

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6339997号
(P6339997)

(45) 発行日 平成30年6月6日 (2018.6.6)

(24) 登録日 平成30年5月18日 (2018.5.18)

(51) Int. Cl.	F I
HO4S 1/00 (2006.01)	HO4S 1/00 200
HO4M 3/56 (2006.01)	HO4M 3/56
HO4S 7/00 (2006.01)	HO4S 7/00 300

請求項の数 38 (全 34 頁)

(21) 出願番号	特願2015-501908 (P2015-501908)	(73) 特許権者	507236292
(86) (22) 出願日	平成25年3月21日 (2013.3.21)		ドルビー ラボラトリーズ ライセンシン
(65) 公表番号	特表2015-515800 (P2015-515800A)		グ コーポレイション
(43) 公表日	平成27年5月28日 (2015.5.28)		アメリカ合衆国 94103 カリフォル
(86) 国際出願番号	PCT/US2013/033270		ニア州 サンフランシスコ マーケット
(87) 国際公開番号	W02013/142668		ストリート 1275
(87) 国際公開日	平成25年9月26日 (2013.9.26)	(74) 代理人	100107766
審査請求日	平成26年9月29日 (2014.9.29)		弁理士 伊東 忠重
審判番号	不服2016-16279 (P2016-16279/J1)	(74) 代理人	100070150
審判請求日	平成28年11月1日 (2016.11.1)		弁理士 伊東 忠彦
(31) 優先権主張番号	61/614,592	(74) 代理人	100091214
(32) 優先日	平成24年3月23日 (2012.3.23)		弁理士 大貫 進介
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 2Dまたは3D会議シーンにおける語り手の配置

(57) 【特許請求の範囲】

【請求項1】

会議参加者に関連付けられた少なくとも一つの上流オーディオ信号（123、173）を、聴取者（211）に対してレンダリングされる2Dまたは3D会議シーン内で配置するよう構成された会議コントローラ（111、175）であって：

・前記会議シーン内のX個の異なる空間的語り手位置（212）をもつX点会議シーンをセットアップする段階であって、Xは整数であり、X>0であり、前記X個の語り手位置（212）は、聴取者の頭部の前方の中央線（215）のまわりの円錐内に位置され、前記円錐の母線（216）と前記中央線（215）は、あらかじめ決定された最大円錐角以下の角度をなす、段階と；

・聴取者の好まれる側を決定する段階であって、前記好まれる側は、前記中央線から左側または右側であり、前記好まれる側は、聴取者に2Dまたは3D会議シーンをレンダリングするために使われる端末（120）の選好設定に基づいて決定される、段階と；

・最初に前記X個の語り手位置の一つに割り当てられる上流オーディオ信号を、前記X個の語り手位置のうち聴取者の前記好まれる側にある一つに割り当てる段階と；

・割り当てられた語り手位置を同定するメタデータを生成し、オーディオ処理ユニット（121、171）が、割り当てられた上流オーディオ信号に基づいて立体音響化されたオーディオ信号を生成できるようにする段階であって、立体音響化されたオーディオ信号を聴取者に対してレンダリングするとき、聴取者は、割り当てられた上流オーディオ信号を、割り当てられた語り手位置から到来するものとして知覚する、段階とを実行するよう構

成されている、
会議コントローラ。

【請求項 2】

前記X個の語り手位置は円または球上に配置され、聴取者はその円または球の中心に配置される、請求項 1 記載の会議コントローラ。

【請求項 3】

・前記X個の語り手位置のうち二つの隣接する語り手位置は、少なくともある最小の角距離だけ離間しており、
・前記最小の角距離は 5° またはそれ以上である、
請求項 1 または 2 記載の会議コントローラ。

10

【請求項 4】

前記X個の語り手位置のうちの隣り合う語り手位置の間の角距離が異なる語り手位置については異なる、請求項 1 ないし 3 のうちいずれか一項記載の会議コントローラ。

【請求項 5】

前記X個の語り手位置のうちの隣り合う語り手位置の間の角距離が、該隣り合う語り手位置の前記中央線からの距離が増すにつれて増大する、請求項 4 記載の会議コントローラ。

【請求項 6】

前記最大円錐角が 30° である、請求項 1 ないし 5 のうちいずれか一項記載の会議コントローラ。

20

【請求項 7】

前記最大円錐角が 20° である、請求項 1 ないし 5 のうちいずれか一項記載の会議コントローラ。

【請求項 8】

前記最大円錐角が 15° である、請求項 1 ないし 5 のうちいずれか一項記載の会議コントローラ。

【請求項 9】

前記X個の語り手位置が前記中央線に対して非対称である、請求項 1 ないし 8 のうちいずれか一項記載の会議コントローラ。

【請求項 10】

30

Xが奇数の場合、前記X個の語り手位置は、該X個の語り手位置のうちの中央語り手位置に対して対称的である、請求項 1 ないし 9 のうちいずれか一項記載の会議コントローラ。

【請求項 11】

・Xが偶数の場合、前記X個の語り手位置の半分が前記中央線の一方の側または前記中央線を通る鉛直面の一方の側にあり、前記X個の語り手位置の残りの半分が前記中央線の反対側または前記中央線を通る前記鉛直面の反対側にある、および／または
・Xが奇数の場合、前記X個の語り手位置のうち $(X-1)/2$ 個が一方の側にあり、 $(X+1)/2$ 個が前記中央線または前記鉛直面の反対側にある、
請求項 1 ないし 10 のうちいずれか一項記載の会議コントローラ。

【請求項 12】

40

・前記X個の語り手位置は、聴取者の頭部からのX本の射線上に位置され、前記中央線とX個の語り手角をなし；
・前記中央線からある方向の語り手角は正であり、前記中央線から反対方向の語り手角は負であり；
・前記X個の語り手角の和は0とは異なる、
請求項 1 ないし 11 のうちいずれか一項記載の会議コントローラ。

