

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7704686号
(P7704686)

(45)発行日 令和7年7月8日(2025.7.8)

(24)登録日 令和7年6月30日(2025.6.30)

(51)国際特許分類		F I	
G 1 0 L	19/22 (2013.01)	G 1 0 L	19/22
G 1 0 L	25/21 (2013.01)	G 1 0 L	25/21
H 0 4 S	7/00 (2006.01)	H 0 4 S	7/00 3 0 0
請求項の数 15 (全43頁)			
(21)出願番号	特願2021-557218(P2021-557218)	(73)特許権者	515076873
(86)(22)出願日	令和2年3月19日(2020.3.19)		ノキア テクノロジーズ オサケユイチア
(65)公表番号	特表2022-528837(P2022-528837		フィンランド国, 0 2 6 1 0 エスプー
	A)		, カラカーリ 7
(43)公表日	令和4年6月16日(2022.6.16)	(74)代理人	100099759
(86)国際出願番号	PCT/FI2020/050174		弁理士 青木 篤
(87)国際公開番号	WO2020/193852	(74)代理人	100123582
(87)国際公開日	令和2年10月1日(2020.10.1)		弁理士 三橋 真二
審査請求日	令和3年11月24日(2021.11.24)	(74)代理人	100092624
審判番号	不服2023-20028(P2023-20028/J		弁理士 鶴田 準一
	1)	(74)代理人	100114018
審判請求日	令和5年11月27日(2023.11.27)		弁理士 南山 知広
(31)優先権主張番号	1904261.3	(74)代理人	100117019
(32)優先日	平成31年3月27日(2019.3.27)		弁理士 渡辺 陽一
(33)優先権主張国・地域又は機関		(74)代理人	100108903
最終頁に続く		最終頁に続く	

(54)【発明の名称】 音場関連のレンダリング

(57)【特許請求の範囲】

【請求項 1】

少なくとも2つのオーディオ信号を取得し、
前記少なくとも2つのオーディオ信号に関連する少なくとも1つのパラメータを取得し、
前記少なくとも2つのオーディオ信号に関連する前記少なくとも1つのパラメータに基づいて、前記少なくとも2つのオーディオ信号のタイプを決定し(201)、
前記少なくとも2つのオーディオ信号の決定された前記タイプに基づいてレンダリングされるように構成された前記少なくとも2つのオーディオ信号を処理する(203)ように構成される手段
を備えた装置。

【請求項 2】

前記少なくとも2つのオーディオ信号は、伝送オーディオ信号(502)、および、以前に処理されたオーディオ信号のうちの1つである、請求項1に記載の装置。

【請求項 3】

前記少なくとも1つのパラメータに基づいて、前記少なくとも2つのオーディオ信号のタイプを決定する(201)ように構成される前記手段は、

前記少なくとも1つのパラメータから少なくとも1つのタイプ信号を抽出して復号すること(501)と、

前記少なくとも1つのパラメータが前記少なくとも2つのオーディオ信号に関連する空間オーディオ態様を表すとき、前記少なくとも2つのオーディオ信号のタイプを決定す

るために、前記少なくとも 1 つのパラメータを解析することと、
のうちの 1 つを実行するように構成される、
請求項 1 または 2 に記載の装置。

【請求項 4】

前記少なくとも 2 つのオーディオ信号のタイプを決定するために、前記少なくとも 1 つのパラメータを解析するように構成された前記手段は、

前記少なくとも 2 つのオーディオ信号に基づいて ブロードバンド左または右チャンネル対トータルエネルギー比率を決定し (5 0 5)、

少なくとも 2 つのオーディオ信号に基づいて 高周波数左または右チャンネル対トータルエネルギー比率を決定し (5 0 7)、

前記少なくとも 2 つのオーディオ信号に基づいて 合計対トータルエネルギー比率を決定し (5 0 9)、

前記少なくとも 2 つのオーディオ信号に基づいて 減算対ターゲットエネルギー比率を決定し (5 1 2)、

および、

前記ブロードバンド左または右チャンネル対トータルエネルギー比率、前記少なくとも 2 つのオーディオ信号に基づいた前記高周波数左または右チャンネル対トータルエネルギー比率、少なくとも 2 つのオーディオ信号に基づいた前記合計対トータルエネルギー比率、および、前記減算対ターゲットエネルギー比率のうちの少なくとも 1 つに基づいて、前記少なくとも 2 つのオーディオ信号のタイプを決定する (5 1 3)

ように構成される、

請求項 3 に記載の装置。

【請求項 5】

前記手段は、前記少なくとも 2 つのオーディオ信号のうちの少なくとも 1 つのタイプに関連する少なくとも 1 つのタイプパラメータを決定する、ように構成される、請求項 1 ないし 4 のいずれか 1 項に記載の装置。

【請求項 6】

前記少なくとも 2 つのオーディオ信号の決定された前記タイプに基づいて レンダリングされるように構成された前記少なくとも 2 つのオーディオ信号を処理するように構成される前記手段は、前記少なくとも 2 つのオーディオ信号の前記タイプに関連する前記少なくとも 1 つのタイプパラメータに基づいて、前記少なくとも 2 つのオーディオ信号を変換するように構成される、請求項 5 に記載の装置。

【請求項 7】

前記少なくとも 2 つのオーディオ信号のタイプは、

間隔を置いたオーディオ信号タイプ、

ダウンミックスオーディオ信号タイプ、

同一オーディオ信号タイプ、

のうちの少なくとも 1 つを含む、

請求項 1 ないし 6 のいずれか 1 項に記載の装置。

【請求項 8】

前記少なくとも 2 つのオーディオ信号を処理するように構成された前記手段は、

アンビソニックオーディオ信号表現に少なくとも 2 つのオーディオ信号を変換することと、

前記少なくとも 2 つのオーディオ信号をアンビソニックオーディオ信号表現に変換すること (9 0 3) と、

前記少なくとも 2 つのオーディオ信号をより少ないオーディオ信号にダウンミックスすること (1 3 0 3) と、

のうちの 1 つを実行するように構成される、

請求項 1 ないし 7 のいずれかに記載の装置。

【請求項 9】

10

20

30

40

50

前記少なくとも2つのオーディオ信号を処理するように構成される前記手段は、

前記少なくとも2つのオーディオ信号および前記少なくとも2つのオーディオ信号のタイプに基づいて、少なくとも1つのプロトタイプ信号を生成するように構成される、請求項1ないし8のいずれか1項に記載の装置。

【請求項10】

少なくとも2つのオーディオ信号を取得するステップ(301)と、

前記少なくとも2つのオーディオ信号に関連する少なくとも1つのパラメータを取得するステップと、

少なくとも1つのパラメータが少なくとも2つのオーディオ信号に関連する前記少なくとも1つのパラメータに基づいて、前記少なくとも2つのオーディオ信号のタイプを決定するステップ(303)と、

前記少なくとも2つのオーディオ信号の前記決定されたタイプに基づいて、レンダリングされるように構成された前記少なくとも2つのオーディオ信号を処理するステップ(305)と、
を含む、方法。

【請求項11】

前記少なくとも2つのオーディオ信号は、伝送オーディオ信号(502)、および、前もって処理されたオーディオ信号、のうちの1つである、請求項10に記載の方法。

【請求項12】

前記少なくとも1つのパラメータに基づいて、前記少なくとも2つのオーディオ信号のタイプを決定するステップは、

前記少なくとも1つのパラメータから少なくとも1つのタイプ信号を抽出して復号するステップ(601)と、

前記少なくとも1つのパラメータが前記少なくとも2つのオーディオ信号に関連する空間オーディオ態様を表すとき、前記少なくとも2つのオーディオ信号のタイプを決定するために、前記少なくとも1つのパラメータを解析するステップと、
のうちの1つを含む、

請求項10または11に記載の方法。

【請求項13】

前記少なくとも2つのオーディオ信号のタイプを決定するために、前記少なくとも1つのパラメータを解析するステップは、

前記少なくとも2つのオーディオ信号に基づいてブロードバンド左または右チャンネル対トータルエネルギー比率を決定するステップ(605)と、

前記少なくとも2つのオーディオ信号に基づいて高周波数左または右チャンネル対トータルエネルギー比率を決定するステップ(607)と、

前記少なくとも2つのオーディオ信号に基づいて合計対トータルエネルギー比率を決定するステップ(609)と、

前記少なくとも2つのオーディオ信号に基づいて減算対ターゲットエネルギー比率を決定するステップ(611)と、

前記ブロードバンド左または右チャンネル対トータルエネルギー比率、

少なくとも2つのオーディオ信号に基づいた前記高周波数左または右チャンネル対トータルエネルギー比率、

少なくとも2つのオーディオ信号に基づいた前記合計対トータルエネルギー比率、
および、

前記減算対ターゲットエネルギー比率

のうちの少なくとも1つに基づいて、前記少なくとも2つのオーディオ信号のタイプを決定するステップ(613)と、

を含む、

請求項12に記載の方法。

【請求項14】

10

20

30

40

50

前記少なくとも2つのオーディオ信号のうちの少なくとも1つタイプに関連する少なくとも1つのタイプパラメータを決定するステップ

をさらに含む、請求項10ないし13のいずれか1項に記載の方法。

【請求項15】

前記少なくとも2つのオーディオ信号の該決定されたタイプに基づいてレンダリングされるように構成された前記少なくとも2つのオーディオ信号を処理するステップは、

前記少なくとも2つのオーディオ信号のタイプに関連する少なくとも1つのタイプパラメータに基づいて、前記少なくとも2つのオーディオ信号を変換するステップ(305)を含む、

請求項10に記載の方法。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音場関連のオーディオ表現およびレンダリングのための装置および方法に関するが、オーディオデコーダのためのオーディオ表現に限るものではない。

【背景技術】

【0002】

イマーシブオーディオコーデックは、低ビットレート動作から透明度まで、多数の動作ポイントをサポートしている。このようなコーデックの一例は、仮想現実(VR)のためのイマーシブ音声およびオーディオのようなイマーシブサービスでの使用を含む3GPP 4G/5Gネットワークのような通信ネットワーク上での使用に適するように設計されているイマーシブ音声およびオーディオサービス(IVAS)コーデックである。この音声コーデックは、音声、音楽、汎用音声の符号化、復号、レンダリングを扱うことが期待される。さらに、音場および音源に関する空間情報を含むチャンネルベースのオーディオおよびシーンベースのオーディオ入力をサポートすることが期待される。また、コーデックは、様々な伝送条件下で高いエラーロバスト性をサポートするだけでなく、会話サービスを可能にするために低い待ち時間で動作することが期待される。

20

【0003】

入力信号は、サポートされている多数のフォーマットのいずれかで(また、可能なフォーマットの組み合わせによって)IVASエンコーダに提示することができる。たとえば、モノラルオーディオ信号(メタデータなし)は、EVS(Enhanced Voice Service)エンコーダを使用してエンコードできる。他の入力フォーマットは、IVAS符号化ツールを利用してもよい。少なくともいくつかの入力は、メタデータ支援空間オーディオ(MASA)ツールまたは任意の適切な空間メタデータベーススキームを利用できる。これは、空間オーディオ処理に適したパラメトリック空間オーディオ形式である。パラメトリック空間音声処理は、音声(または音場)の空間的側面がパラメータの集合を用いて記述される音声信号処理の分野である。例えば、マイクロホンアレイからのパラメトリック空間オーディオキャプチャでは、マイクロホンアレイ信号から、周波数帯域における音の方向、および周波数帯域におけるキャプチャされた音の指向性部分と無指向性部分との間の比率などのパラメータのセットを推定することは、典型的で有効な選択である。これらのパラメータは、マイクロホンアレイの位置で捕捉された音の知覚空間特性をよく記述することが知られている。これらのパラメータは、それに応じて空間音の合成、バイノーラルでのヘッドホン、ラウドスピーカ、またはアンビソニックのような他のフォーマットに利用することができる。

30

40

【0004】

たとえば、オーディオ信号と空間メタデータの2つのチャンネル(ステレオ)がある。空間メタデータは、更に、方向インデックス(時間-周波数パラメータ間隔における音の到着方向を記述する)、方向対トータルエネルギー比率(方向指標、すなわち、時間-周波数サブフレームに対するエネルギー比率を記述する)、拡張コヒーレンス(Spread coherence)(周囲方向に対する無指向性音のエネルギーゼ比率を記述する)、

50

拡散対トータルエネルギー比率 (Diffuse-to-total energy ratio) (周囲方向に対する無指向性音のコヒーレンスを記述する)、サラウンドコヒーレンス (Surround coherence) (周囲方向に対する無指向性音のコヒーレンスを記述する)、残余対トータルエネルギー比率 (Remainder-to-total energy ratio) (エネルギー比率の合計が1であるという要件を満たすための残余 (マイクノイズなど) の音響エネルギーのエネルギー比率を記述する)、および、距離 (対数スケールにおける方向指標 (すなわち、時間 - 周波数サブフレーム) から発する音の距離を記述する)、のようなパラメータを定義することができる。

【0005】

I V A S ストリームは、バイナリ、マルチチャンネル、およびアンビソニック (Ambisonic) (F O A / H O A) 出力など、さまざまな出力形式にデコードおよびレンダリングできる。

10

【0006】

少なくとも2つのオーディオ信号の決定されたタイプに基づいてレンダリングされるように構成された少なくとも2つのオーディオ信号を処理するように構成された手段は、少なくとも2つのオーディオ信号をアンビソニックオーディオ信号表現に変換すること、少なくとも2つのオーディオ信号をマルチチャンネルオーディオ信号表現に変換すること、および少なくとも2つのオーディオ信号をより少ないオーディオ信号にダウンミックスすること、を構成することができる。

【0007】

20

少なくとも2つのオーディオ信号の決定されたタイプに基づいてレンダリングされるように構成された少なくとも2つのオーディオ信号を処理するように構成された手段は、少なくとも2つのオーディオ信号と少なくとも2つのオーディオ信号のタイプに基づいて少なくとも1つのプロトタイプ信号を生成するように構成することができる。

【0008】

第2の態様によれば、少なくとも2つのオーディオ信号を得るステップと、少なくとも2つのオーディオ信号のタイプを決定するステップと、少なくとも2つのオーディオ信号の決定されたタイプに基づいてレンダリングされるように構成された少なくとも2つのオーディオ信号を処理するステップと、を含む方法が提供される。

【0009】

30

少なくとも2つのオーディオ信号は、搬送オーディオ信号、および前に処理されたオーディオ信号のうちの1つであり得る。

【0010】

本方法は、少なくとも2つのオーディオ信号に関連する少なくとも1つのパラメータを取得することをさらに含むことができる。

【0011】

少なくとも2つのオーディオ信号のタイプを決定することは、少なくとも2つのオーディオ信号に関連する少なくとも1つのパラメータに基づいて、少なくとも2つのオーディオ信号のタイプを決定することを含むことができる。

【0012】

40

少なくとも1つのパラメータに基づいて少なくとも2つのオーディオ信号のタイプを決定することは、少なくとも1つのパラメータから少なくとも1つのタイプの信号を抽出してデコードすることと、少なくとも1つのパラメータが少なくとも2つのオーディオ信号に関連する空間オーディオアスペクトを表す場合に、少なくとも1つのパラメータを解析して少なくとも2つのオーディオ信号のタイプを決定することとのうちの1つを含むことができる。

【0013】

少なくとも1つのパラメータを解析して、前記少なくとも2つのオーディオ信号のタイプを決定することは、前記少なくとも2つのオーディオ信号に基づいて、ブロードバンドの左または右のチャンネル対トータルエネルギー比率を決定することと、前記少なくとも2つ

50

のオーディオ信号に基づいて、より高い周波数の左または右のチャンネル対トータルエネルギー比率を決定することと、前記少なくとも2つのオーディオ信号に基づいて、合計対トータルエネルギー比率に対する合計を決定することと、前記少なくとも2つのオーディオ信号に基づいて、減算対ターゲット・エネルギー比率を決定することと、前記少なくとも2つのオーディオ信号のタイプを、前記ブロードバンドの左または右のチャンネル対トータルエネルギー比率、前記少なくとも2つのオーディオ信号に基づいて、より高い周波数の左または右のチャンネル対トータルエネルギー比率、前記少なくとも2つのオーディオ信号に基づいて、合計対トータルエネルギー比率、および前記減算対ターゲット・エネルギー比率のうちの少なくとも1つに基づいて決定することと、を含むことができる。

