップリング係数のみを備え得る。

20

【0123】

したがって、いくつかの実施形態では、所与の音響オブジェクトのための音響カップリングデータは、オーディオ信号生成器 300 から受信したオーディオソース信号からオブジェクトオーディオソース信号をどのように導出すべきかを示す単一の係数を備え得る。たとえば、

【数 1】

$$s_{g,a}[n] = \sum_{i \in D_r} c_a \cdot s_{r,i}[n] = c_a \cdot \sum_{i \in D_r} s_{r,i}[n]$$

30

ここで、 $s_{g,a}$ は、音響オブジェクト a のために生成されたオブジェクトオーディオソース信号、 D_r は、受信したオーディオソース信号のセット、 c_a は、音響オブジェクト a のカップリング係数、 $s_{r,i}$ は、受信した音ソース信号 i 、 n は、時間インデックス（たとえば、時間サンプル）である。この例では、複数のオーディオソースからのカップリングの効果は、個々の寄与の重ね合わせによって考慮される。

40

【0124】

カップリング係数は、広帯域カップリング係数に対応する場合がある。すなわち、周波数応答は、オーディオ帯域全体でフラットであり、レベルの変化のみが考慮される。しかしながら、ほとんどの場合、周波数依存の係数を使用することが有益であり、たとえば、カップリング係数は、低周波数において大幅に高くなる可能性がある。別の例は、周波数依存の係数が、音響オブジェクト、または音響オブジェクトの表面成分における共振を反映するのに特に適している可能性があることである。

【0125】

周波数依存のカップリング係数は、たとえば、FIR (b_i) または IIR (b_i, a_j) フィルタの係数によって、フィルタとして提供され得る。

50

【数 2】

$$C_a(z) = \frac{\sum_{i=1}^I b_i z^{-i}}{\sum_{j=1}^J a_j z^{-j}}$$

10

【0126】

生成されたオブジェクトオーディオソース信号への寄与は、受信されたオーディオソース信号を、フィルタ係数でフィルタリングすることによって取得され得る。たとえば、周波数領域では、

【数 3】

$$S_{g,a}(z) = \sum_{i \in D_r} S_{r,i}(z) \cdot C_a(z)$$

20

であるか、または、時間領域における F I R では、

【数 4】

$$s_{g,a}[n] = \sum_{i \in D_r} (s_{r,i} * c_a)[n] = \left(\sum_{i \in D_r} s_{r,i} \right) * c_a[n]$$

30

であり、ここで、* は、畳込みを表す。

【0127】

あるいは、周波数依存のカップリング係数を、特定の周波数帯域に提供することができ、そのような各パラメータ帯域は、周波数スペクトルの一部で定義される。パラメータ帯域は、FFT 値域、MDCT 値域、またはサブ帯域インデクスで定義できる。

【0128】

そのようなアプローチでは、ほとんどの実施形態において、オブジェクトオーディオソース信号は、各パラメータ帯域のスカラーカップリング係数に、対応するパラメータ帯域におけるすべての FFT 値域またはサブ帯域信号を乗じることによって生成され得る。たとえば、

40

【数 5】

50

$$s_{g,a}[b, n] = \sum_{i \in D_r} s_{r,i}[b, n] \cdot c_{a,B}(b)$$

であり、ここで、 b は、STFT値域（またはサブ帯域インデクス）を示し、 $B(b)$ は、STFT値域（またはサブ帯域インデクス） b が与えられた場合のパラメータ帯域インデクスを返す関数である。

10

【0129】

カップリング係数の値は通常0から1の間であり、ここで、0は、音響オブジェクトを介したカップリングによってエネルギーが伝達されない状況を表現し、値1は、完全な入射オーディオ信号がオブジェクトオーディオソース信号に変換されることを示す。

【0130】

生成されたオブジェクトオーディオソースデータはまた、典型的には、生成されたオブジェクトオーディオソースの空間位置および/または広がりを表現し得る空間データを含み得る。このデータは、通常、音響オブジェクトの空間データ、すなわち、特に音響オブジェクトの位置および/または広がりを反映し得る。しかしながら、いくつかの実施形態およびシナリオでは、元のオーディオソースの位置および/または広がり依存する場合もある。

20

【0131】

多くの実施形態では、オブジェクトオーディオソースは、音響オブジェクトの空間的広がりから決定される空間的広がりを有するように生成され、特に、それは、音響オブジェクトと同じであると判定され得る。たとえば、壁の場合、対応するオブジェクトオーディオソースのオーディオは、壁全体から放射されていると見なすことができる。

