

明 細 書

発明の名称：

音声信号処理方法、コンピュータプログラム、及び、音声信号処理装置
技術分野

[0001] 本開示は、音声信号処理方法等に関する。

背景技術

[0002] 近年、VR (Virtual Reality)、AR (Augmented Reality) 及びMR (Mixed Reality) を含むER (Extended Reality) (XRとも表現され得る) を利用した商品及びサービスの普及が進んでいる。これに伴い、仮想空間又は実空間において、仮想的な音源が発した音に対して当該空間の環境に応じて生じる音響効果を付与してイマーシブオーディオ (Immersive Audio) を受聴者に提供する音声信号処理技術の重要性が向上している。

[0003] なお、受聴者は、リスナ又はユーザとも表現され得る。また、特許文献1、特許文献2、特許文献3及び非特許文献1には、本開示の音声信号処理方法等に関連する技術が示されている。

先行技術文献

特許文献

[0004] 特許文献1：特許第6288100号公報

特許文献2：特開2019-22049号公報

特許文献3：国際公開第2021/180938号

非特許文献

[0005] 非特許文献1：B. C. J. ムーア著、「聴覚心理学概論」、誠信書房、1994/4/20、第6章：空間知覚、p. 225

発明の概要

発明が解決しようとする課題

[0006] ところで、特許文献1が示す技術では、聴覚の感度が考慮された上で、演算量及び演算負荷を適切に削減することが困難な場合がある。

[0007] そこで、本開示は、聴覚の感度が考慮された上で、演算量及び演算負荷を適切に削減することができる音声信号処理方法等を提供することを目的とする。

課題を解決するための手段

[0008] 本開示の一態様に係る音声信号処理方法は、音声信号処理装置が実行する音声信号処理方法であって、音声信号であって、前記音声信号の属性を特定する属性情報を含み前記属性が間接音を示す情報を含む音声信号を取得する取得ステップと、前記間接音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第1補正特性と、取得された前記音声信号とに基づいて、前記間接音が受聴者が居る位置である受聴位置に到来するときの間接音音量に基づく第1音量を算出する第1算出ステップと、前記間接音に係る直接音が前記受聴位置に到来するときの直接音音量に基づく第2音量を算出する第2算出ステップと、算出された前記第2音量と算出された前記第1音量との音量比、及び、前記直接音と前記間接音との到来する時間差に基づいて、取得された前記音声信号に基づく出力信号を出力するか否かを選択する選択処理ステップと、前記出力信号を出力することが選択された場合に、前記出力信号を出力する再生ステップと、を含む。

[0009] また、本開示の一態様に係る音声信号処理方法は、音声信号処理装置が実行する音声信号処理方法であって、音声信号であって、前記音声信号の属性を特定する属性情報を含み前記属性が、直接音が反射物によって反射された音である反射音を示す情報を含む音声信号を取得する取得ステップと、取得された前記音声信号が示す前記反射音が受聴者が居る位置である受聴位置に到来するときの反射音音量と前記直接音が前記受聴位置に到来するときの直接音音量との音量比、前記直接音と前記反射音との到来する時間差、及び、前記反射物の反射係数に基づいて決定される反射係数特徴量に基づいて、取得された前記音声信号に基づく出力信号を出力するか否かを選択する選択処

理ステップと、前記出力信号を出力することが選択された場合に、前記出力信号を出力する再生ステップと、を含む。

[0010] また、本開示の一態様に係る音声信号処理方法は、音声信号処理装置が実行する音声信号処理方法であって、音声信号を取得する取得ステップと、取得された前記音声信号が示す音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された所定補正特性と、取得された前記音声信号とに基づいて、前記音が受聴者が居る位置である受聴位置に到来するときの音量に基づく所定音量を算出する第3算出ステップと、算出された前記所定音量に基づいて、取得された前記音声信号に基づく出力信号を出力するか否かを選択する選択処理ステップと、前記出力信号を出力することが選択された場合に、前記出力信号を出力する再生ステップと、を含む。

[0011] また、本開示の一態様に係る音声信号処理方法は、音声信号処理装置が実行する音声信号処理方法であって、音声信号であって、前記音声信号の属性を特定する属性情報を含む音声信号を取得する取得ステップと、取得された前記音声信号が含む前記属性情報によって特定される前記属性が間接音を示す情報を含む音声信号の所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第1補正特性と、前記間接音のゲイン情報とに基づいて、前記間接音が受聴者が居る位置である受聴位置に到来するときの間接音音量に基づく第1音量を算出する第1算出ステップと、前記間接音に係る直接音の所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第2補正特性と、前記直接音のゲイン情報とに基づいて、前記受聴位置に到来するときの直接音音量に基づく第2音量を算出する第2算出ステップと、算出された前記第2音量と算出された前記第1音量との音量比、及び、前記直接音と前記間接音との到来する時間差に基づいて、取得された前記音声信号に基づく出力信号を出力するか否かを選択する選択処理ステップと、前記出力信号を出力することが選択された場合に、前記出力信号を出力する再生ステップと、を含む。

[0012] また、本開示の一態様に係るコンピュータプログラムは、上記の音声信号処理方法をコンピュータに実行させる。

[0013] また、本開示の一態様に係る音声信号処理装置は、音声信号であって、前記音声信号の属性を特定する属性情報を含み前記属性が間接音を示す情報を含む音声信号を取得する取得部と、前記間接音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第1補正特性と、取得された前記音声信号とに基づいて、前記間接音が受聴者が居る位置である受聴位置に到来するときの間接音音量に基づく第1音量を算出する第1算出部と、前記間接音に係る直接音が前記受聴位置に到来するときの直接音音量に基づく第2音量を算出する第2算出部と、算出された前記第2音量と算出された前記第1音量との音量比、及び、前記直接音と前記間接音との到来する時間差に基づいて、取得された前記音声信号に基づく出力信号を出力するか否かを選択する選択処理部と、前記出力信号を出力することが選択された場合に、前記出力信号を出力する再生部と、を備える。

[0014] なお、これらの包括的又は具体的な態様は、システム、装置、方法、集積回路、コンピュータプログラム、又は、コンピュータ読み取り可能なCD-ROM等の非一時的な記録媒体で実現されてもよく、システム、装置、方法、集積回路、コンピュータプログラム、及び、記録媒体の任意な組み合わせで実現されてもよい。

発明の効果

[0015] 本開示の一態様に係る音声信号処理方法等によれば、聴覚の感度が考慮された上で、演算量及び演算負荷を適切に削減することができる。

図面の簡単な説明

[0016] [図1]図1は、音空間で生成される直接音及び反射音の一例を示す図である。

[図2]図2は、実施の形態1における立体音響再生システムの一例を示す図である。

[図3A]図3Aは、実施の形態1における符号化装置の構成例を示すブロック図である。

[図3B]図3 Bは、実施の形態1における復号装置の構成例を示すブロック図である。

[図3C]図3 Cは、実施の形態1における符号化装置の別の構成例を示すブロック図である。

[図3D]図3 Dは、実施の形態1における復号装置の別の構成例を示すブロック図である。

[図4A]図4 Aは、実施の形態1におけるデコーダの構成例を示すブロック図である。

[図4B]図4 Bは、実施の形態1におけるデコーダの別の構成例を示すブロック図である。

[図5]図5は、実施の形態1における音声信号処理装置の物理的構成の一例を示す図である。

[図6]図6は、実施の形態1における符号化装置の物理的構成の一例を示す図である。

[図7]図7は、実施の形態1におけるレンダリング部の構成例を示すブロック図である。

[図8]図8は、実施の形態1における音声信号処理装置の動作例を示すフローチャートである。

[図9]図9は、受聴者と障害物オブジェクトとが比較的遠い位置関係を示す図である。

[図10]図10は、受聴者と障害物オブジェクトとが比較的近い位置関係を示す図である。

[図11]図11は、直接音と反射音との時間差と、閾値との関係を示す図である。

[図12A]図12 Aは、閾値データの設定方法の例の一部を示す図である。

[図12B]図12 Bは、閾値データの設定方法の例の一部を示す図である。

[図12C]図12 Cは、閾値データの設定方法の例の一部を示す図である。

[図13]図13は、閾値の設定方法の例を示す図である。

[図14]図14は、選択処理の一例を示すフローチャートである。

[図15]図15は、直接音の方向と反射音の方向と時間差と閾値との関係を示す図である。

[図16]図16は、角度差と時間差と閾値との関係を示す図である。

[図17]図17は、レンダリング部の別の構成例を示すブロック図である。

[図18]図18は、選択処理の別の例を示すフローチャートである。

[図19]図19は、選択処理のさらに別の例を示すフローチャートである。

[図20]図20は、実施の形態1における音声信号処理装置の動作の第1変形例を示すフローチャートである。

[図21]図21は、実施の形態1における音声信号処理装置の動作の第2変形例を示すフローチャートである。

[図22]図22は、アバターと音源オブジェクトと障害物オブジェクトとの配置例を示す図である。

[図23]図23は、選択処理のさらに別の例を示すフローチャートである。

[図24]図24は、レンダリング部がパイプライン処理を行うための構成例を示すブロック図である。

[図25]図25は、音の透過及び回折を示す図である。

[図26]図26は、実施の形態1に係る受聴者と障害物オブジェクトとの位置関係の一例を示す図である。

[図27]図27は、実施の形態1に係る受聴者と障害物オブジェクトとの位置関係の他の一例を示す図である。

[図28]図28は、実施の形態1に係るエコー検知限閾値の一例である。

[図29]図29は、実施の形態2に係るレンダリング部の構成例を示すブロック図である。

[図30]図30は、実施の形態2に係る音声信号処理装置の動作例を示すフローチャートである。

[図31]図31は、実施の形態2に係る選択処理の動作例を示すフローチャートである。

[図32]図3 2は、実施の形態2に係る反射音に係る所定の周波数帯域幅ごとのゲイン特性を示す図である。

[図33A]図3 3 Aは、実施の形態2に係る聴覚の感度を示す周波数特性を表す表を示す図である。

[図33B]図3 3 Bは、実施の形態2に係る聴覚の感度を示す周波数特性を示す図である。

[図34]図3 4は、実施の形態2に係る反射音に係るゲイン特性と、聴覚の感度を示す周波数特性（A特性）と、第1補正特性とを示す図である。

[図35]図3 5は、実施の形態2に係る直接音に係る所定の周波数帯域幅ごとのゲイン特性を示す図である。

[図36]図3 6は、実施の形態2に係る直接音に係るゲイン特性と、聴覚の感度を示す周波数特性（A特性）と、第2補正特性とを示す図である。

[図37]図3 7は、実施の形態2に係る受聴者の音量の感度を示す周波数特性の他の第1例である等ラウドネス曲線の逆特性を示す図である。

[図38]図3 8は、実施の形態2に係る受聴者の音量の感度を示す周波数特性の他の第2例である音源位置の最小弁別角度の周波数特性の逆特性を示す図である。

[図39]図3 9は、実施の形態3に係るレンダリング部の構成例を示すブロック図である。

[図40]図4 0は、実施の形態3に係る反射係数特徴量が第1閾値へ与える影響を示す図である。

[図41]図4 1は、実施の形態3に係る音声信号処理装置の動作例を示すフローチャートである。

[図42]図4 2は、実施の形態4に係るレンダリング部の構成例を示すブロック図である。

[図43]図4 3は、実施の形態4に係る第2閾値を示す閾値データを示すグラフである。

[図44]図4 4は、実施の形態4に係る音声信号処理装置の動作例を示すフロ

ーチャートである。

[図45]図45は、実施の形態4に係る選択処理の動作例を示すフローチャートである。

発明を実施するための形態

[0017] (本開示の基礎となった知見)

従来、仮想空間又は実空間において、仮想的な音源が発した音に対して当該空間の環境に応じて生じる音響効果を付与してイマーシブオーディオを聴者に提供する音声信号処理技術が検討されてきた。

[0018] 特許文献1には、このような音声信号処理技術が開示されている。より具体的には、特許文献1には、オーディオ信号（音声信号）の重要度を検出し、検出された重要度が低いオーディオ信号を出力しない技術が開示されている。このように、重要度が低いオーディオ信号を出力しないことにより、当該音声信号処理技術では、演算量及び演算負荷を適切に削減することが期待される。

[0019] ところで、音空間（仮想空間又は実空間）では、反射音が重要になる場合がある。

[0020] 図1は、音空間で生成される直接音及び反射音の一例を示す図である。仮想空間の特徴を音で表現する音響処理においては、空間の広さ及び壁面の材質等を表現するため、並びに、音源の位置（音像の定位）を正確に把握するために、直接音のみならず反射音を再現することが有効である。

[0021] 例えば、図1のような直方体の室内で音を聴く場合、1つの音源について6面の壁面に対応する6本の一次反射音が生じる。これらの反射音を再現することは、空間及び音像に関する適切な理解の手掛かりとなる。さらに各反射音に対し当該反射音を生じさせた反射面以外の面で二次反射音が生じる。これらの反射音も知覚的に有効な手掛かりとなる。

[0022] しかしながら、二次反射までしか考慮しない場合でさえ、1つの音源について、1本の直接音と36本（ $6 + 6 \times 5$ ）の反射音が生じるので、37本の音線が生じることになり、これらの音線を処理するには相当程度の演算量

が必要となる。

[0023] また、バーチャル会議、バーチャルショッピング又はバーチャルコンサート等のような、メタバースに関して想定されている近年の応用商品では、必然的に複数の音源が存在することになるので、さらに膨大な演算量が必要となる。

[0024] また、仮想空間内で音を受聴する受聴者は、ヘッドホン又はVRゴーグルを用いる。このような受聴者に立体音響を提供するため、両耳間に音圧比及び位相差を与えて音の到来方向及び遠近感を再現するバイノーラル（B i n a u r a l）処理が各音線に行われる。したがって、発生する全ての反射音を再現しようとするとその演算量は甚大となる。

[0025] 一方、仮想空間を体験する受聴者によって装着されるVRゴーグルのバッテリーとして、その利便性から小型の蓄電池が用いられる場合がある。その電池寿命を延ばすために、上記のような処理にかかる演算負荷は小さい方がよい。そのために、数百本の規模で発生する音線の数、音の定位及び空間の把握を損なわない範囲で削減することが望ましい。

[0026] また、音響を再生するシステムにおいて、6DoF（6 D e g r e e s o f F r e e d o m）等の自由度が受聴者の位置（つまりは受聴者が居る位置である受聴位置）及び向きに対して許容される場合がある。この場合、受聴者と音源と音を反射する物体との位置関係は、再生時（レンダリング時）でないと確定されない。そのため、反射音も、再生時でないと確定されない。よって、処理対象の反射音を予め定めることは困難である。

[0027] したがって、再生時に音空間において生じる複数の反射音のうち処理対象又は非処理対象の1つ以上の反射音を適切に選択して出力（再生）することは、演算量及び演算負荷の適切な削減に有用である。

[0028] なお、音を選択するか否かを制御することは、音を選択するか否かを判定することに対応し、より具体的には、音を選択して出力（再生）するか否かを判定することに対応する。また、音を選択することは、音を処理対象音として選択することであってもよいし、音を非処理対象音として選択すること

であってもよい。

[0029] ところで、特許文献1においては、オーディオ信号の重要度、より具体的には当該オーディオ信号が示す直接音の重要度を検出するものの、反射音の重要度は検討されていない。このため、図1が示すように反射音等の間接音が発生する場合には、演算量及び演算負荷が増加してしまい、つまりは、演算量及び演算負荷を適切に削減することが困難な場合がある。

[0030] また、特許文献1に開示される技術を含む従来の技術においては、受聴者の聴覚の感度が考慮されずに、オーディオ信号を出力するか否かが選択される。そのようなオーディオ信号が出力され当該オーディオ信号が示す音を受聴者が受聴すると、受聴者は、自身の聴覚の感覚とは異なる音を受聴することとなるため、違和感を覚えてしまう。

[0031] 従って、音空間において、聴覚の感度が考慮された上で、演算量及び演算負荷を適切に削減することができる音声信号処理方法等が求められている。

[0032] そこで、本開示の第1態様に係る音声信号処理方法は、音声信号処理装置が実行する音声信号処理方法であって、音声信号であって、前記音声信号の属性を特定する属性情報を含み前記属性が間接音を示す情報を含む音声信号を取得する取得ステップと、前記間接音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第1補正特性と、取得された前記音声信号とに基づいて、前記間接音が受聴者が居る位置である受聴位置に到来するときの間接音音量に基づく第1音量を算出する第1算出ステップと、前記間接音に係る直接音が前記受聴位置に到来するときの直接音音量に基づく第2音量を算出する第2算出ステップと、算出された前記第2音量と算出された前記第1音量との音量比、及び、前記直接音と前記間接音との到来する時間差に基づいて、取得された前記音声信号に基づく出力信号を出力するか否かを選択する選択処理ステップと、前記出力信号を出力することが選択された場合に、前記出力信号を出力する再生ステップと、を含む。

[0033] これにより、直接音音量に基づく第2音量と間接音音量に基づく第1音量

との音量比、及び、上記時間差に基づいて、間接音を示す音声信号に基づく出力信号を出力するか否かが選択される。すなわち、当該音声信号に基づく出力信号が出力されるか否かが適切に選択される。出力信号が出力されない場合には、演算量及び演算負荷が削減される。つまりは、演算量及び演算負荷を適切に削減することができる音声信号処理方法を実現することができる。

[0034] そしてさらに、聴覚の感度を示す周波数特性が考慮された上で第1音量が算出され、算出された第1音量が用いられた音量比、及び、上記時間差に基づいて、出力信号を出力するか否かが選択される。すなわち、聴覚の感度が考慮された上で、演算量及び演算負荷を適切に削減することができる音声信号処理方法を実現することができる。

[0035] 本開示の第2態様に係る音声信号処理方法は、第1態様に係る音声信号処理方法において、前記第2算出ステップでは、前記直接音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第2補正特性に基づいて、前記第2音量を算出する。

[0036] これにより、聴覚の感度を示す周波数特性が考慮された上で第2音量が算出され、算出された第2音量が用いられた音量比、及び、上記時間差に基づいて、出力信号を出力するか否かが選択される。すなわち、聴覚の感度がより考慮された上で、演算量及び演算負荷を適切に削減することができる音声信号処理方法を実現することができる。

[0037] 本開示の第3態様に係る音声信号処理方法は、第1又は第2態様に係る音声信号処理方法において、前記聴覚の感度を示す前記周波数特性は、前記受聴者の音量の感度を示す周波数特性である。

[0038] これにより、聴覚の感度を示す周波数特性として、受聴者の音量の感度を示す周波数特性を利用することができる音声信号処理方法を実現することができる。

[0039] 本開示の第4態様に係る音声信号処理方法は、第3態様に係る音声信号処理方法において、前記音量の感度を示す前記周波数特性は、等ラウドネス曲

線の逆数に応じた周波数特性である。

[0040] これにより、音量の感度を示す周波数特性として、等ラウドネス曲線の逆数に応じた周波数特性を利用することができる音声信号処理方法を実現することができる。

[0041] 本開示の第5態様に係る音声信号処理方法は、第3態様に係る音声信号処理方法において、前記音量の感度を示す前記周波数特性は、A特性である。

[0042] これにより、音量の感度を示す周波数特性として、A特性を利用することができる音声信号処理方法を実現することができる。

[0043] 本開示の第6態様に係る音声信号処理方法は、第3態様に係る音声信号処理方法において、前記音量の感度を示す前記周波数特性は、音源位置の最小弁別角度の周波数特性の逆特性である。

[0044] これにより、音量の感度を示す周波数特性として、音源位置の最小弁別角度の周波数特性の逆特性を利用することができる音声信号処理方法を実現することができる。

[0045] 本開示の第7態様に係る音声信号処理方法は、音声信号処理装置が実行する音声信号処理方法であって、音声信号であって、前記音声信号の属性を特定する属性情報を含み前記属性が、直接音が反射物によって反射された音である反射音を示す情報を含む音声信号を取得する取得ステップと、取得された前記音声信号が示す前記反射音が受聴者が居る位置である受聴位置に到来するときの反射音音量と前記直接音が前記受聴位置に到来するときの直接音音量との音量比、前記直接音と前記反射音との到来する時間差、及び、前記反射物の反射係数に基づいて決定される反射係数特徴量に基づいて、取得された前記音声信号に基づく出力信号を出力するか否かを選択する選択処理ステップと、前記出力信号を出力することが選択された場合に、前記出力信号を出力する再生ステップと、を含む。

[0046] これにより、上記音量比、及び、上記時間差に基づいて、反射音を示す音声信号に基づく出力信号を出力するか否かが選択される。すなわち、当該音声信号に基づく出力信号が出力されるか否かが適切に選択される。出力信号

が出力されない場合には、演算量及び演算負荷が削減される。つまりは、演算量及び演算負荷を適切に削減することができる音声信号処理方法を実現することができる。

[0047] ここで、先行音効果に着目する。先行音効果は、先行音（例えば直接音）の周波数スペクトルと、後続音（例えば反射音）の周波数スペクトルとが近似している場合に生じる、と言われている。反射音の周波数スペクトルは反射物の反射係数に応じて変化するため、直接音の周波数スペクトルと反射音の周波数スペクトルとが近似しているか否かは、反射係数特徴量に応じて変化する。

[0048] 例えば、反射係数特徴量がある値の場合には直接音の周波数スペクトルと反射音の周波数スペクトルとが近似し、先行音効果が生じやすく、この場合に出力信号が出力され難いように選択されるとよい。また例えば、反射係数特徴量が他のある値の場合には直接音の周波数スペクトルと反射音の周波数スペクトルとが近似せず、先行音効果が生じにくく、この場合に出力信号が出力され易いように選択されるとよい。

[0049] すなわち、反射係数特徴量に基づいて出力信号が出力されるか否かが選択されることは、先行音効果を考慮したうえで、出力信号が出力されるか否かが選択されることに等しい。先行音効果は聴覚の感度の特性の一例であることから、聴覚の感度が考慮された上で、演算量及び演算負荷を適切に削減することができる音声信号処理方法を実現することができる。

[0050] 本開示の第8態様に係る音声信号処理方法は、第7態様に係る音声信号処理方法において、前記反射係数特徴量は、反射係数の周波数特性の平坦さの程度を示す特徴量である。

[0051] これにより、反射係数特徴量として、反射係数の周波数特性の平坦さの程度を示す特徴量を利用することができる音声信号処理方法を実現することができる。

[0052] 本開示の第9態様に係る音声信号処理方法は、音声信号処理装置が実行する音声信号処理方法であって、音声信号を取得する取得ステップと、取得さ

れた前記音声信号が示す音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された所定補正特性と、取得された前記音声信号とに基づいて、前記音が受聴者が居る位置である受聴位置に到来するときの音量に基づく所定音量を算出する第3算出ステップと、算出された前記所定音量に基づいて、取得された前記音声信号に基づく出力信号を出力するか否かを選択する選択処理ステップと、前記出力信号を出力することが選択された場合に、前記出力信号を出力する再生ステップと、を含む。

[0053] これにより、聴覚の感度を示す周波数特性が考慮された上で所定音量が算出され、算出された所定音量に基づいて、音を示す音声信号に基づく出力信号を出力するか否かが選択される。すなわち、聴覚の感度が考慮された上で、当該音声信号に基づく出力信号が出力されるか否かが適切に選択される。出力信号が出力されない場合には、演算量及び演算負荷が削減される。つまりは、聴覚の感度が考慮された上で、演算量及び演算負荷を適切に削減することができる音声信号処理方法を実現することができる。

[0054] 本開示の第10態様に係る音声信号処理方法は、音声信号処理装置が実行する音声信号処理方法であって、音声信号であって、前記音声信号の属性を特定する属性情報を含む音声信号を取得する取得ステップと、取得された前記音声信号が含む前記属性情報によって特定される前記属性が間接音を示す情報を含む音声信号の所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第1補正特性と、前記間接音のゲイン情報とに基づいて、前記間接音が受聴者が居る位置である受聴位置に到来するときの間接音音量に基づく第1音量を算出する第1算出ステップと、前記間接音に係る直接音の所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第2補正特性と、前記直接音のゲイン情報とに基づいて、前記受聴位置に到来するときの直接音音量に基づく第2音量を算出する第2算出ステップと、算出された前記第2音量と算出された前記第1音量との音量比、及び、前記直接音と前記間接音との到来する時間差に基づいて、取得された前記音声信号に基づく出力信号を出力するか否かを選

択する選択処理ステップと、前記出力信号を出力することが選択された場合に、前記出力信号を出力する再生ステップと、を含む。

[0055] これにより、直接音音量に基づく第2音量と間接音音量に基づく第1音量との音量比、及び、上記時間差に基づいて、間接音を示す音声信号に基づく出力信号を出力するか否かが選択される。すなわち、当該音声信号に基づく出力信号が出力されるか否かが適切に選択される。出力信号が出力されない場合には、演算量及び演算負荷が削減される。つまりは、演算量及び演算負荷を適切に削減することができる音声信号処理方法を実現することができる。

[0056] 本開示の第11態様に係る音声信号処理方法は、第10態様に係る音声信号処理方法において、前記間接音に係る所定の周波数帯域幅ごとのゲイン特性は、前記間接音に係る前記属性情報として保持されており、前記間接音に係るゲイン情報は、前記間接音に係る前記属性情報として保持されており、前記直接音に係る所定の周波数帯域幅ごとのゲイン特性は、前記直接音に係る前記属性情報として保持されており、前記直接音に係るゲイン情報は、前記直接音に係る前記属性情報として保持されている。

[0057] これにより、種々の情報が属性情報として保持されている音声信号処理方法を実現することができる。

[0058] 本開示の第12態様に係るコンピュータプログラムは、第1～第11態様のいずれか1の態様に係る音声信号処理方法をコンピュータに実行させるためのコンピュータプログラムである。

[0059] これにより、コンピュータが、コンピュータプログラムに従って、上記の音声信号処理方法を実行することができる。

[0060] 本開示の第13態様に係る音声信号処理装置は、音声信号であって、前記音声信号の属性を特定する属性情報を含み前記属性が間接音を示す情報を含む音声信号を取得する取得部と、前記間接音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第1補正特性と、取得された前記音声信号とに基づいて、前記間接音が受聴者が居る位

置である受聴位置に到来するときの間接音音量に基づく第1音量を算出する第1算出部と、前記間接音に係る直接音が前記受聴位置に到来するときの直接音音量に基づく第2音量を算出する第2算出部と、算出された前記第2音量と算出された前記第1音量との音量比、及び、前記直接音と前記間接音との到来する時間差に基づいて、取得された前記音声信号に基づく出力信号を出力するか否かを選択する選択処理部と、前記出力信号を出力することが選択された場合に、前記出力信号を出力する再生部と、を備える。

[0061] これにより、直接音音量に基づく第2音量と間接音音量に基づく第1音量との音量比、及び、上記時間差に基づいて、間接音を示す音声信号に基づく出力信号を出力するか否かが選択される。すなわち、当該音声信号に基づく出力信号が出力されるか否かが適切に選択される。出力信号が出力されない場合には、演算量及び演算負荷が削減される。つまりは、演算量及び演算負荷を適切に削減することができる音声信号処理装置を実現することができる。

[0062] そしてさらに、聴覚の感度を示す周波数特性が考慮された上で第1音量が算出され、算出された第1音量が用いられた音量比、及び、上記時間差に基づいて、出力信号を出力するか否かが選択される。すなわち、聴覚の感度が考慮された上で、演算量及び演算負荷を適切に削減することができる音声信号処理装置を実現することができる。

[0063] (実施の形態1)

(立体音響再生システムの例)

図2は、立体音響再生システム1000の一例を示す図である。具体的には、図2は、本開示の音響処理又は復号処理が適用可能なシステムの一例である立体音響再生システム1000を示す。立体音響は、イマーシブオーディオ(Immersive Audio)とも表現される。立体音響再生システム1000は、音声信号処理装置1001と音声提示装置1002を含む。

[0064] 音声信号処理装置1001は、音響処理装置とも表現され、仮想音源が発

した音声信号に対して音響処理を施して、受聴者に対して提示される音響処理後の音声信号を生成する。音声信号は声に限らず可聴音であればよい。音響処理とは、例えば、音が、音源で発生してから受聴者に届くまでの間に受ける１つ以上の作用を再現するために、音声信号に対して施される信号処理である。

[0065] 音声信号処理装置１００１は、上述した作用を引き起こす要因を記述した空間情報に基づいて音響処理を実施する。空間情報は、例えば、音源、受聴者、及び、周囲のオブジェクトの位置を示す情報、空間の形状を示す情報、並びに、音の伝搬に関するパラメータ等を含む。音声信号処理装置１００１は、例えば、ＰＣ（Personal Computer）、スマートフォン、タブレット又はゲームコンソール等である。

[0066] 音響処理後の信号は、音声提示装置１００２から受聴者に提示される。音声提示装置１００２は、無線又は有線の通信を介して音声信号処理装置１００１と接続されている。音声信号処理装置１００１で生成された音響処理後の音声信号は、無線又は有線の通信を介して音声提示装置１００２に伝送される。

[0067] 音声提示装置１００２が、例えば、右耳用のデバイス及び左耳用のデバイス等の複数のデバイスで構成されている場合、複数のデバイス間の通信、又は、複数のデバイスのそれぞれと音声信号処理装置１００１との通信によって、複数のデバイスは同期して音を提示する。音声提示装置１００２は、例えば、受聴者の頭部に装着されるヘッドホン、イヤホン、ヘッドマウントディスプレイ、又は、固定された複数のスピーカで構成されたサラウンドスピーカ等である。

[0068] なお、立体音響再生システム１０００は、視覚的にＡＲ／ＶＲを含むＥＲ体験を提供する画像提示装置又は立体映像提示装置と組み合わせて用いられてもよい。例えば、空間情報によって扱われる空間は、仮想空間であって、当該空間における音源、受聴者及びオブジェクトの位置は、仮想空間における仮想的な音源、仮想的な受聴者及び仮想的なオブジェクトの仮想的な位置

である。当該空間は、音空間とも表現され得る。また、空間情報は、音空間情報とも表現され得る。

[0069] また、図2は、音声信号処理装置1001と音声提示装置1002とが別々の装置であるシステム構成例を示しているが、本開示の音響処理方法（音声信号処理方法）又は復号方法が適用可能な立体音響再生システム1000は図2の構成に限定されない。例えば、音声信号処理装置1001が音声提示装置1002に含まれ、音声提示装置1002が音響処理と音の提示との両方を行ってもよい。