【0014】

10

本願方法は、少なくとも1つのオーディオ信号のタイプに関連する少なくとも1つのタイプパラメータを決定することをさらに含むことができる。

【0015】

少なくとも2つのオーディオ信号の決定されたタイプに基づいてレンダリングされるように構成された少なくとも2つのオーディオ信号を処理することは、少なくとも2つのオーディオ信号のタイプに関連する少なくとも1つのタイプパラメータに基づいて少なくとも2つのオーディオ信号を変換することを更に含むことができる。

【0016】

少なくとも2つのオーディオ信号のタイプは、キャプチャマイクロホン配置、キャプチャマイクロホン分離距離、キャプチャマイクロホンパラメータ、トランスポートチャンネル識別子、間隔を置いたオーディオ信号タイプ、ダウンミックスオーディオ信号タイプ、同一オーディオ信号タイプ、およびトランスポートチャンネル配置のうちの少なくとも1つを含むことができる。

20

【0017】

少なくとも2つのオーディオ信号の決定されたタイプに基づいてレンダリングされるように構成された少なくとも2つのオーディオ信号を処理することは、少なくとも2つのオーディオ信号をアンビソニックオーディオ信号表現に変換すること、少なくとも2つのオーディオ信号をマルチチャンネルオーディオ信号表現に変換すること、および少なくとも2つのオーディオ信号をより少ないオーディオ信号にダウンミックスすることのうちの1つを含むことができる。

30

【0018】

少なくとも2つのオーディオ信号の決定されたタイプに基づいてレンダリングされるように構成された少なくとも2つのオーディオ信号を処理することは、少なくとも2つのオーディオ信号および少なくとも2つのオーディオ信号のタイプに基づいて少なくとも1つのプロトタイプ信号を生成することを含むことができる。

【0019】

第3の態様によれば、少なくとも1つのプロセッサと、コンピュータプログラムコードを含む少なくとも1つのメモリとを備える、装置であって、前記少なくとも1つのメモリと、前記少なくとも1つのコンピュータプログラムコードとは、前記少なくとも1つのプロセッサを用いて、前記装置に、少なくとも、前記装置が少なくとも2つのオーディオ信号を取得することと、前記少なくとも2つのオーディオ信号のタイプを決定することと、前記少なくとも2つのオーディオ信号の決定されたタイプに基づいてレンダリングされるように構成された前記少なくとも2つのオーディオ信号を処理することと、をさせるように構成される、装置が提供される。

40

【0020】

少なくとも2つのオーディオ信号は、搬送オーディオ信号、および前に処理されたオーディオ信号のうちの1つであり得る。

【0021】

手段は、少なくとも2つのオーディオ信号に関連する少なくとも1つのパラメータを取得するように構成されることができる。

50

【 0 0 2 2 】

少なくとも2つのオーディオ信号のタイプを決定するようにした装置は、少なくとも2つのオーディオ信号に関連する少なくとも1つのパラメータに基づいて、少なくとも2つのオーディオ信号のタイプを決定することができる。

【 0 0 2 3 】

前記少なくとも1つのパラメータに基づいて前記少なくとも2つのオーディオ信号のタイプを決定する装置は、前記少なくとも1つのパラメータから少なくとも1つのタイプ信号を抽出してデコードすることと、前記少なくとも1つのパラメータが前記少なくとも2つのオーディオ信号に関連する空間オーディオ態様を表すとき、前記少なくとも1つのパラメータを解析して、前記少なくとも2つのオーディオ信号のタイプを決定することとのうちの1つを実行することができる。

10

【 0 0 2 4 】

少なくとも2つの音響信号の種類を決定するための少なくとも1つのパラメータを解析する装置は、少なくとも2つの音響信号に基づいて、ブロードバンド左または右チャンネル対トータルエネルギー比率を決定し、少なくとも2つの音響信号に基づいて、より高い周波または右チャンネル対トータルエネルギー比率を決定し、少なくとも2つの音響信号に基づいて、合計対トータルエネルギー比率を決定し、少なくとも2つの音響信号に基づいて、減算対トータルエネルギー比率を決定しブロードバンド左または右チャンネル対トータルエネルギー比率、少なくとも2つの音響信号に基づく高周波左または右チャンネル対トータルエネルギー比率、少なくとも2つの音響信号に基づく、合計対トータルエネルギー比率、および、減算対ターゲット・エネルギー比率のうちの少なくとも1つに基づいて、少なくとも2つの音響信号のタイプを決定することができる。

20

【 0 0 2 5 】

装置は、少なくとも1つのオーディオ信号のタイプに関連する少なくとも1つのタイプパラメータを決定することができる。

【 0 0 2 6 】

少なくとも2つのオーディオ信号の決定されたタイプに基づいてレンダリングされるように構成された少なくとも2つのオーディオ信号を処理した装置は、少なくとも2つのオーディオ信号のタイプに関連する少なくとも1つのタイプパラメータに基づいて、少なくとも2つのオーディオ信号を変換させることができる。

30

【 0 0 2 7 】

少なくとも2つのオーディオ信号のタイプは、キャプチャマイクロホン配置、キャプチャマイクロホン分離距離、キャプチャマイクロホンパラメータ、トランスポートチャンネル識別子、間隔を置いたオーディオ信号タイプ、ダウンミックスオーディオ信号タイプ、同一オーディオ信号タイプ、およびトランスポートチャンネル配置のうちの少なくとも1つを含むことができる。

【 0 0 2 8 】

装置は、少なくとも2つのオーディオ信号の決定されたタイプに基づいてレンダリングされるように構成された少なくとも2つのオーディオ信号を処理し、少なくとも2つのオーディオ信号をアンビソニックオーディオ信号表現に変換し、少なくとも2つのオーディオ信号をマルチチャンネルオーディオ信号表現に変換し、少なくとも2つのオーディオ信号をより少ないオーディオ信号にダウンミックスすることができる。

40

【 0 0 2 9 】

本願装置は、少なくとも2つのオーディオ信号の決定されたタイプに基づいてレンダリングされるように構成された少なくとも2つのオーディオ信号を処理し、少なくとも2つのオーディオ信号と少なくとも2つのオーディオ信号のタイプに基づいて少なくとも1つのプロトタイプ信号を生成することができる。

【 0 0 3 0 】

第4の態様によれば、少なくとも2つのオーディオ信号を得るように構成された回路を得るステップと、前記少なくとも2つのオーディオ信号のタイプを決定するように構成され

50

た決定回路と、前記少なくとも2つのオーディオ信号の前記決定されたタイプに基づいてレンダリングされるように構成された前記少なくとも2つのオーディオ信号を処理するように構成された処理回路と、を含む装置が提供される。

【0031】

第5の態様によれば、装置に、少なくとも2つのオーディオ信号を取得することと、前記少なくとも2つのオーディオ信号のタイプを決定することと、前記少なくとも2つのオーディオ信号の決定されたタイプに基づいてレンダリングされるように構成された前記少なくとも2つのオーディオ信号を処理することとを少なくとも実行させるための命令を含むコンピュータプログラムが提供される（またはプログラム命令を含むコンピュータ可読メディアが提供される）。

10

【0032】

第6の態様によれば、装置に、少なくとも2つのオーディオ信号を取得することと、前記少なくとも2つのオーディオ信号のタイプを決定することと、前記少なくとも2つのオーディオ信号の決定されたタイプに基づいてレンダリングされるように構成された前記少なくとも2つのオーディオ信号を処理することと、を少なくとも実行させるためのプログラム命令を含む、非一時的コンピュータ可読メディアが提供される。

【0033】

第7の態様によれば、少なくとも2つのオーディオ信号を取得する手段と、少なくとも2つのオーディオ信号のタイプを決定する手段と、前記少なくとも2つのオーディオ信号の決定されたタイプに基づいてレンダリングされるように構成された前記少なくとも2つのオーディオ信号を処理する手段と、を備える装置が提供される。

20

【0034】

第8の態様によれば、少なくとも2つのオーディオ信号を取得することと、前記少なくとも2つのオーディオ信号のタイプを決定することと、前記少なくとも2つのオーディオ信号の決定されたタイプに基づいてレンダリングされるように構成された前記少なくとも2つのオーディオ信号を処理することと、を装置に実行させるためのプログラム命令を含む、コンピュータ可読メディアが提供される。

【0035】

上述の方法の動作を実行する手段を含む装置。

【0036】

上記の方法のアクションを実行するように構成された装置。

30

【0037】

コンピュータに上述の方法を実行させるためのプログラム命令を含むコンピュータプログラム。

【0038】

メディア上に記憶されたコンピュータプログラム製品は、本明細書に記載する方法を装置に実行させることができる。

【0039】

電子デバイスは、本明細書に記載する装置を含むことができる。

【0040】

チップセットは、本明細書に記載する装置を含むことができる。

40

【0041】

本発明の実施形態は、最新技術に関連する課題に対処することを目的としている。

【図面の簡単な説明】

【0042】

本出願の理解を深めるために、ここでは、添付の図面を例として参照することにする。

【図1】図1は、いくつかの実施形態を実施するのに適した装置のシステムを概略的に示す。

【図2】図2は、いくつかの実施形態によるデコーダ/レンダラの例を概略的に示す。

【図3】図3は、いくつかの実施形態による例のデコーダ/レンダラの動作のフロー図を

50

示す。

【図４】図４は、いくつかの実施形態に従う、図２に示されるような一例の搬送オーディオ信号タイプ決定器を概略的に示す。

【図５】図５は、いくつかの実施形態に従う、図２に示されるような第２例の搬送オーディオ信号タイプ決定器を概略的に示す。

【図６】図６は、いくつかの実施例に基づく第２例の搬送オーディオ信号タイプ決定器の動作のフロー図を示す。

【図７】図７は、いくつかの実施形態による、図２に示されるような、アンビソニクスフォーマットコンバータへのメタデータ支援空間オーディオ信号の例を概略的に示す。

【図８】図８は、いくつかの実施形態による、アンビソニクス・フォーマット・変換器へのサンプルメタデータ支援空間オーディオ信号の動作のフロー図を示す。

【図９】図９は、いくつかの実施形態による第２の例のデコーダ/レンダラを概略的に示す。

【図１０】図１０は、いくつかの実施形態によるさらなる例のデコーダ/レンダラの動作のフロー図を示す。

【図１１】図１１は、いくつかの実施形態による、図９に示されるような、マルチチャンネルオーディオ信号フォーマット変換器へのメタデータ支援空間オーディオ信号の例を概略的に示す。

【図１２】図１２は、いくつかの実施形態による、マルチチャンネルオーディオ信号フォーマットコンバータへの、サンプルメタデータ支援空間オーディオ信号の動作のフロー図を示す。

【図１３】図１３は、いくつかの実施形態による第３の例のデコーダ/レンダラを概略的に示す。

【図１４】図１４は、いくつかの実施形態による第３の例のデコーダ/レンダラの動作のフロー図を示す。

【図１５】図１５は、いくつかの実施形態に従った、図１３に示されるような、例示的なメタデータ支援空間オーディオ信号ダウンミキサーを示す。

【図１６】図１６は、いくつかの実施形態による、例のメタデータ支援空間オーディオ信号ダウンミキサーの動作のフロー図を示す。

【図１７】図１７は、図１、２、４、５、７、９、１１、１３および１５に示される装置を実現するのに適した例の装置を示す。

【発明を実施するための形態】

【００４３】

以下では、空間メタデータ支援オーディオ信号の効率的なレンダリングを提供するための、適切な装置および可能なメカニズムをさらに詳細に説明する。

【００４４】

図１に関して、オーディオキャプチャおよびレンダリングを実現するための装置およびシステムの例が示されている。システム１００は、「解析」部１２１と「デマルチプレクサ/デコーダ/シンセサイザ」部１３３とを備えて示されている。「解析」部１２１は、マルチチャンネルラウドスピーカ信号を受信してからメタデータおよび搬送信号を符号化するまでの部分であり、「デマルチプレクサ/デコーダ/シンセサイザ」部１３３は、符号化されたメタデータおよび搬送信号を復号してから、再生成された信号を提示するまでの部分である（たとえば、マルチチャンネルラウドスピーカ形成）。

【００４５】

システム１００および「解析」パート１２１への入力は、マルチチャンネル信号１０２である。以下の例では、マイクロホンチャンネル信号入力が記載されているが、他の実施形態では、任意の適切な入力（または合成マルチチャンネル）フォーマットを実現することができる。例えば、いくつかの実施形態では、空間解析器および空間解析は、エンコーダの外部で実施されてもよい。例えば、ある実施形態では、オーディオ信号に関連する空間メタデータは、別個のビットストリームとしてエンコーダに提供されてもよい。ある実施形態で

10

20

30

40

50

は、空間メタデータは、空間（方向）インデックス値のセットとして提供されてもよい。

【 0 0 4 6 】

マルチチャネル信号は、搬送信号発生器 1 0 3 および解析プロセッサ 1 0 5 に渡される。

【 0 0 4 7 】

いくつかの実施形態では、搬送信号発生器 1 0 3 は、マルチチャネル信号を受信し、決定された数のチャネルを含む適切な搬送信号を発生し、搬送信号 1 0 4 を出力するように構成される。例えば、トランスポート信号発生器 1 0 3 は、マルチチャネル信号の 2 つのオーディオチャネルダウンミックスを生成するように構成することができる。判定されたチャネル数は、任意の適切な数のチャネルとすることができる。いくつかの実施形態における搬送信号発生器は、例えば、ビーム形成技術によって、入力オーディオ信号を決定されたチャネル数に選択または結合し、これらを搬送信号として出力するように構成される。

10

【 0 0 4 8 】

いくつかの実施形態では、搬送信号発生器 1 0 3 は任意であり、マルチチャネル信号は、搬送信号がこの例にあるのと同様に、「エンコーダ / M U X」ブロック 1 0 7 に未処理で渡される。

【 0 0 4 9 】

いくつかの実施形態では、解析プロセッサ 1 0 5 はまた、マルチチャネル信号を受信し、その信号を解析して、マルチチャネル信号に関連し、したがって搬送信号 1 0 4 に関連したメタデータ 1 0 6 を生成するように構成される。解析プロセッサ 1 0 5 は、各時間 - 周波数解析間隔に対して、方向パラメータ 1 0 8 およびエネルギー比率パラメータ 1 1 0（その一例は拡散性パラメータ）およびコヒーレンス・パラメータ 1 1 2 を含むメタデータを生成するように構成することができる。方向、エネルギー比率及びコヒーレンス・パラメータは、実施形態では、空間オーディオパラメータとみなすことができる。言い換えると、空間オーディオパラメータは、マルチチャネル信号（または一般に 2 つ以上の再生オーディオ信号）によって作成された音場を特徴付けることを目的とするパラメータを含む。

20

【 0 0 5 0 】

一部の実施形態では、生成されるパラメータは、周波数帯域ごとに異なる場合がある。したがって、例えばバンド X では、すべてのパラメータが生成されて送信されるのに対し、バンド Y では、生成されて送信されるパラメータは 1 つだけであり、さらにバンド Z では、パラメータは生成されず、送信されない。この実用的な例としては、最高帯域などの一部の周波数帯では、知覚上の理由から一部のパラメータが不要であることが考えられる。トランスポート信号 1 0 4 およびメタデータ 1 0 6 は、「エンコーダ / M U X」ブロック 1 0 7 に渡すことができる。

30

【 0 0 5 1 】

いくつかの実施形態では、空間オーディオパラメータは、方向および非方向（例えば、拡散）パラメータにグループ化されるか、または分離されてもよい。

【 0 0 5 2 】

「エンコーダ / M U X」ブロック 1 0 7 は、トランスポート（例えばダウンミックス）信号 1 0 4 を受信し、これらのオーディオ信号の適切なエンコードを生成するように構成することができる。「エンコーダ / M U X」ブロック 1 0 7 は、ある実施形態では、コンピュータ（メモリ上および少なくとも 1 つのプロセッサ上に記憶された適切なソフトウェアを実行する）、または代替的には、例えば、F P G A または A S I C を利用する特定の装置であり得る。符号化は、任意の適切なスキームを使用して実施することができる。「エンコーダ / M U X」ブロック 1 0 7 は、さらに、メタデータを受信し、情報の符号化または圧縮された形態を生成するように構成されてもよい。ある実施形態では、「エンコーダ / M U X」ブロック 1 0 7 は、図 1 に示す伝送または記憶の前に、単一データストリーム 1 1 1 にインターリーブ、多重化、または符号化ダウンミックス信号内にメタデータを破線によって埋め込むことができる。多重化は、任意の適切なスキームを使用して実施することができる。