【0132】

いくつかの実施形態では、生成されたオブジェクトオーディオソースの広がり、音響オブジェクトの広がりよりも小さい、音響オブジェクトの表面の広がりによって決定され得る。これは、この表面が、音響オブジェクトの他の表面よりも、音響オブジェクトにおけるカップリングまたは振動に反応して、オーディオを生成する傾向があるためであり得る。

30

【0133】

同様に、多くの実施形態では、オブジェクトオーディオソースは、音響オブジェクトの空間位置から決定される空間位置を有するように生成され、特に、音響オブジェクトと同じであると判定され得る。たとえば、家具の場合、対応するオブジェクトオーディオソースのオーディオは、元のオーディオソースの位置からではなく、家具の位置から放射されていると見なすことができる。

【0134】

いくつかの特定の例として、音響オブジェクトは、その中心位置

40

【数6】

$$\bar{p} = (p_x, p_y, p_z)$$

と、中心位置の周りの広がり

【数7】

50

$$\bar{d} = (d_x, d_y, d_z)$$

とによって表現できる。ベクトル

【数 8】

$$\bar{d}$$

10

における3つの値は、音響オブジェクトのボックスまたは楕円体近似の3つの空間的寸法を示す場合がある。寸法

【数 9】

$$\bar{d}$$

20

は、座標軸に対して直接的に寸法を示すことも、寸法における後続するヨー、ピッチ、およびロール回転角を示す回転データ

【数 10】

$$\bar{r} = (r_y, r_p, r_r)$$

30

と組み合わせて示すこともできる。

【0 1 3 5】

あるいは、位置ベクトル

【数 1 1】

$$\bar{p}$$

は、音響オブジェクトの広がり、特に、形状が画定される開始位置を示す。たとえば、ポリゴンを使用して、おのおのがポリゴンのエッジを画定する少なくとも3つの座標ペア $E_{i,j} = \{(e_{x1,i,j}, e_{y1,i,j}, e_{z1,i,j}), (e_{x2,i,j}, e_{y2,i,j}, e_{z2,i,j})\}$ で構成されるポリゴン Q_i を設定することにより、音響オブジェクトの形状を画定できる。座標ペアは、絶対座標で、または位置ベクトルに関して表現され得る。

40

【0 1 3 6】

生成されたオブジェクトオーディオソースの空間特性は、たとえば、中心位置

【数 1 2】

50

$$\bar{p}$$

と、

【数 1 3】

$$\bar{d}$$

10

および

【数 1 4】

$$\bar{r}$$

で表される空間的広がり、またはあるいは、ポリゴン記述 Q_i のような音響要素の空間的広がりから直接引き継ぐことができる。音響オブジェクトがポリゴンによって記載される他の実施形態では、寸法は、ポリゴン定義から導出され、ベクトル

20

【数 1 5】

$$\bar{d}$$

などの異なる表現に変換され得、ここで、

【数 1 6】

30

$$d_x = \max \left(\max_{i,j} e_{x1,i,j}, \max_{i,j} e_{x2,i,j} \right) - \min \left(\min_{i,j} e_{x1,i,j}, \min_{i,j} e_{x2,i,j} \right)$$

$$d_y = \max \left(\max_{i,j} e_{y1,i,j}, \max_{i,j} e_{y2,i,j} \right) - \min \left(\min_{i,j} e_{y1,i,j}, \min_{i,j} e_{y2,i,j} \right)$$

$$d_z = \max \left(\max_{i,j} e_{z1,i,j}, \max_{i,j} e_{z2,i,j} \right) - \min \left(\min_{i,j} e_{z1,i,j}, \min_{i,j} e_{z2,i,j} \right)$$

40

である。

【0 1 3 7】

カップリングデータが各ポリゴンに与えられるいくつかの実施形態では、生成されたオーディオソースの空間的広がり、0より大きいオーディオ伝達への振動を記載するカップリング係数を有するポリゴンによってのみ決定され得る。

【0 1 3 8】

いくつかの実施形態では、生成されたオブジェクトオーディオソースの空間特性は、音響オブジェクトの空間特性にのみ依存し得る。しかしながら、他の実施形態では、空間特性はまた、音響オブジェクトに対するオーディオソースの位置、および/または、オーデ