[0070] また、音声信号処理装置1001と音声提示装置1002とが本開示で説明する音響処理を分担して実施してもよい。また、音声信号処理装置1001又は音声提示装置1002とネットワークを介して接続されたサーバが、本開示で説明する音響処理の一部又は全体を実施してもよい。

[0071] また、音声信号処理装置1001は、音声信号と音響処理に用いる空間情報とのデータの少なくとも一部を符号化することで生成されたビットストリームを復号して、音響処理を実施してもよい。したがって、音声信号処理装置1001は、復号装置と表現されてもよい。

[0072] （符号化装置の例）

図3Aは、符号化装置1100の構成例を示すブロック図である。具体的には、図3Aは、本開示の符号化装置の一例である符号化装置1100の構成を示す。

[0073] 入力データ1101は、エンコーダ1102に入力される空間情報及び／又は音声信号を含む符号化対象データである。空間情報の詳細については後で説明する。

[0074] エンコーダ1102は、入力データ1101を符号化して、符号化データ1103を生成する。符号化データ1103は、例えば、符号化処理によって生成されたビットストリームである。

[0075] メモリ1104は、符号化データ1103を記憶する。メモリ1104は、例えば、ハードディスク又はSSD（Solid State Drive）である。

e)であってもよいし、その他のメモリであってもよい。

[0076] なお、上記説明ではメモリ1104に記憶される符号化データ1103の一例として符号化処理によって生成されたビットストリームが挙げられているが、符号化データ1103は、ビットストリーム以外のデータであってもよい。例えば、符号化装置1100は、ビットストリームを所定のデータフォーマットに変換して生成された変換後のデータをメモリ1104に記憶してもよい。変換後のデータは、例えば、1つ以上のビットストリームに対応するファイル又は多重化ストリームであってもよい。

[0077] ここで、ファイルは、例えばISO BMFF (ISO Base Media File Format) 等のファイルフォーマットを有するファイルである。また、符号化データ1103は、上記のビットストリーム又はファイルを分割して生成された複数のパケットの形式であってもよい。

[0078] 例えば、エンコーダ1102で生成されたビットストリームが、ビットストリームとは異なるデータに変換されてもよい。この場合、符号化装置1100は、図示されていない変換部を備え、変換部で変換処理を行ってもよいし、後述のプロセッサの例であるCPU (Central Processing Unit) で変換処理を行ってもよい。

[0079] (復号装置の例)

図3Bは、復号装置1110の構成例を示すブロック図である。具体的には、図3Bは、本開示の復号装置の一例である復号装置1110の構成を示す。

[0080] メモリ1114は、例えば、符号化装置1100で生成された符号化データ1103と同じデータを記憶している。記憶されているデータは、メモリ1114から読み出され、デコーダ1112に入力データ1113として入力される。入力データ1113は、例えば、復号対象となるビットストリームである。メモリ1114は、例えば、ハードディスク又はSSDであってもよいし、その他のメモリであってもよい。

[0081] なお、復号装置1110は、メモリ1114から読み出されたデータをそ

のまま入力データ 1 1 1 3 としてデコーダ 1 1 1 2 に入力せず、読み出されたデータを変換し、変換後のデータを入力データ 1 1 1 3 としてデコーダ 1 1 1 2 に入力してもよい。変換前のデータは、例えば、1 つ以上のビットストリームを含む多重化データであってもよい。ここで、多重化データは、例えば I S O B M F F 等のファイルフォーマットを有するファイルであってもよい。

[0082] また、変換前のデータは、上記のビットストリーム又はファイルを分割することで生成される複数のパケットであってもよい。メモリ 1 1 1 4 からビットストリームとは異なるデータが読み出され、当該データがビットストリームに変換されてもよい。この場合、復号装置 1 1 1 0 は、図示されていない変換部を備え、変換部で変換処理を行ってもよいし、後述のプロセッサの例である C P U で変換処理を行ってもよい。

[0083] デコーダ 1 1 1 2 は、入力データ 1 1 1 3 を復号して、受聴者に提示される音声を示す音声信号 1 1 1 1 を生成する。

[0084] (符号化装置の別の例)

図 3 C は、符号化装置の別の構成例を示すブロック図である。具体的には、図 3 C は、本開示の符号化装置の別の一例である符号化装置 1 1 2 0 の構成を示す。図 3 C では、図 3 A の構成要素と同じ構成要素に図 3 A の符号と同じ符号を付しており、これらの構成要素については説明を省略する。

[0085] 符号化装置 1 1 0 0 は、符号化データ 1 1 0 3 をメモリ 1 1 0 4 に記憶している。一方、符号化装置 1 1 2 0 は、符号化データ 1 1 0 3 を外部へ送信する送信部 1 1 2 1 を備える点で符号化装置 1 1 0 0 とは異なる。

[0086] 送信部 1 1 2 1 は、符号化データ 1 1 0 3 又は符号化データ 1 1 0 3 から別のデータ形式に変換されたデータに基づいて生成された送信信号 1 1 2 2 を別の装置又はサーバに対して送信する。送信信号 1 1 2 2 の生成に用いられるデータは、例えば、符号化装置 1 1 0 0 で説明したビットストリーム、多重化データ、ファイル又はパケットである。

[0087] (復号装置の別の例)

図3Dは、復号装置の別の構成例を示すブロック図である。具体的には、図3Dは、本開示の復号装置の別の一例である復号装置1130の構成を示す。図3Dでは、図3Bの構成要素と同じ構成要素に図3Bの符号と同じ符号を付しており、これらの構成要素については説明を省略する。

[0088] 復号装置1110は、入力データ1113をメモリ1114から読み出している。一方、復号装置1130は、入力データ1113を外部から受信する受信部1131を備える点で復号装置1110とは異なる。

[0089] 受信部1131は、受信信号1132を受信して受信データを取得し、デコーダ1112に入力される入力データ1113を出力する。受信データは、デコーダ1112に入力される入力データ1113と同じであってもよいし、入力データ1113とは異なるデータ形式のデータであってもよい。

[0090] 受信データのデータ形式が、入力データ1113のデータ形式とは異なる場合、受信部1131が、受信データを入力データ1113に変換してもよい。あるいは、復号装置1130の図示されていない変換部又はCPUが、受信データを入力データ1113に変換してもよい。受信データは、例えば、符号化装置1120で説明したビットストリーム、多重化データ、ファイル又はパケットである。

[0091] (デコーダの例)

図4Aは、デコーダ1200の構成例を示すブロック図である。具体的には、図4Aは、図3B又は図3Dにおけるデコーダ1112の一例であるデコーダ1200の構成を示す。

[0092] 入力データ1113は、符号化されたビットストリームであり、符号化された音声信号である符号化音声データと、音響処理に用いられるメタデータとを含んでいる。

[0093] 空間情報管理部1201は、入力データ1113に含まれるメタデータを取得して、メタデータを解析する。メタデータは、音空間に配置された音に作用する要素を記述した情報を含む。空間情報管理部1201は、メタデータを解析して得られた音響処理に用いられる空間情報を管理し、レンダリン

グ部 1 2 0 3 に対して空間情報を提供する。

- [0094] なお、本開示では音響処理に用いられる情報が、空間情報と表現されているが、その他の表現が用いられてもよい。例えば、音響処理に用いられる情報は、音空間情報と表現されてもよいし、シーン情報と表現されてもよい。また、音響処理に用いられる情報が経時的に変化する場合、レンダリング部 1 2 0 3 に入力される空間情報は、空間状態、音空間状態又はシーン状態等と表現される情報であってもよい。
- [0095] また、空間情報は、音空間毎又はシーン毎に管理されていてもよい。例えば、互いに異なる複数の部屋のそれぞれが仮想空間として表現される場合、複数の部屋が、それぞれ、互いに異なる複数のシーンとして管理されてもよい。また、同じ空間でも表現される状況に応じて異なるシーンとして空間情報が管理されてもよい。
- [0096] したがって、複数の音空間又は複数のシーンに対して、複数の空間情報が管理されてもよい。複数の空間情報の管理において、複数の空間情報のそれぞれを識別する識別子が空間情報に付与されていてもよい。
- [0097] 空間情報のデータは、入力データ 1 1 1 3 の一例であるビットストリームに含まれていてもよい。あるいは、ビットストリームが空間情報の識別子を含み、空間情報のデータは、ビットストリーム以外の情報源から取得されてもよい。具体的には、ビットストリームが空間情報の識別子のみを含む場合、レンダリングにおいて、空間情報の識別子を用いて、装置内のメモリ又は外部のサーバに記憶された空間情報のデータが入力データ 1 1 1 3 として取得されてもよい。
- [0098] なお、空間情報管理部 1 2 0 1 が管理する情報は、ビットストリームに含まれる情報に限定されない。例えば、入力データ 1 1 1 3 は、ビットストリームには含まれないデータとして、VR又はARを提供するソフトウェア又はサーバから取得された空間の特性及び構造を示すデータを含んでいてもよい。
- [0099] また、入力データ 1 1 1 3 は、受聴者又はオブジェクトの特性及び位置等

を示すデータを含んでもよい。また、入力データ1113は、受聴者の位置について復号装置(1110、1130)を含む端末が備えるセンサで取得された情報を含んでもよいし、センサで取得された情報に基づいて推定された端末の位置を示す情報を含んでもよい。

[0100] つまり、空間情報管理部1201は、外部のシステム又はサーバと通信し、空間情報及び受聴者位置(つまりは受聴位置)を取得してもよい。また、空間情報管理部1201は、外部のシステムからクロック同期情報を取得し、レンダリング部1203のクロックと同期する処理を実行してもよい。

[0101] なお、上記の説明における空間は、仮想的に形成された空間、つまりVR空間であってもよいし、実空間又は実空間に対応する仮想空間、つまりAR空間又はMR空間であってもよい。また、仮想空間は、音場又は音空間と表現されてもよい。また、上記の説明における位置を示す情報は、空間内における位置を示す座標値等の情報であってもよいし、所定の基準位置に対する相対位置を示す情報であってもよいし、空間内の位置の動き又は加速度を示す情報であってもよい。

[0102] 音声データデコーダ1202は、入力データ1113に含まれる符号化音声データを復号して、音声信号を取得する。

[0103] 立体音響再生システム1000が取得する符号化音声データは、例えば、MPEG-H 3D Audio (ISO/IEC 23008-3)等の所定の形式で符号化されたビットストリームである。なお、MPEG-H 3D Audioは、あくまでビットストリームに含まれる符号化音声データを生成する際に利用可能な符号化方式の一例である。符号化音声データは、他の符号化方式で符号化されたビットストリームであってもよい。

[0104] 例えば、符号化方式は、MP3 (MPEG-1 Audio Layer-3)、AAC (Advanced Audio Coding)、WMA (Windows Media Audio)、AC3 (Audio Codec-3)又はVorbis等の非可逆コーデックであってもよい。あるいは、符号化方式は、ALAC (Apple Lossless Audi

o Codec) 又はFLAC (Free Lossless Audio Codec) 等の可逆コーデックであってもよい。

- [0105] あるいは、上記以外の任意の符号化方式が用いられてもよい。例えば、PCMデータが符号化音声データの種類であってもよい。この場合、復号処理は、例えば、当該PCMデータの量子化ビット数がNである場合、Nビットの二進数を、レンダリング部1203が処理できる数形式（例えば浮動小数点形式）に変換する処理であってもよい。
- [0106] レンダリング部1203は、音声信号と空間情報とを取得し、空間情報を用いて音声信号に音響処理を施して、音響処理後の音声信号（音声信号1111）を出力する。
- [0107] 空間情報管理部1201は、レンダリングを開始する前に、入力信号のメタデータを読み込み、空間情報で規定されたオブジェクト及び音等のレンダリングアイテムを検出し、レンダリング部1203に送信する。レンダリング開始後、空間情報管理部1201は、空間情報及び受聴者の位置の経時的な変化を把握し、空間情報を更新して管理する。そして、更新された空間情報をレンダリング部1203に送信する。
- [0108] レンダリング部1203は、入力データ1113に含まれる音声信号と、空間情報管理部1201から受信した空間情報とに基づいて、音響処理を付加した音声信号を生成し出力する。
- [0109] 空間情報の更新処理と、音響処理を付加した音声信号の出力処理とは、同じスレッドで実行されてもよい。また、空間情報管理部1201とレンダリング部1203とは、それぞれ独立したスレッドに処理を配分してもよい。空間情報管理部1201とレンダリング部1203とは、空間情報の更新処理と、音響処理を付加した音声信号の出力処理とを異なるスレッドで実行する場合、スレッドの起動頻度を個々に設定してもよいし、並行して処理を実行してもよい。
- [0110] 空間情報管理部1201とレンダリング部1203とが、異なる独立したスレッドで処理を実行する場合、レンダリング部1203に優先的に演算資

源を割り当てることが可能である。これにより、僅かな遅延も許容されないような、例えば、1 サンプル (0.02 msec) の遅延でプチっというノイズが発生するような出音処理を安全に実行することが可能である。

- [0111] その際、空間情報管理部1201には、演算資源の割り当てが制限される。しかし、空間情報の更新は、音声信号の出力処理と比較して、低頻度の処理（例えば、受聴者の顔の向き of 更新のような処理）であるため、音声信号の出力処理のように必ずしも瞬間的に行われなくてもよい。したがって、演算資源の割り当てが制限されても、音響的な品質に大きな影響はない。
- [0112] 空間情報の更新は、予め設定された時間又は期間毎に定期的に実行されてもよいし、予め設定された条件が満たされた場合に実行されてもよい。また、空間情報の更新は、受聴者又は音空間の管理者によって手動で実行されてもよいし、外部システムの変化をトリガとして実行されてもよい。
- [0113] 例えば、受聴者によってコントローラが操作され、自身のアバターの立ち位置が瞬間的にワープしたり、時刻が瞬時に進められたり戻されたりした場合に、空間情報が更新されてもよい。あるいは、仮想空間の管理者によって、突如、場の環境を変更するような演出が施された場合、空間情報が更新されてもよい。これらの場合において、空間情報管理部1201によって管理される空間情報を更新するためのスレッドが、定期的な起動に加えて、単発的な割り込み処理として起動されてもよい。
- [0114] 空間情報の更新処理を実行する情報更新スレッドが担う役割は、例えば、受聴者が装着しているVRゴーグルの位置又は向きに基づいて、仮想空間内に配置された受聴者のアバターの位置又は向きを更新する処理、及び、仮想空間内を移動している物体の位置の更新等であり、数10Hz程度の比較的 low 頻度で起動する処理スレッド内で賄われるものである。そのような、発生頻度の低い処理スレッドで直接音の性質を反映させる処理が行われるようにしてもよい。それは、オーディオ出力のためのオーディオ処理フレームの発生頻度より直接音の性質が変動する頻度が低いためである。むしろそうすることで、当該処理の演算負荷を相対的に小さくすることができるし、不必要

に速い頻度で情報を更新するとパルス状のノイズが発生するリスクが生じるので、そのリスクを回避することもできる。

[0115] 図4 Bは、デコーダの別の構成例を示すブロック図である。具体的には、図4 Bは、図3 B又は図3 Dにおけるデコーダ1 1 1 2の別の一例であるデコーダ1 2 1 0の構成を示す。

[0116] 図4 Bは、入力データ1 1 1 3が、符号化音声データではなく符号化されていない音声信号を含んでいる点で図4 Aと異なる。入力データ1 1 1 3は、メタデータを含むビットストリームと音声信号を含む。

[0117] 空間情報管理部1 2 1 1は、図4 Aの空間情報管理部1 2 0 1と同じであるため説明を省略する。

[0118] レンダリング部1 2 1 3は、図4 Aのレンダリング部1 2 0 3と同じであるため説明を省略する。

[0119] なお、デコーダ1 1 1 2、1 2 0 0及び1 2 1 0は、音響処理を実施する音響処理部と表現されてもよい。また、復号装置1 1 1 0及び1 1 3 0は、音声信号処理装置1 0 0 1であってもよく、音響処理装置と表現されてもよい。

[0120] (音声信号処理装置の物理的構成)

図5は、音声信号処理装置1 0 0 1の物理的構成の一例を示す図である。なお、図5の音声信号処理装置1 0 0 1は、図3 Bの復号装置1 1 1 0又は図3 Dの復号装置1 1 3 0であってもよい。図3 B又は図3 Dに示された複数の構成要素は、図5に示された複数の構成要素によって実装されてもよい。また、ここで説明する構成の一部は音声提示装置1 0 0 2に備えられていてもよい。

[0121] 図5の音声信号処理装置1 0 0 1は、プロセッサ1 4 0 2と、メモリ1 4 0 4と、通信IF (Interface) 1 4 0 3と、センサ1 4 0 5と、スピーカ1 4 0 1とを備える。

[0122] プロセッサ1 4 0 2は、例えば、CPU、DSP (Digital Signal Processor) 又はGPU (Graphics Processor) である。

essing Unit) である。当該CPU、DSP又はGPUが、メモリ1404に記憶されたプログラムを実行することで本開示の音響処理又は復号処理を実施してもよい。また、プロセッサ1402は、例えば、情報処理を行う回路である。プロセッサ1402は、本開示の音響処理を含む音声信号に対する信号処理を行う専用回路であってもよい。

[0123] メモリ1404は、例えば、RAM(Random Access Memory)又はROM(Read Only Memory)で構成される。メモリ1404は、ハードディスクに代表される磁気記録媒体又はSSDに代表される半導体メモリ等を含んでいてもよい。また、メモリ1404は、CPU又はGPUに組み込まれた内部メモリであってもよい。また、メモリ1404には、空間情報管理部1201及び1211で管理される空間情報等が記憶されてもよい。また、後述の閾値データが記憶されてもよい。

[0124] 通信IF1403は、例えば、Bluetooth(登録商標)又はWiGIG(登録商標)等の通信方式に対応した通信モジュールである。音声信号処理装置1001は、例えば、通信IF1403を介して他の通信装置と通信を行い、復号対象のビットストリームを取得する。取得されたビットストリームは、例えば、メモリ1404に記憶される。

[0125] 通信IF1403は、例えば、通信方式に対応した信号処理回路とアンテナで構成される。通信方式は、Bluetooth(登録商標)及びWiGIG(登録商標)に限られず、LTE(Long Term Evolution)、NR(New Radio)又はWi-Fi(登録商標)等であってもよい。

[0126] また、通信方式は、上記のような無線通信方式に限られない。通信方式は、Ethernet(登録商標)、USB(Universal Serial Bus)又はHDMI(登録商標)(High-Definition Multimedia Interface)等の有線通信方式であってもよい。

[0127] センサ1405は、受聴者の位置及び向きを推定するためのセンシングを

行う。具体的には、センサ1405は、身体の一部又は全体の位置、向き、動き、速度、角速度及び加速度等のうちいずれか1つ以上の検出結果に基づいて、受聴者の位置及び／又は向きを推定し、受聴者の位置及び／又は向きを示す位置／向き情報を生成する。

[0128] なお、音声信号処理装置1001の外部の装置がセンサ1405を備えていてもよい。身体の一部は、受聴者の頭部等であってもよい。位置／向き情報は、実空間における受聴者の位置及び／又は向きを示す情報であってもよいし、所定の時点における受聴者の位置及び／又は向きを基準とした受聴者の位置及び／又は向きの変位を示す情報であってもよい。また、位置／向き情報は、立体音響再生システム1000、又は、センサ1405を備える外部装置との相対的な位置及び／又は向きを示す情報であってもよい。

[0129] センサ1405は、例えば、カメラ等の撮像装置又はLiDAR (Laser Imaging Detection and Ranging) 等の測距装置である。センサ1405は、受聴者の頭部の動きを撮像し、撮像された画像を処理することで受聴者の頭部の動きを検知してもよい。また、例えばミリ波等の任意の周波数帯域の無線を用いて位置推定を行う装置がセンサ1405として用いられてもよい。

[0130] また、音声信号処理装置1001は、センサ1405を備える外部の機器から通信IF1403を介して位置情報を取得してもよい。この場合、音声信号処理装置1001は、センサ1405を含んでいなくてもよい。ここで、外部の機器とは、例えば図2で説明した音声提示装置1002、又は、受聴者の頭部に装着される立体映像再生装置等である。このときセンサ1405は、例えば、ジャイロセンサ及び加速度センサ等の各種のセンサを組み合わせ構成される。

[0131] センサ1405は、例えば、受聴者の頭部の動きの速度として、音空間内で互いに直交する3軸の少なくとも1つを回転軸とする回転の角速度を検知してもよいし、上記3軸の少なくとも1つを変位方向とする変位の加速度を検知してもよい。

- [0132] センサ1405は、例えば、受聴者の頭部の動きの量として、音空間内で互いに直交する3軸の少なくとも1つを回転軸とする回転量を検知してもよいし、上記3軸の少なくとも1つを変位方向とする変位量を検知してもよい。具体的には、センサ1405は、受聴者の位置として6Dofの位置（x、y、z）及び角度（yaw、pitch、roll）を検知する。センサ1405は、ジャイロセンサ及び加速度センサ等のような動きの検知に使用される各種のセンサを組み合わせで構成される。
- [0133] なお、センサ1405は、受聴者の位置を検出するためのカメラ又はGPS（Global Positioning System）受信機等により実現されてもよい。LiDAR等をセンサ1405として用いて自己位置推定を実施することで得られる位置情報が用いられてもよい。例えば、センサ1405は、立体音響再生システム1000がスマートフォンにより実現される場合には、スマートフォンに内蔵される。
- [0134] また、センサ1405には、音声信号処理装置1001の温度を検出する熱電対等の温度センサが含まれていてもよい。また、センサ1405には、音声信号処理装置1001が備えるバッテリー、又は、音声信号処理装置1001と接続されたバッテリーの残量を検出するセンサ等が含まれていてもよい。
- [0135] スピーカ1401は、例えば、振動板と、マグネット又はボイスコイル等の駆動機構とアンプを有し、音響処理後の音声信号を音として受聴者に提示する。スピーカ1401は、アンプを介して増幅させた音声信号（より具体的には、音の波形を示す波形信号）に応じて駆動機構を動作させ、駆動機構によって振動板を振動させる。このようにして、音声信号に応じて振動する振動板は、音波を発生させ、音波が空気を伝搬して受聴者の耳に伝達し、受聴者が音を知覚する。
- [0136] なお、ここでは、音声信号処理装置1001がスピーカ1401を備え、スピーカ1401を介して、音響処理後の音声信号を提示する例が挙げられたが、音声信号の提示手段は上記の構成に限定されない。

[0137] 例えば、通信モジュールで接続された外部の音声提示装置 1002 に音響処理後の音声信号が出力されてもよい。通信モジュールで行う通信は有線でも無線でもよい。また別の例として、音声信号処理装置 1001 が音声のアナログ信号を出力する端子を備え、端子にイヤホン等のケーブルを接続してイヤホン等から音声信号を提示してもよい。

[0138] 上記の場合、音声提示装置 1002 は、受聴者の頭部又は体の一部に装着されるヘッドホン、イヤホン、ヘッドマウントディスプレイ、ネックスピーカ又はウェアラブルスピーカ等であってもよい。あるいは、音声提示装置 1002 は、固定された複数のスピーカで構成されたサラウンドスピーカ等であってもよい。そして、音声提示装置 1002 が、音声信号を再生してもよい。

[0139] (符号化装置の物理的構成)

図 6 は、符号化装置 1500 の物理的構成の一例を示す図である。図 6 の符号化装置 1500 は、図 3 A の符号化装置 1100 又は図 3 C の符号化装置 1120 であってもよく、図 3 A 又は図 3 C に示された複数の構成要素が、図 6 に示された複数の構成要素によって実装されてもよい。

[0140] 図 6 の符号化装置 1500 は、プロセッサ 1501 と、メモリ 1503 と、通信 I/F 1502 とを備える。

[0141] プロセッサ 1501 は、例えば、CPU、DSP 又は GPU である。当該 CPU、DSP 又は GPU が、メモリ 1503 に記憶されたプログラムを実行することで本開示の符号化処理を実施してもよい。また、プロセッサ 1501 は、例えば、情報処理を行う回路である。プロセッサ 1501 は、本開示の符号化処理を含む音声信号に対する信号処理を行う専用回路であってもよい。

[0142] メモリ 1503 は、例えば、RAM 又は ROM で構成される。メモリ 1503 は、ハードディスクに代表される磁気記録媒体又は SSD に代表される半導体メモリ等を含んでいてもよい。また、メモリ 1503 は、CPU 又は GPU に組み込まれた内部メモリであってもよい。

[0143] 通信 I F 1 5 0 2 は、例えば、B l u e t o o t h（登録商標）又は W I G I G（登録商標）等の通信方式に対応した通信モジュールである。符号化装置 1 5 0 0 は、例えば、通信 I F 1 5 0 2 を介して他の通信装置と通信を行い、符号化されたビットストリームを送信する。

[0144] 通信 I F 1 5 0 2 は、例えば、通信方式に対応した信号処理回路とアンテナで構成される。通信方式は、B l u e t o o t h（登録商標）及び W I G I G（登録商標）に限られず、L T E、NR 又は W i - F i（登録商標）等であってもよい。また、通信方式は、無線通信方式に限られない。通信方式は、E t h e r n e t（登録商標）、U S B 又は H D M I（登録商標）等の有線通信方式であってもよい。

[0145] 通信モジュールは、例えば、通信方式に対応した信号処理回路とアンテナとで構成される。上記の例では、通信方式として B l u e t o o t h（登録商標）又は W I G I G（登録商標）を例に挙げたが、L T E（L o n g T e r m E v o l u t i o n）、NR（N e w R a d i o）、又は W i - F i（登録商標）等の通信方式に対応していてもよい。また、通信 I F は、上記のような無線通信方式ではなく、E t h e r n e t（登録商標）、U S B（U n i v e r s a l S e r i a l B u s）、H D M I（登録商標）（H i g h - D e f i n i t i o n M u l t i m e d i a I n t e r f a c e）等の有線の通信方式であってもよい。

[0146] [レンダリング部の構成]

図 7 は、レンダリング部 1 3 0 0 の構成例を示すブロック図である。具体的には、図 7 は、図 4 A 及び図 4 B のレンダリング部 1 2 0 3 及び 1 2 1 3 に対応するレンダリング部 1 3 0 0 の詳細な構成の一例を示す。

[0147] レンダリング部 1 3 0 0 は、解析部 1 3 0 1 と、選択部 1 3 0 2 と、再生部 1 3 0 3 とで構成され、入力信号に含まれる音データに対して音響処理を付加し出力する。

[0148] 入力信号は、例えば、空間情報とセンサ情報と音データとで構成される。入力信号は、音データとメタデータ（制御情報）とで構成されるビットスト

リームを含んでいてもよく、その場合、メタデータに空間情報が含まれていてもよい。

[0149] 空間情報は、立体音響再生システム1000が作り出す音空間（三次元音場）に関する情報であって、音空間に含まれるオブジェクトに関する情報と受聴者に関する情報とで構成される。オブジェクトには、音を発し音源となる音源オブジェクトと、音を発しない非発音オブジェクトが存在する。音源オブジェクトは、単に音源とも表現され得る。

[0150] 非発音オブジェクトは、音源オブジェクトが発した音を反射する障害物オブジェクトの役割を果たすが、音源オブジェクトが、別の音源オブジェクトが発した音を反射する障害物オブジェクトの役割を果たす場合もある。障害物オブジェクトは、反射オブジェクトと表現されてもよい。

[0151] 音源オブジェクトと非発音オブジェクトとに共通して付与される情報として、位置情報、形状情報、及び、オブジェクトが音を反射する際の音量の減衰率等がある。

[0152] 位置情報は、ユークリッド空間の例えばX軸、Y軸及びZ軸の3軸の座標値で表されるが、必ずしも三次元情報でなくてもよい。例えば、位置情報は、X軸及びY軸の2軸の座標値で表される二次元情報であってもよい。オブジェクトの位置情報は、メッシュ又はボクセルで表現される形状の代表位置で定められる。

[0153] 形状情報は、表面の素材に関する情報を含んでいてもよい。

[0154] 減衰率は、0以上1以下の実数で表現してもよいし、マイナスのデシベル値で表現してもよい。実空間では反射によって音量が増幅することはないため、減衰率はマイナスのデシベル値が設定されるが、例えば、非現実空間の不気味さを演出するために、あえて1以上の減衰率、すなわち、プラスのデシベル値が設定されてもよい。

[0155] また、減衰率は、複数の周波数帯域を構成する周波数帯域毎に異なる値が設定されてもよいし、周波数帯域毎に独立に値が設定されてもよい。また、減衰率がオブジェクト表面の素材の種類毎に設定されている場合、表面の素

材に関する情報に基づいて対応する減衰率の値が用いられてもよい。

[0156] また、空間情報は、オブジェクトが生物に属するか否かを示す情報、及び、オブジェクトが動体であるか否かを示す情報等を含んでいてもよい。オブジェクトが動体である場合、位置情報で示される位置は、時間とともに移動してもよい。この場合、変化した位置又は変化量の情報がレンダリング部 1300 に伝送される。

[0157] 音源オブジェクトに関する情報は、音源オブジェクトと非発音オブジェクトとに共通して付与される情報に加えて、音データと、音データを音空間内に放射するために必要な情報とを含む。音データは、音の周波数及び強弱に関する情報等を示すデータであって、受聴者に知覚される音が表現されたデータである。

[0158] 音データは、典型的にはPCM信号であるが、MP3等の符号化方式を用いて圧縮されたデータであってもよい。その場合、少なくとも当該信号が再生部 1303 に到来するまでに復号される必要があるため、レンダリング部 1300 は、図示しない復号部を含んでいてもよい。あるいは、当該信号が音声データデコーダ 1202 で復号されてもよい。

[0159] 1つの音源オブジェクトに対して、1つの音データが設定されていてもよいし、複数の音データが設定されていてもよい。また、各音データを識別する識別情報が音データに付与されてもよく、音源オブジェクトに関する情報は、音データの識別情報を含んでいてもよい。

[0160] 音データを音空間内に放射するために必要な情報は、例えば、音データの再生において基準として用いられる基準音量の情報、音データの性質（特性ともいう）を示す情報、音源オブジェクトの位置に関する情報、及び、音源オブジェクトの向きに関する情報（つまり、音源オブジェクトが発する音の指向性に関する情報）等を含んでいてもよい。