40

【 0 0 5 3 】

50

デコーダ側では、受信または検索されたデータ（ストリーム）は、「デマルチプレクサ／デコーダ／シンセサイザ」１３３によって受信されてもよい。「デマルチプレクサ／デコーダ／シンセサイザ」１３３は、符号化されたストリームをデマルチプレクスし、オーディオ信号をデコードして、トランスポート信号を得ることができる。同様に、「デマルチプレクサ／デコーダ／シンセサイザ」１３３は、符号化されたメタデータを受信し、復号するように構成してもよい。一部の実施形態では、「デマルチプレクサ／デコーダ／シンセサイザ」１３３は、コンピュータ（メモリ上および少なくとも１つのプロセッサ上に記憶された適当なソフトウェアを実行する）、または代替的に、例えば、FPGAまたはASICを利用する特定の装置であり得る。

【００５４】

システム１００の「デマルチプレクサ／デコーダ／シンセサイザ」部分１３３は、さらに、トランスポート信号およびメタデータに基づいて、任意の適切なフォーマットで、マルチチャンネル信号１１０の形態の合成空間オーディオを再作成するように構成されてもよい（これらは、マルチチャンネルラウドスピーカフォーマットであり得るし、ある実施形態では、使用ケースに応じて、ヘッドフォンリスニング用のバイノーラル信号またはアンビソニック（Ambisonics）信号のような任意の適切な出力フォーマットであり得る）。

【００５５】

したがって、概要の最初に、システム（解析パート）はマルチチャンネルオーディオ信号を受信するように設定されている。

【００５６】

次に、システム（解析パート）は適切な搬送オーディオ信号を生成するように設定される（たとえば、オーディオ信号チャンネルの一部を選択またはダウンミックスすることで）。

【００５７】

次に、システムは、トランスポート信号およびメタデータを記憶／伝達するために符号化するように構成される。

【００５８】

この後、システムはエンコードされたトランスポートとメタデータを保存／送信することができる。

【００５９】

システムは、符号化された搬送及びメタデータを検索／受信することができる。

【００６０】

次に、システムは、符号化搬送およびメタデータ・パラメータから搬送およびメタデータを抽出し、例えば逆多重化し、符号化搬送およびメタデータ・パラメータを復号するように構成される。

【００６１】

システム（合成部）は、抽出された搬送音声信号とメタデータに基づいて、出力マルチチャンネル音声信号を合成するように構成されている。デコーダ（合成部分）に関しては、空間メタデータを受信し、例えば、５．１信号のダウンミックス、モバイルデバイスからの２つの間隔のあるマイクロホン信号、または一致するマイクロホンアレイからの２つのビームパターンであり得る（潜在的には前処理されたバージョンの）オーディオ信号を転送するように構成される。

【００６２】

デコーダは、空間メタデータおよび搬送オーディオ信号から空間オーディオ（アンビソニックなど）をレンダリングするように構成されてもよい。これは、典型的には、このような入力から空間オーディオをレンダリングするために、線形およびパラメトリックレンダリングの２つのアプローチのうちの１つを採用することによって達成される。

【００６３】

周波数帯域での処理を仮定すると、線形レンダリングは、所望の出力を生成するためにいくつかの静的混合重量を利用することを言う。パラメトリックレンダリングとは、空間メ

10

20

30

40

50

タデータに基づいて搬送オーディオ信号を変更し、目的の出力を生成することである。

【0064】

様々な入力からアンビソニックを生成する方法が提示されている。

【0065】

5.1. 信号からの搬送オーディオ信号と空間メタデータの場合、パラメトリック処理を使用してアンビソニックをレンダリングできる。

【0066】

オーディオ信号や空間的なメタデータを離れたマイクから搬送する場合は、リニア処理とパラメトリック処理の組み合わせを使用することもできる。

【0067】

同時マイクからの搬送音声信号と空間メタデータの場合、線形処理とパラメトリック処理の組合せが使用できる。

【0068】

したがって、様々な種類の入力からアンビソニックをレンダリングするための様々な方法がある。しかし、一定アンビソニックレンダリング方法はすべて、ある種の入力を想定している。以下に説明するいくつかの実施形態は、以下のような問題の発生を防止する装置および方法を示す。

【0069】

線形レンダリングを使用すると、アンビソニックの左向きの1次(8桁)信号であるY信号を、 $Y(f) = S_0(f) - S_1(f)$ により2つの一致する反対のカーディオイドから作成できる。ここで、 f は周波数である。別の例として、Y信号は、 $Y(f) = -i(S_0(f) - S_1(f))g_{eq}(f)$ により作成することができる。ここで、 $g_{eq}(f)$ は、(マイクロホンの距離に依存する)周波数依存イコライザであり、 i 虚数単位である。離間して配置されたマイク(-90度の位相シフトと周波数依存イコライゼーションを含む)の処理は、一致するマイクの処理とは異なり、間違った処理技術を使用すると音質が劣化する可能性がある。

【0070】

一部のレンダリングスキームでパラメトリックレンダリングを使用するには、線形平均を使用して「プロトタイプ」信号を生成する必要がある。これらのプロトタイプ信号は、次に、空間メタデータに基づいて時間周波数領域で適応的に修正される。最適には、プロトタイプ信号はターゲット信号にできるだけ追従する必要がある。これにより、パラメトリック処理の必要性が最小限に抑えられ、したがってパラメトリック処理による潜在的なアーチファクトが最小限に抑えられる。たとえば、プロトタイプ信号には、対応する出力チャンネルに関連するすべての信号成分が十分な範囲で含まれている必要がある。

【0071】

一例として、無指向性信号Wがレンダリングされると(同様の効果が他のアンビソニック信号にも存在する)、プロトタイプは、例えば2つの簡単なアプローチで、ステレオ搬送オーディオ信号から作成することができる。1つのチャンネル(左チャンネルなど)、または、2つのチャンネルの合計を選択する。

【0072】

どちらを選択するかは、搬送オーディオ信号のタイプに大きく依存する。搬送信号が5.1信号から発生する場合、通常、左側の信号は左搬送オーディオ信号のみで、右側の信号は右搬送オーディオ信号のみである(一般的なダウンミックスマトリックスを使用する場合)。したがって、プロトタイプに1つのチャンネルを使用すると、もう1つのチャンネルの信号内容が失われ、明確なアーチファクトが生成される(たとえば、ワーストケースでは、選択された1つのチャンネルに信号がまったく存在しない)。したがって、この場合、Wプロトタイプは、両方のチャンネルの合計として定式化する方が良かった。一方、搬送信号が離れたマイクから発生する場合、W信号のプロトタイプとして搬送オーディオ信号の合計を使用すると、厳しいコムフィルタリングが発生する(信号間に時間遅延があるため)。これにより、上記と同様のアーチファクトが発生する。この場合、少なくとも高い周

10

20

30

40

50

波数範囲で、2つのチャンネルのうちの1つのみをWプロトタイプとして選択した方が良い。

【0073】

したがって、すべての搬送オーディオ信号タイプに適合する適切な選択肢はない。

【0074】

したがって、リニア法とパラメトリック法の両方を用いて、ある搬送オーディオ信号タイプ用に設計された空間オーディオ処理を別の搬送オーディオ信号タイプに適用することは、オーディオ品質の明確な劣化を生み出すことが期待される。

【0075】

以下の実施形態および実施例に関してさらに詳細に論じられるような概念は、デコーダがエンコーダから少なくとも2つの搬送音声信号を受信する場合の音声符号化および復号化に関する。さらに、実施形態は、搬送オーディオ信号が、少なくとも2つのタイプ、例えば、5.1信号のダウンミックス、間隔を置いたマイクロホン信号、または一致するマイクロホン信号であり得る。さらに、いくつかの実施形態では、装置および方法は、搬送オーディオ信号の処理の品質を改善し、決定された出力（例えば、アンビソニック、5.1、モノ）を提供するための解決策を実装する。搬送オーディオ信号のタイプを決定し、決定された搬送オーディオ信号のタイプに基づいてオーディオの処理を実行することにより、品質を改善することができる。

10

【0076】

本明細書でさらに詳細に論じられるいくつかの実施形態では、搬送オーディオ信号タイプは、搬送オーディオ信号の種類を示すメタデータの取得、または搬送オーディオ信号（および利用可能な場合は空間メタデータ）自体に基づいた搬送オーディオ信号の種類の決定のいずれかによって決定される。

20

【0077】

搬送オーディオ信号タイプを記述するメタデータは、例えば、間隔のあるマイク（マイクの位置に付随する場合もある）、一致するマイクまたは連桁は、一致するマイク（マイクの方向パターンを伴う可能性がある）と実質的に似ている、マルチチャンネルオーディオ信号（5.1など）からのダウンミックス、の条件を含むことができる。

【0078】

搬送オーディオ信号自体の解析に基づく搬送オーディオ信号タイプの判定は、（異なる方法で）結合する周波数帯またはスペクトル効果を、期待されるスペクトル効果（利用可能な場合は空間メタデータに部分的に基づいて）と比較することに基づくことができる。

30

【0079】

さらに、いくつかの実施形態においては、オーディオ信号のプロセッシングは、アンビソニック（Ambisonic）信号のレンダリング、マルチチャンネルオーディオ信号（5.1など）のレンダリング、およびオーディオ信号のより少ない数へのダウンミックスのトランスポートを含むことができる：

【0080】

図2は、いくつかの実施形態を実施するのに適したデコーダ例の概要図を示す。この実施形態は、例えば、「デマルチプレクサ/デコーダ/シンセサイザ」ブロック133内で実現することができる。この例では、入力2つのオーディオチャンネルと空間メタデータを含むメタデータ支援空間オーディオ（MASA）ストリームである。しかしながら、本明細書で論じるように、入力フォーマットは、任意の適切なメタデータ支援空間オーディオフォーマットであり得る。

40

【0081】

（MASA）ビットストリームは、搬送オーディオ信号タイプ決定器201に転送される。搬送オーディオ信号タイプ決定器201は、ビットストリームに基づいて搬送オーディオ信号タイプ202、および場合によってはいくつかの追加パラメータ204（マイクロホン距離など）を決定するように構成される。決定されたパラメータは、MASA-アンビソニック信号変換器203に転送される。

50

【 0 0 8 2 】

M A S A - アンビソニック信号変換器 2 0 3 は、ビットストリームおよび搬送オーディオ信号タイプ 2 0 2（および場合によってはいくつかの追加パラメータ 2 0 4）を受信するように構成され、決定された搬送オーディオ信号タイプ 2 0 2（および可能な追加パラメータ 2 0 4）に基づいて、M A S A ストリームをアンビソニック信号に変換するように構成される。

【 0 0 8 3 】

例の動作は、図 3 に示すフロー・ダイアグラムに要約される。

【 0 0 8 4 】

最初の動作は、ステップ 3 0 1 によって図 3 に示すように、ビットストリーム（M A S A ストリーム）を受信または取得することの 1 つである。

10

【 0 0 8 5 】

次の動作は、ステップ 3 0 3 によって図 3 に示されるように、ビットストリームに基づいて搬送オーディオ信号タイプを決定する（そして、タイプ信号またはインジケータおよび可能な他の追加パラメータを生成する）1 つである。

【 0 0 8 6 】

搬送オーディオ信号タイプを決定した次の動作は、ステップ 3 0 5 によって図 3 に示されるように、決定された搬送オーディオ信号タイプに基づいて、ビットストリーム（M A S A ストリーム）をアンビソニック信号に変換することである。

【 0 0 8 7 】

20

図 4 は、一例の搬送オーディオ信号型判定器 2 0 1 の概要図を示す。この例では、搬送オーディオ信号タイプ決定子の例が、搬送オーディオ信号タイプが M A S A ストリームで使用可能な場合に適している。

【 0 0 8 8 】

この例における搬送オーディオ信号タイプ決定器 2 0 1 の例は、搬送オーディオ信号タイプ抽出器 4 0 1 を含む。搬送オーディオ信号タイプ抽出器 4 0 1 は、ビット（M A S A）ストリームを受信し、M A S A ストリームからタイプインジケータを抽出（すなわち、読み出しおよび／またはデコード）するように構成される。この種の情報は、例えば、M A S A ストリームの「チャンネルオーディオフォーマット」フィールドで利用可能である。加えて、追加のパラメータが利用可能であれば、それらも抽出される。この情報は、搬送オーディオ信号タイプ抽出器 4 0 1 から出力される。ある実施形態では、搬送オーディオ信号タイプは、「スペース」、「ダウンミックス」、「一致」を含むことができる。いくつかの他の実施形態では、搬送オーディオ信号タイプは、任意の適切な値を含むことができる。

30

【 0 0 8 9 】

図 5 は、さらなる例としての搬送オーディオ信号タイプ判定器 2 0 1 の概要図を示す。この例では、搬送オーディオ信号タイプを M A S A ストリームから直接抽出またはデコードすることはできない。この例では、M A S A ストリームの解析から搬送オーディオ信号タイプを推定または決定する。いくつかの実施形態におけるこの判定は、異なる搬送オーディオ信号タイプのあるスペクトル効果を明らかにする一組の推定器／エネルギー比較を使用することに基づいている。

40

【 0 0 9 0 】

ある実施形態では、搬送オーディオ信号タイプ決定器 2 0 1 は、搬送オーディオ信号および空間メタデータ抽出器／デコード 5 0 1 を含む。搬送オーディオ信号および空間メタデータ抽出器／復号器 5 0 1 は、M A S A ストリームを受信し、搬送オーディオ信号および空間メタデータを M A S A ストリームから抽出および／または復号するように構成される。得られた搬送オーディオ信号 5 0 2 は、時間／周波数変換器 5 0 3 に転送することができる。得られた空間メタデータ 5 2 2 は、さらに、ターゲット・エネルギーコンパレータ 5 1 1 への減算に転送することができる。

【 0 0 9 1 】

50

いくつかの実施形態では、搬送オーディオ信号タイプ判定器 201 は、時間 / 周波数変換器 503 を含む。時間 / 周波数変換器 503 は、搬送オーディオ信号 502 を受信し、それらを時間 - 周波数領域に変換するように構成される。適切な変換は、例えば、短時間フーリエ変換 (STFT) および錯体変調直交ミラーフィルタバンク (QMF) を含む。結果の信号は、 $S_i(b, n)$ のように表される。ここで、 i は、チャンネル・インデックス、 b は、周波数ビン・インデックス、および、 n は、タイムインデックスである。搬送オーディオ信号 (抽出器および / またはデコーダからの出力) がすでに時間周波数領域にある状況では、これは省略されてもよいし、あるいはある時間周波数領域表現から別の時間周波数領域表現への変換を含むことができる。T / F ドメイン搬送オーディオ信号 504 は、コンパレータに転送することができる。

10

【0092】

ある実施形態では、搬送オーディオ信号タイプ決定器 201 は、ブロードバンド L / R トータルエネルギー比較器 505 を含む。ブロードバンド L / R 対トータルエネルギーコンパレータ 505 は、T / F ドメイン搬送オーディオ信号 504 を受信し、トータル比率パラメータに対してブロードバンド L / R を出力するように構成される。

【0093】

ブロードバンド L / R からトータルエネルギーコンパレータ 505 内で、ブロードバンドの左、右、およびトータルエネルギーが計算される。

【数 1】

$$E_{\text{left,bb}}(n) = \sum_{b=0}^{B-1} |S_0(b, n)|^2;$$

$$E_{\text{right,bb}}(n) = \sum_{b=0}^{B-1} |S_1(b, n)|^2;$$

$$E_{\text{total,bb}}(n) = E_{\text{left,bb}}(n) + E_{\text{right,bb}}(n),$$

20

ここで、 B は、周波数ビンの数である。これらのエネルギーは、例えば、

【数 2】

$$E'_{x,bb}(n) = a_1 E_{x,bb}(n) + b_1 E'_{x,bb}(n-1),$$

30

により平滑化される。ここで、 a_1 および b_1 は平滑化係数である (例えば、 $a_1 = 0.01$ および $b_1 = 1 - a_1$)。次いで、トータルエネルギーコンパレータ 505 に対するブロードバンド L / R は、最小の左および右のエネルギーを選択し、スケーリングするように構成される。