【請求項 13】

語り手角の前記和は 2° 以下である、請求項 12 記載の会議コントローラ。

【請求項 14】

X=3であり、前記X個の語り手位置は、前記中央線から -2° 、 $+6^\circ$ および -10° の語

50

り手角のところに位置される、請求項 1 2 または 1 3 記載の会議コントローラ。

【請求項 1 5】

X=6であり、前記X個の語り手位置は、前記中央線から−2度、+3度、−7度、+8度、−12度および+13度の語り手角のところに位置される、請求項 1 2 または 1 3 記載の会議コントローラ。

【請求項 1 6】

X=6であり、前記X個の語り手位置は、前記中央線から+2度、−5度、+9度、−12度、+16度および−19度の語り手角のところに位置される、請求項 1 2 または 1 3 記載の会議コントローラ。

【請求項 1 7】

複数の会議参加者に関連付けられた複数の上流オーディオ信号を前記X点会議シーン内に配置するよう構成されている、請求項 1 ないし 1 6 のうちいずれか一項記載の会議コントローラであって、当該会議コントローラは、前記複数の上流オーディオ信号を、前記複数の上流オーディオ信号のある逐次順に従って、前記X個の語り手位置に割り当てるよう構成されている、会議コントローラ。

【請求項 1 8】

前記複数の上流オーディオ信号の前記逐次順は、当該会議コントローラによる前記複数の上流オーディオ信号の検出の順序に基づいている、請求項 1 7 記載の会議コントローラ。

【請求項 1 9】

当該会議コントローラは、前記中央線からの絶対的な角距離が増す順に前記X個の語り手位置を割り当てるよう構成されている、請求項 1 7 または 1 8 記載の会議コントローラ。

【請求項 2 0】

前記複数の上流オーディオ信号からの上流オーディオ信号を、前記X個の語り手位置のうちからの最も外側の語り手位置に割り当てることを受けて、当該会議コントローラは、前記複数の上流オーディオ信号のうちからの次の上流オーディオ信号を、前記X個の語り手位置のうちからの最も内側の語り手位置に割り当てるよう構成されている、請求項 1 9 記載の会議コントローラ。

【請求項 2 1】

当該会議コントローラは、上流オーディオ信号の数が語り手位置の数Xより多い場合、前記複数の上流オーディオ信号のうちからのいくつかの上流オーディオ信号を、前記X個の語り手位置のうちの少なくとも一つに割り当てるよう構成されている、請求項 1 7 ないし 2 0 のうちいずれか一項記載の会議コントローラ。

【請求項 2 2】

同じ語り手位置に割り当てられた前記いくつかの上流オーディオ信号の混合を開始するよう構成されている、請求項 2 1 記載の会議コントローラ。

【請求項 2 3】

前記複数の上流オーディオ信号のそれぞれが、前記X個の語り手位置のうちの単一の位置のみに割り当てられる、請求項 1 7 ないし 2 2 のうちいずれか一項記載の会議コントローラ。

【請求項 2 4】

- ・前記複数の上流オーディオ信号からの次の上流オーディオ信号を、前記X個の語り手位置のうちの、まだ割り当てられておらず、かつ、前記中央線にできるだけ近い語り手位置に割り当てる；
- ・割り当てられた諸語り手位置の、前記中央線に対するおよび／または前記X個の語り手位置のうち中心の語り手位置に対する最大の均衡を保証する；
- ・前記X個の語り手位置のうちからの、前記複数の上流オーディオ信号のうちからのすでに割り当てられた上流オーディオ信号の中断によって空にされた空の語り手位置を割り当て直す、

10

20

30

40

50

という配置ルールのうちの一つまたは複数に従って前記複数の上流オーディオ信号を配置するよう構成されている、

請求項 17 ないし 23 のうちいずれか一項記載の会議コントローラ。

【請求項 25】

- ・前記立体音響化されたオーディオ信号をレンダリングするオーディオ・トランシーバ（122）のレンダリング特性を判別し；
- ・前記オーディオ・トランシーバの前記レンダリング特性に依存して前記複数の上流オーディオ信号の一つまたは複数の混合を開始するよう構成されている、

請求項 17 ないし 24 のうちいずれか一項記載の会議コントローラ。

【請求項 26】

前記レンダリング特性が前記オーディオ・トランシーバがモノ・オーディオ信号のレンダリングに限定されていることを示す場合には、前記会議シーンに配置されるべき前記複数の上流オーディオ信号の全部の混合を開始するよう構成されている、請求項 25 記載の会議コントローラ。

【請求項 27】

前記上流オーディオ信号が発話信号を含むことを判別するよう構成されている、請求項 1 ないし 26 のうちいずれか一項記載の会議コントローラ。

【請求項 28】

- ・前記会議シーンが3D会議シーンであり；
- ・前記X個の異なる空間的語り手位置は、前記中央線に対して異なる方位角および／または傾斜角に位置される、

請求項 1 ないし 27 のうちいずれか一項記載の会議コントローラ。

【請求項 29】

- ・前記X個の異なる空間的語り手位置は、異なる傾斜角にある複数の層に配列される；および／または
- ・各鉛直方向層は、異なる方位角にある複数の異なる空間的語り手位置を含む；および／または
- ・第一の鉛直方向層における異なる空間的語り手位置と第二の鉛直方向層における異なる空間的語り手位置は対応する方位角を有する、

請求項 28 記載の会議コントローラ。

【請求項 30】

- ・前記上流オーディオ信号の地位に関する情報を受領する段階と；
- ・前記上流オーディオ信号を、該上流オーディオ信号の地位に基づいて、空間的語り手位置に割り当てる段階とを実行するよう構成されている、