50

ィオがレンダリングされるリスニング位置に依存し得る。

【 0 1 3 9 】

このいくつかの例は、オーディオソースに最も近い音響オブジェクトの部分によって、またはカップリングされたオーディオを生成するのに最も効率的なオーディオソースに面する表面（たとえば、ポリゴン）によって決定される位置を含み得る。より具体的には、振動からオーディオへの伝達を記載するカップリング係数が最も高いポリゴンである。あるいは、位置は、オーディオカップリング係数に対するそれぞれの振動によって重み付けされたポリゴン位置の加重平均によって決定され得る。

【 0 1 4 0 】

ここでのポリゴンの位置は、たとえば、

【数 1 7】

$$\bar{p} = \left(\sum_{j=1}^J e_{x1,i,j}, \sum_{j=1}^J e_{y1,i,j}, \sum_{j=1}^J e_{z1,i,j} \right)$$

10

20

のように、平均的な位置で考慮することができる。

【 0 1 4 1 】

生成されたオーディオソースの空間特性は、リスニング位置に依存し得る。生成されたオブジェクトオーディオソースの位置は、リスニング位置に最も近い音響オブジェクトのポイントとして選択することができる。生成されたオブジェクトオーディオソースの空間的広がり、すなわち、リスニング位置に面する音響オブジェクトの幅と高さに関して表現され得る。同様に、この知覚された空間的広がり、2つの角度で表現され得、リスニング位置の観点から、音響オブジェクトの相対的な高さおよび幅を示す。

【 0 1 4 2 】

あるいは、位置および空間的広がり、リスニング位置に最も直接的に面している表面積によって決定され得る。たとえば、法線ベクトル

【数 1 8】

$$(\bar{n}_j)$$

40

が、ポリゴンの平均位置と、リスニング位置との間の線によって決定されるベクトル

【数 1 9】

$$(\bar{w}_j)$$

と最も一致するポリゴン k の平均位置および広がりであり、

【数 2 0】

50

$$k = \arg \min_j (\bar{n}_j \times \bar{w}_j)$$

ここで、 \times は、2つのベクトルの外積である。

【0143】

他の実施形態では、生成されたオブジェクトオーディオソースの空間特性は、元のオーディオソース位置とリスニング位置との両方に依存する。たとえば、位置は、元のオーディオソースとリスニング位置との間の直接線が、リスニング位置に最も近い音響オブジェクトと交差する点によって決定され得る。次に、生成されたオブジェクトオーディオソースの空間的広がり、その位置に関して画定され得る。

10

【0144】

以前の説明は、生成されるカップリングオーディオ信号を、音響オブジェクトに入射するオーディオに関連付ける、すなわち、オーディオが振動に変換されてからオーディオに戻る効果を反映するカップリング伝達関数が考慮されるシナリオに焦点を当ててきた。

【0145】

しかしながら、他の実施形態では、音響オブジェクトの振動運動へのオーディオの変換を指すカップリング伝達関数が提供され得る。代替的または追加的に、音響オブジェクトの振動運動のオーディオへの変換を指すカップリング伝達関数が提供され得る。

20

【0146】

多くの実施形態では、オーディオから振動へのカップリングを反映する入力カップリング伝達関数と、振動からオーディオへのカップリングを反映する出力カップリング伝達関数との両方を提供することができる。したがって、2つの部分からなるカップリングメタデータを提供することができ、一方の部分は、到来するオーディオ信号から振動信号への変換を記載し、第2の部分は、振動信号からオーディオ信号への変換を記載する。そのような場合、全体的なカップリング効果/伝達関数は、たとえば、

$$c_{a,v} = c_{a,v} * c_{v,a}$$

30

のように、入力伝達関数と出力伝達関数の組合せとして決定でき、ここで、 $*$ は、(時間領域) 畳込みを示し、 $c_{a,v}$ および $c_{v,a}$ は、それぞれ入力カップリング伝達関数および出力カップリング伝達関数を表現する時間領域フィルタである。

【0147】

個々の部分伝達関数を提供するアプローチは、多くの利点を提供し得る。オーディオソース生成器 205 は、音響オブジェクトの振動を判定し、次に、これから、オブジェクトオーディオソースから発せられるオーディオを決定するように構成され得る。これにより、より複雑で詳細な検討が可能になり得る。