40

【数 3】

$$E'_{lr,bb}(n) = 2 \min(E'_{\text{left,bb}}(n), E'_{\text{right,bb}}(n)).$$

ここで、乗算器 2 は、2 つのチャンネルの合計である $E'_{(\text{total}, \text{bb})}(n)$ に関するエネルギーを正規化するものである。

【0094】

次いで、ブロードバンド L / R 対トータルエネルギーコンパレータ 505 は、次のように

50

して、ブロードバンド L / R 対トータルエネルギー割合 506 を生成することができる。

【数 4】

$$\beta(n) = 10 \log_{10} \frac{E'_{lr,bb}(n)}{E'_{total,bb}(n)},$$

これは、次に、比率 506 として出力される。

【0095】

いくつかの実施形態では、搬送オーディオ信号タイプ判定器 201 は、高周波数 L / R - トータルエネルギー比較器 507 を含む。高周波数 L / R - トータルエネルギー比較器 507 は、T / F ドメイン搬送オーディオ信号 504 を受信し、高周波数 L / R - トータル比率パラメータを出力するように構成される。

【0096】

ブロードバンド L / R - トータルエネルギー比較器 507 内では、高周波数帯域の左、右、およびトータルエネルギーが計算される。

【数 5】

$$E_{left,hi}(n) = \sum_{b=B_1}^{B-1} |S_0(b,n)|^2;$$

$$E_{right,hi}(n) = \sum_{b=B_1}^{B-1} |S_1(b,n)|^2;$$

$$E_{total,hi}(n) = E_{left,hi}(n) + E_{right,hi}(n),,$$

ここで、 B_1 は、高周波領域が開始するように定義された第 1 のビンである（その値は適用された T / F 変換に依存し、それは、例えば、6 kHz に対応することができる）。これらのエネルギーは、例えば、

【数 6】

$$E'_{x,hi}(n) = a_2 E_{x,hi}(n) + b_2 E'_{x,hi}(n-1),$$

により平滑化される。ここで、 a_2 および b_2 は平滑化係数である。エネルギー差は、高周波数においてより速いペースで生じ得るので、平滑化係数は、より少ない平滑化を提供するように設定され得る（例えば、 $a_2 = 0.1$ および $b_2 = 1 - a_2$ ）。

【0097】

次いで、高周波 L / R 対トータルエネルギー比較器 507 は、左右のエネルギーから小さい方を選択するように構成することができ、その結果は、2 で乗算される。

【数 7】

$$E'_{lr,hi}(n) = 2 \min(E'_{left,hi}(n), E'_{right,hi}(n))$$

10

20

30

40

50

【 0 0 9 8 】

次いで、高周波 L / R 対トータルエネルギー比較器 5 0 7 は、次いで、高周波 L / R 対トータル比率 5 0 8 を生成することができる。

【 数 8 】

$$\eta(n) = 10 \log_{10} \frac{E'_{lr,hi}(n)}{E'_{total,hi}(n)}$$

として出力される。

10

【 0 0 9 9 】

いくつかの実施形態では、搬送オーディオ信号タイプ判定器 2 0 1 は、トータルエネルギー比較器 5 0 9 を含む。合計対トータルエネルギー比較器 5 0 9 に対する総和は、T / F ドメイン搬送オーディオ信号 5 0 4 を受信し、トータルエネルギー比率パラメータに対する総和を出力するように構成される。合計対トータルエネルギー比較器 5 0 9 への和は、いくつかの周波数において、2 つのチャンネルが、位相がずれている状況を検出するように構成され、この状況は、特に、間隔を置いたマイクロホン録音に対して典型的な現象である。

【 0 1 0 0 】

合計対トータルエネルギー比較器 5 0 9 への総和は、総信号のエネルギーと、各周波数ビンに対する総エネルギーとを計算するように構成される。

20

【 数 9 】

$$E_{\text{sum}}(b, n) = |S_0(b, n) + S_1(b, n)|^2;$$

$$E_{\text{total}}(b, n) = |S_0(b, n)|^2 + |S_1(b, n)|^2.$$

【 0 1 0 1 】

これらのエネルギーは、例えば、

30

【 数 1 0 】

$$E'_x(b, n) = a_3 E_x(b, n) + b_3 E'_x(b, n - 1),$$

により平滑化される。ここで、 a_3 および b_3 は平滑化係数である（例えば、 $a_3 = 0.01$ および $b_3 = 1 - a_3$ ）。

【 0 1 0 2 】

次いで、合計対トータルエネルギー比較器 5 0 9 は、最小合計対トータル比率 5 1 0 を以下のように計算するように構成される。

40

【 数 1 1 】

$$\chi(n) = 10 \log_{10} \min_b \frac{E'_{\text{sum}}(b, n)}{E'_{\text{total}}(b, n)}, 0 \leq b \leq B_2,$$

ここで、 B_2 は、この計算が実行される周波数領域の最も高いビンである（この値は、使用される T / F 変換に依存する。例えば、10 kHz に対応する場合がある）。

50

【 0 1 0 3 】

次いで、合計対トータルエネルギー比較器 5 0 9 への合計は、比率 (n) 5 1 0 を出力するように構成される。

【 0 1 0 4 】

いくつかの実施形態では、搬送オーディオ信号タイプ判定器 2 0 1 は、ターゲット・エネルギー比較器 5 1 1 への減算を含む。ターゲット・エネルギーコンパレータ 5 1 1 への減算は、T / F ドメイン搬送オーディオ信号 5 0 4 および空間メタデータ 5 2 2 を受信し、ターゲット・エネルギー比率パラメータ 5 1 2 への減算を出力するように構成される。

【 0 1 0 5 】

ターゲット・エネルギー比較器 5 1 1 への減算は、左右のチャンネルの差のエネルギーを計算するように構成される。

10

【 数 1 2 】

$$E_{\text{sub}}(b, n) = |S_0(b, n) - S_1(b, n)|^2.$$

【 0 1 0 6 】

これは、少なくともいくつかの入力信号タイプの場合、アンビソニックの Y 信号の「プロトタイプ」と考えることができる (Y 信号は、ダイポールの方向パターンを持ち、左側に正のローブ、右側に負のローブがある) 。

20

【 0 1 0 7 】

次いで、ターゲット・エネルギー比較器 5 1 1 への減算は、Y 信号に対するターゲット・エネルギー $E_{\text{target}}(b, n)$ を計算するように構成することができる。これは、空間メタデータに基づいて、トータルエネルギーが球面調和間でどのように分散されるべきかを推定することに基づいている。例えば、いくつかの実施形態では、ターゲット・エネルギー比較器 5 1 1 への減算は、空間メタデータおよびエネルギー推定値に基づいて目標共分散行列 (チャンネルエネルギーおよび相互相関) を構築するように構成される。しかし、一部の実施形態では、Y 信号のエネルギーのみが推定され、これは目標共分散行列の 1 つのエントリである。したがって、Y のターゲット・エネルギー $E_{\text{target}}(b, n)$ は、2 つの部分から構成される。

30

【 数 1 3 】

$$E_{\text{target}}(b, n) = E_{\text{target,amb}}(b, n) + E_{\text{target,dir}}(b, n),$$

ここで、 $E_{\text{target,amb}}(b, n)$ は、

【 数 1 4 】

40

$$E_{\text{target,amb}}(b, n) = \frac{E_{\text{total}}(b, n)}{3} (1 - r(b, n)) (1 - c_{\text{sur}}(b, n)),$$

で定義される、ターゲット・エネルギーのアンビエンス / 無指向部分である。ここで、は空間メタデータの 0 と 1 の間の直接対トータルエネルギー比率パラメータで、 $c_{\text{sur}}(b, n)$ は、空間メタデータの 0 と 1 の間のサラウンドコヒーレンス・パラメータである (その場合、陽性と負のローブは互いにキャンセルされるため、サラウンドコヒーレント音は Y ダイポールではキャプチャされない) 。 3 による除算は、アンビソニック出力に対して S N 3 D 正規化スキームを仮定しているためであり、Y 成分のアンビエンスエネルギー

50

(a m b i e n c e e n e r g y) は、その場合全オムニエネルギー (o m i n i - e n e r g y) の 3 分の 1 である。

【 0 1 0 8 】

空間メタデータは、パラメータがいくつかの周波数または時間指標に対して同じであり得るように、 b , n 毎よりも低い周波数および / または時間分解能であり得ることに留意されたい。

【 0 1 0 9 】

この $E (t a r g e t , d i r) (b , n)$ は、より指向性の高い部分のエネルギーである。それを定式化するには、空間メタデータのスプレッドコヒーレンス $c_{s p r e a d} (b , n)$ パラメータ 0 ~ 1 の機能としてのスプレッドコヒーレンス分布ベクトルを、

【 数 1 5 】

$$\mathbf{v}_{DISTR,3}(b,n) = \begin{cases} \begin{bmatrix} (2 - 2c_{spread}(b,n)) \\ 1 \\ 1 \end{bmatrix} \frac{1}{\sqrt{(2 - 2c_{spread}(b,n))^2 + 2}} , & \text{when } c_{spread}(b,n) > 0.5 \\ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} (1 - c_{spread}(b,n)) + \begin{bmatrix} \sqrt{1/3} \\ \sqrt{1/3} \\ \sqrt{1/3} \end{bmatrix} c_{spread}(b,n) , & \text{otherwise.} \end{cases} \quad 20$$

のように定義する必要がある。

【 0 1 1 0 】

ターゲット・エネルギー比較器 5 1 1 への減算は、方位角値のベクトル、

【 数 1 6 】

$$\boldsymbol{\theta}(b,n) = \theta(b,n) + \begin{bmatrix} 0 \\ \pi/6 \\ -\pi/6 \end{bmatrix}, \quad 30$$

を決定するように構成することもできる。ここで、 (b, n) は、ラジアン単位の空間メタデータの方位値である。ベクトルエントリベースの $\sin()$ 動作を仮定すると、直接部分ターゲット・エネルギーは、

【 数 1 7 】

$$E_{target,dir}(b,n) = \sin(\boldsymbol{\theta}(b,n))^T \mathbf{v}_{DISTR,3}(b,n) E_{total}(b,n) r(b,n). \quad 40$$

のようになる。

【 0 1 1 1 】

したがって、 $E_{target}(b, n)$ が得られる。これらのエネルギーは、いくつかの実施形態において、例えば、

【 数 1 8 】

$$E'_x(b, n) = a_4 E_x(b, n) + b_4 E'_x(b, n - 1)$$

平滑化されることが可能である。ここで、 a_4 と b_4 とは平滑化係数である（例えば、 $a_4 = 0.0004$ および $b_4 = 1 - a_4$ ）。

【0112】

さらに、ターゲット・エネルギー比較器 511 への減算は、次のように最低周波数ビンにおけるエネルギーを使用してターゲット比率 512 への減算を計算するように構成される。

10

【数19】

$$v(n) = 10 \log_{10} \frac{E'_{\text{sub}}(0, n)}{E'_{\text{target}}(0, n)}$$

これは、出力である。

【0113】

ある実施形態では、搬送オーディオ信号タイプ決定器 201 は、搬送オーディオ信号タイプ（推定メトリックに基づく）決定器 513 を含む。搬送オーディオ信号タイプ決定器 513 は、トータル比率 506 に対するブロードバンド L/R、トータル比率 508 に対する高周波数 L/R、トータル比率 510 に対する分合計、およびターゲット比率 512 に対する減算を受信し、これらの推定されたメトリックに基づいて搬送オーディオ信号タイプを決定するように構成される。

20

【0114】

決定は様々な方法で行うことができ、実際の実装は、使用される T/F 変換のように、多くの側面で異なる可能性がある。限定的でない形式の一例は、搬送オーディオ信号タイプ（推定されたメトリックに基づく）決定器 513 が、まず、非メトリックへの変更を計算することである。

30

【数20】

$$E_s(n) = \frac{(-v(n) - 3)}{3} + \frac{\max(-\chi(n), 0)}{6} + \frac{\beta(n)}{6}, \quad \text{if } v(n) < -3, \text{ else } E_s(n) = 0$$

【0115】

搬送オーディオ信号タイプ（推定メトリックに基づく）決定器 513 は、次に、ダウンミックスメトリックへの変化を計算するように構成することができる。

40

【数21】

$$E_{d1}(n) = \frac{v(n)}{3} + \frac{\chi(n) + 1}{6} + \frac{-\beta(n)}{6}, \quad \text{if } v(n) > 0, \text{ else } E_{d1}(n) = 0$$

$$E_{d2}(n) = \frac{(v(n) + 4)}{3} + \frac{\chi(n)}{6} + \frac{(-\eta(n) - 12)}{3}, \quad \text{if } \eta(n) < -12, \text{ else } E_{d2}(n) = 0.$$

50

【 0 1 1 6 】

搬送オーディオ信号タイプ（推定メトリクスに基づく）決定器 5 1 3 は、次いで、これらのメトリクスに基づいて、搬送オーディオ信号が、間隔を置いたマイクロホンから発生するか、またはサラウンドサウンド信号（5 . 1 など）からのダウンミックスであるかを決定することができる。例えば、

【 数 2 2 】

$$\begin{aligned} & \text{if } \Xi_s(n) > 1, \quad T(n) = \text{"spaced"} \\ & \text{else if } \Xi_{d1}(n) > 1 \vee \Xi_{d2}(n) > 1, \quad T(n) = \text{"downmix"} \\ & \quad \text{else,} \quad T(n) = T(n-1) \end{aligned}$$

10

である。

【 0 1 1 7 】

この例では、搬送オーディオ信号タイプ（推定されたメトリックに基づく）決定器 5 1 3 は、一致するマイクロホンタイプを検出しない。しかしながら、実際には、 $T(n) = \text{"downmix"}$ タイプに従った処理は、一般に、一致したキャプチャの場合（例えば、左右に向けられたカーディオイドを用いた場合）、良好なオーディオを生成することができる。

20

【 0 1 1 8 】

搬送オーディオ信号タイプ（推定メトリックに基づく）決定器 5 1 3 は、次に搬送オーディオ信号タイプを搬送オーディオ信号タイプ 2 0 2 として出力するように構成することができる。いくつかの実施形態では、他のパラメータ 2 0 4 が出力されてもよい。

【 0 1 1 9 】

図 6 は、図 5 に示される装置の動作を要約するものであり、したがって、いくつかの実施形態では、第 1 の動作は、ステップ 6 0 1 によって図 6 に示されるように、M A S A ストリーム（またはビットストリーム）から搬送オーディオ信号およびメタデータを抽出および/または復号する動作である。

【 0 1 2 0 】

次の動作は、ステップ 6 0 3 によって図 6 に示すように、搬送オーディオ信号を時間 - 周波数領域変換することができる。

30

【 0 1 2 1 】

次に、一連の比較を行うことができる。例えば、ブロードバンド L / R エネルギーをトータルエネルギー値と比較することによって、ステップ 6 0 5 によって図 6 に示すようにブロードバンド L / R 対トータルエネルギー比率を生成することができる。

【 0 1 2 2 】

例えば、高周波数 L / R エネルギーをトータルエネルギー値と比較することによって、ステップ 6 0 7 によって、図 6 に示すように、高周波数 L / R 対トータルエネルギー比率を生成することができる。

40

【 0 1 2 3 】

合計エネルギーをトータルエネルギー値と比較することによって、合計対トータルエネルギー比率は、図 6 に示すように、ステップ 6 0 9 によって生成されてもよい。

【 0 1 2 4 】

さらに、ステップ 6 1 1 によって、図 6 に示されるように、減算対ターゲット・エネルギー比率が生成されてもよい。

【 0 1 2 5 】

これらのメトリックを決定した後、本方法は、ステップ 6 1 3 によって図 6 に示すように、これらのメトリック比率を解析することによって、搬送オーディオ信号タイプを決定することができる。

50

【 0 1 2 6 】

図 7 は、M A S A からアンビソニックへの変換器 2 0 3 の例をさらに詳細に示す。M A S A 対アンビソニック変換器 2 0 3 は、M A S A ストリーム（ビットストリーム）および搬送オーディオ信号タイプ 2 0 2 および可能な追加パラメータ 2 0 4 を受信するように構成され、決定された搬送オーディオ信号タイプに基づいて M A S A ストリームをアンビソニック信号に変換するように構成される。