[0161] 基準音量の情報は、例えば、音データを音空間に放射する際の音源位置における音データの振幅値の実効値であって、デシベル（dB）値として浮動小数点で表されてもよい。

- [0162] 例えば、基準音量は、0 d b の場合、音データが示す信号レベルの音量を増減させることなくそのままの音量で、音源オブジェクトの位置に関する情報が指し示す位置から、音空間に対して音を放射することを示していてもよい。また、基準音量は、−6 d b の場合、音データが示す信号レベルの音量を約半分にして、音源オブジェクトの位置に関する情報が指し示す位置から、音空間に対して音を放射することを示していてもよい。
- [0163] 基準音量の情報は、音データ毎に付与されてもよいし、複数の音データに対してまとめて付与されてもよい。
- [0164] 音データの性質を示す情報は、例えば、音源の音量に関する情報であって、音源の音量の時系列的な変動を示す情報であってもよい。
- [0165] 例えば、音空間が仮想会議室であり、音源が話者である場合、音量は短い時間で断続的に遷移する。つまり、有音部分と無音部分とが交互に発生する。また、音空間がコンサートホールであり、音源が演奏者である場合、音量は、一定の時間長において維持される。また、音空間が戦場であり、音源が爆発物である場合、爆発音の音量は、一瞬だけ大きくなり、その後、無音又は小さい状態であり続ける。
- [0166] このように音源の音量の情報は、音の大きさの情報のみならず、音の大きさの遷移の情報を含んでいてもよい。そのような情報が、音データの性質を示す情報として用いられてもよい。
- [0167] 遷移の情報は、周波数特性を時系列に示すデータで表現されてもよい。遷移の情報は、有音区間の継続時間長を示すデータで表現されてもよい。遷移の情報は、有音区間の継続時間長と無音区間の継続時間長との時系列を示すデータで表現されてもよい。遷移の情報は、音信号の振幅が定常的とみなされ得る（概ね一定であるとみなされ得る）継続時間と、その間の当該信号の振幅値との複数組を時系列で列挙したデータ等で表現されてもよい。
- [0168] 遷移の情報は、音信号の周波数特性が定常的とみなされ得る継続時間のデータで表現されてもよい。遷移の情報は、音信号の周波数特性が定常的とみなされ得る継続時間と、その間の当該周波数特性との複数組を時系列で列挙

したデータ等で表現されてもよい。遷移の情報は、例えば、スペクトログラムの概形を示すデータの形式で表現されてもよい。

[0169] また、上記の周波数特性の基準として用いられる音量は、上記の基準音量であってもよい。基準音量の情報と音データの性質を示す情報とは、受聴者に知覚させる直接音又は反射音の音量の算出処理に用いられてもよいし、受聴者に知覚させるか否かの選択処理に用いられてもよい。音データの性質を示す情報の他の例及び利用方法については後述する。

[0170] なお、本実施の形態に係る反射音は、間接音の一例である。間接音は、反射音であってもよく、回折音等であってもよい。本実施の形態においては、間接音の一例である反射音を用いて説明するが、反射音にかえて間接音が用いられても同様の処理が行われる。

[0171] 音源オブジェクトの向きに関する情報（向き情報）は、典型的には、yaw、pitch及びrollで表現される。あるいは、rollの回転が省略され、音源オブジェクトの向き情報は、アジマス（yaw）及びエレベーション（pitch）で表現されてもよい。音源オブジェクトの向き情報は、時間とともに変化してもよく、変化した場合に、レンダリング部1300に伝送される。

[0172] 受聴者に関する情報は、音空間における受聴者の位置及び向きに関する情報である。位置に関する情報（位置情報）は、ユークリッド空間のXYZ軸の位置で表されるが、必ずしも三次元情報でなくてもよく、二次元情報であってもよい。受聴者の向きに関する情報（向き情報）は、典型的には、yaw、pitch及びrollで表現される。あるいは、rollの回転が省略され、受聴者の向き情報は、アジマス（yaw）及びエレベーション（pitch）で表現されてもよい。

[0173] 受聴者の位置情報及び向き情報は、時間とともに変化してもよく、変化した場合に、レンダリング部1300に伝送される。

[0174] センサ情報は、受聴者が装着するセンサ1405で検知された回転量又は変位量等と受聴者の位置及び向きを含む情報である。センサ情報は、レン

ダリング部1300に伝送され、レンジリング部1300は、センサ情報に基づいて受聴者の位置及び向きの情報を更新する。センサ情報は、例えば携帯端末がGPS、カメラ又はLiDAR等を用いて自己位置推定を実施して得られた位置情報を含んでいてもよい。

[0175] また、センサ1405ではなく、通信モジュールを通じて外部から取得された情報がセンサ情報として検出されてもよい。センサ1405から、音声信号処理装置1001の温度を示す情報、及び、バッテリーの残量を示す情報が取得されてもよい。また、音声信号処理装置1001又は音声提示装置1002の演算資源（CPU能力、メモリ資源又はPC性能等）がリアルタイムで取得されてもよい。

[0176] 解析部1301は、入力信号に含まれる音声信号、並びに、空間情報管理部1201及び1211から受け取った空間情報の解析を行い、再生部1303での直接音及び反射音の生成に必要な情報、並びに、反射音を生成するか否かの選択に必要な情報を算出する。

[0177] 直接音及び反射音の生成に必要な情報は、例えば直接音及び反射音のそれぞれについて、受聴位置に到来するまでの経路、到来するまでにかかる時間、及び、到来時の音量等に関する値である。受聴位置に到来するまでの経路、到来するまでにかかる時間、及び、到来時の音量等に関する値は、例えばそれぞれ、受聴位置に到来するまでの経路、到来するまでにかかる時間、及び、到来時の音量を示す値である。

[0178] 出力する反射音の選択に必要な情報は、直接音と反射音との関係を示す情報であって、例えば、直接音と反射音との時間差に関する値、及び、受聴位置における直接音と反射音との音量比に関する値等である。直接音と反射音との時間差に関する値、及び、受聴位置における直接音と反射音との音量比に関する値は、例えばそれぞれ、直接音と反射音との時間差を示す値、及び、受聴位置における直接音と反射音との音量比を示す値である。

[0179] なお、音量が、対数軸のデシベルの単位で表現される場合（音量がデシベル領域で表される場合）、2つの信号の音量比は、デシベル値の差で表され

ることは言うまでもない。具体的には、2つの信号の音量比は、それぞれの信号の振幅値がデシベル領域で表されるときとの差であってもよい。その値がエネルギー値又はパワー値等に基づいて算出されてもよい。また、この差は、デシベル領域において、ゲインの差又は単にゲイン差と呼ばれ得る。

[0180] つまり、本開示における音量比は、実質的に信号の振幅の比であるので、Sound volume ratio、Volume ratio、Amplitude ratio、Sound level ratio、Sound intensity ratio、又は、Gain ratio等と表現されてもよい。また、音量の単位がデシベルである場合、本開示における音量比は、音量差と言い換えられることは言うまでもない。

[0181] 本開示において、「音量比」とは、典型的には2つの音の音量がデシベルの単位で表現されるときとのゲイン差を意味し、実施の形態の例では、閾値データも典型的にはデシベル領域で表されたゲイン差で規定されている。しかし、音量比は、デシベル領域のゲイン差に限定されない。デシベル領域以外で表される音量比が用いられる場合、算出された音量比の単位に、デシベル領域で規定された閾値データが変換されて用いられてもよい。あるいは、予めそれぞれの単位で規定された閾値データがメモリに格納されていてもよい。

[0182] つまり、音量比の代わりに、例えば、エネルギーの値又はパワーの値等の比が用いられても、本開示におけるアルゴリズムを本開示の課題の解決に適用できることは明らかである。

[0183] 直接音と反射音との到来する時間差は、例えば、直接音到来時間（到来時刻）と反射音到来時間（到来時刻）との時間差である。なお、簡単のため、直接音と反射音との到来する時間差を、直接音と反射音との時間差、と記載する場合がある。直接音と反射音との時間差は、直接音と反射音とがそれぞれ受聴位置に到来する時刻の時間差、直接音と反射音とがそれぞれ受聴位置に到来するまでにかかる時間の差、又は、直接音の発音が終了する時刻と反射音が受聴位置に到来する時刻との時間差であってもよい。これらの値の算

出方法については後述する。

- [0184] 選択部 1302 は、解析部 1301 が算出した情報と、閾値データとを用いて、再生部 1303 が反射音を生成するか否かを選択する。言い換えると、選択部 1302 は、反射音を生成対象反射音として選択するか否かを判定する。さらに、言い換えると、選択部 1302 は、再生部 1303 が複数の反射音のうちいずれの反射音を生成するかを選択する。
- [0185] 閾値データは、例えば直接音と反射音との時間差の値を横軸に有し、直接音と反射音との音量比を縦軸に有するグラフで、当該反射音が知覚されるか知覚されないかの境目（閾値）として表される。閾値データは、直接音と反射音との時間差の値を変数として有する近似式で表現されてもよいし、直接音と反射音との時間差の値をインデックスとして有し、対応する閾値を有する配列で表現されてもよい。
- [0186] 選択部 1302 は、例えば直接音の到来時刻と反射音の到来時刻との時間差の値における、直接音の到来時音量と反射音の到来時音量との音量比が、閾値データを参照して設定された閾値よりも大きい値であった場合に、反射音を生成することを選択する。なお、到来時音量は、受聴位置に当該音が到来するときの音量を意味する。
- [0187] 直接音の到来時刻と反射音の到来時刻との時間差は、言い換えると、直接音と反射音とがそれぞれ受聴位置に到来するまでにかかる時間の差である。また、直接音の発音が終了する時点と反射音が受聴位置に到来する時点の時間差が、直接音と反射音との時間差として用いられてもよい。その場合、直接音の到来時刻と反射音の到来時刻との時間差を基準に用いて定められた閾値データとは異なる閾値データが用いられてもよいし、共通の閾値データが用いられてもよい。
- [0188] 閾値データは、音声信号処理装置 1001 のメモリ 1404 から取得してもよいし、外部の記憶装置から通信モジュールを経由して取得してもよい。閾値データの記憶方法及び閾値の設定方法については後述する。
- [0189] 再生部 1303 は、直接音の音声信号と、選択部 1302 が生成すること

を選択した反射音の音声信号とを合成する。

[0190] 具体的には、再生部1303は、解析部1301が算出した直接音到来時刻及び直接音到来時音量の情報に基づいて、入力された音声信号を加工し直接音を生成する。また、再生部1303は、選択部1302が選択した反射音についての反射音到来時刻及び反射音到来時音量の情報に基づいて、入力された音声信号を加工し反射音を生成する。そして、再生部1303は、生成した直接音及び反射音を合成し出力する。

[0191] [レンダリング部の動作例]

図8は、音声信号処理装置1001の動作例を示すフローチャートである。図8には、主に音声信号処理装置1001のレンダリング部1300で実行される処理が示されている。

[0192] 入力信号の解析処理（図8のS101）では、解析部1301が、音声信号処理装置1001に入力された入力信号を解析して、音空間で発生し得る直接音及び反射音を検出する。ここで検出される反射音は、最終的に再生部1303で生成される反射音として選択部1302で選択される反射音の候補である。また、解析部1301は、入力信号を解析して、直接音及び反射音の生成に必要な情報と、生成対象反射音の選択に必要な情報とを算出する。

[0193] まず、直接音及び反射音のそれぞれの特性が算出される。具体的には、直接音及び反射音のそれぞれが受聴者へ到来する際の到来時間及び到来時音量が算出される。反射オブジェクトとして複数のオブジェクトが音空間に存在する場合、複数のオブジェクトのそれぞれに対して反射音の特性が算出される。

[0194] 直接音到来時間（ t_d ）は、直接音到来経路（ p_d ）に基づいて算出される。直接音到来経路（ p_d ）は、音源オブジェクトの位置情報 $S(x_s, y_s, z_s)$ と受聴者の位置情報 $A1(x_a, y_a, z_a)$ とを結ぶ経路である。直接音到来時間（ t_d ）は、位置情報 $S(x_s, y_s, z_s)$ と位置情報 $A1(x_a, y_a, z_a)$ とを結ぶ経路の長さを、音速（約340m/秒

)で割って得られる値である。

[0195] 例えば、経路の長さ(X)は、 $(x_s - x_a)^2 + (y_s - y_a)^2 + (z_s - z_a)^2$ の平方根で求められる。音量は距離に反比例して減衰する。したがって、音源オブジェクトの位置情報S(x_s, y_s, z_s)における音量がNであり、単位距離がUである場合、直接音到来時音量(I_d)は、 $I_d = N * U / X$ で求められる。

[0196] 音源位置での音量Nは、先に説明した基準音量であってもよい。

[0197] 反射音到来時間(t_r)は、反射音到来経路(p_r)に基づいて算出される。反射音到来経路(p_r)は、反射音の音像の位置と位置情報A1(x_a, y_a, z_a)とを結ぶ経路である。

[0198] なお、反射音の音像の位置の導出は、例えば、「鏡像法」又は「レイトレーシング法」が用いられてもよいし、それ以外の任意の音像位置の導出方法が用いられてもよい。鏡像法とは、室内の壁面での反射波が壁面に対して音源と対称な位置に鏡像が存在すると仮定し、その鏡像の位置から音波が放射されると仮定して音像をシミュレートする手法である。レイトレーシング法とは、光線又は音線等のような直線的に伝わる波を追跡することで、ある点において観測される像(音像)等をシミュレートする手法である。

[0199] 図9は、受聴者と障害物オブジェクトとが比較的遠い位置関係を示す図である。図10は、受聴者と障害物オブジェクトとが比較的近い位置関係を示す図である。すなわち、図9及び図10のそれぞれは、音源位置に対して壁を挟んで対称の位置に反射音の音像が形成される例を示している。このような関係に基づいて、反射音の音像の位置をxyz軸で求めることで、反射音到来時間は、直接音到来時間を算出する方法と同様に求められる。

[0200] 反射音到来時間(t_r)は、反射音の音像の位置と位置情報A1(x_a, y_a, z_a)とを結ぶ経路の長さ(Y)を、音速(約340m/秒)で割って得られる値である。音量は距離に反比例して減衰する。したがって、音源位置での音量がNであり、単位距離がUであり、反射における音量の減衰率がGである場合、反射音到来時音量(I_r)は、 $I_r = N * G * U / Y$ で求

められる。

- [0201] 先に説明したように減衰率 G は、0以上1以下の実数で表現されてもよいし、マイナスのデシベル値で表現されてもよい。この場合、信号全体の音量が G の分だけ減衰する。また、減衰率は、複数の周波数帯域を構成する周波数帯域毎に設定されてもよい。この場合、解析部1301は、信号の周波数成分毎に、指定された減衰率を掛ける。また、解析部1301は、演算量の削減のために、複数の周波数帯域の複数の減衰率の代表値又は平均値等を全体の減衰率として用いて、信号全体の音量をその分だけ減衰させてもよい。
- [0202] 次に、解析部1301は、生成対象反射音の選択に必要な、直接音到来時音量(I_d)と反射音到来時音量(I_r)との比である音量比(L)、及び、直接音と反射音との時間差(T)を算出する。
- [0203] 直接音到来時音量(I_d)と上記 I_r との比である音量比(L)は、例えば、反射音到来時音量(I_r)を直接音到来時音量(I_d)で割った値であり、 $L = (N * G * U / Y) / (N * U / X) = G * X / Y$ で求められる。求められる値は、音量の比であるため、 N と U との値は予め定められた任意の値でよい。
- [0204] 直接音と反射音との時間差(T)は、例えば、直接音と反射音とがそれぞれ受聴位置に到来する際にかかる時間の時間差であってもよい。例えば、直接音と反射音とがそれぞれ受聴位置に到来する際にかかる時間差(T)は、 $T = t_r - t_d$ で求められる。
- [0205] また、時間差(T)は、直接音と反射音とがそれぞれ受聴位置に到来する時刻の差であってもよい。また、時間差(T)は、直接音の発話が終了する時刻と反射音が受聴位置に到来する時刻との時間差であってもよい。つまり、時間差(T)は、受聴位置において直接音が終了する時刻と反射音が始まる時刻との時間差であってもよい。
- [0206] 次に、反射音の選択処理(図8のS102)では、選択部1302が、解析部1301で算出された反射音を再生部1303が生成するか否かを選択する。言い換えれば、選択部1302は、反射音を生成対象反射音として選

択するか否かを判定する。複数の反射音が存在する場合、選択部 1302 は、各反射音を生成するか否かを選択する。選択部 1302 は、各反射音を生成するか否かを選択した結果、複数の反射音の中から、1つ以上の生成対象反射音を選択してもよいし、生成対象反射音を1つも選択しなくてもよい。

[0207] なお、選択部 1302 は、生成処理に限らず、他の処理の適用対象の反射音を選択してもよい。例えば、選択部 1302 は、バイノーラル処理の適用対象の反射音を選択してもよい。また、選択部 1302 は、基本的に、処理対象の1つ以上の反射音のみを選択する。しかしながら、選択部 1302 は、処理対象でない1つ以上の反射音のみを選択してもよい。そして、選択されなかった1つ以上の反射音に処理が適用されてもよい。

[0208] 例えば、反射音の選択は、解析部 1301 で算出した音量比（ L ）と時間差（ T ）とに基づいて行われる。直接音と反射音との時間差（ T ）に基づいて選択処理を行うことで、直接音と反射音との音量差のみに基づいて選択処理を行う場合と比較して、受聴者が知覚に影響する度合いが大きい反射音をより適切に選択することが可能である。

[0209] 具体的には、反射音を生成するか否かの選択は、例えば直接音と反射音との時間差に対応する直接音と反射音との音量比と、予め設定された閾値とを比較することで行われる。閾値は、閾値データを参照して設定される。閾値データは、直接音に対する反射音が受聴者に知覚されるか否かの境界を示す指標であって、直接音到来時音量（ I_d ）と反射音到来時音量（ I_r ）との比で定義される。

[0210] なお、閾値は、時間差（ T ）に対応して定められる数値等で表現される値に対応する。閾値データは、時間差（ T ）と閾値との関係に対応し、時間差（ T ）における閾値を特定又は算出するために用いるテーブルデータ又は関係式に対応する。閾値データの形式及び種類は、テーブルデータ又は関係式に限定されない。

[0211] 図 11 は、直接音と反射音との時間差と、閾値との関係を示す図である。例えば、図 11 に示されるような、直接音と反射音との時間差の値毎に予め

定められた音量比の閾値データが参照されてもよい。あるいは、図11に示される閾値データから内挿又は外挿等によって得られる閾値データが参照されてもよい。

[0212] そして、解析部1301で算出された時間差(T)における音量比の閾値が、閾値データから特定される。そして、選択部1302は、解析部1301で算出された直接音と反射音との音量比(L)が当該閾値を上回るか否かで、反射音を生成対象反射音として選択するか否かを決定する。

[0213] 直接音と反射音との時間差の値毎に予め定められた音量比の閾値データを用いて選択処理を行うことで、ポストマスキング又は先行音効果を考慮した選択処理を実現することが可能である。閾値データの種類、形式、記憶方法及び設定方法等について詳細な説明は後述する。

[0214] 次に、直接音及び反射音の生成処理(図8のS103)では、再生部1303が、直接音の音声信号と、選択部1302によって生成対象反射音として選択された反射音の音声信号とを生成し合成する。

[0215] 直接音の音声信号は、入力情報に含まれる音源オブジェクトの音データに対し、解析部1301で算出された直接音到来時間(t_d)及び直接音到来時音量(I_d)を適用することで生成される。具体的には、音データを直接音到来時間(t_d)の分だけ遅延させて、直接音到来時音量(I_d)を乗じる処理が行われる。音データを遅延させる処理は、時間軸上で音データの位置を前後に移動させる処理である。例えば、特許文献2に開示されているような音質を劣化させることなく、音データを遅延させる処理が適用されてもよい。

[0216] 反射音の音声信号は、直接音と同様に、音源オブジェクトの音データに対し、解析部1301で算出された反射音到来時間(t_r)及び反射音到来時音量(I_r)を適用することで生成される。

[0217] ただし、反射音の生成における反射音到来時音量(I_r)は、直接音到来時音量と異なり、反射における音量の減衰率 G が適用された値である。 G は、全周波数帯域に一括して適用される減衰率であってもよい。あるいは、反

射によって生じる周波数成分の偏りを反映させるために、所定の周波数帯域毎に反射率が規定されていてもよい。その場合、反射音到来時音量（ I_r ）を適用する処理は、帯域毎に減衰率を乗じる処理である周波数イコライザーの処理として実施されてもよい。

[0218] 上記の例では、直接音及び反射音候補のそれぞれが受聴者へ到来する際の経路長が算出される。さらに各経路長に基づいて到来時間と到来時音量とが算出される。そして、それらの時間差と音量比とに基づいて、反射音候補の選択処理が行われる。

[0219] なお、別の例として、直接音及び反射音のそれぞれが受聴者へ到来する際の経路長に基づいて選択処理が行われ、直接音及び反射音の到来時間及び到来時音量の算出、並びに、時間差及び音量比の算出が省略されてもよい。この場合、経路長差に応じた閾値が経路長比に対して予め定められてもよい。そして、算出された経路長比が、算出された経路長差に応じた閾値以上であるか否かで選択処理が行われてもよい。これにより、演算量を削減しながら、時間差に対応する経路長差に基づいて選択処理を行うことが可能である。

[0220] また、経路長差に加えて、音の伝搬速度を示すパラメータの値、又は、音の伝搬速度のパラメータに影響を与えるパラメータの値が用いられてもよい。

[0221] （選択処理の詳細）

反射音を生成するか否かの選択処理の詳細について説明する。

[0222] 反射音の選択は、直接音と反射音との時間差（ T ）における、直接音到来時音量と反射音到来時音量との比である音量比を定めた閾値と、解析部 1301 で算出した音量比（ L ）とを比較することで行われる。例えば、直接音と反射音との時間差の値毎に予め定められた音量比の閾値のうち、解析部 1301 で算出した直接音と反射音との時間差（ T ）における、音量比の閾値が参照される。そして、解析部 1301 で算出した音量比（ L ）が閾値を上回るか否かで、反射音を生成対象反射音として選択するか否かが決定される。

[0223] 時間差（ T ）は、例えば、直接音と反射音とがそれぞれ受聴位置に到来する時刻の差、直接音と反射音とがそれぞれ受聴位置に到来する際にかかる時間の時間差、及び、直接音の発音が終了した時点と反射音が受聴位置に到来する時点との時間差のいずれであってもよい。ここで、直接音の終了時刻は、例えば直接音の到来時刻に直接音の継続時間を加算することで求められてもよい。

[0224] 閾値データは、聴覚神経作用又は脳における認知作用によって、より具体的には、後述の先行音効果、後述の経時マスキング現象、又は、それらの組み合わせによって、受聴者の知覚が2つの音のずれを検知することが可能な最小時間差に基づいて定められてもよい。具体的な数値は、既に知られている経時マスキング効果、先行音効果又はエコー検知限等の研究成果から導出されてもよいし、当該仮想空間に適用することを前提とした試聴実験によって求められてもよい。

[0225] 図12A、図12B及び図12Cは、閾値データの設定方法の例を示す図である。閾値データは、図12A、図12B及び図12Cに示されるように、直接音と反射音との時間差を横軸に有し、直接音と反射音との音量比を縦軸に有するグラフで、当該反射音が知覚されるか知覚されないかの境目（閾値）によって表される。

[0226] 閾値データは、直接音と反射音との時間差を変数として有する近似式で表現されてもよい。また、閾値データは、図11に示されるような、直接音と反射音との時間差のインデックスと、インデックスに対応する閾値との配列としてメモリ1404の領域に記憶されていてもよい。

[0227] なお、図12Cの例示4における横軸に平行な線の高さ（最小可聴限）が閾値として用いられる場合、直接音と反射音との音量比（ L ）ではなく、反射音の音量そのものが閾値に対して比較される。なぜなら、当該閾値は、受聴者によって知覚可能か否かの境目の音量を示し、当該閾値よりも音量が小さい音を再現しない音として決定するための閾値であるためである。すなわち、最小可聴限に対応する閾値は、反射音の音量と直接音の音量との比に対

する閾値ではない。

- [0228] 最小可聴限が閾値として用いられる場合、閾値は時間差（ T ）に関係なく一定であるため、時間差（ T ）は算出されなくてもよい。
- [0229] なお、解析処理（図 8 の S 1 0 1）で複数の反射音が生成された場合、全ての反射音に対して選択処理が実行されてもよいし、予め設定された評価方法で各反射音に対して導出された評価値に基づいて評価値が高い反射音に対してのみ選択処理が実行されてもよい。ここで、反射音の評価値は、反射音の知覚的な重要度に対応する。なお、評価値が高いことは、評価値が大きいことに対応し、これらの表現は相互に置き換えられてもよい。
- [0230] 選択部 1 3 0 2 は、例えば、音源の音量、音源の視覚性、音源の定位性、反射オブジェクト（障害物オブジェクト）の視覚性、又は、直接音と反射音との幾何学的関係性等に応じた予め設定された評価方法で、反射音の評価値を算出してもよい。
- [0231] 具体的には、音源の音量が大きいほど、評価値が高くてよい。また、視覚的定位と音響的定位とを一致させるために、受聴者から音源オブジェクト又は反射オブジェクト（障害物オブジェクト）が視認可能である場合、又は、音源オブジェクトの定位性が高い場合、評価値が高くてよい。
- [0232] また、直接音と反射音との到来角の開き、及び、直接音と反射音との到来時間の差は、空間の把握に大きな影響を与える。そのため、直接音と反射音との到来角の開きが大きい場合、又は、直接音と反射音との到来時間の差が大きい場合、評価値が高くてよい。
- [0233] 音源の音量の情報は、コンテンツ毎に定められた基準音量、音量の時間的な遷移、又は、その双方を示していてもよい。
- [0234] 例えば、仮想空間が仮想会議室であり、直接音が会話音声である場合、音量は短い時間で断続的に遷移する。つまり、有音部分と無音部分が交互に発生する。また、仮想空間がコンサートホールであり、直接音が楽曲の演奏である場合、音量は、一定の時間長において維持される。また、仮想空間が戦場であり、直接音が爆発音である場合、音量は、一瞬だけ大きくなり、その

後、無音又は小さい状態であり続ける。

[0235] このように音源の音量の情報は、その音を仮想空間に放射する際のボリューム設定に相当する基準音量の情報のみならず、音の大きさの遷移の情報を含んでいてもよい。

[0236] 遷移の情報は、周波数特性を時系列に示すデータで表現されてもよい。遷移の情報は、有音区間の継続時間長を示すデータで表現されてもよい。遷移の情報は、有音区間の継続時間長と無音区間の継続時間長との時系列を示すデータで表現されてもよい。遷移の情報は、音信号の振幅が定常的とみなされ得る（概ね一定であるとみなされ得る）継続時間と、その間の当該信号の振幅値との複数组を時系列で列挙したデータ等で表現されてもよい。

[0237] 遷移の情報は、音信号の周波数特性が定常的とみなされ得る継続時間のデータで表現されてもよい。遷移の情報は、音信号の周波数特性が定常的であるとみなされ得る継続時間と、その間の当該周波数特性との複数组を時系列で列挙したデータ等で表現されてもよい。

[0238] また、信号の周波数特性の時間的な遷移を仮想空間の音響処理に用いる取り組みが従来から広く行われている（特許文献1等）。そのような従来の技術に鑑みた場合、上記の組が、周波数特性が一定である時間長と、その周波数特性との組であってもよいことは言うまでもない。

[0239] 幾何学的関係性は、仮想空間内の音源、受聴者及び反射オブジェクトの位置の関係性であってもよい。それらの関係性によって、直接音及び反射音がそれぞれ到来する経路長を幾何学的に算出することが可能である。したがって、音量が距離に反比例する関係を利用すれば、直接音の基準音量に対する反射音の基準音量を算出することが可能である。

[0240] 反射音の基準音量の算出には、反射オブジェクトの反射係数が用いられてもよい。また、反射係数として、一般的に用いられている典型的な値が用いられてもよい。一方、反射オブジェクトが吸音材で覆われている等のような特別な条件が存在する場合、反射オブジェクトの反射係数として、特別に付与された反射係数が用いられてもよい。

- [0241] 反射音は、反射音の音量によって評価されてもよい。反射音の音量は、上述したような、直接音と反射音との幾何学的関係性、及び、反射オブジェクトに付与された指標から求められてもよい。その音量と予め定められた閾値とを比較して反射音が評価されてもよい。
- [0242] さらに、音源の音量の時間的な遷移を示す情報が評価に反映されてもよい。例えば、音源の音量の時間的な遷移を示す情報が、有音の区間の継続時間長を示す場合において、時刻が有音の区間内である場合、反射音の評価値がそのままで維持されてもよい。一方、時刻が有音の区間外である場合、反射音の基準音量が閾値を上回っていても、反射音の評価値を小さくする又はゼロにする処理が行われてもよい。
- [0243] あるいは、音源の音量の時間的な遷移を示す情報は、音信号の振幅が概ね一定であるとみなされる継続時間と、その間の当該信号の振幅値との複数組を時系列で列挙したデータであってもよい。この場合、データにおける振幅値の変化に連動して反射音の基準音量を変化させて、反射音を評価する処理が行われてもよい。
- [0244] また、直接音の音量を示す情報として、基準音量の情報と、時間的に遷移する音量の情報との両方が用いられてもよい。例えば、基準音量の情報に基づいて評価値が算出された後に、遷移する音量の情報をを用いて当該評価値が補正されてもよい。
- [0245] 反射音の評価において、上述した全ての方法が実行されてもよいし、一部のみが実行されてもよい。例えば、複数の評価方法で反射音が評価されてもよいし、1つの評価方法で反射音が評価されてもよい。
- [0246] 複数の評価方法で反射音が評価される場合、反射音を選択するか否かは、複数の評価方法で総合的に決定された評価値に基づいて決定されてもよいし、複数の評価方法のそれぞれにおける評価値に基づいて決定されてもよい。
- [0247] 音声信号処理装置1001は、複数の評価方法のそれぞれに基づいて反射音を選択するか否かが決定される場合において、複数の評価方法に基づく複数の評価結果の全てが音を選択することを示す場合、音を選択してもよい。