請求項 1 ないし 29 のうちいずれか一項記載の会議コントローラ。

【請求項 31】

- ・前記地位が、前記上流オーディオ信号が会議の司会者またはまとめ役に対応することを示し；
- ・前記上流オーディオ信号が割り当てられる空間的語り手位置は、最も中央寄りの空間的語り手位置である、

請求項 30 記載の会議コントローラ。

【請求項 32】

あらかじめ決定された地位をもつ上流オーディオ信号のために前記X個の空間的語り手位置のうちの一つまたは複数を予約するよう構成されている、請求項 1 ないし 31 のうちいずれか一項記載の会議コントローラ。

【請求項 33】

前記X個の語り手位置が聴取者の頭部の前方の前記中央線のまわりの前記円錐内に位置されるよう、X個の異なる空間的語り手位置をもつ前記X点会議シーンを計算するよう構成されている、請求項 1 ないし 32 のうちいずれか一項記載の会議コントローラ。

【請求項 34】

X個の異なる空間的語り手位置をもつ前記X点会議シーンを、X個の異なるあらかじめ決定された空間的語り手位置をもつあらかじめ決定されたX点会議シーンの集合から選択するように構成されている、請求項1ないし33のうちいずれか一項記載の会議コントローラ。

【請求項35】

オーディオ会議システム（100、150）であって、
・会議参加者に関連付けられた上流オーディオ信号を生成するよう構成された語り手端末と；
・請求項1ないし34のうちいずれか一項記載の会議コントローラであって、2Dまたは3D会議シーン内の語り手位置に上流オーディオ信号を割り当てよう構成されており、割り当てられた語り手位置を同定するメタデータを生成するよう構成されている会議コントローラと；
・聴取者が上流オーディオ信号を、割り当てられた語り手位置から到来するものとして知覚するよう、前記メタデータを使って聴取者に対して上流オーディオ信号をレンダリングするよう構成された聴取者端末を有する、
オーディオ会議システム。

10

【請求項36】

会議参加者に関連する少なくとも一つの上流オーディオ信号を、聴取者に対してレンダリングされるべき2Dまたは3D会議シーン内で配置する方法であって、
・会議シーン内でX個の異なる空間的語り手位置をもつX点会議シーンをセットアップする段階であって、Xは整数であり、 $X > 0$ であり、前記X個の語り手位置は、聴取者の頭部の前方の中央線のまわりの円錐内に位置され、前記円錐の母線と前記中央線はあらかじめ決定された最大円錐角以下の角度をなす、段階と；
・聴取者の好まれる側を決定する段階であって、前記好まれる側は、前記中央線から左側または右側であり、前記好まれる側は、聴取者に2Dまたは3D会議シーンをレンダリングするために使われる端末（120）の選好設定に基づいて決定される、段階と；
・最初に前記X個の語り手位置の一つに割り当てられる上流オーディオ信号を、前記X個の語り手位置のうち聴取者の前記好まれる側にある一つに割り当てる段階と；
・割り当てられた語り手位置を同定するメタデータを生成し、オーディオ処理ユニット（121、171）が、割り当てられた上流オーディオ信号に基づいて立体音響化されたオーディオ信号を生成できるようにする段階とを含んでおり、立体音響化されたオーディオ信号を聴取者に対してレンダリングするとき、聴取者は、割り当てられた上流オーディオ信号を、割り当てられた語り手位置から到来するものとして知覚する、
方法。

20

30

【請求項37】

プロセッサ上での実行のために、該プロセッサ上で実行されたときに請求項36記載の方法段階を実行するために適応されている、ソフトウェア・プログラム。

【請求項38】

プロセッサ上での実行のために、コンピューティング装置上で実行されたときに請求項36記載の方法段階を実行するために適応されたソフトウェア・プログラムを有している記憶媒体。

40

【発明の詳細な説明】

【技術分野】

【0001】

関連出願への相互参照

本願は2012年3月23日に出願された米国特許仮出願第61/614,592号の優先権を主張するものである。同出願の内容はここに参照によってその全体において組み込まれる。

【0002】

本稿はオーディオ会議に関する。詳細には、本稿はオーディオ会議のために二次元または三次元シーンをセットアップし、管理する方法およびシステムに関する。

50

【背景技術】

【0003】

現行の多者音声会議における欠点の一つは、音声典型的にはみな聴取者に対してモノラル・オーディオ・ストリームとしてレンダリングされる。本質的には互いの上に重畳され、ヘッドホンが使われるときは通例、聴取者に対して「頭の中で」呈示される。ということである。たとえば異なるレンダリングされる位置から話す異なる人々をシミュレートするために使われる立体音響化 (spatialization) 技法は、音声会議において、特に発言している複数の人がいるときに、発話の了解性 (intelligibility) を改善することができる。

【発明の概要】

10

【発明が解決しようとする課題】

【0004】

本稿は、聴取者がオーディオ会議の異なる語り手を簡単に区別できるようにする、オーディオ会議のための適切な二次元 (2D) または三次元 (3D) シーンをデザインするという技術的問題に対処する。さらに、2Dまたは3D会議シーンに参加者および／またはシステム音 (たとえば通知または背景音楽) を分布させるための諸方式であって、新しい参加者を追加するときおよび／または会議シーンにシステム音を挿入するときの進行中のオーディオ会議への影響を軽減できるようにするものが記述される。さらに、さまざまなレンダリング装置 (ヘッドホンおよび／またはラウドスピーカー配位のような) 上で会議シーンをレンダリングするための適切な方式が記述される。