【 0 1 2 7 】

M A S A 対アンビソニック変換器 2 0 3 は、搬送オーディオ信号および空間メタデータ抽出器 / デコーダ 5 0 1 を含む。これは、図 5 に示すように、搬送オーディオ信号タイプ決定器内に見られるのと同じ方法で、M A S A ストリームを受信し、搬送オーディオ信号 5 0 2 および空間メタデータ 5 2 2 を出力するように構成される。いくつかの実施形態では、抽出部材 / 復号器 5 0 1 は、搬送音声信号タイプ判定器からの抽出部材 / 復号器である。得られた搬送オーディオ信号 5 0 2 は、時間 / 周波数変換器 5 0 3 に転送することができる。得られた空間メタデータ 5 2 2 は、さらに、信号ミキサー 7 0 5 に転送することができる。

【 0 1 2 8 】

ある実施形態では、M A S A 対アンビソニックコンバータ 2 0 3 は、時間 / 周波数変換器 5 0 3 を含む。時間 / 周波数変換器 5 0 3 は、搬送オーディオ信号 5 0 2 を受信し、それらを時間 - 周波数領域に変換するように構成される。適切な変換は、例えば、短時間フーリエ変換 (S T F T) および錯体変調直交ミラーフィルタバンク (Q M F) を含む。結果の信号は、 $S_i(b, n)$ のように表される。ここで、 i は、チャンネル・インデックス、 b は、周波数ビン・インデックス、および、 n は時間インデックスである。オーディオ抽出および / または復号化の出力がすでに時間周波数領域にある場合、このブロックは省略されてもよいし、あるいはある時間周波数領域表現から別の時間周波数領域表現への変換を含むことができる。T / F ドメイン搬送オーディオ信号 5 0 4 は、プロトタイプ信号クリエータ 7 0 1 に転送することができる。いくつかの実施形態では、時間 / 周波数変換器 5 0 3 は、搬送音声信号タイプ判定器からの同一時間 / 周波数変換器である。

【 0 1 2 9 】

ある実施形態では、M A S A 対アンビソニック変換器 2 0 3 は、プロトタイプ信号クリエータ 7 0 1 を含む。プロトタイプ信号作成器 7 0 1 は、T / F ドメイン搬送オーディオ信号 5 0 4、搬送オーディオ信号タイプ 2 0 2、および可能な追加パラメータ 2 0 4 を受信するように構成される。次いで、T / F プロトタイプ信号 7 0 2 を信号ミキサー 7 0 5 およびデコレレータ (d e c o r r e l a t o r) 7 0 3 に出力することができる。

【 0 1 3 0 】

ある実施形態では、M A S A 対アンビソニック変換器 2 0 3 は、デコレレータ 7 0 3 を含む。デコレレータ 7 0 3 は、T / F プロトタイプ信号 7 0 2 を受信し、デコレレーション（非相関）を適用し、デコレレーション T / F プロトタイプ信号 7 0 4 を信号ミキサー 7 0 5 に出力するように構成される。いくつかの実施形態において、デコレレータ 7 0 3 はオプションである。

【 0 1 3 1 】

ある実施形態では、M A S A 対アンビソニック変換器 2 0 3 は、信号ミキサー 7 0 5 を含む。信号ミキサー 7 0 5 は、T / F プロトタイプ信号 7 0 2 および非相関 T / F プロトタイプ信号および空間メタデータ 5 2 2 を受信するように構成される。

【 0 1 3 2 】

プロトタイプ信号作成器 7 0 1 は、搬送オーディオ信号タイプに基づいて、アンビソニック (F O A / H O A) の球面調和関数の各々についてプロトタイプ信号を生成するように構成される。

【 0 1 3 3 】

いくつかの実施形態では、プロトタイプ信号作成器 7 0 1 は、以下のように動作するように構成される。もし $T(n) = \text{“ s p a c e d ”}$ であれば、W 信号のプロトタイプを、

10

20

30

40

50

【数 2 3】

$$W_{\text{proto}}(b, n) = \frac{S_0(b, n) + S_1(b, n)}{2}, \quad 0 \leq b \leq B_3$$

$$W_{\text{proto}}(b, n) = S_0(b, n), \quad b > B_3$$

のように作成することができれば実際には、低周波数の搬送オーディオ信号の平均として作成することができる。

10

【0 1 3 4】

実際には、 $W_{\text{proto}}(b, n)$ は、低周波のオーディオ信号を搬送する手段として作成することができる。信号の位相は大まかには同相で、コムフィルタリングは行われない。また、高周波数のチャンネルの1つを選択する。 B_3 の値は、T/F変換とマイク間の距離によって異なる。距離が不明な場合は、一部のデフォルト値が使用されることがある（1 kHz に対応する値など）。 $T(n) = \text{"downmix"}$ または $T(n) = \text{"coincident"}$ ならば、W信号のプロトタイプを次のように作成できる。

【数 2 4】

$$W_{\text{proto}}(b, n) = S_0(b, n) + S_1(b, n)$$

20

【0 1 3 5】

オリジナルのオーディオ信号は、通常、これらの信号タイプとの間に大きな遅延がないと仮定できるため、 $W_{\text{proto}}(b, n)$ は、搬送オーディオ信号を合計することによって作成される。

【0 1 3 6】

Yプロトタイプ信号に関して、もし $T(n) = \text{"spaced"}$ ならば、Y信号のプロトタイプを次のように作成することができる。

30

【数 2 5】

$$Y_{\text{proto}}(b, n) = \frac{S_0(b, n) + S_1(b, n)}{2}, \quad 0 \leq b \leq B_4$$

$$Y_{\text{proto}}(b, n) = -i(S_0(b, n) - S_1(b, n))g_{eq}(b), \quad B_4 < b \leq B_5$$

$$Y_{\text{proto}}(b, n) = S_0(b, n), \quad b > B_5$$

【0 1 3 7】

40

中域周波数（ B_4 および B_5 との間）では、トランスポート信号を差し引いて位相を - 90 度ずらし、イコライジングすることでダイポール信号を作ることができる。したがって、特にマイクロホンの距離が分かっているならば、Y信号の良いプロトタイプとしての役目を果たし、したがってイコライズ係数は適切である。低周波と高周波ではこれは実現不可能であり、プロトタイプ信号は無指向性W信号の場合と同様に生成される。

【0 1 3 8】

マイクロホンの距離が正確に分かっている場合、Yプロトタイプは、それらの周波数（つまり、 $Y(b, n) = Y_{\text{proto}}(b, n)$ ）でYのために直接使用されてもよい。マイクの間隔がわからない場合は、 $g_{eq}(b) = 1$ を使用することができる。

【0 1 3 9】

50

いくつかの実施形態における信号ミキサー 705 は、周波数帯域における利得処理を適用して、潜在的利得平滑化を用いて周波数帯域におけるターゲット・エネルギーに周波数帯域における $W_{proto}(b, n)$ のエネルギーを補正することができる。ある周波数帯域における無指向性信号のターゲット・エネルギーは、その周波数帯域における搬送オーディオ信号エネルギーの合計とすることができる。このプロセッシングの結果、無指向性信号 $W(b, n)$ が得られる。

【0140】

$Y_{proto}(b, n)$ をそのまま $Y(b, n)$ に使用できない Y 信号について、周波数が B_4 と B_5 の間にある場合は、適応ゲイン処理を行う。この場合は、上記の無指向性 W の場合と似ている。プロトタイプシグナルは、潜在的に間違っただスペクトルを除いて、すでに Y ダイポールになっている。シグナルミキサーは、周波数帯域でプロトタイプシグナルのゲイン処理を実行する。(さらに、この特定のコンテキストでは、 Y 信号の非相関処理は必要ない)。利得処理は、空間的メタデータ(方向、比率、他のパラメータ)および周波数帯域における全体的な信号エネルギー推定値(例えば、搬送信号エネルギーの合計)を使用して、 Y 成分のエネルギーが周波数帯域内にあるべきものを決定し、次いで、決定されたエネルギーである周波数帯域内のプロトタイプ信号のエネルギーを利得で補正し、次いで、その結果が出力 $Y(b, n)$ となる。

【0141】

前述の $Y(b, n)$ を生成する手順は、現在のコンテキスト $T(n) = \text{"spaced"}$ ではすべての周波数に対して有効ではない。プロトタイプ信号は異なる周波数で異なるため、信号ミキサーとデコレータは、このトランスポート信号タイプを持つ周波数に応じて異なる構成になる。異なる種類のプロトタイプ信号を説明するために、 Y ダイポールの負のゲイン方向(陽性と負のローブを持つ)から音が到着するシナリオを考えることができる。中周波(B_4 および B_5 の間)では、 Y プロトタイプ信号の位相は、到来する音のその方向のためであるはずであるので、 W プロトタイプ信号の位相とは逆である。他の周波数(B_4 以下および B_5 以上)では、プロトタイプ Y 信号の位相は、 W プロトタイプ信号の位相と同じである。適切な相(およびエネルギーと相関)の合成は、次に、それらの周波数における信号ミキサーおよびデコレータによって説明される。

【0142】

波長が大きい低周波数(B_4 以下)では、間隔を置いたマイク(通常は互いに若干近い)で取り込んだオーディオ信号間の位相差は小さくなる。したがって、プロトタイプ信号の作成者は、 SNR の理由により、 B_4 および B_5 間の周波数と同じ方法でプロトタイプ信号を生成するように設定すべきではない。したがって、典型的には、プロトタイプ信号として代わりにチャンネル合計無指向性信号が使用される。波長が小さい高周波数(B_5 以上)では、空間エイリアシングによってビームパターンがひどく歪む(との B_4 および B_5 間の周波数のような方法が使用される場合)。そのため、チャンネル選択の無指向性プロトタイプ信号を使用する方が良い。

【0143】

次に、これらの周波数(B_4 以下または B_5 以上)での信号ミキサーとデコレータの構成について説明する。単純な例では、空間メタデータ・パラメータ・設定は、周波数帯域の方位と比率 r で構成される。利得 $\sin(\quad) \sqrt{r}$ を信号ミキサー内のプロトタイプ信号に適用して Y ダイポール信号を生成し、その結果がコヒーレント部分信号となる。プロトタイプ信号も(デコレータで)非相関化され、非相関化された結果が信号ミキサーで受信される。ここで、係数 $\sqrt{(1-r)} g_{order}$ で乗算され、結果は非相関部分信号になる。ゲイン g_{order} は、公知の $SN3D$ 正規化方式に従った球調和次数での拡散場ゲインである。例えば、第1オーダーの場合(この場合は Y ダイポールの場合は $\sqrt{1/3}$)、第2オーダーの場合は $\sqrt{1/5}$ 、第3の場合は $\sqrt{1/7}$ というようになる。コヒーレント部分信号とインコヒーレント部分信号を加算した。その結果、プロトタイプ信号エネルギーが間違っている可能性があるため、誤ったエネルギーを除いて、合成された Y 信号が得られる。中周波数(B_4 と B_5 と

10

20

30

40

50

の間)の文脈で説明されている周波数帯における同じエネルギー補正手順を適用して、周波数帯におけるエネルギーを所望の目標に補正することができ、出力は信号 $Y(b, n)$ である。

【0144】

X、Z成分や2次以上の成分など、他の球面調和に関して、方位に関するゲイン(及び他の潜在的パラメータ)がどの球面調和信号が合成されているかに依存することを除いて、上述の手順を適用することができる。例えば、WプロトタイプからXダイポールコヒーレント部分に対して生成するゲインは $\cos(\theta) \sqrt{r}$ である。非相関、割合-処理、エネルギー補正は、 B_4 と B_5 との間の周波数以外のY成分に対して上記で決定されたものと同じにすることができる。

10

【0145】

高度、スプレッドコヒーレンス、サラウンドコヒーレンスなどのその他のパラメータは、上記の手順で考慮できる。スプレッドコヒーレンス・パラメータには、0~1の値を指定できる。コヒーレンス拡散値0は点音源を示す。言い換えれば、マルチラウドスピーカシステムを使用してオーディオ信号を再生する場合、サウンドはできるだけ少ないラウドスピーカ(例えば、方向が中央の場合は中央のラウドスピーカのみ)で再生する必要がある。拡散コヒーレンスの値が増加するにつれて、値0.5になるまで、センターラウドスピーカの周囲の他のラウドスピーカにより多くのエネルギーが拡散され、エネルギーはセンターと隣接するラウドスピーカの間に均等に拡散される。拡散コヒーレンスの値が0.5以上に増加すると、センターラウドスピーカのエネルギーは値1になるまで減少し、センターラウドスピーカにはエネルギーはなく、エネルギーはすべて近隣のラウドスピーカにある。周囲のコヒーレンス・パラメータの値は0~1である。値が1の場合、すべての(またはほぼすべての)ラウドスピーカチャンネル間にコヒーレンスがあることを意味する。値が0の場合、すべての(またはほぼすべての)ラウドスピーカチャンネル間にコヒーレンスがないことを意味する。これについては、GB出願第1718341.9、加えて、PCT出願PCT/FI2018/050788でさらに説明されている。

20

【0146】

例えば、増加したサラウンドコヒーレンスは、球面調和成分における合成アンビエンスエネルギーの減少によって実施することができ、エレベーションは、コヒーレント部分の生成におけるアンビソニックパターンの定義にしたがってエレベーション関連利得を加えることによって追加することができる。

30

【0147】

$T(n) = \text{"downmix"}$ または $T(n) = \text{"coincident"}$ であれば、Y信号のプロトタイプを、

【数26】

$$Y_{\text{proto}}(b, n) = S_0(b, n) - S_1(b, n).$$

40

のように作成できる。

【0148】

この状況では、オリジナルのオーディオ信号は通常これらの信号タイプとの間に有意な遅延を持たないと仮定できるため、位相シフトの必要はない。「混合信号」ブロックに関して、 $T(n) = \text{"coincident"}$ の場合、YとWのプロトタイプは、(実際の方向性パターンに応じて)場合によってはゲイニング後に、YとWの出力に直接使用されることがある。 $T(n) = \text{"downmix"}$ の場合、 $y_{\text{proto}}(b, n)$ と $w_{\text{proto}}(b, n)$ は、 $Y(b, n)$ と $W(b, n)$ に直接使用することはできない。ただし、 $T(n) = \text{"spaced"}$ の場合に決定された望ましいターゲットへの周波数帯でのエネルギー補正が必要な場合がある(無指向性成分は到来する音の角度に関わらず、空間ゲイン1と

50

なることに留意する)。

【0149】

他の球面調和関数(XやZなど)では、対象信号をうまく再現するプロトタイプを作成することはできない。典型的なダウンミックス信号は、フロントバックX軸やトップボトムZ軸ではなく、左右軸に向いているからである。したがって、いくつかの実施形態では、アプローチは、例えば、無指向性(omnidirectional)信号のプロトタイプを利用することである。

【数27】

$$X_{\text{proto}}(b, n) = W_{\text{proto}}(b, n)$$

10

$$Z_{\text{proto}}(b, n) = W_{\text{proto}}(b, n)$$

【0150】

同様に、 $W_{\text{proto}}(b, n)$ も同じ理由で高次の高調波に使用される。このような状況での信号ミキサーとデコレータは、これらの球状調和成分に対して、 $T(n) = \text{"spaced"}$ の場合と同様の方法で信号を処理することができる。

【0151】

場合によっては、オーディオ再生中に搬送オーディオ信号のタイプが変わることがある(例えば、実際の信号タイプの変更や自動タイプ検出の不完全さなどによる)。急激に変化するタイプによるアーチファクトを避けるために、一部の実施形態におけるプロトタイプ信号を補間することができる。これは、例えば、旧型に応じたプロトタイプ信号から、新型に応じたプロトタイプ信号に単純に直線補間することによって実現されてもよい。

20

【0152】

信号ミキサーの出力は、得られた時間-周波数領域アンビソニック信号であり、逆T/F変圧器707に転送される。

【0153】

いくつかの実施形態では、MASA-アンビソニック信号変換器203は、信号を時間領域に変換するように構成された逆数T/F変圧器707を含む。時間領域アンビソニック信号906は、MASA-アンビソニック信号変換器からの出力である。

30

【0154】

図8に関して、図7に示す装置の動作の概要を示す。

【0155】

したがって、ある実施形態では、第1の動作は、ステップ801によって図8に示すように、MASAストリーム(またはビットストリーム)から搬送オーディオ信号およびメタデータを抽出および/または復号する動作である。