あるいは、音声信号処理装置 1001 は、複数の評価方法に基づく複数の評価結果のいずれか 1 つが音を選択することを示す場合、音を選択してもよい。

[0248] また、例えば、第 1 から第 3 の評価方法に、優先度が設けられてもよい。そして、音声信号処理装置 1001 は、第 1 の評価方法で音を選択しないと判定された場合、第 2 及び第 3 の評価方法における判定結果に依存することなく音を選択しないと最終的に判定してもよい。また、音声信号処理装置 1001 は、第 2 及び第 3 の評価方法のいずれか一方で音を選択しないと判定されても他方で音を選択すると判定された場合、音を選択すると最終的に判定してもよい。

[0249] また、選択処理と評価処理とは独立して実行されてもよく、いずれか一方だけが実行されてもよい。また、選択処理で選択されると決定された反射音に対してのみ評価処理が実行され、反射音を選択するか否かが評価処理で再決定されてもよい。あるいは、選択処理で選択されないと決定された反射音に対してのみ評価処理が実行され、反射音を選択するか否かが評価処理で再決定されてもよい。

[0250] 上述された選択処理は、直接音の性質に応じて反射音を選択する処理と解釈され得る。例えば、直接音の性質に応じて反射音を選択する処理において、反射音の選択に用いられる閾値が、直接音の性質に応じて設定又は調整される。あるいは、反射音の選択に用いられる評価値が、音源の音量、音源の視覚性、音源の定位性、反射オブジェクト（障害物オブジェクト）の視覚性、及び、直接音と反射音との幾何学的関係性等のいずれか 1 つ以上に基づいて算出される。

[0251] また、直接音の性質に応じて反射音を選択する処理において、直接音の性質に応じて閾値を設定又は調整する処理、及び、処理対象の反射音の選択に用いられる評価値を算出する処理に限られず、他の処理が行われてもよい。また、直接音の性質に応じて閾値を設定又は調整する処理、あるいは、処理対象の反射音の選択に用いられる評価値を算出する処理が行われる場合であ

っても、処理が部分的に変更されてもよいし、新たな処理が追加されてもよい。

[0252] なお、閾値を設定することは、閾値を調整すること、及び、閾値を変更すること等を含んでいてもよい。

[0253] [閾値の設定方法]

選択処理に用いられる閾値データは、例えば既に知られている先行音効果に基づくエコー検知限の値、又は、ポストマスキング効果に基づくマスキング閾値を参考に設定されてもよい。

[0254] 先行音効果とは、2ヶ所から音が聴こえてきた時、時間的に先に聞こえた方に音源があると認識する現象である。もし、2つの短音が融合して1つの音に聞こえるならば、全体の音の聞こえる位置（定位位置）は、最初の音の位置によって概ね支配的に決定される。エコー検知限とは、先行音効果によって生じる現象であり、2つの音のずれを受聴者の知覚が検知する最小時間差である。

[0255] 図12Cの例示2において、横軸は、反射音（エコー）到来時間に対応し、具体的には、直接音の到来時刻から反射音の到来時刻までの遅れ時間に対応する。縦軸は、直接音に対する検知可能な反射音の音量比に対応し、具体的には、遅れ時間を伴って到来した反射音が検知可能であるか否かの閾値に対応する。

[0256] 図13は、閾値の設定方法の例を示す図である。図13における横軸は、反射音到来時間に対応し、具体的には、直接音と反射音との時間差（T）に対応する。図13における縦軸は、反射音の音量に対応する。具体的には、図13における縦軸は、直接音の音量に対して相対的に定められる反射音の音量（音量比）に対応していてもよいし、直接音の音量によらず絶対的に定められる反射音の音量に対応していてもよい。

[0257] 例えば、図9に示されるように受聴者と障害物オブジェクトとが比較的遠い場合、反射音到来時間が遅くなり、図13のCに示されるように、閾値は低く設定される。結果、図9の場合、反射音が生成される。一方、図10に

示すように受聴者と障害物オブジェクトとが比較的近い場合、反射音到来時間が図9の場合よりも早くなり、図13のBに示されるように、閾値は高く設定される。結果、図10の場合、反射音が生成されない。

- [0258] また、閾値データは、メモリ1404に記憶され、選択処理時にメモリ1404から取得され、選択処理に用いられてもよい。
- [0259] 図14は、選択処理の一例を示すフローチャートである。まず、選択部1302は、解析部1301で検出された反射音を指定する(S201)。そして、選択部1302は、直接音と反射音との音量比(L)、及び、直接音と反射音との時間差(T)を検出する(S202及びS203)。
- [0260] 時間差(T)は、例えば直接音と反射音とがそれぞれ受聴位置に到来する際にかかる時間の時間差、直接音到来時刻と反射音到来時刻との時間差、及び、直接音の発音が終了した時点と反射音を受聴位置に到来する時点との時間差のいずれであってもよい。ここでは直接音到来時刻と反射音到来時刻との時間差に基づく例を説明する。
- [0261] 具体的には、選択部1302は、音源オブジェクト及び受聴者の位置情報、並びに、障害物オブジェクトの位置情報及び形状情報から、直接音のパスの長さ、及び、反射音のパスの長さとの差を算出する。そして、選択部1302は、当該長さの差を音速で除すことで、直接音が受聴者位置に到来する時刻と反射音が受聴者位置に到来する時刻との時間差(T)を検出する。
- [0262] 受聴者に到来する音量は、音源の音量に対して、受聴者までの距離に比例して減衰(距離に反比例)する。したがって、直接音の音量は、音源の音量を直接音のパスの長さで除することで得られる。反射音の音量は、音源の音量を反射音のパスの長さで除して、さらに、仮想的な障害物オブジェクトに付与された減衰率を乗じることで得られる。選択部1302は、それらの音量の比を算出することで音量比を検出する。
- [0263] また、選択部1302は、閾値データを用いて、時間差(T)に対応する閾値を特定する(S204)。そして、選択部1302は、検出した音量比(L)が閾値以上であるか否かを判定する(S205)。

[0264] 音量比（L）が閾値以上である場合（S205でYes）、選択部1302は、当該反射音を生成対象の反射音として選択する（S206）。音量比（L）が閾値よりも小さい場合（S205でNo）、選択部1302は、当該反射音を生成対象の反射音として選択しない（S207）。すなわち、この場合、選択部1302は、当該反射音を生成対象外の反射音として決定する。

[0265] その後、選択部1302は、未指定の反射音があるか否かを判定する（S208）。未指定の反射音があれば（S208でYes）、選択部1302は、上述の処理（S201～S207）を繰り返す。未指定の反射音がなければ（S208でNo）、選択部1302は、処理を終了する。

[0266] この選択処理は、解析処理で生成した全ての反射音に対して実行されてもよいし、先述した評価値の高い反射音に対してのみ実行されてもよい。

[0267] [閾値の記憶方法の詳細]

本実施の形態に係る閾値データは、音声信号処理装置1001のメモリ1404に記憶される。記憶しておく閾値データの形式及び種類は、任意の形式及び任意の種類であってよい。複数の形式及び複数の種類の閾値が記憶される場合、選択処理において、いずれの形式及びいずれの種類の閾値を反射音の選択処理に用いるかが決定されてもよい。いずれの閾値データを選択処理に用いるかを決定する方法については、後述する。

[0268] また、複数の形式及び複数の種類の閾値データが組み合わせて記憶されていてもよい。組み合わせた閾値データが空間情報管理部1201及び1211から読み出され、選択処理に用いられる閾値が設定されてもよい。なお、メモリ1404に記憶される閾値データは、空間情報管理部1201及び1211に記憶されてもよい。

[0269] 閾値データは、例えば、図12Cの[例示1]及び[例示2]に示す閾値の線を描くように、各時間差における閾値で記憶されていてもよい。

[0270] また、閾値データは、図11に示すように閾値と時間差（T）とが対応付けられたテーブルデータとして記憶されていてもよい。つまり、閾値データ

は、時間差（T）をインデックスとして有するテーブルデータとして記憶されていてもよい。勿論、図11に示された閾値は一例であって、閾値は、図11の例に限定されない。また、閾値自体が記憶されず、時間差（T）を変数として有する関数で閾値が近似され、当該関数の係数が記憶されてもよい。また、複数の近似式が組み合わせて記憶されてもよい。

[0271] 例えば、時間差（T）を $timeDiff$ 、閾値を $gainThresh$ として、下記の様な式で閾値データが表現されてもよい。

[0272] [数1]

$$gainThresh = \begin{cases} -0.10 * timeDiff - 17.5 & , if \ 1.0 < timeDiff < 10.0 \\ -0.55 * timeDiff - 13.0 & , if \ 10.0 \leq timeDiff < 20.0 \\ -0.60 * timeDiff - 12.0 & if \ 20.0 \leq timeDiff < 30.0 \\ -0.80 * timeDiff - 6.0 & , if \ 30.0 \leq timeDiff \leq 40.0 \end{cases}$$

[0273] 閾値は、先行音効果が生じるとされる時間範囲だけで定義される。その時間範囲以外（上記式で言えば1ms以下又は40msより大きい値）の時間差の場合、 $gainThresh$ による判定は行わず、後述する、仮想空間で再現する最小の音量を示す閾値だけで判定が行われてもよい。

[0274] 先行音効果が生じるとされる時間範囲においては、閾値は上に凸である関数で近似されることが望ましいことが、本発明者らによる実験で明らかになった。上記式は、その実験に基づいて生成した近似式の一例である。

[0275] メモリ1404には、時間差（T）と閾値との関係を示す関係式に関する情報が記憶されていてもよい。つまり、時間差（T）を変数として有する式が記憶されていてもよい。各時間差（T）の閾値が直線又は曲線で近似され、直線又は曲線の幾何学的な形状を示すパラメータが記憶されていてもよい。例えば、当該幾何学的な形状が直線である場合、当該直線を表現するための始点及び傾きが記憶されていてもよい。

[0276] また、直接音の性質毎に閾値データの種類及び形式が定められて記憶されていてもよい。また、直接音の性質に応じて閾値を調整して選択処理に用いるためのパラメータが記憶されていてもよい。直接音の性質に応じて閾値を調整して選択処理に用いる処理は、閾値の設定方法の変形例として後述され

る。

[0277] 複数種類の閾値データが組み合わせられて記憶される例として、図12Cの〔例示3〕に示されるように、時間差（ T ）毎に、マスキング閾値とエコー検知限閾値とのうち大きい方の値が記憶されていてもよい。図12Cの〔例示4〕に示されるように、時間差（ T ）毎に、仮想空間で再現する最小の音量と、エコー検知限閾値とのうち大きい方の値が記憶されていてもよい。

[0278] 複数種類の閾値データの組み合わせは、これに限定されない。例えば、複数の閾値データにおいて時間差（ T ）毎に最大値の情報が記憶されていてもよい。

[0279] また、上記では、閾値に関する情報は、時間の項目を一次元のインデックスとして有する。閾値に関する情報は、さらに到来方向に関する変数を含む二次元又は三次元のインデックスを有していてもよい。

[0280] 図15は、直接音の方向と反射音の方向と時間差と閾値との関係を示す図である。例えば、図15に示すように、直接音の方向（ θ ）と、反射音の方向（ γ ）と、時間差（ T ）と、音量比（ L ）との関係に従って予め算出された閾値が記憶されていてもよい。

[0281] 直接音の方向（ θ ）は、受聴者に対する直接音の到来方向の角度に対応する。反射音の方向（ γ ）は、受聴者に対する反射音の到来方向の角度に対応する。ここでは、受聴者が向いている方向が0度と定められる。時間差（ T ）は、受聴位置への直接音到来時間と反射音到来時間との差に対応する。音量比（ L ）は、直接音到来時音量と反射音到来時音量との音量比に対応する。

[0282] 勿論、図15に示された閾値は一例であって、閾値は、図15の例に限定されない。また、図15では、主に直接音の到来方向の角度（ θ ）が0度である場合の閾値が例示されている。しかしながら、直接音の到来方向（ θ ）が0度以外である場合の閾値もメモリ1404に記憶される。

[0283] また、上記では、直接音の方向（ θ ）（より具体的には、直接音の到来方向の角度（ θ ））と、反射音の方向（ γ ）（より具体的には、反射音の到来

方向の角度 (γ)) とを独立した変数又はインデックスとして有する配列に、閾値が記憶される。しかし、直接音の到来方向の角度 (θ) と、反射音の到来方向の角度 (γ) とが独立した変数として用いられなくてもよい。

[0284] 例えば、直接音の到来方向の角度 (θ) と反射音の到来方向の角度 (γ) との角度差が用いられてもよい。この角度差は、直接音の到来方向と反射音の到来方向との成す角度に対応し、直接音と反射音との到来角と表現されてもよい。

[0285] 図 1 6 は、角度差と時間差と閾値との関係を示す図である。例えば、直接音の到来方向の角度 (θ) と反射音の到来方向の角度 (γ) との角度差 (ϕ) を変数として用いて予め算出された閾値が、図 1 6 に示される例のように記憶されていてもよい。勿論、図 1 6 に示された閾値は一例であって、閾値は、図 1 6 の例に限定されない。

[0286] 図 1 6 の例では、閾値の導出に用いられる変数の数を減らすことが可能である。そのため、メモリ 1 4 0 4 に記憶される閾値の数を減らすことが可能である。したがって、メモリ 1 4 0 4 に記憶されるデータ量を削減することが可能である。

[0287] また、直接音の到来方向の角度 (θ) と反射音の到来方向の角度 (γ) との角度差 (ϕ) が用いられる場合、二次元の配列で閾値データが記憶されてもよい。また、選択処理において、三次元の配列を用いて、直接音の到来方向の角度 (θ) と反射音の到来方向の角度 (γ) との差が算出されてもよい。

[0288] 到来方向に応じた閾値を用いて反射音を選択する方法については後述する。

[0289] [閾値の設定方法の第 1 変形例]

図 1 2 A、図 1 2 B 及び図 1 2 C の例において、複数の形式及び複数の種類の閾値が空間情報管理部 1 2 0 1 及び 1 2 1 1 に記憶されてもよい。そして、複数の形式及び複数の種類の閾値のうち、いずれの形式及びいずれの種類の閾値を反射音の選択処理に用いるかが決定されてもよい。具体的には、

図 1 2 C の例示 3 に示すように、反射音到来時刻に対応する時間差 (T) において、最も高い閾値が採用されてもよい。

[0290] また、例示 4 に示されるように、マスキング閾値、エコー検知限の閾値、及び、仮想空間で再現する最小の音量を示す閾値が記憶されていてもよい。そして、反射音到来時刻に対応する時間差 (T) において、最も高い閾値が採用されてもよい。

[0291] [閾値の設定方法の第 2 変形例]

閾値の設定方法の別の例として、直接音の性質に応じて閾値を設定する方法について説明する。

[0292] 図 1 7 は、図 7 に示されたレンダリング部 1 3 0 0 の別の構成例を示すブロック図である。図 1 7 のレンダリング部 1 3 0 0 は、図 7 のレンダリング部 1 3 0 0 とは、閾値調整部 1 3 0 4 を含む点で異なる。閾値調整部 1 3 0 4 以外の説明は、図 7 で説明した内容と同じであるため省略する。

[0293] 閾値調整部 1 3 0 4 は、音声信号の性質を示す情報に基づいて、閾値データから選択部 1 3 0 2 で用いられる閾値を選択する。あるいは、閾値調整部 1 3 0 4 は、音声信号の性質を示す情報に基づいて、閾値データに含まれる閾値を調整してもよい。

[0294] 音声信号の性質を示す情報は、入力信号に含まれていてもよい。そして、閾値調整部 1 3 0 4 は、入力信号から、音声信号の性質を示す情報を取得してもよい。あるいは、解析部 1 3 0 1 が、受け付けた入力信号に含まれる音声信号を解析することにより、音声信号の性質を導出し、音声信号の性質を示す情報を閾値調整部 1 3 0 4 に出力してもよい。

[0295] 音声信号の性質を示す情報は、レンダリング処理を開始する前に取得されてもよいし、レンダリング中に都度取得されてもよい。

[0296] また、閾値調整部 1 3 0 4 は、音声信号処理装置 1 0 0 1 に含まれていなくてもよく、他の通信装置が閾値調整部 1 3 0 4 の役割を有していてもよい。その場合、解析部 1 3 0 1 又は選択部 1 3 0 2 は、通信 I F 1 4 0 3 を介して他の通信装置から、音声信号の性質を示す情報、性質に応じた閾値デー

タ、又は、性質に応じて閾値データを調整するための情報を取得してもよい。

[0297] 図18は、選択処理の別の例を示すフローチャートである。図19は、選択処理のさらに別の例を示すフローチャートである。図18及び図19では、直接音の性質に応じて閾値が設定される。具体的には、図18では、閾値調整部1304が、時間差(T)と、音声信号の性質とに基づいて、閾値データから閾値を特定する。図19では、閾値調整部1304が、時間差(T)に基づいて閾値データから特定される閾値を音声信号の性質に基づいて調整する。

[0298] 以下、各例の動作を説明する。なお、図14の例と共通の処理については説明を省略する。

[0299] まず、図18で示す処理の例を説明する。ここでは、直接音の性質毎に閾値データがメモリ1404に予め記憶される。これにより、複数の性質に対応する複数の閾値データがメモリ1404に予め記憶される。そして、閾値調整部1304が、複数の閾値データの中から、反射音の選択処理に用いられる閾値データを特定する。

[0300] 例えば、閾値調整部1304は、入力信号に基づいて直接音の性質を取得する(S211)。閾値調整部1304は、入力信号に紐づいている直接音の性質を取得してもよい。そして、閾値調整部1304は、時間差(T)と、直接音の性質とに対応する閾値を特定する(S212)。

[0301] また、図19に示すように、閾値調整部1304は、直接音の性質に基づいて、選択部1302で特定された閾値を調整してもよい(S221)。

[0302] いずれの場合も、入力信号に、音声信号の性質を示す情報、音声信号の性質に応じて閾値を調整するための情報、又は、それらの両方が含まれていてもよい。閾値調整部1304は、それらの一方又は両方を用いて閾値を調整してもよい。

[0303] また、音声信号の性質を示す情報、閾値を調整するための情報、又は、それらの両方が、音声信号が含まれる入力信号とは別の入力信号で伝送されて

もよい。その場合、音声信号が含まれる入力信号に、当該入力信号とは別の入力信号を関連付ける情報が含まれていてもよいし、入力信号に別の入力信号を関連付ける情報が、閾値に関する情報とともにメモリ 1404 に記憶されていてもよい。

[0304] 図 18 及び図 19 の例では、反射音の選択に用いられる閾値が、直接音の性質、つまり音声信号の性質に応じて設定される。図 18 のように、予め性質毎に設定された閾値データが用いられてもよいし、図 19 のように、音声信号の性質に応じて閾値が調整されてもよい。また、音声信号の性質に応じて閾値データのパラメータが調整されてもよい。

[0305] また、閾値調整部 1304 によって行われる動作は、解析部 1301 又は選択部 1302 によって行われてもよい。例えば、解析部 1301 が、音声信号の性質を取得してもよい。また、選択部 1302 が、音声信号の性質に応じて閾値を設定してもよい。

[0306] 次に、音声信号の性質と閾値との関係について説明する。

[0307] 受聴者の耳に連続して到来する 2 つの短音は、その時間間隔が十分に短ければ 1 つの音として聞こえる。この現象は、先行音効果と呼ばれる。先行音効果は、不連続な、すなわち、過渡的な音に対してのみ生じることが知られている（非特許文献 1）。したがって、音声信号が定常音を示す場合に、音声信号が非定常音を示す場合よりも、エコー検知限が低く設定されてもよい。

[0308] すなわち、このような先行音効果の特性に従って、例えば、直接音が定常的な音である場合、閾値が小さく設定される。また、定常性が高ければ高いほど閾値が小さく設定されてもよい。

[0309] 音声信号の性質が定常的である場合の処理の例について説明する。まず閾値調整部 1304 又は解析部 1301 が、時間の経過に伴う音声信号の周波数成分の変動量に基づいて、定常性を判定する。例えば、当該変動量が少ない場合、定常性が高いと判定される。逆に当該変動量が多い場合、定常性が低いと判定される。判定の結果、定常性のレベルを示すフラグが設定されて

もよいし、当該変動量に応じて定常性を示すパラメータが設定されてもよい。

[0310] 次に、閾値調整部 1304 は、音声信号の定常性を示すフラグ又はパラメータ等のような定常性を示す情報に基づいて、閾値データ又は閾値を調整し、調整された閾値データ又は閾値を選択部 1302 で用いられる閾値データ又は閾値として設定してもよい。

[0311] あるいは、直接音の定常性を示す情報に応じて閾値データを設定するためのパラメータが、予めメモリ 1404 に記憶されていてもよい。この場合、閾値調整部 1304 は、音声信号の定常性を判定し、定常性を示す情報及びパラメータに基づいて、反射音の選択に用いられる閾値データを設定してもよい。

[0312] あるいは、閾値データの複数のパラメータが、直接音の定常性の複数のパターンに対応して、予めメモリ 1404 に記憶されていてもよい。この場合、閾値調整部 1304 は、音声信号の定常性を判定し、直接音の定常性のパターンに基づいて閾値データのパラメータを選択し、閾値データのパラメータに基づいて、反射音の選択に用いられる閾値データを設定してもよい。

[0313] なお、音声信号の定常性は、音声信号が入力される度に、音声信号の周波数成分の変動量に基づいて判定されてもよい。

[0314] あるいは、音声信号の定常性は、音声信号に予め紐づけられた、定常性を示す情報に基づいて、判定されてもよい。つまり、音声信号の定常性を示す情報が、音声信号に紐づけられて、予めメモリ 1404 に記憶されていてもよい。解析部 1301 は、音声信号が入力される度に、音声信号に紐づけられた定常性を示す情報を取得してもよい。そして、閾値調整部 1304 は、音声信号に紐づけられた定常性を示す情報に基づいて、閾値を調整してもよい。

[0315] 音声信号の性質に応じて閾値が設定される他の例として、音声信号が短い音（クリック音等）を示す場合に、音声信号が長い音を示す場合よりも、エコー検知限の適用範囲が短く設定されてもよい。この処理は、先行音効果の

特性に基づく。

[0316] 先行音効果によって、受聴者の耳に連続して到来する2つの短音は、その時間間隔が十分に短ければ1つの音として聞こえることが知られている。この時間間隔の上限は、音の長さに依存する。例えば、この時間間隔の上限は、クリック音では約5 msであるが、人の声又は音楽のような複雑な音では40 msになることもある（非特許文献1）。

[0317] このような先行音効果の特性に従って、例えば、直接音の継続時間長が短い音の場合、短い時間長の閾値が設定される。また、直接音の継続時間長が短いほど、より短い時間長の閾値が設定される。

[0318] 短い時間長の閾値が設定されることは、直接音と反射音との時間差（ T ）が小さい範囲において、先行音効果の特性に基づくエコー検知限に対応する閾値が設けられることを意味する。この範囲外では、先行音効果の特性に基づくエコー検知限に対応する閾値が設けられない。すなわち、この範囲外では、閾値が小さい。したがって、短い音に対して短い時間長の閾値が設定されることは、短い音に対して小さい閾値が設定されることに対応し得る。

[0319] 直接音の性質に応じて閾値が設定される他の例として、直接音が断続的な音（スピーチ等）である場合に、直接音が継続的な音（音楽等）である場合よりも、閾値が低く設定されてもよい。

[0320] 例えば直接音がスピーチに対応する場合、有音部と無音部とが繰り返され、マスキング効果として、無音部では、ポストマスキング効果のみが生じる。一方、直接音が音楽コンテンツのように継続的な音である場合、ポストマスキング効果と、その時に発生している音による同時マスキング効果との両方のマスキング効果が生じる。したがって、総合的なマスキング効果は、スピーチ等の場合よりも音楽等の場合に高い。

[0321] 上記のようなマスキング効果の特性に従って、スピーチ等の場合よりも音楽等の場合に、閾値が高く設定されてもよい。逆に、音楽等の場合よりもスピーチ等の場合に、閾値が低く設定されてもよい。つまり、直接音に断続的な部分が多い場合、閾値が小さく設定されてもよい。

- [0322] 上述した通り、直接音の性質を示す情報は、直接音の定常性、断続性及び継続時間等を示す情報であってもよい。また、直接音の性質を示す情報は、それらの任意の組み合わせであってもよい。また、直接音の性質を示す情報は、それらのいずれかの時間変動を示す情報であってもよいし、それらの任意の組み合わせの時間変動を示す情報であってもよい。つまり、直接音の性質を示す情報は、直接音の時間変動を示す情報であってもよい。
- [0323] 例えば、定常性の判定の説明において示したように、直接音の性質を示す情報は、周波数特性の時系列データであってもよい。ここで、周波数特性は、周波数バンド毎のゲイン値、時間軸信号に対するフーリエ級数、あるいは、周波数包絡線を求めるためのLPC係数又はケプストラム係数等のような、慣用されている形式で表現されてもよい。
- [0324] さらに、直接音の性質を示す情報は、直接音の断続性を示す情報として、信号の振幅が定常的である継続時間と、その間の当該信号の振幅値との複数組を時系列で列挙した情報（振幅の包絡線の概形）であってもよい。ここで、振幅値は、基準音量に対する比で表現されてもよい。
- [0325] さらに、直接音の性質を示す情報は、直接音の周波数特性に関する情報であってもよい。例えば、直接音の性質を示す情報は、直接音の周波数特性の定常性を示す情報であってもよい。具体的には、直接音の性質を示す情報は、周波数特性の変動が小さい状態である継続時間と、その間の当該信号の周波数特性との複数組を時系列で列挙した情報（スペクトログラムの概形）であってもよい。ここで、上記の周波数特性の基準として用いられる音量は、上記の基準音量であってもよい。
- [0326] 例えば、直接音の時間変動を示す情報は、直接音の包絡線を示す情報である。直接音の時間変動を示す情報は、図12Cの[例示4]に記載の「最小可聴限」が閾値である場合に用いられてもよい。最小可聴限と比較する信号は、反射音の音量である。
- [0327] 反射音の音量は、音源、受聴者及び反射オブジェクトの位置の情報から幾何学的な計算によって得られる。具体的には、音源の基準音量に対する反射

音の基準音量が得られる。直接音の性質を示す情報として音源の音の大きさの遷移の情報を用いて反射音の基準音量を増減させることで、時々刻々の反射音の音量を正確に求めることが可能である。この理由は、音源の音量の変動が反射音の音量の変動に反映されるためである。

[0328] 反射音の音量の調整後、反射音の音量を閾値と比較することで、より正確に聴感上必要な反射音を適切に選択することが可能である。

[0329] 勿論、反射音の基準音量を調整せずに、音源の音の大きさの遷移の情報の逆数に基づいて閾値を調整し、調整された閾値と反射音の基準音量とを比較することによっても同じ結果が得られることは言うまでもない。つまり、音源の音の大きさの遷移の情報を用いて反射音の基準音量が調整されてもよいし、音源の音の大きさの遷移の情報を用いて閾値が調整されてもよい。反射音の基準音量の調整と閾値の調整とは、互いに対応する。

[0330] 音を反射する物体の表面の組成によって、音の反射率（反射に伴う音の減衰率）は、周波数帯域毎に異なる。そこで、後述するように、音を反射する物体に対し、周波数帯域毎に、音の反射率（減衰率）が紐づけられていてもよい。そのような反射率の情報と、スペクトログラムの情報とによって、当該反射音を選択するか否かをより正確に判定することが可能である。例えば、下記のような処理が行われる。

[0331] 具体的には、例えば、スペクトログラムの情報によって、ある時間区間において、高域の周波数成分が、低域の周波数成分よりも優勢であることが示されている。また、例えば、音の反射率の情報によって、高域の周波数成分において、低域の周波数成分よりも、反射率が極めて小さいことが示されている。

[0332] この場合、音源の信号の時間軸での振幅が大きくても、スペクトログラムの情報によって示される周波数成分と、反射率の情報によって示される周波数帯域毎の減衰率とを乗じることで得られる反射音の音量は小さくなり、反射音は選択されない可能性がある。

[0333] 上述した通り、直接音の性質を示す情報は、直接音の時間変動を示す情報

であってもよい。例えば、直接音の性質を示す情報は、直接音を予め定められた時間長で分析することで得られる値を示してもよい。

[0334] 具体的には、直接音の性質を示す情報は、予め定められた時間長毎に直接音の平均エネルギー又は平均振幅を算出することで得られる情報であってもよい。また、直接音の性質を示す情報は、短時間分析長毎に直接音のエネルギー又は平均振幅を算出し、短時間分析長よりも長い長時間分析長毎にエネルギー又は平均振幅の加重平均を算出することで得られる情報であってもよい。

[0335] より具体的には、例えば、直接音の時間変動を示す情報は、予め定められた短い時間長（例えば5 m s、以後この時間長のフレームは分析フレームと表現される）毎に、直接音のエネルギー又は平均振幅を算出することで得られる情報であってもよい。また、直接音の時間変動を示す情報は、過去N－1個の分析フレームで算出されたエネルギー又は平均振幅の加重平均で表される情報であってもよい。

[0336] 仮に、第n分析フレームのエネルギーがE（n）で表現される場合、直接音の性質を示す情報I（n）は、次式に従って求められる。

[0337] [数2]

$$I(n) = a(0) \cdot E(n) + \sum_{i=1}^{N-1} a(i) \cdot E(n-i) \\ \dots (式1)$$

[0338] ここで、パラメータa（i）は、重み係数を表す。一般的には、a（i） ≥ 0 、かつ、a（i）の総和が1になるように、a（i）が設定される。しかしながら、a（i）の設定方法は、この限りではない。

[0339] なお、直接音が5 m s 取り込まれる度に、直接音の性質を示す情報I（n）が算出される。すなわち、直接音の性質を示す情報I（n）の時間変動を低遅延で算出することが可能である。そのため、この方法は、リアルタイム性の必要なアプリケーションへの適用に適している。

[0340] また、次式に従って、直接音の性質を示す情報I（n）が求められてもよ

い。

[0341] [数3]

$$I(n) = b(0) \cdot E(n) + \sum_{i=1}^{N-1} b(i) \cdot I(n-i)$$

・・・(式2)

[0342] ここで、パラメータ $b(i)$ は、重み係数を表す。一般的には、 $b(i) \geq 0$ 、かつ、 $b(i)$ の総和が1になるように、 $b(i)$ が設定される。しかしながら、 $b(i)$ の設定方法は、この限りではない。