20

【課題を解決するための手段】

【0005】

ある側面によれば、会議参加者に関連付けられた上流オーディオ信号を、2Dまたは3D会議シーン内で配置するよう構成された会議コントローラが記述される。会議シーンは、聴取者に対してレンダリングされるものである。典型的には、聴取者は、会議シーンの中央位置に (たとえば、会議シーンが円または球としてモデル化される場合、円または球の中心に) 位置される。上流オーディオ信号は、会議参加者の端末 (たとえばコンピューティング装置または電話装置) において生成されたオーディオ信号であってもよい。よって、上流オーディオ信号は典型的には会議参加者の発話信号を含む。この理由により、上流オーディオ信号は語り手オーディオ信号と称されてもよい。会議コントローラは、 (たとえばいわゆる中央集中式の会議アーキテクチャでは) 通信ネットワーク内に (中央位置において) 位置されてもよく、および／または会議コントローラは (たとえばいわゆる分散式の会議アーキテクチャでは) 会議参加者の端末のところに位置されてもよい。会議コントローラは、2Dまたは3Dレンダリング・システムを使う場合、シーン・マネージャと称されてもよい。会議コントローラは、コンピューティング装置 (たとえばサーバー) を使って実装されてもよい。

30

【0006】

会議コントローラは、会議シーン内のX個の異なる空間的な語り手位置をもつX点会議シーンをセットアップするよう構成されていてもよい。Xは整数であり、 $X > 0$ である (たとえば、 $X > 1$ 、特に $X = 1, 2, 3, 4, 5, 6, 7, 8$ または10)。このコンテキストにおいて、会議コントローラは、本稿に記載される会議シーン・デザイン・ルールの一つまたは複数に基づいてX個の異なる空間的語り手位置をもつX点会議シーンを計算するよう構成されていてもよい。一つのそのようなデザイン・ルールは、たとえば、X個の語り手位置は、聴取者の頭部の前方の中央線のまわりの円錐内に位置されるというものであってもよい。他のデザイン・ルールは、X個の語り手位置の角度的な分離に関係していてもよい。代替的または追加的に、会議コントローラは、X個の異なる空間的語り手位置をもつX点会議シーンを、あらかじめ決定された話者位置を含むあらかじめ決定された会議シーンの集合から選択するよう構成されていてもよい。例として、前記集合は、X個の異なるあらかじめ決定された空間的語り手位置をもつ一つまたは複数のあらかじめ決定されたX点会議シーンを含んでいてもよい。よって、X点会議シーンは、X個のあらかじめ決定された話者位置をもつあら

40

50

かじめ決定されたX点会議シーンであってもよい。

【0007】

会議コントローラは、種々の会議シーン（たとえば、異なる配置にされた語り手位置をもつ異なるX点会議シーンおよび／またはXの異なる値をもつ会議シーン）をセットアップするよう構成されていてもよい。X点会議シーンのX個の語り手位置は、聴取者の頭部の正面の中央線のまわりの円錐内に位置される。中央線は、聴取者の両耳の間の架空の線の中点に始まり、聴取者の両耳の間の前記架空の線に垂直に、聴取者の頭部の前方に延びる架空の直線であってもよい。円錐の母線と前記中央線は、あらかじめ決定された最大円錐角以下の角度（絶対値）をなす。最大円錐角は、円錐の分布数に依存して、好ましくは30°または20°またはさらには15°などより狭い角であってもよい。

10

【0008】

会議コントローラはさらに、上流オーディオ信号をX個の語り手位置の一つに割り当てるよう構成されている。上流オーディオ信号を会議シーン内の特定の語り手位置に割り当てることによって、会議コントローラは、レンダリング装置（たとえば会議シーンの聴取者の端末）が上流オーディオ信号を、あたかも該上流オーディオ信号がその特定の語り手位置から発しているかのようにレンダリングできるようにする。この目的のために、会議コントローラは、割り当てられた語り手位置を同定するメタデータを生成し、（聴取者の端末における）オーディオ処理ユニットが上流オーディオ信号に基づいて立体音響化されたオーディオ信号を生成できるようにするよう構成されている。立体音響化されたオーディオ信号を聴取者に対してレンダリングするとき、聴取者は上流オーディオ信号を、割り

20

【0009】

X個の語り手位置は、円または球状に配置されてもよく、該円または球の中心に聴取者が配置される。代替的な会議シーンは、楕円または楕円体上には位置される語り手位置を有していてもよい。聴取者は必ずしも中心に配置される必要はない。例として、テーブルのまわりの、会議参加者および聴取者を含む会合をシミュレートするために、聴取者は会議シーンをなす幾何学形状のエッジに、たとえば円または球または楕円または楕円体の端に位置されてもよい。後者の場合（および聴取者が楕円または楕円体の中心に位置される場合）、X個の語り手位置と聴取者の間の距離は語り手位置に依存して異なることになる。

30

【0010】

X個の語り手位置のうち二つの隣接する語り手位置は、少なくともある最小の角距離だけ離間していてもよい。最小の角距離は5°またはそれ以上であってもよい。上述した条件は、X個の語り手位置の隣接する語り手位置のすべての対によって充足されてもよい。最小の角距離は、聴取者が、異なる語り手位置からレンダリングされる上流オーディオ信号を明瞭に区別できるようにする。X個の語り手位置のうちの隣り合う語り手位置の間の角距離は、異なる語り手位置については異なってもよい。例として、X個の語り手位置のうちの隣り合う語り手位置の間の角距離は、該隣り合う語り手位置の中央線からの距離が増すにつれて増大してもよい。こうすることによって、聴取者が異なる角度から到来する音の源を区別する能力の変化が考慮に入れられうる。

40

【0011】

X個の語り手位置は中央線に対して非対称であってもよい。これは、会議シーンを聴取者に対してレンダリングするときに像形成〔イメージング〕効果を回避するために有益でありうる。しかしながら、Xが奇数の場合、X個の語り手位置は、該X個の語り手位置のうちの中央語り手位置に対して対称的であってもよい。ここで、中央語り手位置は、典型的