【0156】

次の動作は、ステップ803によって図8に示されるように、搬送オーディオ信号を時間-周波数領域変換することができる。

40

【0157】

次いで、本方法は、時間-周波数領域の搬送信号に基づいてプロトタイプのオーディオ信号を作成し、更に、ステップ805によって図8に示されるように、搬送オーディオ信号のタイプ(更に、付加的なパラメータに基づいて)に基づいて、プロトタイプのオーディオ信号を作成することを含む。

【0158】

いくつかの実施形態では、本方法は、ステップ807によって図8に示されるように、時間-周波数プロトタイプオーディオ信号上に非相関化を適用するステップを含む。

【0159】

次いで、ステップ809によって、図8に示されるように、空間メタデータおよび搬送オ

50

オーディオ信号タイプに基づいて、相関のない時間 - 周波数プロトタイプオーディオ信号および時間 - 周波数プロトタイプオーディオ信号を混合することができる。

【0160】

次いで、混合信号は、ステップ811によって、図8に示されるように、逆時間 - 周波数変換されてもよい。

【0161】

次いで、ステップ813によって、図8に示されるように、時間領域信号を出力することができる。

【0162】

図9は、いくつかの実施形態を実施するのに適したデコーダ例の概要図を示す。この実施例は、例えば、図1に示す「デマルチプレクサ/デコーダ/シンセサイザ」ブロック133内に実装することができ、この例では、入力は、2つのオーディオチャネルおよび空間メタデータを含むメタデータ支援空間オーディオ(MASA)ストリームである。しかしながら、本明細書で論じるように、入力フォーマットは、任意の適切なメタデータ支援空間オーディオフォーマットであり得る。

10

【0163】

(MASA)ビットストリームは、搬送オーディオ信号タイプ決定器201に転送される。搬送オーディオ信号タイプ決定器201は、ビットストリームに基づいて搬送オーディオ信号タイプ202、および場合によってはいくつかの追加パラメータ204(マイクロホン距離など)を決定するように構成される。決定されたパラメータは、MASAからマルチチャネルオーディオ信号変換器903に転送される。いくつかの実施形態における搬送オーディオ信号タイプ決定器201は、図2に関して上述したのと同じ搬送オーディオ信号タイプ決定器201であるか、または、図2に示す例に関して上述したように搬送オーディオ信号タイプ決定器201と同様に動作するように構成された搬送オーディオ信号タイプ決定器201の別個のインスタンスであり得る。

20

【0164】

MASA対マルチチャネルオーディオ信号変換器903は、ビットストリームおよび搬送オーディオ信号タイプ202(および場合によってはいくつかの追加パラメータ204)を受信するように構成され、決定された搬送オーディオ信号タイプ202(および可能な追加パラメータ204)に基づいて、MASAストリームをマルチチャネルオーディオ信号(5.1など)に変換するように構成される。

30

【0165】

図9に示す例の動作は、図10に示すフロー・ダイアグラムにまとめられている。

【0166】

最初の動作は、ステップ301によって図10に示すように、ビットストリーム(MASAストリーム)を受信または取得することの1つである。

【0167】

次の動作は、ステップ303によって図10に示されるように、ビットストリームに基づいて搬送オーディオ信号タイプを決定する(およびタイプ信号またはインジケータおよび可能な他の追加パラメータを生成する)1つの動作である。

40

【0168】

搬送オーディオ信号タイプを決定したら、次の操作は、ステップ1005によって図10に示されるように、決定された搬送オーディオ信号タイプに基づいて、ビットストリーム(MASAストリーム)をマルチチャネルオーディオ信号(5.1など)に変換することである。

【0169】

図11は、例示的なMASA - マルチチャネルオーディオ信号変換器903を更に詳細に示す。MASA対マルチチャネルオーディオ信号変換器903は、MASAストリーム(ビットストリーム)および搬送オーディオ信号タイプ202および可能な追加パラメータ204を受信するように構成され、決定された搬送オーディオ信号タイプに基づいてMA

50

S A ストリームをマルチチャネルオーディオ信号に変換するように構成される。

【 0 1 7 0 】

M A S A 対マルチチャネル音声信号変換器 9 0 3 は、搬送音声信号および空間メタデータ抽出器 / デコーダ 5 0 1 を含む。これは、図 5 に示すように、また、議論されるように、搬送オーディオ信号タイプ決定器内に見られるのと同じ方法で、M A S A ストリームを受信し、搬送オーディオ信号 5 0 2 および空間メタデータ 5 2 2 を出力するように構成される。ある実施形態では、抽出器 / 復号器 5 0 1 は、先に説明した搬送オーディオ信号タイプ決定器からの抽出器 / 復号器、または抽出器 / 復号器の別個のインスタンスである。得られた搬送オーディオ信号 5 0 2 は、時間 / 周波数変換器 5 0 3 に転送することができる。得られた空間メタデータ 5 2 2 は、さらに、ターゲット信号特性決定器 1 1 0 1 に転送することができる。

10

【 0 1 7 1 】

いくつかの実施形態では、M A S A - マルチチャネル音声信号変換器 9 0 3 は、時間 / 周波数変換器 5 0 3 を含む。時間 / 周波数変換器 5 0 3 は、搬送オーディオ信号 5 0 2 を受信し、それらを時間 - 周波数領域に変換するように構成される。適切な変換は、例えば、短時間フーリエ変換 (S T F T) および錯体変調直交ミラーフィルタバンク (Q M F) を含む。その結果、得られた信号を $S_i(b, n)$ とする。ここで、 i はチャンネルインデックス、 b は周波数ビン・インデックス、 n は時間インデックスを表す。ここで、は、チャンネル・インデックス、周波数ビン・インデックス、および時間インデックスである。オーディオ抽出および / または復号化の出力がすでに時間周波数領域にある場合、このブロックは省略されてもよいし、あるいはある時間周波数領域表現から別の時間周波数領域表現への変換を含むことができる。T / F ドメイン搬送オーディオ信号 5 0 4 は、プロトタイプ信号クリエータ 1 1 1 1 に転送することができる。いくつかの実施形態では、時間 / 周波数変換器 5 0 3 は、搬送音声信号タイプ決定器または M A S A - アンビソニック変換器または別個のインスタンスからの同一時間 / 周波数変換器である。ある実施形態では、M A S A 対マルチチャネルオーディオ信号変換器 9 0 3 は、プロトタイプ信号クリエータ 1 1 1 1 を含む。

20

【 0 1 7 2 】

プロトタイプ信号作成者 1 1 1 1 は、T / F ドメイン搬送オーディオ信号 5 0 4、搬送オーディオ信号タイプ 2 0 2、および可能な追加パラメータ 2 0 4 を受信するように構成される。次いで、T / F プロトタイプ信号 1 1 1 2 を信号ミキサー 1 1 0 5 およびデコレレータ 1 1 0 3 に出力することができる。

30

【 0 1 7 3 】

プロトタイプ信号作成者 1 1 1 1 a の動作に関する一例として、5 . 1 マルチチャネルオーディオ信号構成へのレンダリングについて説明する。この例では、左側 (左フロントおよび左サラウンド) 出力チャンネルのプロトタイプ信号を

【 数 2 8 】

$$L_{f, \text{proto}}(b, n) = L_{s, \text{proto}}(b, n) = S_0(b, n)$$

40

のように作成でき、右サイドの出力 (右フロントと右サラウンド) チャンネルを

【 数 2 9 】

$$R_{f, \text{proto}}(b, n) = R_{s, \text{proto}}(b, n) = S_1(b, n).$$

のように作成できる。

50

【 0 1 7 4 】

したがって、中央平面の両側への出力チャンネルに対して、プロトタイプ信号は、対応する搬送オーディオ信号を直接利用することができる。センター出力チャンネルの場合、プロトタイプのオーディオ信号には左右からのエネルギーが含まれている必要がある。これは、どちらのサイドへのパンにも使用できるからである。したがって、プロトタイプ信号は、アンビソニックレンダリングの場合、全方向チャンネルと同じように作成できる。つまり、 $T(n) = \text{"spaced"}$ の場合、

【 数 3 0 】

$$C_{\text{proto}}(b, n) = \frac{S_0(b, n) + S_1(b, n)}{2}, \quad 0 \leq b \leq B_3$$

$$C_{\text{proto}}(b, n) = S_0(b, n), \quad b > B_3$$

10

ある実施形態では、プロトタイプのオーディオ信号は、プロトタイプのセンターオーディオチャンネルを生成することができる。

$T(n) = \text{"downmix"}$ または $T(n) = \text{"coincident"}$ である場合、

【 数 3 1 】

$$C_{\text{proto}}(b, n) = S_0(b, n) + S_1(b, n)$$

20

【 0 1 7 5 】

ある実施形態では、M A S A 対マルチチャンネル音声信号変換器 9 0 3 は、デコレレータ 1 1 0 3 を含む。デコレレータ 1 1 0 3 は、T / F プロトタイプ信号 1 1 1 2 を受信し、デコレレーションを適用し、デコレレーション T / F プロトタイプ信号 1 1 0 4 を信号ミキサー 1 1 0 5 に出力するように構成される。いくつかの実施形態において、デコレレータ 1 1 0 3 はオプションである。

30

【 0 1 7 6 】

ある実施形態では、M A S A 対マルチチャンネルオーディオ信号変換器 9 0 3 は、ターゲット信号特性決定器 1 1 0 1 を含む。一部の実施形態における目標信号特性決定器 1 1 0 1 は、空間メタデータおよび周波数帯域内の信号エネルギーの全体推定に基づいて、周波数帯域内の目標共分散行列（目標信号特性）を生成するように構成される。いくつかの実施形態では、このエネルギー推定値は、周波数帯における搬送信号エネルギーの合計とすることができる。このターゲット共分散行列（ターゲット信号特性）判定は、特許出願 G B 1 7 1 8 3 4 1 . 9 によって提供されるのと同様の方法で実行することができる。

【 0 1 7 7 】

次に、ターゲット信号特性 1 1 0 2 を信号ミキサー 1 1 0 5 に渡すことができる。

40

【 0 1 7 8 】

ある実施形態では、M A S A 対マルチチャンネルオーディオ信号変換器 9 0 3 は、信号ミキサー 1 1 0 5 を含む。信号ミキサー 1 1 0 5 は、プロトタイプ信号の共分散行列を測定するように構成され、推定された（プロトタイプ信号）共分散行列および目標共分散行列に基づいてミクシングソリューションを定式化する。いくつかの実施形態において、ミクシングソリューションは、G B 1 7 1 8 3 4 1 . 9 に記載されているものと同様であり得る。ミクシングソリューションをプロトタイプ信号と非相関プロトタイプ信号に適用し、得られた信号を目標信号特性に基づいて周波数帯域特性で得た。つまり、決定された目標共分散行列に基づいている。いくつかの実施形態では、M A S A - マルチチャンネル音声信号変換器 9 0 3 は、信号を時間領域に変換するように構成された逆数 T / F 変圧器 7 0 7 を

50

含む。時間領域マルチチャネルオーディオ信号は、M A S A からマルチチャネルオーディオ信号変換器への出力である。

【 0 1 7 9 】

図 1 2 に関して、図 1 1 に示す装置の動作の概要を示す。

【 0 1 8 0 】

したがって、ある実施形態では、第 1 の動作は、ステップ 8 0 1 によって図 1 2 に示すように、M A S A ストリーム（またはビットストリーム）から搬送オーディオ信号およびメタデータを抽出および/または復号する動作である。

【 0 1 8 1 】

次の動作は、ステップ 8 0 3 によって図 1 2 に示されるように、搬送オーディオ信号を時間 - 周波数領域変換することができる。

【 0 1 8 2 】

次いで、本方法は、時間 - 周波数領域の搬送信号に基づいてプロトタイプのオーディオ信号を作成し、さらに、ステップ 1 2 0 5 によって、図 1 2 に示されるように、搬送オーディオ信号のタイプ（さらに、追加のパラメータに基づいて）に基づいて、プロトタイプのオーディオ信号を作成するステップを含む。

【 0 1 8 3 】

いくつかの実施形態では、本方法は、ステップ 1 2 0 7 によって、図 1 2 に示されるように、時間 - 周波数プロトタイプオーディオ信号上に非相関化を適用するステップを含む。

【 0 1 8 4 】

次いで、ステップ 1 2 0 8 によって、図 1 2 に示されるように、時間 - 周波数領域搬送オーディオ信号および空間メタデータ（ターゲット信号の共分散行列を生成するために）に基づいて、ターゲット信号特性を決定することができる。

【 0 1 8 5 】

プロトタイプのオーディオ信号の共分散行列は、ステップ 1 2 0 9 までに図 1 2 に示すように測定することができる。

【 0 1 8 6 】

次いで、ステップ 1 2 0 9 によって、図 1 2 に示されるように、非相関時間 - 周波数プロトタイプオーディオ信号および時間 - 周波数プロトタイプオーディオ信号を、目標信号特性に基づいて混合することができる。

【 0 1 8 7 】

次いで、混合信号は、ステップ 1 2 1 1 によって、図 1 2 に示されるように、逆時間 - 周波数変換されてもよい。

【 0 1 8 8 】

次いで、時間領域信号は、ステップ 1 2 1 3 によって、図 1 2 に示されるように出力され得る。

【 0 1 8 9 】

図 1 3 は、いくつかの実施形態を実現するのに適したさらなる例のデコーダの概要図を示す。他の実施形態では、同様の方法は、例えばエンコーダの一部として、デコーダ以外の装置で実施することができる。この実施例は、例えば、図 1 に示すように、（ I V A S ）デマルチプレクサ/デコーダ/シンセサイザブロック 1 3 3 内に実装することができ、この例では、入力、2 つのオーディオチャネルおよび空間メタデータを含むメタデータ支援空間オーディオ（M A S A）ストリームである。しかしながら、本明細書で論じるように、入力フォーマットは、任意の適切なメタデータ支援空間オーディオフォーマットであり得る。

【 0 1 9 0 】

（M A S A）ビットストリームは、搬送オーディオ信号タイプ決定器 2 0 1 に転送される。搬送オーディオ信号タイプ決定器 2 0 1 は、搬送オーディオ信号タイプ 2 0 2、および場合によってはいくつかの追加パラメータ 2 0 4（このような追加パラメータの一例は、マイクロホン距離）をビットストリームに基づいて決定するように構成される。決定され

10

20

30

40

50

たパラメータは、ダウンミキサー 1303 に転送される。いくつかの実施形態における搬送オーディオ信号タイプ決定器 201 は、上述したように同じ搬送オーディオ信号タイプ決定器 201 であるか、上述したように搬送オーディオ信号タイプ決定器 201 と同様に動作するように構成された搬送オーディオ信号タイプ決定器 201 の別個のインスタンスであり得る。

【0191】

ダウンミキサー 1303 は、ビットストリームおよび搬送オーディオ信号タイプ 202（および場合によってはいくつかの追加パラメータ 204）を受信するように構成され、決定された搬送オーディオ信号タイプ 202（および可能な追加パラメータ 204）に基づいて、2つの搬送オーディオ信号から1つの搬送オーディオ信号にMASAストリームをダウンミックスするように構成される。次に、出力MASAストリーム1306が出力される。

10

【0192】

図13に示す例の動作は、図14に示すフロー・ダイアグラムにまとめられている。

【0193】

最初の動作は、ステップ301によって図14に示されるように、ビットストリーム（MASAストリーム）を受信または取得することである。

【0194】

次の動作は、ステップ303によって図14に示されるように、ビットストリームに基づいて搬送オーディオ信号タイプを決定する（そして、タイプ信号またはインジケータおよび可能な他の追加パラメータを生成する）ことである。

20

【0195】

搬送オーディオ信号のタイプを決定した後、次の動作は、ステップ1405によって図14に示されるように、決定された搬送オーディオ信号のタイプ202（および可能な追加パラメータ204）に基づいて、2つの搬送オーディオ信号から1つの搬送オーディオ信号へのMASAストリームをダウンミックスする。

【0196】

図15は、ダウンミキサー1303の一例をさらに詳細に示す。ダウンミキサー1303は、MASAストリーム（ビットストリーム）および搬送オーディオ信号タイプ202および可能な追加パラメータ204を受信するように構成され、決定された搬送オーディオ信号タイプに基づいて、2つの搬送オーディオ信号を1つの搬送オーディオ信号にダウンミックスするように構成される。