【0148】

たとえば、多くの実施形態では、オーディオソース生成器 205 は、音響オブジェクトの物理モデルを考慮するように構成され得る。物理モデルは、音響オブジェクトの機械的性能および/または動作のモデルであり得る。たとえば、音響オブジェクトの構造と構成、およびこれが振動にどのように反応するかをモデル化できる。たとえば、モデルは、振動が音響オブジェクト内でどのように分布および伝播するかをモデル化するように構成され得る。

40

【0149】

モデルは、たとえば、振動がシステム内でどのように機械的に減衰されるかを含む、音響オブジェクトにおける振動を判定するために評価され得る。次に、出力カップリング伝達関数を使用して、判定された振動から発せられるオーディオを決定することにより、結果として生じるオーディオを生成することができる。別の例として、モデルは、共振動作

50

を反映でき、たとえば、振動のソースが終了した後でさえも振動が続く可能性があるリングング効果をモデル化できる。

【0150】

このアプローチは、特に、振動の空間分布と、その結果として発せられるオーディオの空間分布とを決定するために使用され得る。たとえば、特定の振動は、音響オブジェクトの1つの小さな領域において、たとえば、音響オブジェクトの小さな領域に焦点を合わせた指向性の高いオーディオソース（または、たとえば、音響オブジェクトの一部に非常に近いオーディオソース）によって誘発され得る。次に、音響オブジェクト全体の振動効果は、物理モデルを評価することによって決定することができる。音響オブジェクトはより小さな部分に分割され得、各部分について平均振動が決定され得、対応するオーディオ信号は、この振動に出力カップリング伝達関数を適用することによって決定され得る。このようにして、カップリングによって生成されたオーディオを反映する空間的に分散されたオーディオソースを生成することができる。

10

【0151】

この例では、オーディオが生成させる振動は、オーディオが、音響オブジェクトに到達した結果である。しかしながら、他の実施形態では、振動は、他のソースから生じ得る。特に、多くのシナリオでは、振動は、音響オブジェクトに加えられる時間変動する力に起因する振動であり得る。

【0152】

実際、物理的相互作用（たとえば、ノッキングなど）も、カップリング効果を介して音響オブジェクトを音ソースに変える可能性がある。たとえば、壁は、ドリル自体ではないノイズの大部分を発生させるため、壁のドリルは聞こえる。同様に、物理的に接続されたシーンオブジェクト間でもカップリングが発生する（たとえば、建物全体でドリルが聞こえるようになる）。

20

【0153】

これは、カップリング係数が、到来するエネルギーのどの部分が振動エネルギーに変換されるかを示す係数と、振動エネルギーとオーディオ信号との間の関係を記載する係数に分割されることによってモデル化され得る。

【0154】

音響オブジェクトのこの振動モデリングは、（物理的に）接続された音響オブジェクト間の振動の伝達を記載する追加の係数に基づく場合があり、したがって、このアプローチは、異なる音響オブジェクト間の機械的相互作用を評価するためにも使用できる。

30

【0155】

したがって、カップリング係数は、（音響オブジェクトに到達する音波を記載するオーディオ信号を、音響オブジェクトによって放射されるオーディオ信号に直接変換するという意味で）入力オーディオから出力オーディオへの伝達を表現できるが、2つの構成要素に分割され、一方の係数成分が、到来する音波のオーディオ信号を、音響オブジェクトにおける振動信号に変換し、他方の係数成分が、それらの振動信号を、音響オブジェクトによって放射されるオーディオ信号に変換する。

【0156】

これは、音響オブジェクトが互いに接触し、それによって、音響オブジェクトを捕捉する要素から、音波に直接曝されていない別の音響オブジェクトに振動信号を伝達できる、より高度な実施形態において有用であり得る。たとえば、ある部屋で大音量の音システムが音楽を再生している場合、2つの部屋の間の壁を通過して、隣の部屋で音楽が聞こえる場合がある。しかしながら、振動は通常、建物構造の他の部分に伝達され、壁、天井、および床が、音楽のある部屋に直接隣接していなくても、他の部屋のリスナにも音楽の低音が聞こえる。

40

【0157】

別の例は、（仮想的な）ポータブルスピーカまたは電話が、モデル化されたシーンにおいてオーディオを再生していることである。スピーカまたは電話がテーブル表面に接触す

50

るとすぐに、デバイスとテーブル表面との間のカップリングがはるかに強くなる。これは、音ソース（ラウドスピーカ）から、ケーシングへのカップリングが高いためである。しかしながら、ケーシング自体の振動はほとんど聞こえないが、これらの振動は、接触によって非常に効果的にテーブル表面に伝達される。テーブル表面は、その後、その振動をより大きな放射音に変換するサウンドボードとして機能する。