50

には中央線上にはない。中央語り手位置に対する語り手位置の対称性を使うことにより、聴取者の左耳と右耳の間の会議シーンの均衡が達成できる。

【0012】

Xが偶数であれば、X個の語り手位置の半分が中央線の一方の側または中央線を通る鉛直面の一方の側にあってもよく、X個の語り手位置の残りの半分が中央線の反対側または中央線を通る鉛直面の反対側にあってもよい。Xが奇数なら、X個の語り手位置のうち $(X-1)/2$ 個が一方の側にあり、 $(X+1)/2$ 個が前記中央線または前記鉛直面の反対側にあってもよい。この条件は、均衡した会議シーンを目標とする。

【0013】

X個の語り手位置は、聴取者の頭部からのX本の射線上に位置されてもよい。X本の射線はそれぞれ中央線とともにX個の語り手角をなす。中央線からある方向の諸語り手角は正であってもよく、中央線から反対方向の諸語り手角は負であってもよい。X個の語り手角の和は0と異なってもよい。しかしながら、語り手角の和は 2° 以下であってもよい。よって、会議シーンは、会議シーンの中央線に対してわずかに非対称的であり、同時に十分均衡していてもよい。

10

【0014】

会議シーンの負荷はわずかに不均衡であってもよく、中央線のどちら側が最初に分布させられるかに依存して、シーンの一方の側に（すなわち、聴取者の頭部の正面の中央線から一方の方向に）「傾く」傾向があってもよいことを注意しておくべきである。聴取者が電話通話のために自然に使うであろう側でシーンに分布させるのを開始するのがより自然であることがある。この側は、典型的には、聴取者が書くときに使う手と反対側である。たとえば、右利きの人は電話を左手で保持し、左側で（左側の耳を用いて）モノ・ストリームを聞くことに慣れている。ユーザー選好がシーン・マネージャ（たとえば、端末またはエンドポイントに位置される）に通知するために使われてもよい。さらに、聴取者の利き手を決定するために、追加的なアクセサリが問い合わせされることができる。たとえば、コンピュータのマウス設定である。

20

【0015】

よって、会議コントローラは、聴取者の好まれる側を決定するよう構成されていてもよい。ここで、好まれる側は、聴取者の頭部の正面の中央線から左側または右側であってもよい。この目的のために、会議コントローラは、聴取者の端末から前記好まれる側に関する指示を受領するよう構成されていてもよい。前記好まれる側に関する指示は、聴取者の端末における選好設定（たとえばマウス設定）に基づいていてもよい。聴取者の前記好まれる側は、聴取者の左利き／右利きと相関付けられてもよい。上記のように、左利きの聴取者は典型的には好まれる右側をもち、右利きの聴取者は典型的には好まれる左側をもつ。会議コントローラは、到来するオーディオ信号を、X個の語り手位置のうち聴取者の前記好まれる側にある一つに、たとえば、聴取者の前記好まれる側の最も中央寄りの語り手位置に割り当てるよう構成されていてもよい。

30

【0016】

ある実施形態では、 $X=3$ であり、X個の語り手位置は中央線から実質的に -2° 、 $+6^\circ$ および -10° の語り手角のところに位置される。あるさらなる実施形態では、 $X=6$ であり、X個の語り手位置は、中央線から実質的に -2° 、 $+3^\circ$ 、 -7° 、 $+8^\circ$ 、 -12° および $+13^\circ$ の語り手角のところに位置される。ある好ましい実施形態では、 $X=6$ であり、X個の語り手位置は中央線から実質的に $+2^\circ$ 、 -5° 、 $+9^\circ$ 、 -12° 、 $+16^\circ$ および -19° の語り手角のところに位置される。

40

【0017】

典型的には、会議コントローラは、複数の会議参加者に関連付けられた複数の上流オーディオ信号をX点会議シーン内に配置するよう構成されている。よって、会議シーンの聴取者は、会議シーン内の異なる（X個までの異なる）空間的位置から到来する前記複数の上流オーディオ信号を知覚することができるようにされてもよい。会議コントローラは、前記複数の上流オーディオ信号を、前記複数の上流オーディオ信号のある逐次順に従って

50

、前記X個の語り手位置に割り当てるよう構成されていてもよい。前記逐次順は、上流オーディオ信号を会議内に配置するための会議コントローラの待ち行列を参照してもよい。代替的または追加的に、前記複数の上流オーディオ信号の前記逐次順は、会議コントローラによる前記複数の上流オーディオ信号の検出の順序に基づいていてもよい。換言すれば、前記複数の上流オーディオ信号の前記逐次順は、異なる会議参加者がオーディオ会議にダイヤルしてきて、それにより対応する上流オーディオ信号が会議コントローラによって検出される逐次順に影響する順序に関連していてもよい。第一の、最も中央の位置は、源エンドポイントからの（すなわち、会合まとめ役または司会者の端末からの）メタデータを通じて識別される、当該会議通話の会合まとめ役または司会者のために予約されていてもよい。したがって、たとえ他の会議参加者が司会者より前に会合に到着するとしても、それらの会議参加者はシーン内のプレミアム位置に割り振られなくてもよい。よって、会議コントローラは、ある特定の端末から（たとえば、まとめ役／司会者の端末から）到来する特定の流オーディオ信号について、X個の語り手位置のうちの一つを予約するよう構成されていてもよい。予約される語り手位置は、最も中央寄りの語り手位置であってもよい。