[0343] この式では、再帰的に、直接音の性質を示す情報 $I(n)$ が求められている。そのため、少ない演算量で、長い時間長の平均エネルギーを算出することが可能である。

[0344] 上記の式1及び式2は、 $E(n)$ が入力信号であり $I(n)$ が出力信号であるフィルタとみなされ得る。その場合、式1は、移動平均(MA)モデルのフィルタであり、式2は、自己回帰(AR)モデルのフィルタであり、いずれも、低域通過型(ローパス)フィルタの特性を有する。また、両者が組み合わされたARMAモデルのフィルタが用いられてもよい。

[0345] なお、直接音の時間変動を示す情報を導出する方法には、上述の計算式又はフィルタに限られず、他の公知の手法が用いられてもよい。上述した通り、直接音の時間変動を示す情報は、直接音を予め定められた時間長で分析することで得られる値を示す。平均エネルギー以外の観点で直接音が分析されてもよい。

[0346] また、上述した通り、直接音の性質を示す情報は、直接音の周波数特性に関する情報であってもよい。直接音の周波数特性に関する情報は、直接音の周波数特性を用いて算出された情報であってもよい。例えば、直接音の周波数特性に関する情報は、直接音の低域成分を予め定められた分析長で平均化することで低域成分の平均エネルギーとして得られる情報であってもよい。

[0347] 具体的には、分析フレーム長に含まれる直接音に対して、ローパス特性を有するフィルタを適用することで、直接音の低域成分が求められる。この低

域成分のエネルギー又は平均振幅から、上述した式 1 と同様に、直接音の性質を示す情報が導出される。

[0348] 仮に、第 n 分析フレームの低域成分のエネルギーが $E_L(n)$ で表現される場合、直接音の性質を示す情報 $I(n)$ は、次式に従って求められる。

[0349] [数4]

$$I(n) = c(0) \cdot E_L(n) + \sum_{i=1}^{N-1} c(i) \cdot E_L(n-i) \\ \dots (式 3)$$

[0350] ここで、パラメータ $c(i)$ は、重み係数を表す。一般的には、 $c(i) \geq 0$ 、かつ、 $c(i)$ の総和が 1 になるように、 $c(i)$ が設定される。しかしながら、 $c(i)$ の設定方法は、この限りではない。

[0351] なお、直接音が 5 ms 取り込まれる度に、直接音の性質を示す情報 $I(n)$ が算出される。すなわち、直接音の性質を示す情報 $I(n)$ の時間変動を低遅延で算出することが可能である。そのため、この方法は、リアルタイム性の必要なアプリケーションへの適用に適している。

[0352] また、式 2 と同様に、次式に従って、直接音の性質を示す情報 $I(n)$ が求められてもよい。

[0353] [数5]

$$I(n) = d(0) \cdot E_L(n) + \sum_{i=1}^{N-1} d(i) \cdot I(n-i) \\ \dots (式 4)$$

[0354] ここで、パラメータ $d(i)$ は、重み係数を表す。一般的には、 $d(i) \geq 0$ 、かつ、 $d(i)$ の総和が 1 になるように、 $d(i)$ が設定される。しかしながら、 $d(i)$ の設定方法は、この限りではない。

[0355] この式では、再帰的に、直接音の性質を示す情報 $I(n)$ が求められている。そのため、少ない演算量で、長い時間長の平均エネルギーを算出することが可能である。

[0356] 上記の式 3 及び式 4 は、 $E(n)$ が入力信号であり $I(n)$ が出力信号で

あるフィルタとみなされ得る。その場合、式3は、移動平均(MA)モデルのフィルタであり、式4は、自己回帰(AR)モデルのフィルタであり、いずれも、低域通過型(ローパス)フィルタの特性を有する。また、両者が組み合わせられたARMAモデルのフィルタが用いられてもよい。

[0357] 上記では、直接音の低域成分を求める方法に、ローパス特性を有するフィルタが用いられているが、直接音の低域成分を求める方法は、これに限られない。また、直接音の時間変動を示す情報を導出する方法には、上述の計算式又はフィルタに限られず、他の公知の手法が用いられてもよい。例えば、直接音に対して周波数変換を施すことにより、直接音のスペクトルが算出されてもよい。そして、当該スペクトルの低域成分のエネルギー又は平均振幅が算出されてもよい。

[0358] また、上記では、直接音の時間変動を示す情報の導出に、MAモデル又はARモデルが用いられている。これらモデルの係数は、予め定められた固定値であってもよいし、時間的に可変な値である可変値であってもよい。

[0359] また、上記の分析フレーム長と、上記の情報更新スレッドの発生インターバルとの関係は、下記の様なものであってもよい。

[0360] 例えば、分析フレームの時間長が T_A (m s e c)であり、情報更新スレッドの発生インターバルが T_U (m s e c)である場合、MAフィルタにおける上記の(式1)及び(式3)の N の値は、 T_U / T_A によって与えられる値程度であってもよい。また、ARフィルタにおける上記(式2)及び(式4)の $b(i)$ 及び $d(i)$ ($1 \leq i < N$)は、当該フィルタの時定数が T_U (m s e c)程度になるような値であってもよい。

[0361] 上記の設定の理由は、情報更新のインターバル期間内で、当該フィルタが収束することが期待されるためである。

[0362] 一方、もし上記の設定で直接音の時間変動を示す情報において値が急峻に変動し過ぎるような場合は、 $l(n)$ が予め算出されてもよい。そして、予め算出された $l(n)$ が反射音の選択処理に適用されてもよい。例えば、 t 時刻目のフレームの処理において、 $l(t + t_{au})$ が用いられてもよい。

ここで、 τ は、フィルタの収束の特性に応じて定められる値である。収束が遅い場合、収束が速い場合に比べて、 τ の値は大きい。

[0363] また、直接音の特性を示す情報として、直接音から算出される聴覚マスキング（周波数マスキング）の情報が用いられてもよい。聴覚マスキングの情報は、直接音によってマスクされる周波数領域での振幅値の閾値を示す。同じ周波数領域での反射音の振幅値を閾値と比較して、閾値よりも振幅値の小さい反射音を選択しない処理が行われてもよい。周波数領域における反射音の振幅値は、反射音の特性を示す情報として解析部1301で取得されてもよい。

[0364] このように直接音の性質に応じて反射音の選択に用いられる閾値を設定することで、聴感上必要な反射音を適切に選択することが可能になり、聴覚の感度の特性を効果的に立体音響再生システム1000に反映することが可能になる。直接音の性質を検出する処理、性質に応じて閾値を決定する処理、及び、性質に応じて閾値を調整する処理は、レンダリング処理時に行われてもよいし、レンダリング処理を開始する前に行われてもよい。

[0365] 例えば、これらの処理は、仮想空間作成時（ソフトウェアの作成時）、仮想空間の処理の開始時（ソフトウェアの起動時又はレンダリング開始時）、又は、仮想空間の処理において定期的に発生する情報更新スレッドが発生したタイミング等で行われてもよい。また、仮想空間作成時は、音響処理の開始前に仮想空間を構築するタイミングであってもよいし、仮想空間の情報（空間情報）の取得時であってもよいし、ソフトウェアの取得時であってもよい。

[0366] ここで、情報更新スレッドでは、空間情報管理部1201及び1211によって管理される空間情報を更新するための処理が行われる。

[0367] 情報更新スレッドが担う役割は、例えば、受聴者が装着しているVRゴーグルの位置及び向きに基づいて、仮想空間内に配置された受聴者のアバターの位置及び向きを更新する処理、又は、仮想空間内を移動している物体の位置の更新等である。このような処理は、数10Hz程度の比較的低頻度で起

動する処理スレッド内で賄われる。

[0368] このような発生頻度の低い処理スレッドで、直接音の性質を示す情報を更新する処理が行われてもよい。この理由は、オーディオ出力のためのオーディオ処理フレームの発生頻度よりも直接音の性質が変動する頻度が低いためである。これにより、当該処理の演算負荷を相対的に小さくすることが可能になる。また、不必要に速い頻度で情報を更新するとパルシブなノイズが発生するリスクが生じる。低い頻度で情報が更新されることにより、このようなリスクを回避することも可能である。

[0369] [閾値の設定方法の第3変形例]

閾値の設定方法の別の例として、当該仮想空間の再現を処理する演算資源（CPU能力、メモリ資源、PC性能又はバッテリー残量等）に応じて、閾値が設定されてもよい。より具体的には、音声信号処理装置1001のセンサ1405が演算資源の量を検知し、演算資源量が少ない場合、閾値が高く設定される。これにより、より多くの反射音の音量が閾値よりも小さくなるため、バイノーラル処理が行われる反射音を減らすことが可能になり、演算量を減らすことが可能になる。

[0370] あるいは、信号処理が、スマートフォン又はVRゴーグル等のような、蓄電池で駆動する機器で行われる場合、処理を長時間持続させることを優先し演算資源を節約することが期待される。そのような場合、演算資源の量又は残量を検知するまでもなく、閾値が高く設定されてもよい。

[0371] [閾値の設定方法の第4変形例]

閾値の設定方法の別の例として、図示しない閾値設定部を音声信号処理装置1001又は音声提示装置1002が備えることで、当該仮想空間の管理者又は受聴者によって閾値が設定されてもよい。

[0372] 例えば、音声提示装置1002を装着する受聴者によって、受聴対象反射音が少なく演算量が少ない「省エネモード」か、受聴対象反射音が多く演算量が多い「高性能モード」かが選択可能であってもよい。もしくは、立体音響再生システム1000を管理する管理者又は立体音響コンテンツの作成者

によって、モードが選択可能であってもよい。また、モードではなく、閾値又は閾値データが直接選択可能であってもよい。

[0373] [レンダリング部の動作の第1変形例]

図20は、音声信号処理装置1001の動作の第1変形例を示すフローチャートである。図20には、主に音声信号処理装置1001のレンダリング部1300で実行される処理が示されている。本変形例では、レンダリング部1300の動作に音量補償処理が追加される。

[0374] 例えば、解析部1301は、データ（入力信号）を取得する（S301）。次に、解析部1301は、データを解析する（S302）。次に、選択部1302は、解析結果に基づいて、反射音を選択するか否かの判定を行う（S303）。次に、再生部1303は、選択されない反射音に基づいて音量補償処理を行う（S304）。次に、再生部1303は、直接音及び反射音の音響処理を行う（S305）。そして、再生部1303は、直接音及び反射音をオーディオとして出力する（S306）。

[0375] 上記の処理（S301～S306）のうち音量補償処理（S304）以外は、上述された他の例と共通する処理であるため、それらの説明を省略する。

[0376] 音量補償処理は、選択処理において選択されなかった反射音に応じて実行される。例えば、選択処理において反射音を選択しないことによって音量感の欠如が生じる。音量補償処理によって、このような音量感の欠如に伴う違和感が抑制される。音量感を補償する方法の例として、以下の2つの方法を開示する。2つの方法のうちいずれの方法が用いられてもよい。

[0377] まず、直接音の音量を上昇させることで音量感を補償する方法について説明する。再生部1303は、選択されなかった反射音の音量の分だけ直接音の音量を上昇させて、直接音を生成する。これによって、反射音を生成しないことで失われる音量感が補償される。

[0378] 再生部1303は、当該音量を上昇させる際、反射音の周波数特性に従って、周波数成分毎に音量を上昇させてもよい。このような処理を可能にする

ため、反射オブジェクトが減衰させる音量の減衰率が所定の周波数帯域毎に付与されていてもよい。これにより、反射音の周波数特性を導出することが可能になる。

[0379] 次に、直接音に反射音を合成させることで、音量感を補償する方法について説明する。この方法において、再生部1303は、選択されなかった反射音を直接音に加算し、直接音を生成することによって、反射音を生成しないことによる音量感を補償する。生成された直接音には、選択されなかった反射音の音量（振幅）、周波数及び遅延等が反映される。

[0380] 直接音の音量を上昇させる方法の場合、補償処理の演算量は極めて軽微であるが、音量のみが補償される。直接音に反射音を合成させる方法の場合、補償処理の演算量は、直接音の音量を上昇させる方法に比べれば大きい、反射音の特性がより正確に補償される。

[0381] いずれの場合も、反射音を生成せず、直接音のみが生成されるため、全体の演算量が削減される。特にHRTFを畳み込む処理を含むバイノーラル処理にかかる演算量が削減されるため、全体の演算量は大きく削減される。この理由は、バイノーラル処理にかかる演算量が上記の補償処理にかかる演算量よりも遥かに大きいためである。

[0382] なお、反射音が選択されない理由が、反射音の音量がマスキング閾値を下回ることである場合、音量感が失われないため、補償処理を行うことなく、単に反射音が除去されてもよい。

[0383] [レンダリング部の動作の第2変形例]

図21は、音声信号処理装置1001の動作の第2変形例を示すフローチャートである。図21には、主に音声信号処理装置1001のレンダリング部1300で実行される処理が示されている。本変形例では、レンダリング部1300の動作に左右音量差調整処理が追加される。

[0384] 例えば、解析部1301は、入力信号を解析する（S401）。次に、解析部1301は、音の到来方向を検出する（S402）。次に、選択部1302は、左右の耳で知覚される音の音量の差を調整する（S403）。また

、選択部1302は、左右の耳で知覚される音の到来時間の差（遅延）を調整する（S404）。選択部1302は、調整後の音の情報に基づいて、反射音を選択するか否かの判定を行う（S405）。

[0385] 上記の処理（S401～S405）のうち左右音量差調整処理（S403）及び遅延調整処理（S404）以外は、上述された他の例と共通する処理であるため、それらの説明を省略する。

[0386] 図22は、アバターと音源オブジェクトと障害物オブジェクトとの配置例を示す図である。例えば、受聴者の正面方向が0度である場合において、図22のように、直接音の到来方向（ θ ）と反射音の到来方向（ γ ）（反射音の方向（ γ ））との極性（例えば正負）が異なる場合、両耳間で生じる音量差が補正される。

[0387] 具体的には、 θ と γ との極性が異なる場合、直接音と反射音とで主に（先に）音を知覚する耳が異なる。この場合、選択部1302は、左右音量差調整処理（S403）として、反射音を主に知覚する耳の位置に合わせて直接音の音量を調整する。例えば、選択部1302は、直接音が受聴者に到来する時の音量に、 $(1 - 0.3 \sin(\theta))$ （ $0 \leq \theta \leq 180$ ）を乗じることによって、直接音が受聴者に到来する時の音量を減衰させる。

[0388] 選択部1302は、上記のように補正された直接音の音量と、反射音の音量との音量比を算出し、算出された音量比と閾値とを比較することによって、反射音を選択するか否かの判定を行う。これにより、両耳間で生じる音量差が補正され、反射音に影響する直接音の音量がより正確に導出され、反射音を選択するか否かの判定がより正確に行われる。

[0389] また、選択部1302は、左右音量差調整処理（S403）に加えて、遅延調整処理（S404）として、反射音を知覚する耳の位置に合わせて、直接音到来時間を遅延させてもよい。具体的には、選択部1302は、直接音到来時間に $(a(\sin \theta + \theta) / c)$ ms（ a は頭部の半径、 c は音速）を加えることにより、直接音到来時間を遅延させてもよい。

[0390] [レンダリング部の動作の第3変形例]

到来方向に応じた閾値を設定する方法について説明する。

- [0391] 図23は、選択処理のさらに別の一例を示すフローチャートである。図14の例と共通の処理については説明を省略する。図23の例では、選択部1302が、到来方向に応じた閾値を用いて反射音を選択する。
- [0392] 具体的には、選択部1302は、解析部1301で算出された直接音到来経路(p d)、反射音到来経路(p r)、及び、アバターの向き情報D1から、アバターの向きを基準に用いて定められる直接音到来方向(θ)及び反射音到来方向(γ) (反射音の方向(γ))を算出する。すなわち、選択部1302は、直接音到来方向(θ)及び反射音到来方向(γ)を検出する(S231)。アバターの向きは、受聴者の向きに対応する。アバターの向き情報D1は、入力信号に含まれていてもよい。
- [0393] 選択部1302は、直接音到来方向(θ)及び反射音到来方向(γ)に加えて時間差(T)を含む3つのインデックスを用いて、図15に示されるような三次元配列から、選択処理に用いられる閾値を特定する(S232)。
- [0394] 例として、図22のように、アバターと音源オブジェクトと障害物オブジェクトとが配置されている場合に、選択処理に用いられる閾値の設定方法を説明する。
- [0395] 入力信号から、アバター、音源オブジェクト及び障害物オブジェクトの位置情報と、アバターの向き情報D1とが取得される。これらの位置情報及び向き情報D1を用いて、アバターの向きが0度と定められる場合における直接音の方向(θ)と反射音の音像の方向(γ)とが算出される。図22の場合、直接音の方向(θ)は、20度程度であり、反射音の音像の方向(γ)は、265度(−95度)程度である。
- [0396] 次に、図15に示される三次元の配列で記憶された閾値データを参照して、2つの方向(θ)及び(γ)の値と、解析部1301で算出された時間差(T)の値とに対応する配列領域から閾値が特定される。算出された(θ)、(γ)、(T)の値に対応するインデックスが存在しない場合、最も近いインデックスに対応する閾値が特定されてもよい。

- [0397] 別の方法として、算出された (θ) 、 (γ) 、 (T) の値に近い1つ以上のインデックスに対応する1つ以上の閾値に基づいて、補間、内挿又は外挿等の処理を行うことにより、閾値が特定されてもよい。例えば、 $(0$ 度、 225 度、 $T)$ 、 $(0$ 度、 270 度、 $T)$ 、 $(45$ 度、 225 度、 $T)$ 及び $(45$ 度、 270 度、 $T)$ の4つのインデックスに対応する4つの閾値に基づいて、 $(20$ 度、 265 度、 $T)$ に対応する閾値が特定されてもよい。
- [0398] 直接音の到来方向の角度 (θ) と反射音の到来方向の角度 (γ) との差に基づく選択処理について説明する。
- [0399] 例えば、図16に示されるような、直接音の到来方向 (θ) と反射音の到来方向 (γ) との角度差 (Φ) と、時間差 (T) とを二次元インデックス配列として有する閾値データが、予め作成され、設定されていてもよい。この場合、選択処理で角度差 (Φ) と時間差 (T) とが参照される。あるいは、選択処理で直接音の到来方向の角度 (θ) と反射音の到来方向の角度 (γ) との角度差 (Φ) が算出され、算出された角度差 (Φ) が閾値の特定に用いられてもよい。
- [0400] あるいは、当該角度差 (Φ) と直接音の到来方向 (θ) と時間差 (T) との組み合わせ、又は、角度差 (Φ) と反射音の到来方向 (γ) と時間差 (T) との組み合わせをインデックス配列として有する閾値データが設定されてもよい。
- [0401] あるいは、図15に示されるような、 (θ) 、 (γ) 及び (T) の値を三次元インデックス配列として有する閾値データが設定されてもよい。
- [0402] [レンダリング部の動作の第4変形例]
上述の解析部1301、選択部1302及び再生部1303で行われる処理は、例えば特許文献3で説明されているようなパイプライン処理として行われてもよい。
- [0403] 図24は、レンダリング部1300がパイプライン処理を行うための構成例を示すブロック図である。
- [0404] 図24のレンダリング部1300は、残響処理部1311、初期反射処理

部 1 3 1 2、距離減衰処理部 1 3 1 3、選択部 1 3 1 4、生成部 1 3 1 5 及びバイノーラル処理部 1 3 1 6 を備える。これらの複数の構成要素は、図 7 に示されたレンダリング部 1 3 0 0 の複数の構成要素で構成されていてもよいし、図 5 に示された音声信号処理装置 1 0 0 1 の複数の構成要素の少なくとも一部で構成されていてもよい。

[0405] パイプライン処理とは、音響効果を付与するための処理を複数の処理に分割し、複数の処理を 1 つずつ順番に実行することを指す。複数の処理のそれぞれでは、例えば、音声信号に対する信号処理、又は、信号処理に用いられるパラメータの生成等が実行される。

[0406] レンダリング部 1 3 0 0 は、パイプライン処理として、残響処理、初期反射処理、距離減衰処理及びバイノーラル処理等を行ってもよい。ただし、これらの処理は一例であり、パイプライン処理は、これら以外の処理を含んでもよいし、一部の処理を含んでいなくてもよい。例えば、パイプライン処理は、回折処理及びオクルージョン処理を含んでもよい。また、例えば、残響処理が、不要な場合、省略されてもよい。

[0407] また、各処理がステージと表現されてもよい。また、各処理の結果、生成された反射音等の音声信号は、レンダリングアイテムと表現されてもよい。パイプライン処理における複数のステージ、及び、それらの順番は、図 2 4 に示された例に限られない。

[0408] ここでは、選択処理で用いられるパラメータ（直接音及び反射音に関する到来経路、到来時間及び音量比）が、レンダリングアイテムを生成するための複数のステージのいずれかで算出される。つまり、反射音の選択に用いられるパラメータは、レンダリングアイテムを生成するためのパイプライン処理の一部で算出される。なお、全てのステージがレンダリング部 1 3 0 0 で行われなくてもよい。例えば、一部のステージが、省略されてもよいし、レンダリング部 1 3 0 0 以外で行われてもよい。

[0409] パイプライン処理にステージとして含まれ得る残響処理、初期反射処理、距離減衰処理、選択処理、生成処理及びバイノーラル処理について説明する

。各ステージにおいて、入力信号に含まれるメタデータを解析して、反射音の生成に用いられるパラメータが算出されてもよい。

[0410] 残響処理では、残響処理部1311が、残響音を示す音声信号、又は、音声信号の生成に用いられるパラメータを生成する。残響音とは、直接音の後に残響として受聴者に到来する音である。一例として、残響音は、後述する初期反射音が受聴者へ到来した後の比較的后期（例えば、直接音の到来時から百数十ms程度）の段階で、初期反射音よりも多くの回数（例えば、数十回）の反射を経て受聴者へ到来する音である。

[0411] 残響処理部1311は、入力信号に含まれる音声信号及び空間情報を参照し、残響音を生成するための関数として事前に準備された所定の関数を用いて残響音を計算する。

[0412] 残響処理部1311は、入力信号に含まれる音声信号に、公知の残響生成方法を適用して、残響音を生成してもよい。公知の残響生成方法の例は、シュレーダー法であるが、公知の残響生成方法は、シュレーダー法に限られない。また、残響処理部1311は、公知の残響生成方法の適用において、空間情報が示す音再生空間の形状及び音響特性を用いる。これにより、残響処理部1311は、残響音を生成するためのパラメータを算出することができる。

[0413] 初期反射処理では、初期反射処理部1312が、空間情報に基づいて、初期反射音を生成するためのパラメータを算出する。初期反射音は、音源オブジェクトから直接音が受聴者へ到来した後の比較的前期（例えば、直接音の到来時から数十ms程度）の段階で、1回以上の反射を経て受聴者へ到来する反射音である。

[0414] 初期反射処理部1312は、例えば音声信号及びメタデータを参照し、音源オブジェクトから反射オブジェクトで反射して受聴者に到来する反射音の経路を計算する。例えば、経路の計算において、三次元音場（空間）の形状、三次元音場の大きさ、構造物等の反射オブジェクトの位置、及び、反射オブジェクトの反射率等が用いられてもよい。

- [0415] また、初期反射処理部 1 3 1 2 が、直接音の経路も算出してもよい。当該経路の情報は、初期反射処理部 1 3 1 2 が初期反射音を生成するためのパラメータとして用いられてもよいし、選択部 1 3 1 4 が反射音を選択するためのパラメータとして用いられてもよい。
- [0416] 距離減衰処理では、距離減衰処理部 1 3 1 3 が、直接音及び反射音の経路の長さに基づいて、受聴者に到来する直接音及び反射音の音量を算出する。受聴者に到来する直接音及び反射音の音量は、音源の音量に対して、受聴者までの経路の距離に比例して減衰（距離に反比例）する。したがって、距離減衰処理部 1 3 1 3 は、音源の音量を直接音の経路の長さで除して直接音の音量を算出することができ、音源の音量を反射音のパスの長さで除して反射音の音量を算出することができる。
- [0417] 選択処理では、選択部 1 3 1 4 が、選択処理の前に算出されたパラメータに基づいて、生成対象反射音を選択する。生成対象反射音の選択には、本開示のいずれかの選択方法が用いられてもよい。
- [0418] 選択処理は、全ての反射音に対して実行されてもよいし、前述したように評価処理に基づいて評価値の高い反射音に対してのみ実行されてもよい。つまり、評価値の低い反射音は、選択処理を行うまでもなく非選択と判定されてもよい。例えば、音量が非常に小さい反射音は、反射音の評価値が低いとみなされ、非選択と判定されてもよい。
- [0419] また、例えば、全ての反射音に対して選択処理が実施されてもよい。そして、選択処理で選択された反射音の評価値が判定され、判定された評価値が低い反射音は、非選択と判定し直されてもよい。
- [0420] 選択処理と評価処理とは、それぞれ独立で実行されてもよく、組み合わせて実行されてもよい。選択処理と評価処理とが組み合わせて実行する場合、2つの処理のうちのいずれが先に実行されてもよい。
- [0421] 生成処理において、生成部 1 3 1 5 は、直接音及び反射音を生成する。例えば、生成部 1 3 1 5 は、入力信号に含まれる音声信号から、直接音の到来時刻及び到来時音量に基づいて直接音を生成する。また、生成部 1 3 1 5 は

、選択処理において選択された反射音について、入力信号に含まれる音声信号から、反射音の到来時刻及び到来時音量に基づいて反射音を生成する。

[0422] バイノーラル処理において、バイノーラル処理部1316は、直接音の音声信号が音源オブジェクトの方向から受聴者に到来する音として知覚されるように、信号処理を実行する。さらに、バイノーラル処理部1316は、選択部1314で選択された反射音が反射オブジェクトから受聴者に到来する音として知覚されるように、信号処理を実行する。

[0423] 例えば、バイノーラル処理部1316は、音空間における受聴者の位置及び向きに基づいて、音源オブジェクトの位置又は障害物オブジェクトの位置から音が受聴者に到来するように、H R I R DBを適用する処理を実行する。

[0424] なお、H R I R (Head-Related Impulse Responses) は、1個のインパルスが発生させたときの応答特性である。具体的には、H R I Rは、耳殻、人頭及び肩まで含めた周辺物によって生じる音の変化が伝達関数として表現された頭部伝達関数をフーリエ変換によって周波数領域における表現から時間領域における表現に変換することで得られる応答特性である。H R I R DBは、このような情報を含むデータベースである。

[0425] また、音空間における受聴者の位置及び向きは、例えば、仮想的な音空間における仮想的な受聴者の位置及び向きである。受聴者の頭部の動きに合わせて、仮想的な音空間における仮想的な受聴者の位置及び向きが変化してもよい。また、仮想的な音空間における仮想的な受聴者の位置及び向きは、センサ1405から取得された情報に基づいて定められてもよい。

[0426] 上記の処理に用いられるプログラム、空間情報、H R I R DB、閾値データ又はその他のパラメータ等は、音声信号処理装置1001に備わるメモリ1404又は音声信号処理装置1001の外部から取得される。

[0427] また、パイプライン処理は、他の処理を含んでもよい。そして、レンダリング部1300は、パイプライン処理に含まれる他の処理を行うための

図示されない処理部を含んでいてもよい。例えば、レンダリング部1300は、回折処理部及びオクルージョン処理部を含んでいてもよい。

[0428] 回折処理部は、三次元音場（空間）における受聴者と音源オブジェクトとの間の障害物オブジェクトに起因する回折音を含む音を示す音声信号を生成する処理を実行する。回折音は、音源オブジェクトと受聴者との間に障害物オブジェクトが存在する場合に、当該障害物オブジェクトを回り込むようにして音源オブジェクトから受聴者へ到来する音である。

[0429] 回折処理部は、例えば音声信号及びメタデータを参照し、音源オブジェクトから障害物オブジェクトを迂回して受聴者へと到来する回折音の経路を計算し、当該経路に基づいて回折音を生成する。経路の計算において、三次元音場（空間）における音源オブジェクト、受聴者及び障害物オブジェクトの位置、並びに、障害物オブジェクトの形状及び大きさ等が用いられてもよい。

[0430] オクルージョン処理部は、障害物オブジェクトの向こう側に音源オブジェクトが存在する場合に、空間情報及び障害物オブジェクトの材質等の情報に基づいて、音源オブジェクトから障害物オブジェクトを通り抜けて漏れ聞こえる音の音声信号を生成する。

[0431] [音源オブジェクトの例]

上記では、音源オブジェクトに付与される位置情報は、仮想空間内における「点」を音源オブジェクトの位置として示す。すなわち、上記では、音源は、「点音源」として定義されている。

[0432] 一方で、仮想空間における音源は、長さ、大きさ及び形状等を有する物体として、すなわち、点音源でない、空間的に拡張された音源として定義されてもよい。この場合、受聴者と音源との距離、及び、音の到来方向が確定しない。したがって、このような音源に起因する反射音は、解析部1301で解析を行うまでもなく、又は、解析結果の如何にかかわらず、選択部1302で選択されることに限定されてもよい。これにより、反射音を選択しないことによって生じるかもしれない音質の劣化を避けることが可能である。

[0433] あるいは、当該物体の重心等の代表点が定められて、その代表点から音が発生していると仮定して本開示の処理が適用されてもよい。この場合、音源の空間的な拡張の情報に応じて、閾値が調整されてもよい。

[0434] [直接音及び反射音の例]

例えば、直接音は、反射オブジェクトによって反射されていない音であり、反射音は、反射オブジェクトによって反射された音である。直接音は、音源から反射オブジェクトによって反射することなく受聴者に到来した音であってもよいし、反射音は、音源から反射オブジェクトによって反射して受聴者に到来した音であってもよい。

[0435] また、直接音及び反射音のそれぞれは、受聴者に到来した音に限られず、受聴者に到来する前の音であってもよい。例えば、直接音は、音源から出力された音、さらに言い換えれば、音源の音であってもよい。

[0436] 図25は、音の透過及び回折を示す図である。図25に示されるように、障害物オブジェクトが音源オブジェクトと受聴者との間に存在することによって、直接音が受聴者に到来しない場合がある。この場合、音源オブジェクトから発せられ、障害物オブジェクトを透過し、受聴者に到来した音が、直接音とみなされてもよい。そして、音源オブジェクトから発せられ、障害物オブジェクトによって回折し、受聴者に到来した音が、反射音とみなされてもよい。