30

【0197】

ダウンミキサー1303は、搬送オーディオ信号および空間メタデータ抽出器/デコーダ501を含む。これは、MASAストリームを受信し、そこで議論されている搬送オーディオ信号タイプ決定器内に見られるのと同じ方法で搬送オーディオ信号502および空間メタデータ522を出力するように構成される。ある実施形態では、抽出器/復号器501は、先に説明した抽出器/復号器、または抽出器/復号器の別個のインスタンスである。得られた搬送オーディオ信号502は、時間/周波数変換器503に転送することができる。得られた空間メタデータ522は、さらに、信号マルチプレクサ1507に転送することができる。

40

【0198】

いくつかの実施形態では、ダウンミキサー1303は、時間/周波数変換器503を含む。時間/周波数変換器503は、搬送オーディオ信号502を受信し、それらを時間-周波数領域に変換するように構成される。適切な変換は、例えば、短時間フーリエ変換（STFT）および錯体変調直交ミラーフィルタバンク（QMF）を含む。結果の信号は、 $S_i(b, n)$ のように表される。ここで、 i は、チャンネル・インデックス、周波数ビン・インデックス、および時間インデックスである。オーディオ抽出および/または復号化の出力がすでに時間周波数領域にある場合、このブロックは省略されてもよいし、あるいはある時間周波数領域表現から別の時間周波数領域表現への変換を含むことができる。T/F

50

ドメイン搬送オーディオ信号 5 0 4 は、プロトタイプ信号作成器 1 5 1 1 に転送することができる。いくつかの実施形態では、時間 / 周波数変換器 5 0 3 は、先に説明したものと同一時間 / 周波数変換器、または別個のインスタンスである。

【 0 1 9 9 】

いくつかの実施形態において、ダウンミキサー 1 3 0 3 は、プロトタイプ信号作成器 1 5 1 1 を含む。プロトタイプ信号作成器 1 5 1 1 は、T / F ドメイン搬送オーディオ信号 5 0 4、搬送オーディオ信号タイプ 2 0 2、および可能な追加パラメータ 2 0 4 を受信するように構成される。次いで、T / F プロトタイプ信号 1 5 1 2 をプロトエネルギー決定器 1 5 0 3 に出力し、プロトタイプ信号をターゲット・エネルギーコライザ 1 5 0 5 に整合させることができる。

10

【 0 2 0 0 】

一部の実施形態におけるプロトタイプ信号作成器 1 5 1 1 は、受信した搬送オーディオ信号タイプに基づいて、2 つの搬送オーディオ信号を使用して、モノ搬送オーディオ信号のプロトタイプ信号を作成するように構成される。例えば、以下を使用することができる。
 $T(n) = \text{"spaced"}$ である場合、

【 数 3 2 】

If $T(n) = \text{"spaced"}$,

$$M_{\text{proto}}(b, n) = S_0(b, n).$$

20

である。

$T(n) = \text{"downmix"}$ または $T(n) = \text{"coincident"}$ である場合、

【 数 3 3 】

If $T(n) = \text{"downmix"}$ or $T(n) = \text{"coincident"}$,

$$M_{\text{proto}}(b, n) = S_0(b, n) + S_1(b, n).$$

30

である。

【 0 2 0 1 】

いくつかの実施形態において、ダウンミキサー 1 3 0 3 は、ターゲット・エネルギー決定器 1 5 0 1 を含む。ターゲット・エネルギー決定器 1 5 0 1 は、T / F ドメイン搬送オーディオ信号 5 0 4 を受信し、搬送オーディオ信号のエネルギーの合計としてターゲット・エネルギー値

【 数 3 4 】

$$E_{\text{target}}(b, n) = |S_0(b, n)|^2 + |S_1(b, n)|^2.$$

40

を生成するように構成される。

【 0 2 0 2 】

ターゲット・エネルギー値は、次に、対象イコライザ 1 5 0 5 に一致するようにプロトにパスすることができる。

【 0 2 0 3 】

いくつかの実施形態では、ダウンミキサー 1 3 0 3 は、プロトエネルギー決定器 1 5 0 3 を含む。プロトエネルギー決定器 1 5 0 3 は、T / F プロトタイプ信号 1 5 1 2 を受信し

50

、例えば、
【数 3 5】

$$E_{\text{proto}}(b, n) = |M_{\text{proto}}(b, n)|^2.$$

のようにエネルギー値を決定するように構成される。

【0 2 0 4】

次に、プロトエネルギー値をプロトに渡して、対象イコライザ 1 5 0 5 に一致させることができる。

10

【0 2 0 5】

いくつかの実施形態におけるダウンミキサー 1 3 0 3 は、ターゲット・エネルギーコライザ 1 5 0 5 に一致するプロトを含む。いくつかの実施形態におけるターゲット・エネルギーコライザ 1 5 0 5 に一致させるためのプロトは、T / F プロトタイプ信号 1 5 0 2、プロトエネルギー値、およびターゲット・エネルギー値を受信するように構成される。いくつかの実施形態におけるイコライザ 1 5 0 5 は、まず、例えば、

【数 3 6】

$$E'_x(b, n) = a_5 E_x(b, n) + b_5 E'_x(b, n - 1)$$

20

を用いて、時間の経過とともにエネルギーを平滑化するように構成される。ここで、 a_5 と b_5 とは平滑化係数である（例えば、 $a_5 = 0.1$ および $b_5 = 1 - a_5$ ）。次に、イコライザ 1 5 0 5 は、

【数 3 7】

$$g_{\text{eq}}(b, n) = \sqrt{\frac{E'_{\text{target}}(b, n)}{E'_{\text{proto}}(b, n)}}.$$

30

のようにしてイコライゼーションゲインを決定するように構成される。

【0 2 0 6】

次いで、プロトタイプ信号は、次のようなこれらの利得を用いてイコライズすることができる。

【数 3 8】

$$M(b, n) = g_{\text{eq}}(b, n) M_{\text{proto}}(b, n),$$

40

イコライズされたプロトタイプ信号は逆 T / F 変圧器 7 0 7 に渡される。

【0 2 0 7】

いくつかの実施形態では、ダウンミキサー 1 3 0 3 は、イコライザの出力を時間領域バージョンに変換するように構成された逆数 T / F 変圧器 7 0 7 を含む。次いで、時間領域イコライズオーディオ信号（モノラル信号）1 5 1 0 は、搬送オーディオ信号および空間メタデータマルチプレクサ 1 5 0 7（またはマルチプレクサ）に渡される。

【0 2 0 8】

50

いくつかの実施形態では、ダウンミキサー 1303 は、搬送オーディオ信号および空間メタデータマルチプレクサ 1507（またはマルチプレクサ）を含む。搬送オーディオ信号および空間メタデータマルチプレクサ 1507（またはマルチプレクサ）は、空間メタデータ 522 およびモノオーディオ信号 1510 を受信し、それらを多重化して、適切な出力フォーマット（たとえば、1つの搬送オーディオ信号のみを有する MASA ストリーム）1506 を再生成するように構成される。一部の実施形態では、入力モノラルオーディオ信号は、パルス符号変調（PCM）形式である。このような実施形態では、信号は、多重化されるだけでなく、符号化されてもよい。いくつかの実施形態では、多重化は省略されてもよく、モノ搬送オーディオ信号および空間メタデータは、オーディオエンコーダで直接使用される。

10

【0209】

ある実施形態では、図 15 に示す装置の出力は、空間メタデータが破棄されるモノ PCM オーディオ信号 1510 である。

【0210】

いくつかの実施形態では、他のパラメータを実装することができ、例えば、いくつかの実施形態では、タイプが「間隔を置いて」いる場合に、間隔を置いたマイクロホン距離を推定することができる。

【0211】

図 16 に関して、図 15 に示される装置の一例の動作が示される。

【0212】

したがって、ある実施形態では、第 1 の動作は、ステップ 1601 によって図 16 に示すように、MASA ストリーム（またはビットストリーム）から搬送オーディオ信号およびメタデータを抽出および/または復号する動作である。

20

【0213】

次の動作は、ステップ 1603 によって図 16 に示されるように、搬送オーディオ信号の時間 - 周波数領域変換であり得る。

【0214】

次いで、本方法は、時間 - 周波数領域の搬送信号に基づいてプロトタイプのオーディオ信号を作成し、さらに、ステップ 1605 によって、図 16 に示されるように、搬送オーディオ信号のタイプ（さらに、追加のパラメータに基づいて）に基づいて、プロトタイプのオーディオ信号を作成するステップを含む。

30

【0215】

さらに、いくつかの実施形態では、方法は、ステップ 1604 によって図 16 に示されるように、変換された搬送オーディオ信号に基づいて、ターゲット・エネルギー値を生成、決定、または計算するように構成される。

【0216】

さらに、いくつかの実施形態では、方法は、ステップ 1606 によって、図 16 に示されるように、プロトタイプのオーディオ信号エネルギー値に基づいて、プロトタイプのオーディオ信号エネルギー値を生成、決定、または計算するように構成される。

【0217】

エネルギーを決定した後、本方法は、ステップ 1607 によって、図 16 に示されるように、ターゲットオーディオ信号エネルギーに一致するように、プロトタイプオーディオ信号をさらにイコライズすることができる。

40

【0218】

次いで、イコライズされたプロトタイプ信号（モノ信号）は、ステップ 1609 によって、図 16 に示されるように、時間領域モノ信号を生成するために、逆時間 - 周波数領域変換されてもよい。

【0219】

次いで、ステップ 1610 によって、図 16 に示すように、時間領域モノラルオーディオ信号を空間メタデータと（任意に符号化し、多重化してもよい）。

50

【0220】

次に、ステップ1611によって図16に示すように、多重化されたオーディオ信号を(MASAデータストリームとして)出力することができる。

【0221】

上述したように、示されたブロック図は、可能な実装の一例にすぎない。他の実用的な実装は、上記の例とは異なる可能性がある。例えば、実装は、個別のT/F変換器を持たないことがある。

【0222】

さらに、上に示したような入力MASAストリームを有するのではなく、いくつかの実施形態では、オーディオチャネルおよび(空間)メタデータを利用する任意の適切なビットストリームを使用することができる。さらに、いくつかの実施形態では、IVASコーデックは、任意の他の適切なコーデック(例えば、オーディオチャネルおよび空間メタデータの動作モードを有するもの)に置き換えることができる。

10

【0223】

いくつかの実施形態では、搬送オーディオ信号タイプ決定器を使用して、搬送オーディオ信号タイプ以外のパラメータを推定することができる。例えば、マイクロホンの間隔を推定することができる。マイクロホンの間隔は、可能な追加パラメータ204の一例である。これは、いくつかの実施形態において、 $E_{sum}(b, n)$ および $E_{sub}(b, n)$ の極大と極小の周波数を検査し、それらに基づいてマイクロホン間の時間遅延を決定し、遅延と推定到着方向(空間メタデータにおいて利用可能)に基づいて間隔を推定することによって、実現することができる。また、2つの信号間の遅延を推定する方法もある。

20

【0224】

図17に関して、解析装置または合成装置として使用され得る電子装置の例が示されている。この装置は、任意の適切な電子装置または装置とすることができる。例えば、ある実施形態では、装置1700は、モバイル装置、ユーザ装置、タブレットコンピュータ、コンピュータ、オーディオ再生装置等である。

【0225】

ある実施形態では、装置1700は、少なくとも1つのプロセッサまたは中央処理ユニット1707を含む。プロセッサ1707は、本明細書に記載するような方法のような様々なプログラムコードを実行するように構成することができる。

30

【0226】

ある実施形態では、装置1700はメモリ1711を含む。ある実施形態では、少なくとも1つのプロセッサ1707は、メモリ1711に結合される。メモリ1711は、任意の適切な記憶手段とすることができる。ある実施形態では、メモリ1711は、プロセッサ1707上に実装可能なプログラムコードを格納するためのプログラムコード・セクションを含む。さらに、いくつかの実施形態では、メモリ1711は、例えば、本明細書に記載する実施形態にしたがって処理された、または処理されるべきデータを記憶するための記憶データ・セクションをさらに含むことができる。プログラムコード・セクション内に記憶された実施されたプログラムコードおよび記憶されたデータ・セクション内に記憶されたデータは、メモリ・プロセッサ結合を介して必要なときにいつでもプロセッサ1707によって検索することができる。

40

【0227】

ある実施形態では、装置1700は、ユーザインターフェース1705を含む。ユーザインターフェース1705は、いくつかの実施形態において、プロセッサ1707に結合することができる。ある実施形態では、プロセッサ1707は、ユーザインターフェース1705の動作を制御し、ユーザインターフェース1705から入力を受信することができる。ある実施形態では、ユーザインターフェース1705は、ユーザが、例えばキーパッドを介して、装置1700にコマンドを入力することを可能にすることができる。ある実施形態では、ユーザインターフェース1705は、ユーザが装置1700から情報を取得することを可能にすることができる。例えば、ユーザインターフェース1705は、装置

50

１７００からユーザに情報を表示するように構成されたディスプレイを含むことができる。ユーザインターフェース１７０５は、ある実施形態では、情報を装置１７００に入力することを可能にし、装置１７００のユーザに情報をさらに表示することの両方が可能なタッチスクリーンまたはタッチインターフェースを備えることができる。いくつかの実施形態において、ユーザインターフェース１７０５は、本明細書に記載するように、位置決定器と通信するためのユーザインターフェースであり得る。

【０２２８】

ある実施形態では、装置１７００は、入出力ポート１７０９を含む。いくつかの実施形態における入出力ポート１７０９は、トランシーバを含む。このような実施形態のトランシーバは、プロセッサ１７０７に結合され、例えば無線通信ネットワークを介して、他の装置または電子装置との通信を可能にするように構成されることができる。トランシーバまたは任意の適切なトランシーバまたはトランスミッタおよび／またはレシーバ手段は、一部の実施形態では、ワイヤまたは有線結合を介して他の電子装置または装置と通信するように構成することができる。

10

【０２２９】

トランシーバは、任意の適切な既知の通信プロトコルによって、さらなる装置と通信することができる。例えば、いくつかの実施形態では、トランシーバは、適切なユニバーサル移動通信システム（UMTS）プロトコル、例えばIEEE 802.Xのような無線ローカルエリアネットワーク（WLAN）プロトコル、ブルートゥース（Bluetooth）（登録商標）のような適切な短距離無線周波数通信プロトコル、または赤外線データ通信経路（IRDA）を使用することができる。

20

【０２３０】

トランシーバ入出力ポート１７０９は、信号を受信するように、および、いくつかの実施形態では、適切なコードを実行するプロセッサ１７０７を使用することによって、本明細書に記載するようにパラメータを決定するように構成されてもよい。

【０２３１】

いくつかの実施形態では、装置１７００は、合成装置の少なくとも一部として採用されてもよい。入出力ポート１７０９は、任意の適切なオーディオ出力、例えば、マルチチャンネルスピーカシステムおよび／またはヘッドホン（これは、ヘッドトラッキングされたヘッドホンまたは追跡されていないヘッドホンであり得る）または同様のものに結合することができる。

30

【０２３２】

一般に、本発明の様々な実施形態は、ハードウェアまたは特殊目的回路、ソフトウェア、ロジック、またはそれらの任意の組み合わせで実現することができる。例えば、いくつかの態様は、ハードウェアで実施されてもよいが、本発明はこれに限定されないが、コントローラ、マイクロプロセッサまたは他の計算装置によって実行されてもよいファームウェアまたはソフトウェアで実施されてもよい。本発明の様々な態様は、ブロック図、フロー図、または何らかの他の絵表示として図示および説明することができるが、本明細書に記載するこれらのブロック、装置、システム、技術または方法は、非限定的な例として、ハードウェア、ソフトウェア、ファームウェア、特殊目的回路または論理、汎用ハードウェアまたはコントローラ、または他の計算装置、またはそれらの組み合わせで実装することができるが、よく理解される。

40

【０２３３】

本発明の実施形態は、プロセッサエンティティ内などのモバイル装置のデータプロセッサによって実行可能なコンピュータソフトウェアによって、またはハードウェアによって、あるいはソフトウェアとハードウェアの組み合わせによって実行可能なコンピュータソフトウェアによって実現することができる。さらに、図のような論理フローの任意のブロックは、プログラムステップ、または相互接続された論理回路、ブロックおよび機能、またはプログラムステップおよび論理回路、ブロックおよび機能の組み合わせを表すことができることに留意されたい。このソフトウェアは、メモリチップなどの物理メディア、また