10

【0018】

会議コントローラは、中央線からの絶対的な角距離が増す順にX個の語り手位置を割り当てるよう構成されていてもよい。換言すれば、会議コントローラは第一の上流オーディオ信号を最も中央寄りの語り手位置に割り当て、第二の上流オーディオ信号を二番目に中央寄りの語り手位置に割り当てる、などとして最も外側の語り手位置に到達するまで続けてもよい。前記複数の上流オーディオ信号からの上流オーディオ信号を、前記X個の語り手位置のうちからの最も外側の語り手位置に割り当てることを受けて、会議コントローラは、前記複数の上流オーディオ信号のうちからの次の上流オーディオ信号を、前記X個の語り手位置のうちからの最も内側の（最も中央寄りとも称される）語り手位置に割り当てるよう構成されていてもよい。

20

【0019】

よって、会議コントローラは、前記複数の上流オーディオ信号のうちからの複数の上流オーディオ信号を、前記X個の語り手位置のうち少なくとも一つに割り当てるよう構成されていてもよい。会議コントローラは、特に、会議シーン内に配置されるべき上流オーディオ信号の数(M-1)が語り手位置の数Xよりも多い場合にそうするよう構成されていてもよい。円錐の分布数がX個の語り手位置を超過する場合、3D会議シーン（および聴取者の端末における3Dレンダリング・エンジン）が使われているならば、語り手の追加的な鉛直方向の諸層がシーン内に配置されてもよい。たとえば、最初の6人の参加者は仰角0度（耳の水準）にある6個の語り手位置に割り当てられてもよく、次の6人の参加者は最初の6人の参加者と同じ方位角だがたとえば8°の仰角に割り当てられてもよい。数Mはアクティブな会議参加者の総数であってもよく、そのうち1人は特定の会議シーンの聴取者であり、他の(M-1)人はX点会議シーン内でX個の語り手位置に配置されるべき語り手である。複数の上流オーディオ信号が同じ語り手位置に割り当てられる場合、会議コントローラは、その同じ語り手位置に割り当てられた複数の上流オーディオ信号の混合を開始するよう構成されていてもよい。それにより、その語り手位置においてレンダリングされる混合されたオーディオ信号が生成される。上流オーディオ信号の実際の混合は、（たとえばデジタル信号プロセッサを有する）オーディオ・サーバーによって、あるいは聴取者のエンドポイントにおける（すなわち端末における）オーディオ・レンダリング・エンジン内で実行されてもよい。オーディオ・サーバーは、会議コントローラとは別個であってもよい。

30

40

【0020】

特に、会議シーン内に配置される上流オーディオ信号の数(M-1)がXより小さい場合、会議コントローラは、前記複数の上流オーディオ信号のそれぞれを、前記X個の語り手位置のうち異なる位置に割り当てるよう構成されていてもよい。典型的には、上流オーディオ信号は、前記X個の語り手位置のうちの単一の位置に割り当てられるのみである。場合によっては、空間的な広がりを作り出すために、単一のストリームが二つの隣接する

50

語り手位置に割り当てられてもよい。

【0021】

会議コントローラは、以下の配置ルールの一つまたは複数に従って前記複数の上流オーディオ信号を配置するよう構成されていてもよい。第一の配置ルールは、前記複数の上流オーディオ信号からの次の上流オーディオ信号を、前記X個の語り手位置のうちの、まだ割り当てられていない語り手位置であって、中央線にできるだけ近い語り手位置に割り当てるというものであってもよい。換言すれば、上流オーディオ信号に（すなわち、会議参加者に）専用の語り手位置を提供することが好ましい、および／または上流オーディオ信号を会議シーンの中央線のできるだけ近くに割り当てることが好ましい。第二の配置ルールは、割り当てられた諸語り手位置の、中央線に対するおよび／または前記X個の語り手位置のうち最も中央寄りの語り手位置に対する最大の均衡を保証するというものであってもよい。換言すれば、割り当てられた上流オーディオ信号が会議シーンの中央線のまわりに均衡した仕方でも割り当てられることを保証することが好ましいことがある。第三の配置ルールは、前記X個の語り手位置のうちからの空の語り手位置を割り当て直すというものであってもよい。ここで、前記空の語り手位置は、前記複数の上流オーディオ信号のうちからのすでに割り当てられた上流オーディオ信号の中断によって空にされたものである。中断は、たとえば、会議参加者が会議シーンを去ったこと、あるいは該参加者が話していないまたは聞くだけの状態に後退したことであってもよい。これは、会議コントローラによって、会議シーンを去った会議参加者の端末から受領される適切な信号によって検出されてもよい。

10

20

【0022】

会議コントローラは、立体音響化されたオーディオ信号をレンダリングするオーディオ・トランシーバのレンダリング特性を決定するよう構成されていてもよい。例として、会議コントローラは、聴取者の端末におけるオーディオ・トランシーバが、バイノーラル・オーディオ信号、ステレオ・オーディオ信号またはサラウンドサウンド・オーディオ信号またはモノ信号のみをレンダリングできることを判別するよう構成されていてもよい。会議コントローラは、下流オーディオ信号の組と、聴取者の端末におけるオーディオ・トランシーバが会議シーンを適切にレンダリングできるようにする適切なメタデータとを生成するよう構成されていてもよい。下流オーディオ信号の組は典型的には前記上流オーディオ信号または前記複数の上流オーディオ信号を含む。メタデータは典型的には、前記上流オーディオ信号または前記複数の上流オーディオ信号の、X点会議シーン内のその配置に従った立体音響化されたレンダリングを許容する情報を含む。

30

【0023】

例として、会議コントローラは、オーディオ・トランシーバのレンダリング特性に依存して前記複数の上流オーディオ信号の一つまたは複数の混合を開始するよう構成されていてもよい。特に、会議コントローラは、レンダリング特性がオーディオ・トランシーバがモノ・オーディオ信号のレンダリングに限定されていることを示す場合には、会議シーンに配置されるべき前記複数の上流オーディオ信号の全部の混合を開始するよう構成されていてもよい。オーディオ信号の実際の混合は、たとえばデジタル信号プロセッサを有するオーディオ・サーバーによって実行されてもよい。