[0437] また、選択処理において比較される2つの音は、1つの音源が発する音に基づく直接音及び反射音に限定されない。例えば、1つの音源が発する音に基づく2つの反射音の間で比較を行って音の選択が行われてもよい。この場合、本開示における直接音は先に受聴者に届く音と読み替えられてもよく、本開示における反射音は後から受聴者に届く音と読み替えられてもよい。

[0438] [ビットストリームの構造例]

ビットストリームには、例えば、音声信号とメタデータとが含まれる。音声信号は、音が表現された音データであって、音の周波数及び強弱に関する情報等を示す。また、メタデータは、音場の空間である音空間に関する空間

情報を含む。

- [0439] 例えば、空間情報は、音声信号に基づく音を聞く受聴者が位置する空間に関する情報である。具体的には、空間情報は、音空間（例えば三次元音場）における所定の位置に音像を定位させるための、つまり、所定の位置に対応する方向から到来する音を受聴者に知覚させるための、当該所定の位置（定位位置）に関する情報である。空間情報には、例えば、音源オブジェクト情報と、受聴者の位置を示す位置情報とが含まれる。
- [0440] 音源オブジェクト情報は、音声信号に基づく音を発生させる音源オブジェクトの情報である。つまり、音源オブジェクト情報は、音声信号を再生する物体（音源オブジェクト）に関する情報であって、仮想的な音空間に配置される仮想的な音源オブジェクトに関する情報である。ここで、仮想的な音空間は、音を発生させる物体が配置される実空間に対応していてもよく、仮想的な音空間における音源オブジェクトは、実空間において音を発生させる物体に対応していてもよい。
- [0441] 音源オブジェクト情報は、音空間に配置される音源オブジェクトの位置、音源オブジェクトの向き、音源オブジェクトが発する音の指向性、音源オブジェクトが生物に属するか否か、及び、音源オブジェクトが動体であるか否か等を示してもよい。例えば、音声信号は、音源オブジェクト情報によって示される1以上の音源オブジェクトに対応付けられている。
- [0442] ビットストリームは、例えばメタデータ（制御情報）と音声信号とで構成されるデータ構造を有する。
- [0443] 音声信号及びメタデータは、1つのビットストリームに含まれていてもよいし、複数のビットストリームに別々に含まれていてもよい。また、音声信号及びメタデータは、1つのファイルに含まれていてもよいし、複数のファイルに別々に含まれていてもよい。
- [0444] ビットストリームは、音源毎に存在してもよいし、再生時間毎に存在してもよい。再生時間毎にビットストリームが存在する場合であっても、同時に複数のビットストリームが並列で処理されてもよい。

- [0445] メタデータは、ビットストリーム毎に付与されてもよく、複数のビットストリームを制御するための情報として複数のビットストリームにまとめて付与されてもよい。この場合、複数のビットストリームがメタデータを共有してもよい。また、メタデータは、再生時間毎に付与されてもよい。
- [0446] 複数のビットストリーム又は複数のファイルが存在する場合、1つ以上のビットストリーム又は1つ以上のファイルに、関連ビットストリーム又は関連ファイルを示す情報が含まれていてもよい。あるいは、全てのビットストリームのそれぞれ又は全てのファイルのそれぞれに、関連ビットストリーム又は関連ファイルを示す情報が含まれていてもよい。
- [0447] ここで、関連ビットストリーム又は関連ファイルとは、例えば、音響処理の際に同時に用いられる可能性のあるビットストリーム又はファイルである。また、関連ビットストリーム又は関連ファイルを示す情報をまとめて記述したビットストリーム又はファイルが含まれていてもよい。
- [0448] ここで、関連ビットストリーム又は関連ファイルを示す情報は、例えば、関連ビットストリーム又は関連ファイルを示す識別子であってもよい。また、関連ビットストリーム又は関連ファイルを示す情報は、例えば、関連ビットストリーム又は関連ファイルを示すファイル名、URL (Uniform Resource Locator) 又はURI (Uniform Resource Identifier) 等であってもよい。
- [0449] この場合、取得部は、関連ビットストリーム又は関連ファイルを示す情報に基づいて、関連ビットストリーム又は関連ファイルを特定し取得する。また、ビットストリーム又はファイルに、関連ビットストリーム又は関連ファイルを示す情報が含まれ、かつ、別のビットストリーム又は別のファイルに、関連ビットストリーム又は関連ファイルを示す情報が含まれていてもよい。
- [0450] ここで、関連ビットストリーム又は関連ファイルを示す情報を含むファイルは、例えばコンテンツの配信に用いられるマニフェストファイル等のような制御ファイルであってもよい。

[0451] なお、全てのメタデータ又は一部のメタデータは、音声信号のビットストリーム以外から取得されてもよい。例えば、音響を制御するためのメタデータと、映像を制御するためのメタデータとのうち、いずれか一方のメタデータが、ビットストリーム以外から取得されてもよいし、両方のメタデータが、ビットストリーム以外から取得されてもよい。

[0452] また、映像を制御するためのメタデータが、立体音響再生システム1000で取得されるビットストリームに含まれていてもよい。この場合、立体音響再生システム1000は、画像を表示する表示装置、又は、立体映像を再生する立体映像再生装置へ、映像を制御するためのメタデータを出力してもよい。

[0453] [メタデータに含まれる情報の例]

メタデータは、音空間で表現されるシーンの記述に用いられる情報であってもよい。ここで、シーンとは、メタデータを用いて立体音響再生システム1000でモデリングされる音空間における三次元映像及び音響イベントを表す全ての要素の集合体を指す用語である。

[0454] つまり、メタデータは、音響処理を制御するための情報だけでなく、映像処理を制御するための情報も含んでいてもよい。メタデータには、音響処理を制御するための情報と、映像処理を制御するための情報とのうち、いずれか一方のみが含まれていてもよいし、両方が含まれていてもよい。

[0455] 立体音響再生システム1000は、ビットストリームに含まれるメタデータ、及び、追加で取得されるインタラクティブな受聴者の位置情報等を用いて、音声信号に音響処理を行うことで、仮想的な音響効果を生成する。音響効果のうち、初期反射処理と、障害物処理と、回折処理と、遮断処理と、残響処理とが行われてもよいし、メタデータを用いて他の音響処理が行われてもよい。例えば、距離減衰効果、ローカリゼーション又はドップラー効果等の音響効果が付加されてもよい。

[0456] また、音響効果の全て又は一部のオンオフを切り替える情報、又は、音響効果の複数の処理に対する優先度情報が、メタデータに付加されてもよい。

- [0457] また、一例として、メタデータは、音源オブジェクト及び障害物オブジェクトを含む音空間に関する情報と、音空間内において所定位置に音像を定位させる（つまり所定方向から到来する音を受聴者に知覚させる）ための、定位位置に関する情報を含む。
- [0458] ここで、障害物オブジェクトは、音源オブジェクトが発する音が受聴者へと到来するまでの間において、例えば音を遮ったり、音を反射したりして、受聴者が知覚する音に影響を及ぼし得るオブジェクトである。障害物オブジェクトは、静止物体の他に、動物又は機械等の動体を含み得る。動物は、人等であってもよい。
- [0459] また、音空間に複数の音源オブジェクトが存在する場合、任意の音源オブジェクトにとっては、他の音源オブジェクトは障害物オブジェクトとなり得る。つまり、建材又は無生物等の音を発しないオブジェクトである非発音オブジェクトも、音を発する音源オブジェクトも、障害物オブジェクトとなり得る。
- [0460] メタデータには、音空間の形状、音空間における障害物オブジェクトの形状及び位置、音空間における音源オブジェクトの形状及び位置、並びに、音空間における受聴者の位置及び向きの全て又は一部を表す情報が含まれる。
- [0461] 音空間は、閉空間及び開空間のいずれであってもよい。また、メタデータには、音空間において音を反射し得る障害物オブジェクトの反射率を表す情報が含まれていてもよい。例えば、音空間の境界を構成する床、壁又は天井等も障害物オブジェクトを構成し得る。
- [0462] 反射率は、反射音と入射音とのエネルギーの比であって、音の周波数帯域毎に設定されていてもよい。勿論、反射率は、音の周波数帯域に依らず、一律に設定されていてもよい。なお、音空間が開空間の場合、例えば一律で設定された減衰率、回折音及び初期反射音等のパラメータが用いられてもよい。
- [0463] メタデータは、障害物オブジェクト又は音源オブジェクトに関するパラメータとして、反射率以外の情報を含んでいてもよい。例えば、メタデータは

、音源オブジェクト及び非発音オブジェクトの両方に関するパラメータとして、オブジェクトの素材に関する情報を含んでいてもよい。具体的には、メタデータは、拡散率、透過率及び吸音率等の情報を含んでいてもよい。

[0464] 音源オブジェクトに関する情報は、音量、放射特性（指向性）、再生条件、1つのオブジェクトにおける音源の数及び種類、並びに、オブジェクトにおける音源領域を示す情報等を含んでいてもよい。再生条件では、例えば、継続的に流れ続ける音なのかイベント発動する音なのかが定められてもよい。オブジェクトにおける音源領域は、受聴者の位置とオブジェクトの位置との相対的な関係で定められてもよいし、オブジェクトを基準に用いて定められてもよい。

[0465] 例えば、受聴者の位置とオブジェクトの位置との相対的な関係で音源領域が定められる場合、受聴者から見て、オブジェクトの右側から音E、及び、オブジェクトの左側から音Fを受聴者に知覚させることが可能である。

[0466] また、オブジェクトを基準に用いて音源領域が定められる場合、オブジェクトを基準に用いて、オブジェクトのどの領域からどの音を出すかを固定することが可能である。例えば、受聴者がオブジェクトを正面から見た場合、オブジェクトの右側から高い音、及び、オブジェクトの左側から低い音を受聴者に知覚させることが可能である。そして、受聴者がオブジェクトを背面から見た場合、オブジェクトの右側から低い音、及び、オブジェクトの左側から高い音を受聴者に知覚させることが可能である。

[0467] 空間に関するメタデータは、初期反射音までの時間、残響時間、及び、直接音と拡散音との比率等を含んでいてもよい。直接音と拡散音との比率がゼロの場合、直接音のみを受聴者に知覚させることが可能である。

[0468] [簡単なまとめ]

ここで本実施の形態を簡単にまとめる。

[0469] 直接音と反射音との関係性が分析され、直接音を先行音とし反射音を後続音とした場合に、先行音効果が生じる関係性であるとき、言い換えると、反射音がエコー検知限を下回るとき、その反射音は知覚されないため、その反

射音が削除されても受聴者の聴感上の影響は少ない。

[0470] 図 2 6 は、本実施の形態に係る受聴者と障害物オブジェクトとの位置関係の一例を示す図である。図 2 7 は、本実施の形態に係る受聴者と障害物オブジェクトとの位置関係の他の一例を示す図である。なお、図 2 6 は図 9 が示す位置関係に等しく、図 2 7 は図 1 0 が示す位置関係に等しい。また、図 2 8 は、本実施の形態に係るエコー検知限閾値の一例である。なお、図 2 8 が示すエコー検知限閾値は、図 1 2 C など示された閾値データの一例である。

[0471] 例えば、図 2 6 の位置関係と、図 2 7 の位置関係とを比較すると、図 2 6 の位置関係での受聴者が受聴する反射音の音量は、図 2 7 の位置関係での受聴者が受聴する反射音の音量よりも小さい。これは、図 2 6 の位置関係での反射音が到来する経路長が、図 2 7 の位置関係での反射音が到来する経路長よりも長いためである。

[0472] 従って、反射音の音量だけで判定した場合、図 2 6 で示される反射音の方が図 2 7 で示される反射音よりも聴感的な影響が小さい。しかしながら、図 2 6 で示される反射音の受聴位置への到来時刻と、図 2 7 で示される反射音の受聴位置への到来時刻を比較すると、図 2 6 に示した場合の方が遅い。

[0473] このため、エコー検知限の観点から判断が行われると、図 2 8 に示したように、図 2 7 で示される反射音はエコー検知限を下回っているため反射音として受聴者に知覚されず、図 2 6 で示される反射音はエコー検知限を上回っているため反射音として受聴者に知覚される。

[0474] 本実施の形態においては、これが利用され、反射音の聴感的な重要度を判定して、重要でない反射音は再生しないようにすることで、反射音の処理に係る演算量が削減される。

[0475] 本実施の形態を簡単にまとめると以上の通りである。

[0476] ところで、音量比 L と閾値とに着目する。

[0477] 音量比 L と比較される閾値（エコー検知限閾値）は、聴覚の先行音効果に基づいているので、上記のように、音量比 L が物理的な特性だけで計算され

たものである場合、当該反射音を再生するか再生しないかを選択した結果（選択結果）が、受聴者の実際の感覚と整合しない場合がある。

[0478] すなわち、受聴者の聴覚の感度が考慮されずに、反射音を示す音声信号（より具体的には、当該音声信号に基づく出力信号）を出力するか否かが選択され、そのような出力信号が出力され当該出力信号が示す音（反射音）を受聴者が受聴してしまう場合がある。このような場合には、受聴者は、自身の聴覚の感覚とは異なる音を受聴することとなるため、違和感を覚えてしまう。

[0479] 従って、以下では、音空間において、聴覚の感度が考慮された上で、演算量及び演算負荷を適切に削減することができる音声信号処理方法について、さらに詳細に説明をする。

[0480] （実施の形態２）

以下、実施の形態２について説明する。以下では、実施の形態１との相違点を中心に説明し、共通点の説明を省略又は簡略化する。

[0481] [レンダリング部の構成]

まず、本実施の形態に係るレンダリング部２３００の構成について説明する。図２９は、本実施の形態に係るレンダリング部２３００の構成例を示すブロック図である。

[0482] レンダリング部２３００は、解析部２３０１と、選択部２３０２と、再生部２３０３とを備える。なお、上記の通り、本実施の形態に係る音声信号処理装置は復号装置の一例であり、当該復号装置はデコーダを含み、当該デコーダがレンダリング部２３００を含む。すなわち、本実施の形態に係る音声信号処理装置が、解析部２３０１と、選択部２３０２と、再生部２３０３とを備えている、と言える。レンダリング部２３００は、入力信号に含まれる音データに対して音響処理を付加し出力する。

[0483] 実施の形態１と同じく、入力信号は、例えば、空間情報とセンサ情報と音データとで構成される。空間情報は、非発音オブジェクト（障害物オブジェクト）の反射係数、透過係数、回折係数等の物理情報も含む。

- [0484] なお、本実施の形態においては、主に、間接音の一例である反射音を用いて説明するが、反射音にかえて間接音が用いられても同様の処理が行われる。また、間接音は、一例として反射音又は回折音等である。
- [0485] 解析部2301は、実施の形態1に係る解析部1301が行う全ての処理又は一部の処理を行えるとよい。
- [0486] 解析部2301は、実施の形態1に係る解析部1301と同様に、入力信号に含まれる音声信号、並びに、空間情報管理部1201及び1211から受け取った空間情報の解析を行う。これにより、解析部2301は、再生部2303での直接音及び反射音の生成に必要な情報、並びに、反射音を生成するか否かの選択に必要な情報を算出する。解析部2301がこれらの情報を算出する方法は、実施の形態1で説明されたとおりである。
- [0487] また、解析部2301は、実施の形態1に係る解析部1301が図8のS101で行ったように、入力信号の解析処理を行う。つまり、解析部2301は、本実施の形態に係る音声信号処理装置に入力された入力信号を解析して、音空間で発生し得る直接音及び反射音を検出する。
- [0488] このような直接音及び反射音が検出された場合に、解析部2301は、空間情報と音データとに基づいて、当該反射音を示す音声信号、及び、当該直接音を示す音声信号を作成する。
- [0489] より具体的には、解析部2301は、空間情報に含まれる音源オブジェクトの位置情報、非発音オブジェクト（障害物オブジェクト）の位置情報、受聴者の位置情報及び物理情報と、音データとに基づいて、反射音を示す音声信号及び直接音を示す音声信号を作成する。
- [0490] すなわち、解析部2301は、空間情報と音データとに基づいて、仮想空間内に発生する音声信号を作成し、作成された音声信号に、当該音声信号を特定する属性を示す属性情報を付与し、属性情報を含む音声信号を作成する。属性情報を含む音声信号は、仮想空間内で発生する音ごとに作成される。属性は、当該音声信号が示す音声が直接音又は反射音（間接音）であるかを示す情報を含む。本実施の形態においては、一例として、属性は、当該音声

信号が示す音声信号が直接音又は反射音であることを示す情報である。さらに属性情報は、音声信号を音空間内に放射するために必要な情報、例えば、ゲイン情報、周波数帯域幅ごとのゲイン特性の情報、位置の情報及び指向性の情報等を含んでいてもよい。すなわち、当該必要な情報は、属性情報に保持されていてもよい。また、属性情報はメタデータとして音声信号に紐づけられてもよい。属性情報に含まれる、音声信号の周波数帯域幅ごとのゲイン特性は、例えば、入力情報に含まれる空間情報に基づいて特定されてもよい。聴覚の感度を示す周波数特性を示す情報は、例えば、入力情報に含まれる空間情報に基づいて特定されてもよく、特に受聴者のアバターに紐づく情報として特定されてもよい。

- [0491] なお簡単のため、属性が反射音（間接音）を示す情報である音声信号を、反射音（間接音）を示す音声信号と記載し、属性が直接音を示す情報である音声信号を、直接音を示す音声信号と記載する場合がある。
- [0492] 1つの音源から受聴者の頭部へ直接届く音が直接音であり、当該1つの音源から出力された後、非発音オブジェクトに反射して又は非発音オブジェクトにより回折して受聴者の頭部へ届く音が間接音（反射音又は回折音）である。
- [0493] なお、本実施の形態においては、解析部2301は、反射音（間接音）を示す音声信号と、当該反射音（間接音）に係る直接音を示す音声信号とを作成する。
- [0494] また、間接音に係る直接音とは、当該間接音と同じ音源に由来する直接音を意味する。直接音に係る間接音とは、当該直接音と同じ音源に由来する間接音を意味する。さらに言えば、反射音とは、当該反射音に係る直接音が反射物によって反射された音である。
- [0495] 属性が反射音（間接音）を示す情報である音声信号には、当該反射音（間接音）に係る直接音の音声信号を示す情報が含まれている。
- [0496] 解析部2301は、作成した音声信号を、解析部2301が有するメモリに記憶させるとよい。また、解析部2301は、複数の音声信号を作成し、

当該メモリに記憶させる。

[0497] また、解析部2301は、実施の形態1と同じく、直接音及び反射音のそれぞれについて、受聴位置に到来するまでの経路、到来するまでにかかる時間、及び、到来時の音量等に関する値を算出しているとよい。同様に解析部2301は、直接音と反射音との関係を示す情報であって、例えば、直接音と反射音との到来する時間差（直接音と反射音との時間差）に関する値等を算出しているとよい。

[0498] なお、到来時の音量等の一例が、反射音到来時音量（ I_r ）及び直接音到来時音量（ I_d ）である。直接音到来時音量（ I_d ）は、直接音が、仮想空間における受聴者が居る位置である受聴位置に到来するときの音量を意味し、また換言すると、受聴位置における直接音音量である。反射音到来時音量（ I_r ）は、間接音の一例である反射音が受聴位置に到来するときの音量を意味し、また換言すると、受聴位置における間接音音量（反射音音量）である。

[0499] 本実施の形態においては、属性が反射音を示す情報である音声信号及び属性が直接音を示す情報である音声信号のそれぞれは、受聴位置における音声信号が示す音の音量を示す情報を含んでもよい。つまり本実施の形態においては、属性が反射音を示す情報である音声信号は、間接音音量（反射音音量）として、上記反射音到来時音量（ I_r ）を示す情報を含む。同様に、属性が直接音を示す情報である音声信号は、直接音音量として、上記直接音到来時音量（ I_d ）を示す情報を含む。

[0500] なお、属性が反射音を示す情報である音声信号は、間接音音量（反射音音量）と、当該反射音に係る直接音の直接音音量とを示す情報を含んでもよい。同様に、属性が直接音を示す情報である音声信号は、直接音音量と、当該直接音に係る間接音（反射音）の間接音音量（反射音音量）とを示す情報を含んでもよい。

[0501] 選択部2302は、実施の形態1に係る選択部1302が行う全ての処理又は一部の処理を行えるとよい。また、選択部2302は、解析部2301

によって作成された音声信号に基づく出力信号が、再生部2303によって出力（再生）されるか否かを判定する。すなわち、選択部2302は、まず、解析部2301によって作成された複数の音声信号（例えば反射音を示す音声信号）のうち1つの音声信号を指定し、さらに、再生部2303が指定された当該1つの音声信号に基づく出力信号を生成し出力するか否かを選択する。

[0502] 選択部2302は、取得部2302aと、第1算出部2302bと、第2算出部2302cと、選択処理部2302dとを有する。

[0503] 取得部2302aは、解析部2301によって作成され解析部2301のメモリに記憶された属性情報を含む音声信号を取得する。取得部2302aは、例えば、反射音（間接音）を示す音声信号と、当該反射音（間接音）に係る直接音を示す音声信号とを取得する。また、取得部2302aは、解析部2301によって算出された直接音と反射音との時間差に関する値を取得する。

[0504] 第1算出部2302bは、取得部2302aによって取得された間接音を示す音声信号に基づいて、第1音量を算出する。ここでは、第1算出部2302bは、当該音声信号が示す間接音が、受聴者が居る位置である受聴位置に到来するときの音量である間接音音量に基づく第1音量を算出する。

[0505] より具体的には、第1算出部2302bは、当該音声信号が示す反射音が、受聴者が居る位置である受聴位置に到来するときの音量である反射音音量（つまりは反射音到来時音量（ I_r ））に基づく第1音量を算出する。

[0506] 第2算出部2302cは、取得部2302aによって取得された直接音を示す音声信号に基づいて、第2音量を算出する。ここでは、第2算出部2302cは、当該音声信号が示す直接音が、受聴者が居る位置である受聴位置に到来するときの音量である直接音音量（つまりは直接音到来時音量（ I_d ））に基づく第2音量を算出する。

[0507] なお、本実施の形態においては、第2音量は、直接音音量とは異なる音量であるが、第2音量として直接音音量がそのまま用いられてもよい。

- [0508] また、選択処理部2302dは、算出された第2音量と算出された第1音量との音量比に基づいて、再生部2303が、取得された反射音（間接音）を示す音声信号に基づく出力信号を出力するか否かを選択する。
- [0509] 選択処理部2302dが、再生部2303が当該出力信号を出力することを選択した場合に、選択部2302は、取得された音声信号を再生部2303に出力する。
- [0510] 再生部2303は、実施の形態1に係る再生部1303が行う全ての処理又は一部の処理を行えるとよい。また、再生部2303は、選択部2302から出力された音声信号を取得し、取得された音声信号に基づく出力信号を出力する。
- [0511] 再生部2303は、取得された音声信号に、バイノーラルフィルタ処理等の処理を行うことで、出力信号を生成して出力する。バイノーラルフィルタ処理は、例えば、取得された音声信号に頭部伝達関数を用いた処理が行われることで実現される。
- [0512] また、再生部2303は、取得部2302aによって取得された直接音を示す音声信号と、生成された出力信号とを合成して出力してもよい。
- [0513] また、再生部2303は、選択部2302から出力された音声信号に、バイノーラルフィルタ処理、及び、拡散フィルタ処理の両方を行うことで、出力信号を生成して出力してもよい。拡散フィルタ処理は、例えば、取得された音声信号に、当該音声信号が示す反射音（間接音）を拡散することで間接音のリアリティを向上させる処理である。また、拡散フィルタ処理は、取得された音声信号が示す音の拡散の聴感的な強弱を、それらしく模擬（すなわち受聴者が感じる音の拡散の聴感的な強弱を模擬）するフィルタが用いられる処理である。拡散フィルタ処理では、有限インパルスフィルタ及び／又は無限インパルスフィルタが用いられる。
- [0514] 以下、本実施の形態に係る音声信号処理装置（より具体的には、レンダリング部2300）によって行われる音声信号処理方法の動作例について説明する。

[0515] [レンダリング部の動作例]

図30は、本実施の形態に係る音声信号処理装置の動作例を示すフローチャートである。図30には、主に本実施の形態に係る音声信号処理装置が備えるレンダリング部2300で実行される処理が示されている。なお、ここでは、実施の形態1に係る図8との共通点の説明を省略又は簡略化する。

[0516] まず、解析部2301は、入力信号を解析する解析処理を行う（S101a）。より具体的には、解析部2301は、入力信号を解析して、音空間で発生し得る直接音及び反射音を検出する。このような直接音及び反射音を検出された場合に、解析部2301は、空間情報と音データとに基づいて、属性情報を含む音声信号、すなわち当該反射音を示す音声信号及び当該直接音を示す音声信号を作成する。解析部2301は、作成した音声信号を、解析部2301のメモリに記憶させる。

[0517] また、解析部2301は、入力信号を解析して、直接音及び反射音のそれぞれについて受聴位置に到来するまでの経路、到来するまでにかかる時間、及び、到来時の音量等に関する値と、直接音と反射音との時間差に関する値とを算出する。

[0518] まず、解析部2301は、作成された音声信号が示す直接音、及び、作成された音声信号が示す反射音のそれぞれの特性を算出する。具体的には、直接音及び反射音のそれぞれが受聴者（受聴位置）へ到来する際の到来時間及び到来時音量が算出される。なお、これらの到来時間及び到来時音量の算出方法は、実施の形態1で示された方法を用いることができる。

[0519] なお上記の通り、反射音を示す音声信号は反射音音量として反射音到来時音量（ I_r ）を示す情報を含み、直接音を示す音声信号は直接音音量として直接音到来時音量（ I_d ）を示す情報を含む。

[0520] そして、解析部2301は、直接音と反射音との時間差（ T ）（直接音と反射音との到来する時間差（ T ））を算出する。なお、時間差（ T ）の算出方法は、実施の形態1で示された方法を用いることができる。

[0521] なお、実施の形態1に係るステップS101とは異なり、ステップS10

1 aでは、音量比（L）が算出される必要はない。

[0522] 選択部2302（より具体的には、選択処理部2302d）は、反射音の選択（選択処理）を行う（S102a）。つまりは、選択部2302が、解析部2301で作成された反射音を示す音声信号に基づく出力信号を、再生部2303が再生するか否かを選択する。

[0523] まず、取得部2302aは、解析部2301によって作成されメモリに記憶された属性情報を含む音声信号を取得する。取得部2302aは、例えば、反射音（間接音）を示す音声信号、及び、当該反射音（間接音）に係る直接音を示す音声信号の少なくとも一方を取得し、ここでは、両方を取得する。また、取得部2302aは、解析部2301によって算出された直接音と反射音との時間差（T）に関する値を取得する。

[0524] 上記の通り、反射音と、当該反射音に係る直接音とは、同じ音源に由来する音である。

[0525] 第1算出部2302bは、取得部2302aによって取得された反射音を示す音声信号に基づいて、反射音音量に基づく第1音量を算出する。

[0526] 第2算出部2302cは、取得部2302aによって取得された直接音を示す音声信号に基づいて、直接音音量に基づく第2音量を算出する。

[0527] また、選択処理部2302dは、算出された第2音量と算出された第1音量との音量比、及び、直接音と反射音（間接音）との時間差（T）に基づいて、再生部2303が、取得された反射音（間接音）を示す音声信号に基づく出力信号を出力するか否かを選択する。

[0528] より具体的には、選択処理部2302dは、当該音量比が、直接音と反射音（間接音）との時間差（T）に応じて決定される第1閾値以上である場合に、再生部2303が、取得された反射音（間接音）を示す音声信号に基づく出力信号を出力することを選択する。

[0529] 当該音量比は、反射音到来時音量（ I_r ）に基づく第1音量を、直接音到来時音量（ I_d ）に基づく第2音量で割った値である。

[0530] また、当該時間差（T）は、取得された音声信号が示す反射音に係る直接

音と、取得された音声信号が示す反射音との時間差（ T ）であり、ステップ $S101a$ で、解析部 2301 によって算出された直接音と反射音との時間差（ T ）である。直接音と反射音との時間差（ T ）は、実施の形態 1 で説明された通り、例えば、直接音到来時間（到来時刻）と反射音到来時間（到来時刻）との時間差であるがこれに限られない。

[0531] 第 1 閾値は、反射音（間接音）に係る直接音と当該反射音（間接音）との時間差（ T ）に応じて決定される値であり、換言すると時間差（ T ）に依存する値であり、実施の形態 1 の閾値データに示される値である。閾値データは、例えば直接音と反射音との時間差（ T ）の値を横軸に有し、直接音と反射音との音量比を縦軸に有するグラフで、当該反射音が知覚されるか知覚されないかの閾値（第 1 閾値）として表される。

[0532] より具体的には、第 1 閾値を示す閾値データは、図 $11 \sim 13$ 等で示されるデータである。

[0533] また、ステップ $S102a$ で、取得部 $2302a$ は、間接音（反射音）に係る所定の周波数帯域幅ごとのゲイン特性と、聴覚の感度を示す周波数特性とを取得する。同様に、ステップ $S102a$ で、取得部 $2302a$ は、直接音に係る所定の周波数帯域幅ごとのゲイン特性を取得する。

[0534] 反射音（間接音）に係る所定の周波数帯域幅ごとのゲイン特性と、聴覚の感度を示す周波数特性と、直接音に係る所定の周波数帯域幅ごとのゲイン特性とは、例えば解析部 2301 のメモリに記憶されている。取得部 $2302a$ は、当該メモリから反射音（間接音）に係る所定の周波数帯域幅ごとのゲイン特性と、聴覚の感度を示す周波数特性と、直接音に係る所定の周波数帯域幅ごとのゲイン特性とを取得する。

[0535] なお、反射音（間接音）に係る所定の周波数帯域幅ごとのゲイン特性と、聴覚の感度を示す周波数特性と、直接音に係る所定の周波数帯域幅ごとのゲイン特性とは、通信回線等を介して取得部 $2302a$ によって取得されてもよい。