【0158】

したがって、この中間ステップを使用すると、オーディオ環境のより現実的なシミュレーションのために、音響オブジェクト間の振動信号の伝達をモデル化できる。

【0159】

データ信号で提供される正確なデータ、および使用される特定の構文は、異なる実施形態では異なる場合がある。

10

【0160】

たとえば、パラメータ帯域データの場合、パラメータ帯域定義は、事前定義されているか、または送信で示され得る。ビットストリームフィールドは、いくつかの事前定義されたバンディング定義の1つを示し得るか、または、パラメータ帯域定義は、周波数、FFT値域、またはサブ帯域インデクスの観点から、帯域境界または帯域幅としてビットストリームで明示的に伝送される。

【0161】

帯域周波数係数とフィルタ係数との両方をサポートするビットストリーム構文の例を次に示す。

20

【0162】

30

40

50

【表 1】

構文	#ビット
acousticElement()	
{	
posX;	8
posY;	8
posZ;	8
oriYaw;	6
oriPitch;	6
oriRoll;	6
sizeX;	6
sizeY;	6
sizeZ;	6
if (parBands)	1
{	
bandDefIdx;	3
for b = 0:nrBands(bandDefIdx)	
{	
couplingCoeff[b];	5
reflCoeff[b];	5

10

20

30

40

50

<code>dissCoeff[b];</code>	5	
<code>}</code>		
<code>}</code>		
<code>else /* Filter coefficients */</code>		
<code>{</code>		10
<code>fltOrder:</code>	2	
<code>for c = 1:fltOrder+1</code>		
<code>{</code>		
<code>couplingCoeffB[c];</code>	7	
<code>reflCoeffB[c];</code>	7	
<code>dissCoeffB[c];</code>	7	20
<code>}</code>		
<code>if (iirFilter)</code>	1	
<code>{</code>		
<code>for c = 1:fltOrder+1</code>		
<code>{</code>		
<code>couplingCoeffA[c];</code>	7	30
<code>reflCoeffA[c];</code>	7	
<code>dissCoeffA[c];</code>	7	
<code>}</code>		
<code>}</code>		
<code>}</code>		
<code>}</code>		40

【0163】

あるいは、音響オブジェクトメタデータ、より具体的には、カップリング係数メタデータ、および任意選択でバンディング情報を、テキストベースのファイルで記載することができる。たとえば、XMLファイルまたはJSON形式として。

【0164】

例として、音響要素3は、中心位置

【数21】

$$\bar{p}_{a,3} = (\text{pos}X, \text{pos}Y, \text{pos}Z) = (0, 4, 1.5)$$

、向き $\text{ori}Yaw = 0$ 、 $\text{ori}Pitch = 0$ 、 $\text{ori}Roll = 0$ 、および寸法
【数 2 2】

$$\bar{d} = (\text{size}X, \text{size}Y, \text{size}Z) = (20, 0.1, 3)$$

10

によって画定され得る。ビットストリームフィールド $\text{par}Bnads$ の対応する値は 1
であり、 bandDefIdx は 0 であり得、これは、以下の表に示されているバンディ
ング定義を参照することができる。

【0 1 6 5】

【表 2】

帯域インデクス	帯域開始周波数 (Hz)	帯域停止周波数 (Hz)
0	0	30
1	30	100
2	100	500
3	500	1000
4	1000	2500
5	2500	8000
6	8000	$f_s/2$

20

30

【0 1 6 6】

couplingCoef の 7 つの値は、次の表に示す通りである。

【0 1 6 7】

40

50

【表 3】

帯域インデクス(すなわち、b)	couplingCoeff[b]	$c_3[b]$ (dB)
0	9	-9
1	18	-20
2	20	-25
3	25	-50
4	27	-Inf
5	27	-Inf
6	27	-Inf

10

【 0 1 6 8 】

逆量子化された値には、以下の表からの逆量子化ベクトルが使用される。

20

【 0 1 6 9 】

30

40

50

【表 4】

量子化インデクス	カップリング係数値 (dB)
0	0
1	-1
2	-2
3	-3
4	-4
5	-5
6	-6
7	-7
8	-8
9	-9
10	-10
11	-11
12	-12
13	-13
14	-14
15	-15
16	-16
17	-18
18	-20
19	-22.5
20	-25
21	-30
22	-35
23	-40
24	-45
25	-50
26	-60
27	-Inf
28-31	予備