40

【0024】

本稿で概説された諸側面は、2D会議シーンおよび／または3D会議シーンに関係していてもよい。よって、語り手位置の角度配列に関する諸側面は、方位角および／または傾斜角に関係していてもよい。方位角は、中央線を含む水平面内で定義されてもよい。傾斜角は、中央線を含む鉛直面内で定義されてもよい。2D会議シーンでは、X個の異なる空間的語り手位置は、中央線に対して異なる方位角のところに位置されてもよい。3D会議シーンでは、X個の異なる空間的語り手位置の少なくともいくつかは、中央線に対して異なる方位角および／または傾斜角のところに位置されてもよい。

【0025】

3D会議シーンでは、X個の異なる空間的語り手位置は、異なる傾斜角にある複数の層に

50

配列されてもよい。各層は、前記X個の空間的語り手位置のうちからの、異なる方位角にある複数の異なる空間的語り手位置を含んでいてもよい。特に、第一の層における異なる空間的語り手位置と第二の層における異なる空間的語り手位置は対応する方位角を有していてもよく、それにより規則的な層構成にされた構造を作り出してもよい。

【0026】

会議コントローラは、上流オーディオ信号の地位〔ステータス〕に関する情報を受領するよう構成されていてもよい。地位は、上流オーディオ信号に関連したメタデータとして伝送されてもよい。地位は、たとえば、その上流オーディオ信号が会議の司会者またはまとめ役に対応することを指示してもよい。換言すれば、地位は、その上流オーディオ信号が司会者またはまとめ役の発話信号を含むことを示してもよい。代替的または追加的に、地位は、プレミアムまたは基本地位（たとえば、異なる会議参加者によって支払われる異なる料金に係する）に係してもよい。

10

【0027】

会議コントローラは、上流オーディオ信号を、該上流オーディオ信号の地位に基づいて、空間的語り手位置に割り当てるよう構成されていてもよい。例として、地位は、会議のまとめ役の上流オーディオ信号を示してもよく、その上流オーディオ信号が割り当てられる空間的語り手位置は、最も中央寄りの空間的語り手位置であってもよい。もう一つの例では、地位は、基本地位（すなわち、基本的な会議参加者）を示してもよい。基本的な会議参加者は、それほど顕著でない語り手位置（たとえば、最も外側の語り手位置または異なる層の語り手位置）に割り当てられてもよい。

20

【0028】

会議コントローラは、あらかじめ決定された地位をもつ上流オーディオ信号のために前記X個の空間的語り手位置のうちの一つまたは複数を予約するよう構成されていてもよい。例として、最も中央寄りの語り手位置が、会議のまとめ役のために、あるいはプレミアム地位をもつ会議参加者のために予約されてもよい。

【0029】

あるさらなる側面によれば、オーディオ会議システムが記述される。本オーディオ会議システムは、（たとえばマイクロホンを使って会議参加者の発話信号を記録することによって）会議参加者に関連付けられた上流オーディオ信号を生成するよう構成された語り手端末を有する。本システムはさらに、本稿で概説される諸側面のいずれかに基づく会議コントローラを有する。会議コントローラは、2Dまたは3D会議シーン内の語り手位置に上流オーディオ信号を割り当てるよう構成されている。さらに、会議コントローラは、割り当てられた語り手位置を同定するメタデータを生成するよう構成されている。さらに、本システムは、メタデータを使ってオーディオ会議の聴取者に上流オーディオ信号をレンダリングするよう構成された聴取者端末を有する。レンダリングは、聴取者が上流オーディオ信号を、割り当てられた語り手位置から到来するものとして知覚するように実行されてもよい。

30

【0030】

もう一つの側面によれば、会議参加者に関連する上流オーディオ信号を、聴取者に対してレンダリングされるべき2Dまたは3D会議シーン内で配置する方法が記述される。本方法は、会議シーン内でX個の異なる空間的語り手位置をもつX点会議シーンをセットアップすることを含む。Xは整数であり、 $X > 0$ である。X個の語り手位置は、聴取者の頭部の前方の中央線のまわりの円錐内に位置される。円錐の母線と前記中央線が、あらかじめ決定された最大円錐角以下の角度をなす。さらに、本方法は、上流オーディオ信号をX個の語り手位置の一つに割り当てることを含む。さらに、本方法は、割り当てられた語り手位置を同定するメタデータを生成し、オーディオ処理ユニットが上流オーディオ信号に基づいて立体音響化されたオーディオ信号を生成できるようにすることを含む。立体音響化されたオーディオ信号を聴取者に対してレンダリングするとき、聴取者は、上流オーディオ信号が、割り当てられた語り手位置から到来するものとして知覚する。

40

【0031】

50

あるさらなる側面によれば、ソフトウェア・プログラムが記述される。ソフトウェア・プログラムは、プロセッサ上での実行のために、該プロセッサで実行されたときに本稿で概説される方法段階を実行するために適応されていてもよい。

【0032】

もう一つの側面によれば、記憶媒体が記述される。本記憶媒体は、プロセッサ上での実行のために、コンピューティング装置上で実行されたときに本稿で概説される方法段階を実行するために適応されたソフトウェア・プログラムを有していてもよい。

【0033】

あるさらなる側面によれば、コンピュータ・プログラム・プロダクトが記述される。コンピュータ・プログラムは、コンピュータ上で実行されたときに本稿で概説される方法段階を実行するための実行可能命令を有していてもよい。