[0536] 選択部 2302 は、以上のように、選択処理を行う。

- [0537] そして、図 3 1 を用いて選択処理について、特に第 1 音量及び第 2 音量の算出についてより詳細に説明する。
- [0538] 図 3 1 は、本実施の形態に係る選択処理の動作例を示すフローチャートである。なお、ここでは、実施の形態 1 に係る図 1 4 との共通点の説明を省略又は簡略化する。
- [0539] まず、選択部 2 3 0 2 は、解析部 2 3 0 1 で検出された反射音を指定する (S 2 0 1)。すなわち、選択部 2 3 0 2 の取得部 2 3 0 2 a は、解析部 2 3 0 1 によって作成されメモリに記憶された属性情報を含む音声信号を指定し、指定された音声信号を取得する。取得部 2 3 0 2 a は、例えば、反射音を示す音声信号を指定し、指定された音声信号を取得する。このとき、取得部 2 3 0 2 a は、当該反射音（間接音）に係る直接音を示す音声信号も取得するとよい。
- [0540] そして、選択部 2 3 0 2 は、反射音音量に基づく第 1 音量を算出する (S 2 1 0)。
- [0541] より具体的には、選択部 2 3 0 2 の第 1 算出部 2 3 0 2 b は、反射音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第 1 補正特性と、反射音を示す音声信号とに基づいて、反射音音量に基づく第 1 音量を算出する。
- [0542] また、選択部 2 3 0 2 は、直接音音量に基づく第 2 音量を算出する (S 2 2 0)。
- [0543] より具体的には、選択部 2 3 0 2 の第 2 算出部 2 3 0 2 c は、直接音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第 2 補正特性に基づいて、直接音音量に基づく第 2 音量を算出する。ここでは、第 2 算出部 2 3 0 2 c は、上記第 2 補正特性と、直接音を示す音声信号とに基づいて、直接音音量に基づく第 2 音量を算出する。なお、反射音を示す音声信号が当該反射音に係る直接音の直接音音量を示す情報を含む場合には、第 2 算出部 2 3 0 2 c は、第 2 補正特性と、反射音を示す音声信号とに基づいて、直接音音量に基づく第 2 音量を算出してもよい。

- [0544] 以下、第1音量及び第2音量の算出について説明する。第1音量及び第2音量は、以下の式5及び式6により算出される。
- [0545] 第1音量 = $\text{GlobalGain_R} * \text{EqualiserGain_R}$ (式5)
- [0546] 第2音量 = $\text{GlobalGain_D} * \text{EqualiserGain_D}$ (式6)
- [0547] なお、GlobalGain_Rは、反射音の全周波数帯域の信号のGainであり、EqualiserGain_Rは、反射音に掛かるイコライザーの周波数帯域ごとのGainの代表値である。GlobalGain_Dは、直接音の全周波数帯域の信号のGainであり、EqualiserGain_Dは、直接音に掛かるイコライザーの周波数帯域ごとのGainの代表値である。
- [0548] 全周波数帯域の信号のGainは、実施の形態1で説明されたように、当該音の音源の音量と音源から受聴者までの距離とによって算出されるとよい。より具体的には、反射音の全周波数帯域の信号のGainは、間接音音量（反射音音量）に相当し、すなわち、反射音を示す音声信号が含む反射音到来時音量（ I_r ）に相当する。また、直接音の全周波数帯域の信号のGainは、直接音音量に相当し、すなわち、直接音を示す音声信号が含む直接音到来時音量（ I_d ）に相当する。
- [0549] なお、取得部2302aが反射音を示す音声信号のみを取得した場合には、直接音の全周波数帯域の信号のGainとして、反射音を示す音声信号が含む直接音の直接音音量が利用されるとよい。
- [0550] 上記の通り、EqualiserGain_Rが示す代表値は、反射音に掛かるイコライザーの周波数帯域ごとのGainの代表値である。反射音に掛かるイコライザーの周波数帯域ごとのGainが、上記の第1補正特性に相当する。すなわち、EqualiserGain_Rは、反射音に係る所定の周波数帯域幅ごとのゲイン特性が、聴覚の感度を示す周波数特性によって補正された第1補正特性の代表値である。
- [0551] 反射音（間接音）に係る所定の周波数帯域幅ごとのゲイン特性は、簡単のため、反射音（間接音）に係るゲイン特性と記載される場合がある。
- [0552] 反射音（間接音）に係るゲイン特性は、例えば、音波が物体（非発音オブジェクト）に衝突した際、その表面の形状、材質又は硬度に応じて、周波数

成分ごとに反射率が異なることに起因して生じる周波数特性の増減を実現するイコライザーの各バンドのGainであってもよい。図32は、本実施の形態に係る反射音に係る所定の周波数帯域幅ごとのゲイン特性を示す図である。反射音に係るゲイン特性は、図32で示されるように反射係数で示されるが、これに限られない。

[0553] 図32には、3種類の壁面のそれぞれに対応するゲイン特性の例が示されている。図32には、表面が固く凹凸の多い壁面、表面が固く凹凸の少ない壁面及び表面が柔軟な壁面のそれぞれのゲイン特性が示されている。反射音を生じさせる物体（非発音オブジェクト）に対応するゲイン特性が、用いられる。なお、図32に示されたゲイン特性は、一例に過ぎない。

[0554] また、ここでは、所定の周波数帯域幅として、1/3オクターブバンドが用いられているが、これに限るものではない。

[0555] 聴覚の感度を示す周波数特性は、受聴者の音量の感度を示す周波数特性であり、例えばA特性(A-weighting)が用いられてもよい。A特性は、人間の聴覚を考慮した周波数重み付け特性である。図33Aは、本実施の形態に係る聴覚の感度を示す周波数特性を表す表を示す図である。図33Bは、本実施の形態に係る聴覚の感度を示す周波数特性を示す図である。図33Bは、縦軸がデシベル(dB)表記で示された周波数特性を示す図である。なお、図33Bが示す聴覚の感度を示す周波数特性(A特性)は、1/3オクターブバンドごとの値が示されているが、これに限るものではない。

[0556] そして、EqualiserGain_Rは、以下のように算出される。ここでは、図34を用いて説明する。

[0557] 図34は、本実施の形態に係る反射音に係るゲイン特性と、聴覚の感度を示す周波数特性(A特性)と、第1補正特性とを示す図である。図34では、1/3オクターブバンドごとの値が示されている。

[0558] 上記聴覚の感度を示す周波数特性(A特性)を用いて、反射音に係るゲイン特性が補正される。具体的には、聴覚の感度を示す周波数特性の1/3オクターブバンドごとの値と、当該バンドに対応する反射音に係るゲイン特性

の値とを乗じることで、当該バンドに対応する第1補正特性の値が算出される。図34では、縦軸が対数軸（dB）で表現されているので、周波数帯域（バンド）ごとに、聴覚の感度を示す周波数特性の値と反射音に係るゲイン特性の値とを加算することで、第1補正特性の値が算出される。また、図34の縦軸がリニア軸（倍率）で表現されている場合は、乗算により算出されればよいことは言うまでもない。

- [0559] このように算出された第1補正特性の代表値を、EqualiserGain_Rとする。なお、第1補正特性の代表値は、第1補正特性の値の平均値、最大値又は最小値等であってもよく、第1補正特性の所定の周波数帯域内の平均値、最大値又は最小値等であってもよい。
- [0560] さらに、EqualiserGain_Dについて説明する。
- [0561] 上記の通り、EqualiserGain_Dが示す代表値は、直接音に掛かるイコライザーの周波数帯域ごとのGainの代表値である。直接音に掛かるイコライザーの周波数帯域ごとのGainが、上記の第2補正特性に相当する。すなわち、EqualiserGain_Dは、直接音に係る所定の周波数帯域幅ごとのゲイン特性が、聴覚の感度を示す周波数特性によって補正された第2補正特性の代表値である。
- [0562] 直接音に係る所定の周波数帯域幅ごとのゲイン特性は、簡単のため、直接音に係るゲイン特性と記載される場合がある。
- [0563] 直接音に係るゲイン特性は、例えば、音が伝搬する空間の状態による周波数成分のゲインを示すイコライザーの各バンドのGainであってもよい。より具体的には、直接音に係るゲイン特性は、音が伝搬する空間の温度若しくは湿度、又は、塵若しくは花粉等の微粒子の密度等の影響で高い周波数成分程ゲインが減衰する、というような傾向を示す周波数特性を実現するイコライザーの各バンドのGainであってもよい。図35は、本実施の形態に係る直接音に係る所定の周波数帯域幅ごとのゲイン特性を示す図である。直接音に係るゲイン特性は、図35に示されたものに限られない。
- [0564] 図35が示すように、直接音に係るゲイン特性は、所定の周波数帯域幅ごと、つまりは、1/3オクターブバンドごとの値が示されているが、これに

限るものではない。

- [0565] なお、EqualiserGain_Dの算出に用いられる聴覚の感度を示す周波数特性も、図33Bで示されるA特性であるとよい。
- [0566] そして、EqualiserGain_Dは、以下のように算出される。ここでは、図36を用いて説明する。
- [0567] 図36は、本実施の形態に係る直接音に係るゲイン特性と、聴覚の感度を示す周波数特性（A特性）と、第2補正特性とを示す図である。図36では、1/3オクターブバンドごとの値が示されている。
- [0568] 上記聴覚の感度を示す周波数特性（A特性）を用いて、直接音に係るゲイン特性が補正される。具体的には、聴覚の感度を示す周波数特性の1/3オクターブバンドごとの値と、当該バンドに対応する直接音に係るゲイン特性の値とを乗じることで、当該バンドに対応する第2補正特性の値が算出される。図36では、縦軸が対数軸（dB）で表現されているので、周波数帯域（バンド）ごとに、聴覚の感度を示す周波数特性の値と直接音に係るゲイン特性の値とを加算することで、第2補正特性の値が算出される。また、図36の縦軸がリニア軸（倍率）で表されている場合は、乗算によって算出されればよいことは言うまでもない。
- [0569] このように算出された第2補正特性の代表値を、EqualiserGain_Dとする。なお、第2補正特性の代表値は、第2補正特性の値の平均値、最大値又は最小値等であってもよく、第2補正特性の所定の周波数帯域内の平均値、最大値又は最小値等であってもよい。
- [0570] このように算出されたEqualiserGain_D、及び、EqualiserGain_Rと、上記式5及び式6により、第1音量及び第2音量が算出される。
- [0571] すなわち第1音量は、反射音を示す音声信号が含む反射音到来時音量（ I_r ）（反射音音量）に、算出されたEqualiserGain_Rが乗算されることで算出される。同様に、第2音量は、直接音を示す音声信号が含む直接音到来時音量（ I_d ）（直接音音量）に、算出されたEqualiserGain_Dが乗算されることで算出される。

[0572] なお、本実施の形態においては、聴覚の感度を示す周波数特性、すなわち、受聴者の音量の感度を示す周波数特性としてA特性が用いられたが、これに限られるものではない。例えば、受聴者の音量の感度を示す周波数特性として、等ラウドネス曲線の逆特性、又は、音源位置の最小弁別角度の周波数特性の逆特性等が用いられてもよい。

[0573] 図37は、本実施の形態に係る受聴者の音量の感度を示す周波数特性の他の第1例である等ラウドネス曲線の逆特性を示す図である。A特性が示された図33Bにおいては、A特性は、上に凸の特性を示す。等ラウドネス曲線の逆特性も同様に、上に凸の特性を示す。また、等ラウドネス曲線の逆特性が、上記の補正に用いられる方法は、上記と同じでよい。

[0574] 図38は、本実施の形態に係る受聴者の音量の感度を示す周波数特性の他の第2例である音源位置の最小弁別角度の周波数特性の逆特性を示す図である。また、音源位置の最小弁別角度の周波数特性の逆特性が、上記の補正に用いられる方法は、上記と同じでよい。

[0575] このように、本実施の形態においては、聴覚の感度を示す周波数特性として、受聴者の音量の感度を示す周波数特性を利用することができる。さらに、音量の感度を示す周波数特性として、等ラウドネス曲線の逆数に応じた周波数特性又は音源位置の最小弁別角度の周波数特性の逆特性を利用することができる。

[0576] 再度図31を用いて説明する。

[0577] 次に、選択処理部2302dは、算出された第2音量と算出された第1音量との音量比を算出する(S202a)。

[0578] そして、選択処理部2302dは、直接音と反射音との時間差(T)を検出する(S203)。時間差(T)は、ステップS101aで解析部2301によって既に算出されている。例えば、時間差(T)を示すデータが解析部2301のメモリに記憶されており、選択部2302が当該データを取得することで、時間差(T)を検出する。

[0579] また、選択処理部2302dは、閾値データを用いて、時間差(T)に対

応する第1閾値を特定する(S204)。そして、選択処理部2302dは、算出した音量比が第1閾値以上であるか否かを判定する(S205a)。

[0580] 音量比が第1閾値以上である場合(S205aでYes)、選択処理部2302dは、当該反射音を生成対象の反射音として選択する(S206)。すなわち、この場合に、選択処理部2302dは、解析部2301で作成された反射音を示す音声信号に基づく出力信号を、再生部2303が再生することを選択する。

[0581] 音量比が第1閾値よりも小さい場合(S205aでNo)、選択処理部2302dは、当該反射音を生成対象の反射音として選択しない(S207)。すなわち、この場合に、選択処理部2302dは、解析部2301で作成された反射音を示す音声信号に基づく出力信号を、再生部2303が再生しないことを選択し、当該反射音を生成対象外の反射音として決定する。

[0582] その後、選択処理部2302dは、未指定の反射音があるか否かを判定する(S208)。すなわち、解析部2301で作成された複数の音声信号のうち、選択処理が行われていない音声信号があるか否かを判定する。未指定の反射音があれば(S208でYes)、選択処理部2302dは、上述の処理(S201～S207、S210及びS220)を繰り返す。未指定の反射音がなければ(S208でNo)、選択処理部2302dは、処理を終了する。

[0583] このように、選択処理が行われ、ステップS206で、選択処理部2302dが、反射音を示す音声信号に基づく出力信号を再生部2303が再生することを選択すると、選択部2302は、当該音声信号を再生部2303に出力する。

[0584] 再度図30を用いて説明する。

[0585] 再生部2303は、選択部2302から出力された音声信号を取得して、当該音声信号に基づく出力信号を出力する(S103a)。ここでは、再生部2303は、取得部2302aによって取得された直接音を示す音声信号と、生成された音声信号(反射音を示す音声信号)とを合成して出力する。

- [0586] このように、ステップS 2 0 5 aで音量比が第1 閾値以上である場合に、すなわち、選択処理部2 3 0 2 dが、反射音を示す音声信号に基づく出力信号を再生部2 3 0 3が再生することを選択した場合に、再生部2 3 0 3は、当該音声信号に基づく出力信号を出力する。
- [0587] なお、ステップS 2 0 5 aで音量比が第1 閾値未満である場合に、すなわち、選択処理部2 3 0 2 dが、反射音を示す音声信号に基づく出力信号を再生部2 3 0 3が再生することを選択しなかった場合には、再生部2 3 0 3は、当該音声信号に基づく出力信号を出力しない。このような場合には、再生部2 3 0 3は、当該出力信号を出力しないため、演算量及び演算負荷が削減される。
- [0588] 以上のように、本実施の形態に係る音声信号処理方法は、音声信号処理装置（レンダリング部2 3 0 0）が実行する音声信号処理方法であって、取得ステップと、第1 算出ステップと、第2 算出ステップと、選択処理ステップと、再生ステップとを含む。
- [0589] 取得ステップは、音声信号であって、音声信号の属性を特定する属性情報を含み属性が間接音（例えば反射音）を示す情報を含む音声信号を取得する。第1 算出ステップは、間接音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第1 補正特性と、取得された音声信号とに基づいて、間接音が受聴者が居る位置である受聴位置に到来するときの間接音音量（反射音音量）に基づく第1 音量を算出する。第2 算出ステップは、間接音に係る直接音が受聴位置に到来するときの直接音音量に基づく第2 音量を算出する。選択処理ステップは、算出された第2 音量と算出された第1 音量との音量比、及び、直接音と間接音との到来する時間差（T）に基づいて、取得された音声信号に基づく出力信号を出力するか否かを選択する。再生ステップは、出力信号を出力することが選択された場合に、出力信号を出力する。
- [0590] これにより、直接音音量に基づく第2 音量と間接音音量に基づく第1 音量との音量比、及び、上記時間差（T）に基づいて、間接音を示す音声信号に

基づく出力信号を出力するか否かが選択される。すなわち、当該音声信号に基づく出力信号が出力されるか否かが適切に選択される。出力信号が出力されない場合には、演算量及び演算負荷が削減される。つまりは、演算量及び演算負荷を適切に削減することができる音声信号処理方法を実現することができる。

[0591] そしてさらに、聴覚の感度を示す周波数特性が考慮された上で第1音量が算出され、算出された第1音量が用いられた音量比、及び、上記時間差（T）に基づいて、出力信号を出力するか否かが選択される。すなわち、聴覚の感度が考慮された上で、演算量及び演算負荷を適切に削減することができる音声信号処理方法を実現することができる。

[0592] また、本実施の形態に係る音声信号処理方法においては、第2算出ステップでは、直接音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第2補正特性に基づいて、第2音量を算出する。

[0593] これにより、聴覚の感度を示す周波数特性が考慮された上で第2音量が算出され、算出された第2音量が用いられた音量比、及び、上記時間差に基づいて、出力信号を出力するか否かが選択される。すなわち、聴覚の感度がより考慮された上で、演算量及び演算負荷を適切に削減することができる音声信号処理方法を実現することができる。

[0594] また、本実施の形態に係る音声信号処理方法においては、聴覚の感度を示す周波数特性は受聴者の音量の感度を示す周波数特性であり、音量の感度を示す周波数特性はA特性である。

[0595] これにより、聴覚の感度を示す周波数特性として受聴者の音量の感度を示す周波数特性を利用することができ、かつ、音量の感度を示す周波数特性としてA特性を利用することができる音声信号処理方法を実現することができる。

[0596] （実施の形態3）

以下、実施の形態3について説明する。以下では、実施の形態1及び2と

の相違点を中心に説明し、共通点の説明を省略又は簡略化する。

[0597] 実施の形態2では、第2音量と第1音量との音量比が、第1閾値以上である場合に出力信号が出力されることが選択された。実施の形態3では、音量比(L)が第1閾値以上である場合に出力信号が出力されることが選択される点は実施の形態2と同じであるが、第1閾値が決定される方法が実施の形態2とは異なる。

[0598] [レンダリング部の構成]

まず、本実施の形態に係るレンダリング部3300の構成について説明する。図39は、本実施の形態に係るレンダリング部3300の構成例を示すブロック図である。

[0599] レンダリング部3300は、解析部3301と、選択部3302と、再生部2303とを備える。

[0600] 解析部3301は、受聴位置における直接音と反射音との音量比(L)に関する値等を算出する点が、実施の形態2に係る解析部2301とは異なる。

[0601] すなわち、解析部3301は、音空間で発生し得る直接音及び反射音を検出し、このような直接音及び反射音が検出された場合に、空間情報と音データとに基づいて、当該反射音を示す音声信号、及び、当該直接音を示す音声信号を作成する。

[0602] 本実施の形態においては、直接音と、当該直接音が反射物（例えば、非発音オブジェクト）によって反射された音である反射音とが、解析部3301によって検出される。当該直接音とは、当該反射音に係る音でもある。解析部3301は、反射音を示す音声信号と、当該反射音に係る直接音を示す音声信号とを作成する。

[0603] また、解析部3301は、実施の形態1及び2と同じく、直接音及び反射音のそれぞれについて、受聴位置に到来するまでの経路、到来するまでにかかる時間、及び、到来時の音量等に関する値を算出しているとよい。そして、解析部3301は、直接音と反射音との関係を示す情報であって、例えば

、直接音と反射音との時間差（ T ）（直接音と反射音との到来する時間差（ T ））に関する値、及び、受聴位置における直接音と反射音との音量比（ L ）に関する値等を算出する。解析部 3301 がこれらの情報を算出する方法は、実施の形態 1 で説明されたとおりである。

[0604] なお、実施の形態 2 で説明された通り、反射音音量（間接音音量）は反射音到来時音量（ I_r ）であり、直接音音量は直接音到来時音量（ I_d ）である。つまり、解析部 3301 によって算出される受聴位置における直接音と反射音との音量比（ L ）とは、直接音音量と反射音音量との音量比（ L ）である。

[0605] なお、解析部 3301 は、実施の形態 1 に係る解析部 1301 が行う全ての処理又は一部の処理を行えるとよい。

[0606] 選択部 3302 は、取得部 3302a と選択処理部 3302d とを有する。選択部 3302 は、実施の形態 1 に係る選択部 1302 が行う全ての処理又は一部の処理を行えるとよい。

[0607] 取得部 3302a は、解析部 3301 によって作成された、反射音を示す音声信号と、当該反射音に係る直接音を示す音声信号とを取得する。また、取得部 3302a は、解析部 3301 によって算出された直接音と反射音との時間差（ T ）に関する値、及び、受聴位置における直接音と反射音との音量比（ L ）に関する値等を取得する。

[0608] 選択処理部 3302d は、取得部 3302a によって取得された音量比（ L ）及び時間差（ T ）と、反射係数特徴量とに基づいて、再生部 2303 が、取得された反射音を示す音声信号に基づく出力信号を出力するか否かを選択する。

[0609] 例えば、選択処理部 3302d は、実施の形態 1 における図 8 のステップ S102 及び実施の形態 2 における図 30 のステップ S102a で説明されたような選択処理を行う。選択処理部 3302d は、解析部 3301 で算出された音量比（ L ）が、直接音と反射音との時間差（ T ）に応じて決定される第 1 閾値以上である場合に、再生部 2303 が、取得された反射音を示す

音声信号に基づく出力信号を出力することを選択する。

[0610] また、当該時間差（ T ）は、取得された音声信号が示す反射音に係る直接音と、取得された前声信号が示す反射音との時間差（ T ）であり、解析部 3301 によって算出された直接音と反射音との時間差（ T ）である。直接音と反射音との時間差（ T ）は、実施の形態 1 で説明された通り、例えば、直接音到来時間（到来時刻）と反射音到来時間（到来時刻）との時間差であるがこれに限られない。

[0611] 第 1 閾値は、反射音に係る直接音と当該反射音との時間差（ T ）に応じて決定される値であり、換言すると時間差（ T ）に依存する値であり、実施の形態 1 の閾値データに示される値である。

[0612] そして、反射係数特徴量は、反射物（例えば、非発音オブジェクト）の反射係数に基づいて決定される特徴量である。反射係数特徴量は、例えば、図 32 で例示されるような反射係数の周波数特性の平坦さの程度を示す特徴量、より具体的には、反射係数で示されるゲイン特性の周波数スペクトルの平坦さの程度を示す特徴量である。図 32 が示す表面が固く凹凸の多い壁面及び表面が固く凹凸の少ない壁面では、反射係数の周波数特性の平坦さの程度が高く、反射係数特徴量が大きい。図 32 が示す表面が柔軟な壁面では、反射係数の周波数特性の平坦さの程度が低く、反射係数特徴量が小さい。

[0613] 反射係数特徴量とは、また例えば、図 32 で例示されるような反射係数で示される減衰量のバラツキの程度を示す特徴量であるともいえる。

[0614] 反射係数特徴量は、通信回線等を介して取得部 3302a によって取得されてもよい。また、反射係数特徴量は解析部 3301 が有するメモリに記憶されていてもよく、取得部 3302a は、当該メモリに記憶されている反射係数特徴量を取得してもよい。

[0615] 選択処理部 3302d が選択処理を行う際に、反射係数特徴量は、以下のように利用される。

[0616] 本実施の形態においては、反射係数特徴量の大きさに応じて、第 1 閾値が変更される。例えば、反射係数特徴量が大きい場合には、第 1 閾値が高くな

るように変更され、反射係数特徴量が小さい場合には、第1閾値が低くなるように変更される。

[0617] 図40は、本実施の形態に係る反射係数特徴量が第1閾値へ与える影響を示す図である。図40には、図28と同じくエコー検知限閾値（第1閾値）の一例が示されている。反射係数特徴量が大きい場合には、第1閾値が高くなり、例えば図40が示す横軸の全範囲にわたって第1閾値が高くなるようにシフトする。反射係数特徴量が小さい場合には、第1閾値が低くなり、例えば図40が示す横軸の全範囲にわたって第1閾値が低くなるようにシフトする。

[0618] ここで、先行音効果と反射係数特徴量とについて検証する。

[0619] 上記の通り実施の形態1等に記載された技術は、先行音効果が利用されている。この先行音効果は、先行音（例えば直接音）の周波数スペクトルと、後続音（例えば反射音）の周波数スペクトルとが近似している場合に生じる、とされている。つまり、先行音の周波数スペクトルと、後続音の周波数スペクトルとが近似していない場合には、先行音効果は生じないと考えられる。

[0620] ところで、反射音は直接音が反射物（例えば、非発音オブジェクト）によって反射された音であるため、後続音（反射音）の周波数スペクトルは、反射物の反射係数、より具体的には、反射係数特徴量に依存する。従って、直接音の周波数スペクトルと反射音の周波数スペクトルとが近似しているか否かは、反射係数特徴量に応じて変化する。

[0621] 例えば、反射物の反射係数が図32の表面が固く凹凸の多い壁面及び表面が固く凹凸の少ない壁面の反射係数である場合には、反射係数特徴量が大きく、先行音の周波数スペクトルと後続音の周波数スペクトルとが近似するため、先行音効果が生じ易い。この場合には、第1閾値が高くなるように変更され、すなわち、音声信号に基づく出力信号が出力され難いように選択される。これは、先行音効果が生じ易い場合ほど、反射音の聴感的な価値が低下するためである。

[0622] また例えば、反射物の反射係数が図32の表面が柔軟な壁面の反射係数である場合には、反射係数特徴量が小さく、先行音の周波数スペクトルと後続音の周波数スペクトルとが近似しないため、先行音効果が生じ難い。この場合には、第1閾値が低くなるように変更され、すなわち、音声信号に基づく出力信号が出力され易いように選択される。

[0623] このように、反射係数特徴量に基づいて出力信号が出力されるか否かが選択されることは、先行音効果を考慮したうえで、出力信号が出力されるか否かが選択されることに等しい。先行音効果は聴覚の感度の特性の一例であることから、本実施の形態に係る音声信号処理方法は、聴覚の感度が考慮された上で、演算量及び演算負荷を適切に削減することができる。

[0624] 以上説明されたように、選択処理部3302dは、取得部3302aによって取得された音量比(L)及び時間差(T)と、反射係数特徴量とに基づいて、再生部2303が、取得された反射音を示す音声信号に基づく出力信号を出力するか否かを選択する。

[0625] 選択処理部3302dが再生部2303が当該出力信号を出力することを選択した場合に、選択部3302は、取得された音声信号を再生部2303に出力する。

[0626] 再生部2303は、選択部3302から出力された音声信号を取得し、取得された音声信号に基づく出力信号を出力する。

[0627] 以下、本実施の形態に係る音声信号処理装置（より具体的には、レンダリング部3300）によって行われる音声信号処理方法の動作例について説明する。

[0628] [レンダリング部の動作例]

図41は、本実施の形態に係る音声信号処理装置の動作例を示すフローチャートである。図41には、主に本実施の形態に係る音声信号処理装置が備えるレンダリング部3300で実行される処理が示されている。

[0629] まず、解析部3301は、入力信号を解析する解析処理を行う（S101b）。より具体的には、解析部3301は、入力信号を解析して、音空間で

発生し得る直接音及び反射音を検出する。このような直接音及び反射音が検出された場合に、解析部3301は、空間情報と音データとに基づいて、属性情報を含む音声信号、すなわち当該反射音を示す音声信号及び当該直接音を示す音声信号を作成する。解析部3301は、作成した音声信号を、解析部3301のメモリに記憶させる。

[0630] また、解析部3301は、入力信号を解析して、直接音及び反射音のそれぞれについて受聴位置に到来するまでの経路、到来するまでにかかる時間、及び、到来時の音量等に関する値と、直接音と反射音との時間差（ T ）に関する値と、受聴位置における直接音と反射音との音量比（ L ）に関する値とを算出する。

[0631] すなわち、解析部3301は、直接音到来時音量（ I_d ）と反射音到来時音量（ I_r ）との比である音量比（ L ）、及び、直接音と反射音との時間差（ T ）を算出する。音量比（ L ）は、受聴位置における直接音と反射音との音量比（ L ）である。なお、これらの音量比（ L ）及び時間差（ T ）の算出方法は、実施の形態1で示された方法を用いることができる。

[0632] なお、実施の形態2に係るステップS101aとは異なり、本実施の形態に係るステップS101bでは、音量比（ L ）が算出される。

[0633] 選択部3302（より具体的には、選択処理部3302d）は、反射音の選択（選択処理）を行う（S102b）。つまりは、選択部3302が、解析部3301で作成された反射音を示す音声信号に基づく出力信号を、再生部2303が再生するか否かを選択する。

[0634] まず、取得部3302aは、解析部3301によって作成されメモリに記憶された属性情報を含む音声信号を取得する。取得部3302aは、例えば、反射音を示す音声信号と、当該反射音に係る直接音を示す音声信号とを取得する。また、取得部3302aは、解析部3301によって算出された直接音と反射音との時間差（ T ）に関する値、及び、受聴位置における直接音と反射音との音量比（ L ）に関する値等を取得する。また、取得部3302aは、解析部3301が有するメモリに記憶されている反射係数特徴量を取

得する。

[0635] 選択処理部3302dは、取得された音量比(L)が、第1閾値以上である場合に、再生部2303が、取得された反射音を示す音声信号に基づく出力信号を出力することを選択する。

[0636] 第1閾値は、直接音と反射音との時間差(T)に応じて決定され、かつ、取得された反射係数特徴量に応じて決定される値である。すなわち、選択処理部3302dは、直接音と反射音との時間差(T)及び反射係数特徴量に応じて、第1閾値を決定する。

[0637] 以上のように、選択処理部3302dは、反射音音量と直接音音量との音量比(L)、直接音と反射音との時間差(T)、及び、反射係数特徴量に基づいて、再生部2303が取得された音声信号に基づく出力信号を出力するか否かを選択する。

[0638] このように、選択処理が行われ、選択処理部3302dが、反射音を示す音声信号に基づく出力信号を再生部2303が再生することを選択すると、選択部3302は、当該音声信号を再生部2303に出力する。