10

20

30

40

【 0 1 7 0 】

$S_{r,0}$ を、唯一のアクティブなオーディオソース信号の STFT 表現とし、リスニング位置が

【数 2 4】

$$\bar{p}_l = (5, 10, 1.8)$$

50

である間、位置

【数 2 3】

$$\bar{p}_{s,0} = (1, -2, 0.5)$$

における無指向性オーディオソースを表現する。生成されたオーディオオブジェクトソースの信号は、

10

【数 2 5】

$$S_{g,3}[n, k] = S_{r,i}[n, k] \cdot 10^{c_3[b]/20}$$

によって、パラメータ帯域 b におけるすべての STFT 値域 k について計算される。

【0 1 7 1】

生成されたオーディオオブジェクトソースに関連付けられた空間データは、音響オブジェクトから得られる中心位置

20

【数 2 6】

$$\bar{p}_{g,3} = (0, 4, 1.5)$$

および寸法

【数 2 7】

30

$$\bar{d}_{g,3} = (20, 0.1, 3)$$

である。上記の例では、元のオーディオソースと、音響オブジェクトとの間の距離による減衰、および、生成されたオーディオオブジェクトソースと、リスニング位置との間の距離による減衰は無視される。実際には、信号のエネルギーは、波面のますます広い領域に分散し、音ソースからの距離 r に対して、通常

40

【数 2 8】

$$\frac{1}{r^3}$$

に比例する減衰係数を引き起こす。そのような減衰ゲインは、音ソースの位置および寸法、音響オブジェクト、およびリスニング位置に基づいて計算できる。

【0 1 7 2】

50

最後に、レンダラでは、生成された信号は、たとえば、左右の時間領域フィルタにFFTを適用することによって、周波数領域HRTF ($H_{LEFT}[k]$, $H_{RIGHT}[k]$)でレンダリングされる。

$$Y_{LEFT}[n, k] = S_{g, 3}[n, k] \cdot H_{LEFT}[k]$$

$$Y_{RIGHT}[n, k] = S_{g, 3}[n, k] \cdot H_{RIGHT}[k]$$

【0173】

ここで、 Y_{LEFT} および Y_{RIGHT} は、左右のヘッドフォンスピーカの複雑なSTFT領域信号である。

【0174】

このアプローチにより、多くのシナリオにおいて、性能が向上し、レンダリングと表現とが向上し得る。性能の向上は、たとえば、多くの場合、次の特定のシナリオで達成され得る。

10

- 第1のオーディオソースは、音響オブジェクトによって直接経路においてリスナに遮られるが、空気中に有意な間接経路があり、それを介して、第1のオーディオソースのエネルギーのかなりの部分がまだリスナに到達する。

○ この場合、音響オブジェクトは、追加のオーディオソースになり、送信されたオーディオおよび反射とともに、オーディオ環境のより現実的なレンダリングを作成する。

○ 例は、リスナの視点から背後にソースが配置された部屋の仕切りである。

- 第1のオーディオソースは、リスナへの1つまたは複数の音響オブジェクトによって完全に遮られており、空中からリスナへの有意な経路はない。

20

○ この場合、音響オブジェクトは、追加のオーディオソースになり、1つまたは複数のオーディオソースを置き換えることが多いので、オーディオ環境のより現実的なレンダリングの作成と、計算の複雑さの低減との両方を行う。

○ 例は、壁と閉じたドアで区切られた2つの隣接する部屋があるシナリオである。

- 第1のオーディオソースは、音響オブジェクトと直接接触している。

○ この場合、カップリング係数は、音波とのカップリングとは異なり、オーディオソースの振動は、音響オブジェクトによってより直接的に引き継がれるため、ラウドネスの増幅を引き起こす可能性がある。

○ 例は、オーディオを再生する電話のようなスピーカがテーブルに配置され、テーブルがサウンドボードとして機能するシナリオである。

30

- 第1のオーディオソースは、音響オブジェクトによってリスナに遮られないが、音響オブジェクトにカップリングエネルギーを導入する。

○ この場合、音響オブジェクトは、第1のオーディオソースの知覚された広がりを生成する追加のオーディオソースになる。

【0175】

明確化のための上記の記載は、異なる機能回路、ユニット、およびプロセッサを参照して本発明の実施形態を記載していることが理解されよう。しかしながら、本発明を損なうことなく、異なる機能回路、ユニット、またはプロセッサ間での機能の任意の適切な分担が使用され得ることは明らかであろう。たとえば、別々のプロセッサまたはコントローラによって実行されるように示されている機能は、同じプロセッサまたはコントローラによって実行され得る。したがって、特定の機能ユニットまたは回路への言及は、厳密な論理的または物理的構造または体系を示すものではなく、記載された機能を提供するための適切な手段への言及としてのみ理解されるべきである。