50

はプロセッサ内に実装されたメモリブロック、ハードディスクまたはフロッピー（登録商標）ディスクなどの磁気メディア、およびたとえばDVDやそのデータ変異体などの光学メディアに格納することができる。

【0234】

メモリは、ローカル技術環境に適した任意のタイプでよく、半導体ベースのメモリデバイス、磁気メモリデバイスおよびシステム、光メモリデバイスおよびシステム、固定メモリおよび取り外し可能メモリなどの任意の適切なデータ記憶技術を使用して実施することができる。データプロセッサは、ローカル技術環境に適した任意のタイプでよく、限定されない例として、汎用コンピュータ、特殊目的コンピュータ、マイクロプロセッサ、デジタル信号プロセッサ、特定用途向け集積回路（ASIC）、ゲートレベル回路、およびマルチコアプロセッサキテクチャに基づくプロセッサのうちの1つ以上を含むことができる。

10

【0235】

本発明の実施形態は、集積回路モジュールなどの様々な部品において実施可能である。集積回路の設計は、高度に自動化された処理によるものであり、大規模である。論理レベルの設計を、エッチングされ、半導体基板上に形成される準備ができているの整った半導体回路設計に変換するための、複雑で強力なソフトウェアツールが利用可能である。

【0236】

カリフォルニア州マウンテンビューにあるシノプシス社（Synopsys, Inc）およびカリフォルニア州サンノゼにあるケイデンスデザイン社（Cadence Design）から提供されているようなプログラムは、設計の十分に確立されたルール、及び予め記憶された設計モジュールのライブラリを用いて、導体を自動的にルーティングし、半導体チップ上の部品の位置を特定する。半導体回路の設計が完了すると、結果として得られた設計は、標準化された電子フォーマット（例えば、Opus、GDSII等）で、半導体製造設備または製造のための「fab」に伝送され得る。

20

【0237】

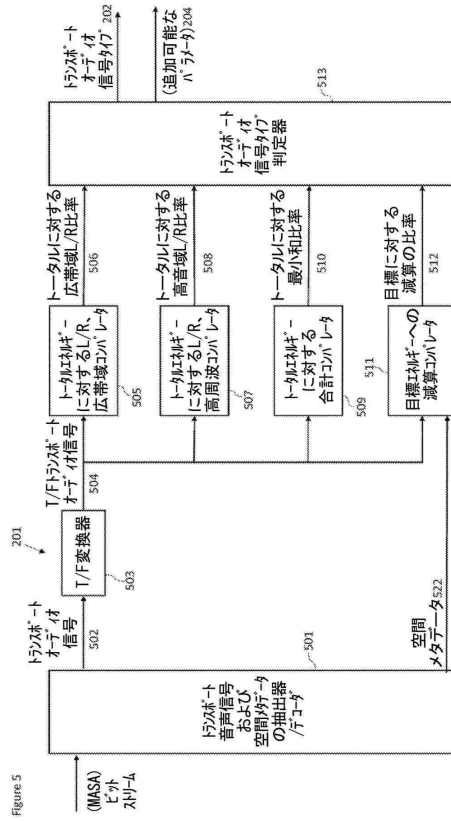
上述の説明は、本発明の例示的な実施形態の完全かつ参考的な説明を例示的な例および非限定的な例によって提供したものである。しかしながら、添付の図面および付随の請求項を熟読する際に、前述の説明を考慮して、種々の修正および適合が、当業者に明白になるであろう。しかしながら、この発明の教示のこのような変更および類似した変更のすべては、引き続き、添付のクレームに定義されている本発明の範囲内に収まるであろう。

30

40

50

【 図 5 】



【 図 6 】

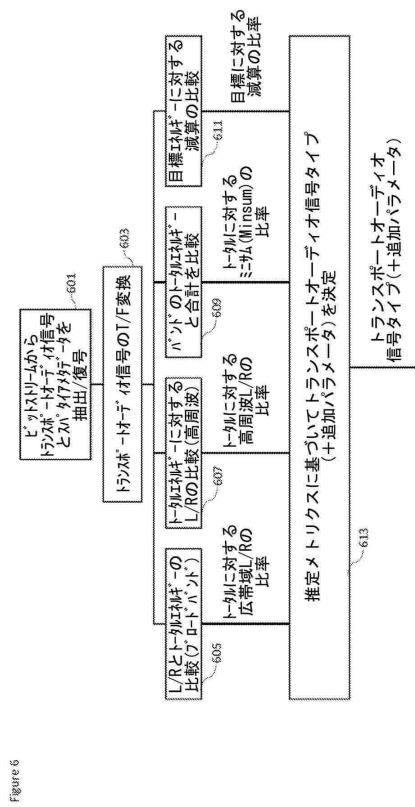


Figure 6

【圖 7】

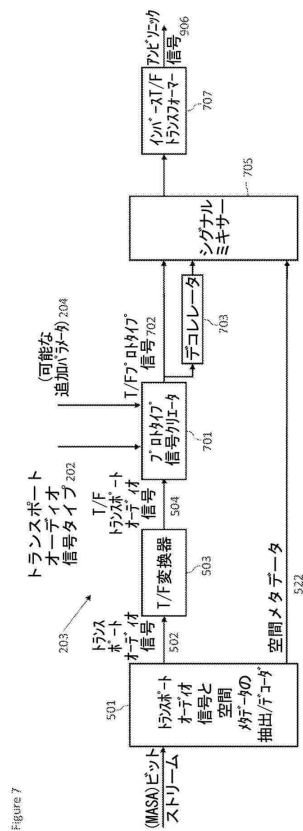


Figure 7

【 図 8 】

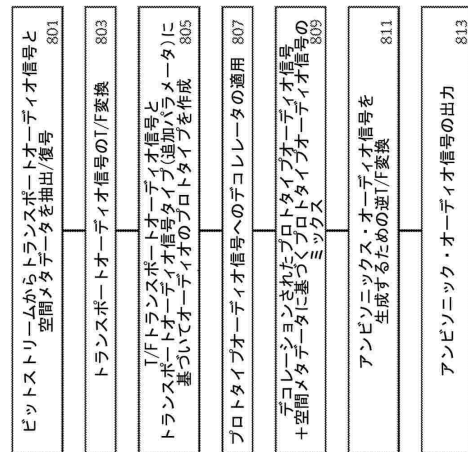
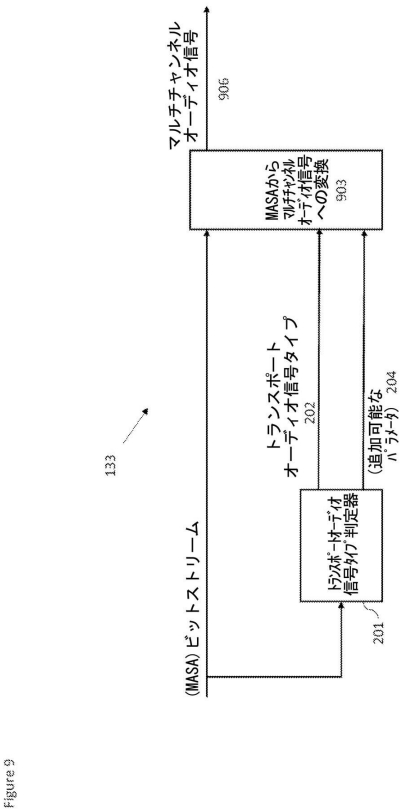
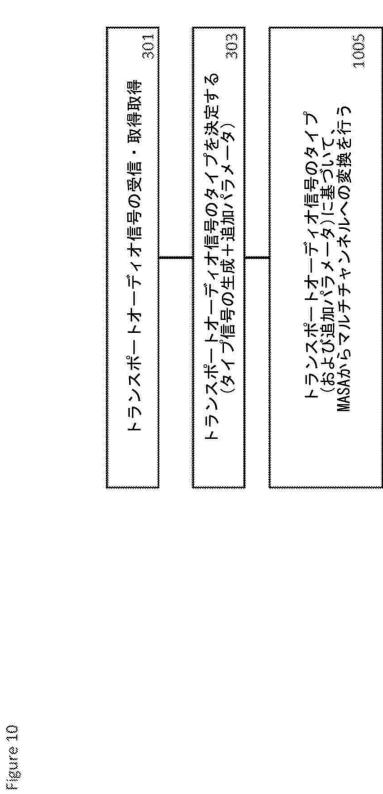


Figure 8

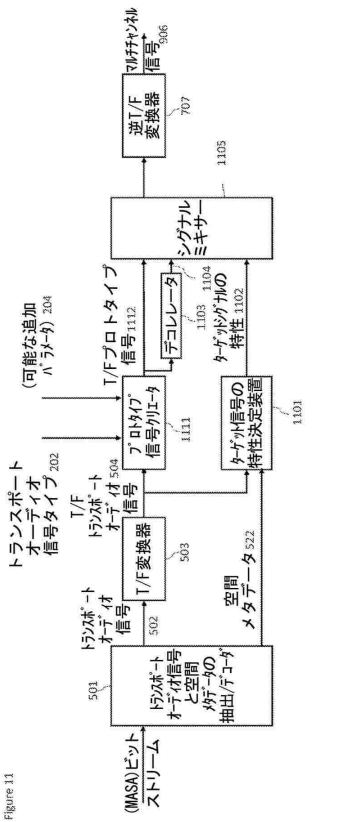
【図 9】



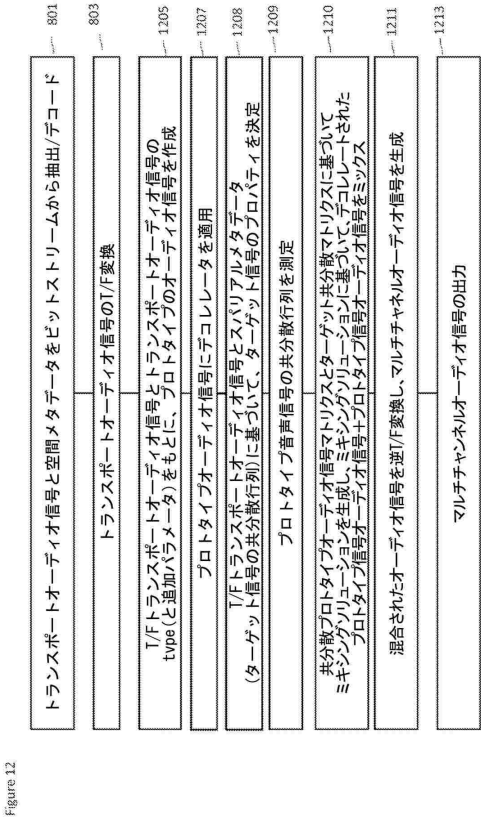
【図 10】



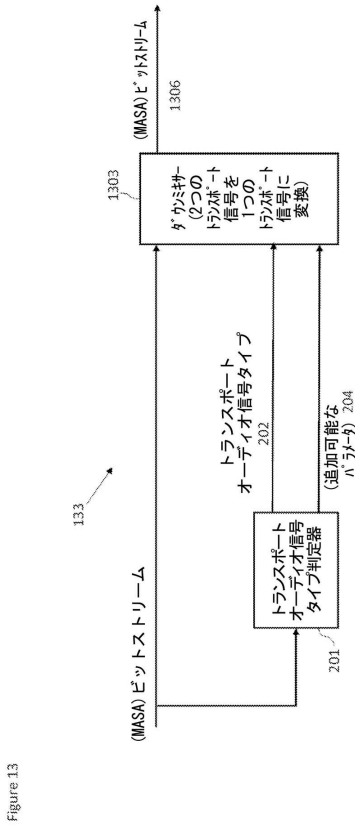
【図 11】



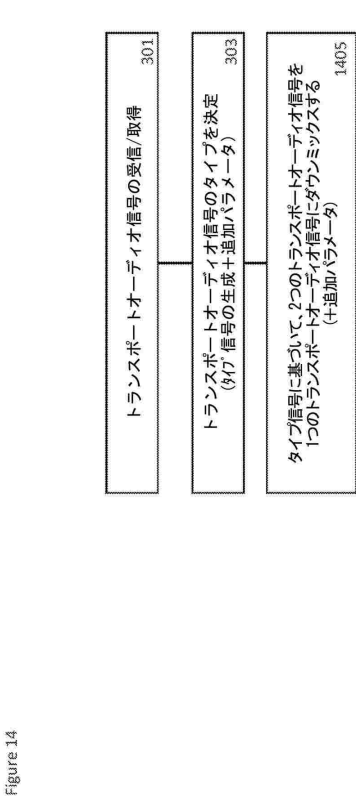
【図 12】



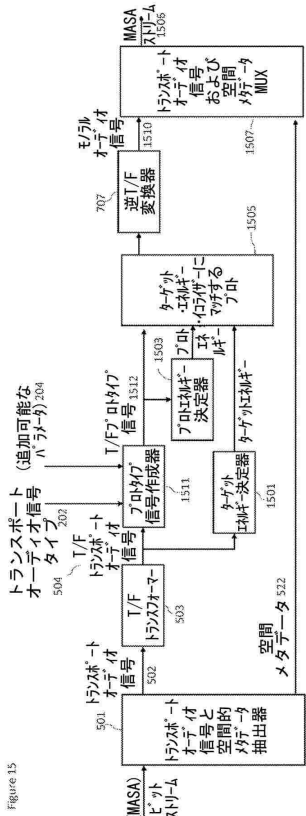
【図 1 3】



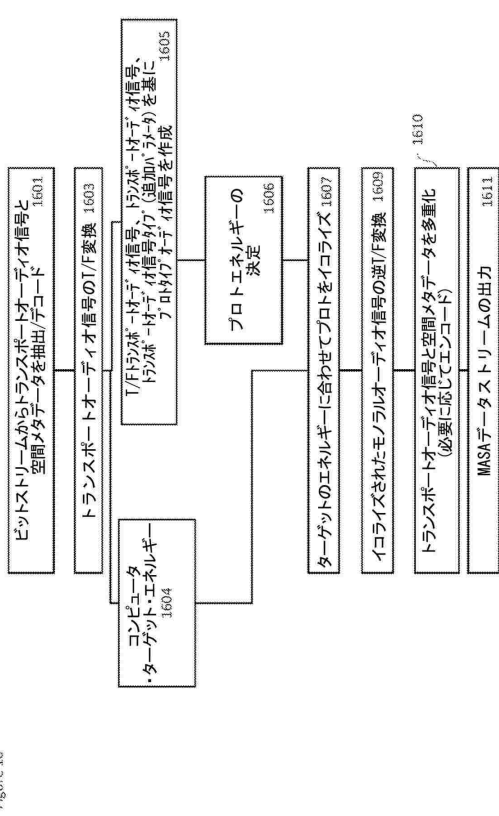
【図 1 4】



【図 1 5】



【図 1 6】



10

20

30

40

50

【図 17】

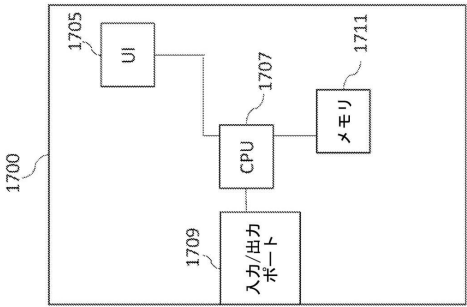


Figure 17

10

20

30

40

50

フロントページの続き

英国(GB)
弁理士 中村 和広

(74)代理人

森 啓

(72)発明者 ミッコ - ビッレ ライティネン

フィンランド国, 0 2 1 3 0 エスポー, メツツァピルティンティエ 2 セー 1 5

(72)発明者 ユハ ビルカモ

フィンランド国, 0 0 1 2 0 ヘルシンキ, アンナンカトゥ 4 アー 1 8

(72)発明者 ラッセ ラークソネン

フィンランド国, 3 3 2 1 0 タンペレ, ナシリナンカトゥ 2 3 ベー 2 8

合議体

審判長 高橋 宣博

審判官 宮下 誠

審判官 榎本 剛

(56)参考文献 特表 2 0 1 9 - 5 3 3 4 0 4 (J P , A)

特表 2 0 1 8 - 5 3 4 6 1 7 (J P , A)

米国特許出願公開第 2 0 1 7 / 0 1 6 2 2 1 0 (U S , A 1)

国際公開第 2 0 1 8 / 0 5 6 7 8 0 (W O , A 1)

(58)調査した分野 (Int.Cl., D B 名)

G10L 19/22