[0639] そして、再生部2303は、選択部3302から出力された音声信号を取得して、当該音声信号に基づく出力信号を出力する(S103b)。ここでは、例えば、再生部2303は、取得部3302aによって取得された直接音を示す音声信号と、生成された音声信号(反射音を示す音声信号)とを合成して出力する。

[0640] 以上のように、本実施の形態に音声信号処理方法は、音声信号処理装置が実行する音声信号処理方法であって、取得ステップと、選択処理ステップと、再生ステップとを含む。

[0641] 取得ステップは、音声信号であって、音声信号の属性を特定する属性情報を含み属性が、直接音が反射物によって反射された音である反射音を示す情報を含む音声信号を取得する。すなわちこの属性は、反射音を示す情報を含む。選択処理ステップは、取得された音声信号が示す反射音が受聴者が居る位置である受聴位置に到来するときの反射音音量と直接音が受聴位置に到来

するときの直接音音量との音量比、直接音と反射音との到来する時間差、及び、反射物の反射係数に基づいて決定される反射係数特徴量に基づいて、取得された音声信号に基づく出力信号を出力するか否かを選択する。再生ステップは、出力信号を出力することが選択された場合に、出力信号を出力する。

[0642] これにより、上記音量比、及び、上記時間差に基づいて、反射音を示す音声信号に基づく出力信号を出力するか否かが選択される。すなわち、当該音声信号に基づく出力信号が出力されるか否かが適切に選択される。出力信号が出力されない場合には、演算量及び演算負荷が削減される。つまりは、演算量及び演算負荷を適切に削減することができる音声信号処理方法を実現することができる。

[0643] ここで、先行音効果に着目する。先行音効果は、先行音（例えば直接音）の周波数スペクトルと、後続音（例えば反射音）の周波数スペクトルとが近似している場合に生じる、と言われている。反射音の周波数スペクトルは反射物の反射係数に応じて変化するため、直接音の周波数スペクトルと反射音の周波数スペクトルとが近似しているか否かは、反射係数特徴量に応じて変化する。

[0644] このように、反射係数特徴量に基づいて出力信号が出力されるか否かが選択されることは、先行音効果を考慮したうえで、出力信号が出力されるか否かが選択されることに等しい。先行音効果は聴覚の感度の特性の一例であることから、聴覚の感度が考慮された上で、演算量及び演算負荷を適切に削減することができる音声信号処理方法を実現することができる。

[0645] また、本実施の形態においては、反射係数特徴量は、反射係数の周波数特性の平坦さの程度を示す特徴量である。

[0646] これにより、反射係数特徴量として反射係数の周波数特性の平坦さの程度を示す特徴量を利用することができる音声信号処理方法を実現することができる。

[0647] （実施の形態4）

以下、実施の形態４について説明する。以下では、実施の形態２との相違点を中心に説明し、共通点の説明を省略又は簡略化する。

[0648] 実施の形態２では、第２音量と第１音量との音量比が、第１閾値以上である場合に出力信号が出力されることが選択された。実施の形態４では、音が受聴者に到来するときの音量に基づく所定音量に応じて、出力信号が出力されるか否かが選択される点の実施の形態２とは異なる。

[0649] なお、実施の形態２等では、間接音（反射音）と直接音とが識別されていたが、本実施の形態においては、間接音（反射音）と直接音とを識別する必要がない場合には、間接音（反射音）及び直接音のそれぞれを、単に、音、と記載する場合がある。

[0650] [レンダリング部の構成]

まず、本実施の形態に係るレンダリング部４３００の構成について説明する。図４２は、本実施の形態に係るレンダリング部４３００の構成例を示すブロック図である。

[0651] レンダリング部４３００は、解析部４３０１と、選択部４３０２と、再生部２３０３とを備える。

[0652] 解析部４３０１は、音空間で発生し得る音を検出し、このような音が検出された場合に、空間情報と音データとに基づいて、当該音を示す音声信号を作成する。

[0653] より具体的には、解析部４３０１は、音空間で発生し得る直接音及び反射音を検出し、このような直接音及び反射音が検出された場合に、空間情報と音データとに基づいて、当該反射音を示す音声信号、及び、当該直接音を示す音声信号を作成する。つまり、音を示す音声信号は、反射音を示す音声信号、又は、直接音を示す音声信号である。

[0654] 本実施の形態においては、直接音と、当該直接音が反射物（例えば、非発音オブジェクト）によって反射された音である反射音とが、解析部４３０１によって検出される。従って、解析部４３０１は、当該反射音を示す音声信号と、当該反射音に係る直接音を示す音声信号とを作成する。

- [0655] また、解析部4301は、上記音について、受聴位置に到来するまでの経路、到来するまでにかかる時間、及び、到来時の音量等に関する値を算出しているとよい。すなわち、実施の形態1及び2と同じく、直接音及び反射音のそれぞれについて、受聴位置に到来するまでの経路、到来するまでにかかる時間、及び、到来時の音量等に関する値を算出しているとよい。
- [0656] なお、本実施の形態においては、解析部4301は、直接音と反射音との関係を示す情報であって、例えば、直接音と反射音との時間差（ T ）（直接音と反射音との到来する時間差（ T ））に関する値、及び、受聴位置における直接音と反射音との音量比（ L ）に関する値等を算出しなくてもよい。
- [0657] 本実施の形態に係る音声信号は、当該音声信号が示す音が受聴位置に到来するときの音量を示す情報を含む。すなわち、実施の形態2と同じく、属性が反射音を示す情報である音声信号は、当該反射音が受聴位置に到来するときの反射音音量として、反射音到来時音量（ I_r ）を示す情報を含む。属性が直接音を示す情報である音声信号は、当該直接音が受聴位置に到来するときの直接音音量として、直接音到来時音量（ I_d ）を示す情報を含む。
- [0658] なお、解析部4301は、実施の形態1に係る解析部1301が行う全ての処理又は一部の処理を行えるとよい。
- [0659] 選択部4302は、取得部4302aと、第3算出部4302eと、選択処理部4302dとを有する。
- [0660] 取得部4302aは、解析部4301によって作成された、音を示す音声信号を取得する。すなわち、取得部4302aは、反射音を示す音声信号と、当該反射音に係る直接音を示す音声信号とを取得する。
- [0661] なお、上記の通り、音を示す音声信号は、当該音声信号が示す音が受聴位置に到来するときの音量を示す情報を含む。すなわち、反射音を示す音声信号は反射音音量（反射音到来時音量（ I_r ））を示す情報を含み、直接音を示す音声信号は直接音音量（直接音到来時音量（ I_d ））を示す情報を含む。
- [0662] 第3算出部4302eは、音を示す音声信号に基づいて、当該音声信号が

示す音が受聴位置に到来するときの音量に基づく所定音量を算出する。より具体的には、第3算出部4302eは、取得部4302aによって取得された音声信号が示す音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された所定補正特性と、取得された音声信号とに基づいて、上記所定音量を算出する。

- [0663] 本実施の形態においては、音を示す音声信号は、反射音を示す音声信号、又は、直接音を示す音声信号である。
- [0664] 音を示す音声信号が反射音を示す音声信号である場合には、第3算出部4302eでは、以下の処理が行われ、所定音量が算出される。
- [0665] この場合、音声信号が示す音に係る所定の周波数帯域幅ごとのゲイン特性は、実施の形態2で説明された反射音に係る所定の周波数帯域幅ごとのゲイン特性と同じである。同様に、聴覚の感度を示す周波数特性は、例えばA特性である。そして、所定補正特性は、実施の形態2で説明された第1補正特性である。そして、第3算出部4302eは、所定補正特性（第1補正特性）と反射音を示す音声信号とに基づいて、実施の形態2と同様の方法で、所定音量として第1音量を算出する。つまり、この場合には、所定音量は、第1音量である。
- [0666] また、音を示す音声信号が直接音を示す音声信号である場合には、第3算出部4302eでは、以下の処理が行われ、所定音量が算出される。
- [0667] この場合、音声信号が示す音に係る所定の周波数帯域幅ごとのゲイン特性は、実施の形態2で説明された直接音に係る所定の周波数帯域幅ごとのゲイン特性と同じである。同様に、聴覚の感度を示す周波数特性は、例えばA特性である。そして、所定補正特性は、実施の形態2で説明された第2補正特性である。そして、第3算出部4302eは、所定補正特性（第2補正特性）と直接音を示す音声信号とに基づいて、実施の形態2と同様の方法で、所定音量として第2音量を算出する。つまり、この場合には、所定音量は、第2音量である。
- [0668] 選択処理部4302dは、第3算出部4302eによって算出された所定

音量に基づいて、再生部 2303 が、取得された音声信号に基づく出力信号を出力するか否かを選択する。

[0669] 選択処理部 4302d は、算出された所定音量が第 2 閾値以上である場合に、再生部 2303 が、取得された音声信号に基づく出力信号を出力することを選択する。

[0670] 第 2 閾値は、上記第 1 閾値とは異なり、反射音に係る直接音と反射音との時間差 (T) に依存しない値であり、一定の値である。第 2 閾値は、取得された音量に係る値であり、換言すると振幅値に係る値である。さらに言えば、第 2 閾値は、受聴者によって知覚可能か否かの境目の音量を示し、当該閾値よりも音量が小さい音を再現しない音として決定するための閾値である。

[0671] 図 43 は、本実施の形態に係る第 2 閾値を示す閾値データを示すグラフである。例えば、第 2 閾値は -70 dB であり、図 43 における直接音を示す音声信号の所定音量 (第 2 音量) が第 2 閾値以上であるため、当該音声信号に基づく出力信号が出力されることが選択される。また、図 43 における反射音を示す音声信号の所定音量 (第 1 音量) が第 2 閾値未満であるため、当該音声信号に基づく出力信号が出力されないことが選択される。

[0672] なお、選択部 4302 は、実施の形態 1 に係る選択部 1302 が行う全ての処理又は一部の処理を行えるとよい。

[0673] 以下、本実施の形態に係る音声信号処理装置 (より具体的には、レンダリング部 4300) によって行われる音声信号処理方法の動作例について説明する。

[0674] [レンダリング部の動作例]

図 44 は、本実施の形態に係る音声信号処理装置の動作例を示すフローチャートである。図 44 には、主に本実施の形態に係る音声信号処理装置が備えるレンダリング部 4300 で実行される処理が示されている。

[0675] まず、解析部 4301 は、入力信号を解析する解析処理を行う (S101c)。より具体的には、解析部 4301 は、入力信号を解析して、音空間で発生し得る音 (直接音及び反射音) を検出する。このような音が検出された

場合に、解析部4301は、空間情報と音データとに基づいて、音を示す音声信号、より具体的には当該反射音を示す音声信号及び当該直接音を示す音声信号を作成する。解析部4301は、作成した音声信号を、解析部4301のメモリに記憶させる。

[0676] また、解析部4301は、入力信号を解析して、音（直接音及び反射音）について受聴位置に到来するまでの経路、到来するまでにかかる時間、及び、到来時の音量等に関する値を算出する。

[0677] なお、本実施の形態に係るステップS101cにおいては、直接音と反射音との時間差（T）に関する値、及び、受聴位置における直接音と反射音との音量比（L）に関する値等を算出しなくてもよい。

[0678] 選択部4302（より具体的には、選択処理部4302d）は、音の選択（選択処理）を行う（S102c）。つまりは、選択部4302が、解析部4301で作成された音を示す音声信号に基づく出力信号を、再生部2303が再生するか否かを選択する。

[0679] まず、取得部4302aは、解析部4301によって作成されメモリに記憶された属性情報を含む音声信号（反射音を示す音声信号、及び、直接音を示す音声信号）を取得する。

[0680] そして、第3算出部4302eは、取得部4302aによって取得された音声信号が示す音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された所定補正特性と、取得された音声信号とに基づいて、所定音量を算出する。

[0681] 音を示す音声信号が反射音を示す音声信号である場合には、所定音量は第1音量である。また、音を示す音声信号が直接音を示す音声信号である場合には、所定音量は第2音量である。

[0682] 次に、選択処理部4302dは、第3算出部4302eによって算出された所定音量が第2閾値以上である場合に、再生部2303が、取得された音声信号に基づく出力信号を出力することを選択する。

[0683] 選択部4302は、以上のように、選択処理を行う。

- [0684] そして、図45を用いて選択処理についてより詳細に説明する。
- [0685] 図45は、本実施の形態に係る選択処理の動作例を示すフローチャートである。
- [0686] まず、選択部4302は、解析部4301で検出された音を指定する（S201c）。すなわち、選択部4302の取得部4302aは、解析部4301によって作成されメモリに記憶された音声信号を指定し、指定された音声信号を取得する。
- [0687] そして、第3算出部4302eは、所定音量（第1音量又は第2音量）を算出する（S230）。
- [0688] 次に、選択処理部4302dは、算出された所定音量が第2閾値以上であるか否かを判定する（S205c）。
- [0689] 所定音量が第2閾値以上である場合（S205cでYes）、選択処理部4302dは、取得された音声信号が示す音を生成対象の音として選択する（S206c）。すなわち、この場合に、選択処理部4302dは、解析部4301で作成された音を示す音声信号に基づく出力信号を、再生部2303が再生することを選択する。
- [0690] 所定音量が第2閾値よりも小さい場合（S205cでNo）、選択処理部4302dは、取得された音声信号が示す音を生成対象の音として選択しない（S207c）。すなわち、この場合に、選択処理部4302dは、解析部4301で作成された音を示す音声信号に基づく出力信号を、再生部2303が再生しないことを選択し、当該音を生成対象外の音として決定する。
- [0691] その後、選択処理部4302dは、未指定の音があるか否かを判定する（S208c）。すなわち、解析部4301で作成された複数の音声信号のうち、選択処理が行われていない音声信号があるか否かを判定する。未指定の音があれば（S208cでYes）、選択処理部4302dは、上述の処理（S201c～S207c及びS230）を繰り返す。未指定の反射音がなければ（S208cでNo）、選択処理部4302dは、処理を終了する。
- [0692] このように、選択処理が行われ、ステップS206cで、選択処理部43

02dが、音を示す音声信号に基づく出力信号を再生部2303が再生することを選択すると、選択部4302は、当該音声信号を再生部2303に出力する。

[0693] そして、再生部2303は、選択部4302から出力された音声信号を取得して、当該音声信号に基づく出力信号を出力する（S103c）。

[0694] 以上のように、本実施の形態に係る音声信号処理方法は、音声信号処理装置が実行する音声信号処理方法であって、取得ステップと、第3算出ステップと、選択処理ステップと、再生ステップとを含む。

[0695] 取得ステップは、音声信号を取得する。第3算出ステップは、取得された音声信号が示す音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された所定補正特性と、取得された音声信号とに基づいて、音が受聴者が居る位置である受聴位置に到来するときの音量に基づく所定音量を算出する。選択処理ステップは、算出された所定音量に基づいて、取得された音声信号に基づく出力信号を出力するか否かを選択する。再生ステップは、出力信号を出力することが選択された場合に、出力信号を出力する。

[0696] これにより、聴覚の感度を示す周波数特性が考慮された上で所定音量が算出され、算出された所定音量に基づいて、音を示す音声信号に基づく出力信号を出力するか否かが選択される。すなわち、聴覚の感度が考慮された上で、当該音声信号に基づく出力信号が出力されるか否かが適切に選択される。出力信号が出力されない場合には、演算量及び演算負荷が削減される。つまりは、聴覚の感度が考慮された上で、演算量及び演算負荷を適切に削減することができる音声信号処理方法を実現することができる。

[0697] （補足）

なお、本開示に基づいて把握される態様は、実施の形態に限定されず、種々変更して実施されてもよい。

[0698] 例えば、実施の形態において特定の構成要素によって実行される処理を特定の構成要素の代わりに別の構成要素が実行してもよい。また、複数の処理

の順序が変更されてもよいし、複数の処理が並行して実行されてもよい。

[0699] また、説明に用いられた第1及び第2等の序数は、適宜、付け替えられてもよいし、取り除かれてもよいし、新たに付与されてもよい。これらの序数は、意味のある順序に必ずしも対応せず、要素の識別に用いられてもよい。

[0700] また、例えば、閾値に対する比較において、閾値以上であることと、閾値よりも大きいこととが相互に読み替えられてもよい。同様に、閾値以下であることと、閾値よりも小さいこととが相互に読み替えられてもよい。また、例えば、時間及び時刻は、相互に読み替えられる場合がある。

[0701] また、複数の音から1つ以上の処理対象音を選択する処理において、条件を満たす音が存在しなければ、どの音も処理対象音として選択されなくてもよい。つまり、複数の音から1つ以上の処理対象音を選択する処理には、処理対象音が選択されないケースが含まれていてもよい。

[0702] また、例えば、第1要素、第2要素及び第3要素の少なくとも1つという表現は、第1要素、第2要素、第3要素、又は、これらの任意の組み合わせに対応し得る。

[0703] なお、実施の形態2で説明された反射音（間接音）に係る所定の周波数帯域幅ごとのゲイン特性、直接音に係る所定の周波数帯域幅ごとのゲイン特性、及び、聴覚の感度を示す周波数特性を示す情報は、例えば、入力情報に含まれる空間情報に基づいて特定されてもよい。メタデータには、例えば床、壁、又は天井等の音空間において音を反射し得る構造物の反射率、及び音空間に存在する障害物オブジェクトの反射率または反射面の反射率を表す情報が含まれる。ここで、反射率は、反射音と入射音とのエネルギーの比或いは、振幅の比であって、音の周波数帯域ごとに設定されている。また反射率は反射係数と呼ばれてもよい。例えば、図32で示されるようなデータが反射率を示す情報としてメタデータに設定されていてもよい。図32では3種類の壁面のそれぞれの反射係数が例示されている。非発音オブジェクトには、1つの反射係数が設定されていてもよいし複数の反射係数が設定されていてもよい。反射音に掛かるイコライザの周波数帯域ごとのGain（すなわち第1

補正特性)は、当該反射音に係る反射面に設定された上述の例の反射係数に基づいて算出されてもよい。

[0704] また、反射音(間接音)に係る所定の周波数帯域幅ごとのゲイン特性、直接音に係る所定の周波数帯域幅ごとのゲイン特性、及び、聴覚の感度を示す周波数特性は、通信回線等を用いて供給されてもよいし、予め解析部2301のメモリに記憶されていてもよい。

[0705] また、例えば、実施の形態では、本開示に基づいて把握される態様を音声信号処理装置、符号化装置又は復号装置として実施する場合について説明している。しかし、本開示に基づいて把握される態様は、これらに限られず、音声信号処理方法、符号化方法又は復号方法を実行するためのソフトウェアとして実施されてもよい。

[0706] 例えば、上述の音声信号処理方法、符号化方法又は復号方法を実行するためのプログラムが、予めROMに記憶されていてもよい。そして、そのプログラムに従って、CPUが動作してもよい。

[0707] また、上述の音声信号処理方法、符号化方法又は復号方法を実行するためのプログラムが、コンピュータ読み取り可能な記録媒体に記憶されてもよい。そして、コンピュータは、記録媒体に記憶されたプログラムをコンピュータのRAMに記録し、そのプログラムに従って動作してもよい。

[0708] そして、上記の各構成要素は、典型的には、入力端子及び出力端子を有する集積回路であるLSIとして実現されてもよい。これらは、個別に1チップ化されてもよいし、実施の形態の全ての構成要素又は一部の構成要素を含むように1チップ化されてもよい。LSIは、集積度の違いにより、IC、システムLSI、スーパーLSI又はウルトラLSIと表現されてもよい。

[0709] また、LSIに限られず、専用回路又は汎用プロセッサが利用されてもよい。また、LSI製造後に、プログラミングが可能なFPGA、又は、LSI内部の回路セルの接続又は設定の再構成が可能なリプログラマブル・プロセッサが利用されてもよい。さらに、半導体技術の進歩又は派生する別技術によりLSIに置き換わる集積回路化の技術が登場すれば、当然、その

技術を用いて構成要素の集積化が行われてもよい。バイオ技術の適応等が可能性としてあり得る。

[0710] また、FPGA又はCPU等が、本開示において説明した音声信号処理方法、符号化方法又は復号方法を実現するためのソフトウェアの全部又は一部を無線通信又は有線通信によりダウンロードしてもよい。さらに、更新のためのソフトウェアの全部又は一部が無線通信又は有線通信によりダウンロードされてもよい。そして、FPGA又はCPU等が、ダウンロードされたソフトウェアをメモリに記憶し、記憶されたソフトウェアに基づいて動作することにより、本開示において説明したデジタル信号処理を実行してもよい。

[0711] このとき、FPGA又はCPU等を具備する機器は、信号処理装置と無線又は有線で接続してもよいし、信号処理サーバとネットワークを介して接続してもよい。そして、この機器と、信号処理装置又は信号処理サーバとは、本開示において説明した音声信号処理方法、符号化方法又は復号方法を行ってもよい。

[0712] 例えば、本開示における音声信号処理装置、符号化装置又は復号装置が、FPGA又はCPU等を具備してもよい。さらに、音声信号処理装置、符号化装置又は復号装置は、FPGA又はCPU等を動作させるためのソフトウェアを外部から入手するためのインタフェース、及び、入手されたソフトウェアを記憶するためのメモリを具備してもよい。そして、FPGA又はCPU等は、記憶されたソフトウェアに基づいて動作することで、本開示において説明した信号処理を実行してもよい。

[0713] 本開示の音響処理、符号化処理又は復号処理に関するソフトウェアをサーバが提供してもよい。そして、端末又は機器は、当該ソフトウェアをインストールすることで、本開示で記載した音声信号処理装置、符号化装置又は復号装置として動作してもよい。なお、端末又は機器が、ネットワークを介しサーバと接続し、ソフトウェアをインストールしてもよい。

[0714] また、端末又は機器とは異なる別の装置が、ネットワークを介しサーバと接続してソフトウェアのインストール用のデータを取得し、当該別の装置が

端末又は機器にソフトウェアのインストール用のデータを提供することで端末又は機器にソフトウェアがインストールされてもよい。なお、ソフトウェアの例は、実施の形態を用いて説明した音声信号処理方法を端末又は機器に実行させるためのVR用ソフトウェア又はAR用ソフトウェアであってもよい。

[0715] なお、上記実施の形態において、各構成要素は、専用のハードウェアで構成されるか、各構成要素に適したソフトウェアプログラムを実行することによって実現されてもよい。各構成要素は、CPU又はプロセッサ等のプログラム実行部が、ハードディスク又は半導体メモリ等の記録媒体に記録されたソフトウェアプログラムを読み出して実行することによって実現されてもよい。

[0716] 以上、1つ以上の態様に係る装置等について、実施の形態に基づいて説明したが、本開示に基づいて把握される態様は、実施の形態に限定されない。本開示の趣旨を逸脱しない限り、当業者が思いつく各種変形を実施の形態に施すことで得られる形態、及び、異なる変形例における構成要素を組み合わせることで構築される形態も、1つ以上の態様の範囲内に含まれてもよい。

産業上の利用可能性

[0717] 本開示は、例えば、音声信号処理装置、符号化装置、復号装置、又は、これらのいずれかの装置を備える端末又は機器において適用可能な態様を含む。

符号の説明

- [0718]
- 1000 立体音響再生システム
 - 1001 音声信号処理装置（音響処理装置）
 - 1002 音声提示装置
 - 1100、1120、1500 符号化装置
 - 1101、1113 入力データ
 - 1102 エンコーダ
 - 1103 符号化データ

1104、1114、1404、1503 メモリ
1110、1130 復号装置
1111 音声信号
1112、1200、1210 デコーダ
1121 送信部
1122 送信信号
1131 受信部
1132 受信信号
1201、1211 空間情報管理部
1202 音声データデコーダ
1203、1213、1300、2300、3300、4300 レン
ダリング部
1301、2301、3301、4301 解析部
1302、1314、2302、3302、4302 選択部
1303、2303 再生部
1304 閾値調整部
1311 残響処理部
1312 初期反射処理部
1313 距離減衰処理部
1315 生成部
1316 バイノーラル処理部
1401 スピーカ
1402、1501 プロセッサ
1403、1502 通信ＩＦ
1405 センサ
2302a、3302a、4302a 取得部
2302b 第1算出部
2302c 第2算出部

2 3 0 2 d、3 3 0 2 d、4 3 0 2 d 選択処理部

4 3 0 2 e 第3算出部

請求の範囲

- [請求項1] 音声信号処理装置が実行する音声信号処理方法であって、
- 音声信号であって、前記音声信号の属性を特定する属性情報を含み前記属性が間接音を示す情報を含む音声信号を取得する取得ステップと、
- 前記間接音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第1補正特性と、取得された前記音声信号とに基づいて、前記間接音が受聴者が居る位置である受聴位置に到来するときの間接音音量に基づく第1音量を算出する第1算出ステップと、
- 前記間接音に係る直接音が前記受聴位置に到来するときの直接音音量に基づく第2音量を算出する第2算出ステップと、
- 算出された前記第2音量と算出された前記第1音量との音量比、及び、前記直接音と前記間接音との到来する時間差に基づいて、取得された前記音声信号に基づく出力信号を出力するか否かを選択する選択処理ステップと、
- 前記出力信号を出力することが選択された場合に、前記出力信号を出力する再生ステップと、を含む、
- 音声信号処理方法。
- [請求項2] 前記第2算出ステップでは、前記直接音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第2補正特性に基づいて、前記第2音量を算出する、
- 請求項1に記載の音声信号処理方法。
- [請求項3] 前記聴覚の感度を示す前記周波数特性は、前記受聴者の音量の感度を示す周波数特性である、
- 請求項1又は2に記載の音声信号処理方法。
- [請求項4] 前記音量の感度を示す前記周波数特性は、等ラウドネス曲線の逆数に応じた周波数特性である、

- 請求項 3 に記載の音声信号処理方法。
- [請求項5] 前記音量の感度を示す前記周波数特性は、A 特性である、
請求項 3 に記載の音声信号処理方法。
- [請求項6] 前記音量の感度を示す前記周波数特性は、音源位置の最小弁別角度
の周波数特性の逆特性である、
請求項 3 に記載の音声信号処理方法。
- [請求項7] 音声信号処理装置が実行する音声信号処理方法であって、
音声信号であって、前記音声信号の属性を特定する属性情報を含み
前記属性が、直接音が反射物によって反射された音である反射音を示
す情報を含む音声信号を取得する取得ステップと、
取得された前記音声信号が示す前記反射音が受聴者が居る位置であ
る受聴位置に到来するときの反射音音量と前記直接音が前記受聴位置
に到来するときの直接音音量との音量比、前記直接音と前記反射音と
の到来する時間差、及び、前記反射物の反射係数に基づいて決定され
る反射係数特徴量に基づいて、取得された前記音声信号に基づく出力
信号を出力するか否かを選択する選択処理ステップと、
前記出力信号を出力することが選択された場合に、前記出力信号を
出力する再生ステップと、を含む、
音声信号処理方法。
- [請求項8] 前記反射係数特徴量は、反射係数の周波数特性の平坦さの程度を示
す特徴量である、
請求項 7 に記載の音声信号処理方法。
- [請求項9] 音声信号処理装置が実行する音声信号処理方法であって、
音声信号を取得する取得ステップと、
取得された前記音声信号が示す音に係る所定の周波数帯域幅ごとの
ゲイン特性が聴覚の感度を示す周波数特性によって補正された所定補
正特性と、取得された前記音声信号とに基づいて、前記音が受聴者が
居る位置である受聴位置に到来するときの音量に基づく所定音量を算

出する第3算出ステップと、

算出された前記所定音量に基づいて、取得された前記音声信号に基づく出力信号を出力するか否かを選択する選択処理ステップと、

前記出力信号を出力することが選択された場合に、前記出力信号を出力する再生ステップと、を含む、

音声信号処理方法。

[請求項10]

音声信号処理装置が実行する音声信号処理方法であって、

音声信号であって、前記音声信号の属性を特定する属性情報を含む音声信号を取得する取得ステップと、

取得された前記音声信号が含む前記属性情報によって特定される前記属性が間接音を示す情報を含む音声信号の所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第1補正特性と、前記間接音のゲイン情報とに基づいて、前記間接音が受聴者が居る位置である受聴位置に到来するときの間接音音量に基づく第1音量を算出する第1算出ステップと、

前記間接音に係る直接音の所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第2補正特性と、前記直接音のゲイン情報とに基づいて、前記受聴位置に到来するときの直接音音量に基づく第2音量を算出する第2算出ステップと、

算出された前記第2音量と算出された前記第1音量との音量比、及び、前記直接音と前記間接音との到来する時間差に基づいて、取得された前記音声信号に基づく出力信号を出力するか否かを選択する選択処理ステップと、

前記出力信号を出力することが選択された場合に、前記出力信号を出力する再生ステップと、を含む、

音声信号処理方法。

[請求項11]

前記間接音に係る所定の周波数帯域幅ごとのゲイン特性は、前記間接音に係る前記属性情報として保持されており、

前記間接音に係るゲイン情報は、前記間接音に係る前記属性情報として保持されており、

前記直接音に係る所定の周波数帯域幅ごとのゲイン特性は、前記直接音に係る前記属性情報として保持されており、

前記直接音に係るゲイン情報は、前記直接音に係る前記属性情報として保持されている、

請求項 10 記載の音声信号処理方法。

[請求項12] 請求項 1 ～ 11 のいずれか 1 項に記載の音声信号処理方法をコンピュータに実行させるためのコンピュータプログラム。

[請求項13] 音声信号であって、前記音声信号の属性を特定する属性情報を含み前記属性が間接音を示す情報を含む音声信号を取得する取得部と、

前記間接音に係る所定の周波数帯域幅ごとのゲイン特性が聴覚の感度を示す周波数特性によって補正された第 1 補正特性と、取得された前記音声信号とに基づいて、前記間接音が受聴者が居る位置である受聴位置に到来するときの間接音音量に基づく第 1 音量を算出する第 1 算出部と、

前記間接音に係る直接音が前記受聴位置に到来するときの直接音音量に基づく第 2 音量を算出する第 2 算出部と、

算出された前記第 2 音量と算出された前記第 1 音量との音量比、及び、前記直接音と前記間接音との到来する時間差に基づいて、取得された前記音声信号に基づく出力信号を出力するか否かを選択する選択処理部と、

前記出力信号を出力することが選択された場合に、前記出力信号を出力する再生部と、を備える、

音声信号処理装置。