40

【0176】

本発明は、ハードウェア、ソフトウェア、ファームウェア、またはこれらの任意の組合せを含む任意の適切な形態で実施することができる。本発明は、任意選択で、1つまたは複数のデータプロセッサおよび/またはデジタル信号プロセッサ上で実行するコンピュータソフトウェアとして少なくとも部分的に実施され得る。本発明の実施形態の要素および構成要素は、任意の適切な手法で物理的、機能的、および論理的に実施され得る。実際、機能は、単一のユニット、複数のユニット、または他の機能ユニットの一部として実施さ

50

れ得る。したがって、本発明は、単一のユニットで実施することができ、または異なるユニット、回路、およびプロセッサ間で物理的および機能的に分散され得る。

【0177】

本発明は、いくつかの実施形態に関連して説明されてきたが、本明細書に記載の特定の形態に限定されることは意図されていない。むしろ、本発明の範囲は、付随する特許請求の範囲によってのみ限定される。それに加えて、特徴は特定の実施形態に関連して記載されているように見え得るが、当業者は、記載された実施形態の様々な特徴が、本発明にしたがって組み合わせられ得ることを認識するであろう。請求項において、備えるという用語は、他の要素またはステップの存在を排除しない。

【0178】

さらに、個々にリストされているが、複数の手段、要素、回路、または方法のステップは、たとえば、単一の回路、ユニット、またはプロセッサによって実施され得る。それに加えて、個々の特徴は、異なる請求項に含まれ得るが、これらはおそらく、有利に組み合わせることができ、異なる請求項に含まれることは、特徴の組合せが、実現可能および/または有利ではないことを意味しない。また、請求項の1つのカテゴリに特徴が含まれていることは、このカテゴリに限定されることを意味せず、その特徴が、必要に応じて他の請求項のカテゴリにも同様に適用可能であることを示す。さらに、請求項における特徴の順序は、特徴が機能しなければならない特定の順序を意味するものではなく、特に、方法の請求項における個々のステップの順序は、ステップがこの順序で実行されなければならないことを意味しない。むしろ、これらのステップは、任意の適切な順序で実行することができる。それに加えて、単数の参照は、複数を除外しない。したがって、「第1」、「第2」などへの言及は、複数を排除するものではない。請求項における参照記号は、単に明確な例として提供されており、いずれにせよ、請求項の範囲を限定するものとして解釈されないものとする。

10

20

30

40

50

【図面】

【図 1】

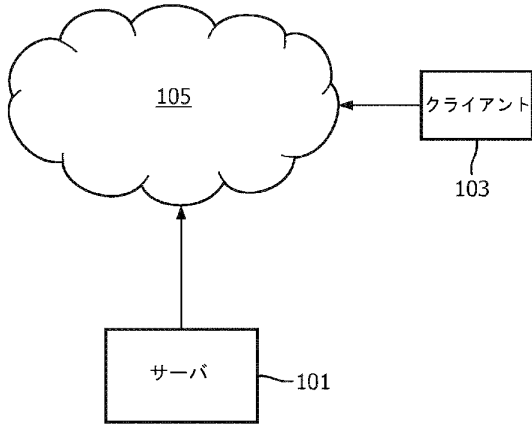


図 1

【図 2】

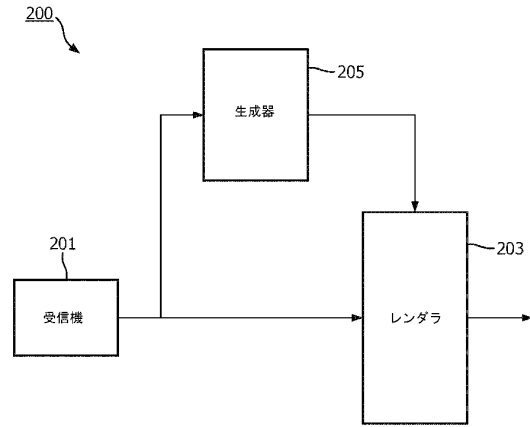


図 2

【図 3】

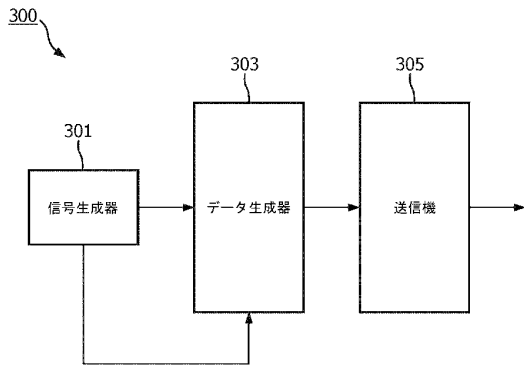


図 3

【図 4】

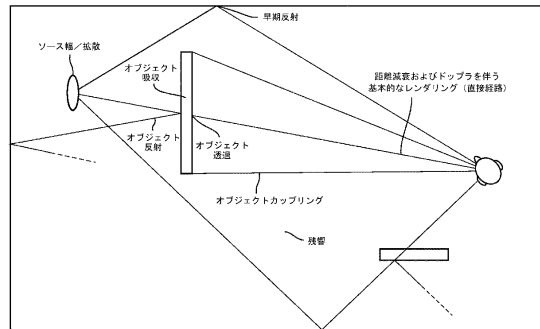


図 4

10

20

30

40

50

【図5】

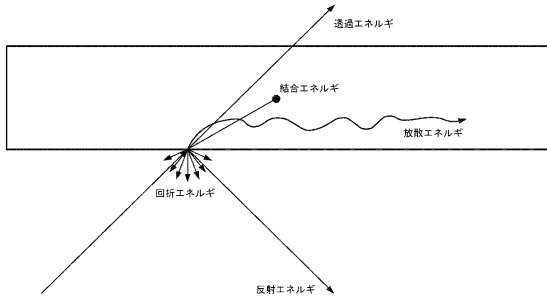


図5

【図6】

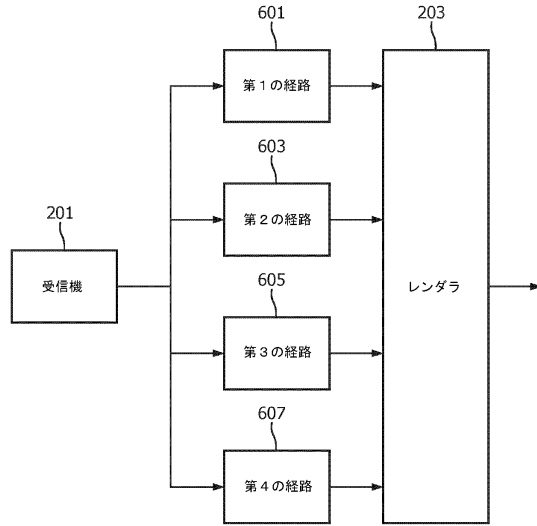


図6

【図7】

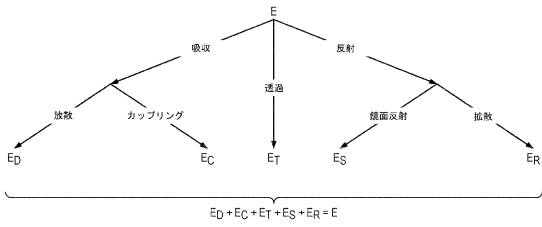


図7

【図8】

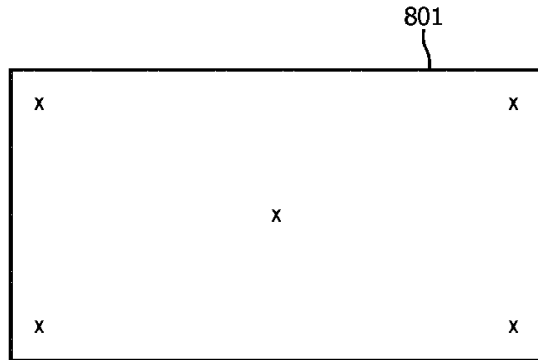


FIG. 8

10

20

30

40

50

フロントページの続き

ドーフェン ハイ テック キャンパス 5

審査官 大石 剛

(56)参考文献 特表2012-524434(JP,A)

米国特許出願公開第2015/0057083(US,A1)

(58)調査した分野 (Int.Cl., DB名)

H04S 7/00

H04R 3/00