【0034】

本特許出願において概説される好ましい実施形態を含む方法およびシステムは単独で、または本稿で開示される他の方法およびシステムとの組み合わせにおいて使用されうることを注意しておくべきである。さらに、本特許出願において概説される方法およびシステムのあらゆる側面は任意に組み合わせられうる。特に、請求項の特徴は、互いに、任意の仕方

【図面の簡単な説明】

【0035】

本発明について、付属の図面を参照しつつ例示的な仕方

【図1】 aは、例示的な中央集中式のオーディオ会議システムのブロック図であり、bは、例示的な分散式のオーディオ会議システムのブロック図である。

【図2】 オーディオ会議システムのシーン・マネージャのための例示的なグラフィカル・ユーザー・インターフェース (GUI) を示す図である。

【図3 a】 例示的なオーディオ会議シーンを示す図である。

【図3 b】 例示的なオーディオ会議シーンを示す図である。

【図4】 オーディオ会議シーンの例示的なクラスターを示す図である。

【発明を実施するための形態】

【0036】

導入部で概説したように、現行の多者オーディオ会議システムは典型的にはオーディオ会議に参加している複数の当事者のオーディオ信号をモノラル・オーディオ信号に重畳してしまい、それが各参加者に対して単一のオーディオ・ストリームとして与えられる。このため、参加者にとって（聞いているとき）、（他の複数の当事者が話しているとき）他の参加者を互いに区別することが難しくなる。本稿では、オーディオ会議の複数の当事者の空間化を許容する多者オーディオ会議システムであって、オーディオ会議の異なる参加者を、二次元 (2D) または三次元 (3D) オーディオ・シーン内の異なる空間的位置に配置することを許容するものが記述される。結果として、聴取者は、他の参加者を、異なるそれぞれの空間的位置から話しているものとして知覚する。それにより、聴取者は他の参加者をよりよく区別できるようになる。

【0037】

図1のaは、中央集中式のアーキテクチャをもつ例示的な多者オーディオ会議システム100を示している。中央集中式の会議サーバー110が、複数の上流オーディオ信号123をそれぞれの複数の端末120から受領する。上流オーディオ信号123は典型的にはオーディオ・ストリーム、たとえばビットストリームとして伝送される。例として、上流オーディオ信号123はG.711、G722.2 (AMR-WB)、MPEG2またはMPEG4オーディオ・ビットストリームとしてエンコードされてもよい。典型的には、上流オーディオ信号123はモノ・オーディオ信号である。よって、中央集中式会議サーバー110（たとえば会議サーバー110内に含まれる諸オーディオ・サーバー112）は、（上流オーディオ信号123を表わす）上流オーディオ・ストリームをデコードし、上流オーディオ・ストリームに付随する任意的なメタデータを抽出するよう構成されていてもよい。

【0038】

会議サーバー110はたとえば、遠隔通信ネットワーク内のオーディオ会議サービス・プロバイダーのアプリケーション・サーバーであってもよい。端末120はたとえば、ラップトップ・コンピュータ、デスクトップ・コンピュータ、タブレット・コンピュータおよび／またはスマートフォンといったコンピューティング装置や、携帯電話、コードレス電話、デスクトップ・ハンドセットなどといった電話であってもよい。会議サーバー110は、オーディオ会議への／からの複数の上流オーディオ信号123を組み合わせるよう構成されている中央会議コントローラ111を有する。中央会議コントローラ111は、複数の上流オーディオ信号123を2Dまたは3D会議シーン内の特定の諸位置に配置し、会議シーン内での前記複数の上流オーディオ信号123の配列（すなわち諸位置）に関する情報を生成するよう構成されていてもよい。

10

【0039】

さらに、会議サーバー110は、それぞれ前記複数の端末120のための複数のオーディオ・サーバー112を有する。前記複数のオーディオ・サーバー112は単一のコンピューティング装置／デジタル信号プロセッサ内に設けられてもよいことを注意しておくべきである。複数のオーディオ・サーバー112はたとえば、それぞれの複数の端末120のためのオーディオ信号にサービスするための、前記サーバー内の専用の処理モジュールまたは専用のソフトウェア・スレッドであってもよい。よって、オーディオ・サーバー112は、それぞれの端末120の必要性に従ってオーディオ信号を処理する「論理的な」エンティティであってもよい。オーディオ・サーバー112（または組み合わされたサーバー内の等価な処理モジュールまたはスレッド）は（たとえばオーディオ・ストリームの形の）複数の上流オーディオ信号123の一部または全部ならびに会議シーン内での前記複数の上流オーディオ信号123の配列に関する前記情報を受領する。会議シーン内での前記複数の上流オーディオ信号123の配列に関する前記情報は、典型的には、会議コントローラ111によって与えられる。それにより会議コントローラ111はオーディオ・サーバー112（または処理モジュール／スレッド）に、オーディオ信号をどのように処理すべきかについて通知する。この情報を使って、オーディオ・サーバー112は、一組の下流オーディオ信号124および対応するメタデータを生成し、それが、それぞれの端末120に伝送される。それぞれの端末120が参加者のオーディオ信号を、会議コントローラ111内で確立された会議シーンに従ってレンダリングできるようにするためである。前記一組の下流オーディオ信号124は典型的には、一組の下流オーディオ・ストリーム、たとえばビットストリームとして伝送される。例として、前記一組の下流オーディオ信号124は、G.711、G722.2（AMR-WB）、MPEG2またはMPEG4または独自のオーディオ・ビットストリームとしてエンコードされてもよい。会議シーン内の下流オーディオ信号124の配置に関する前記情報は、たとえば前記一組の下流オーディオ・ストリーム内で、メタデータとしてエンコードされてもよい。よって、会議サーバー110（特にオーディオ・サーバー112）は、前記一組の下流オーディオ信号124を、端末120において会議シーンをレンダリングするためのメタデータを含む一組の下流オーディオ・ストリームにエンコードするよう構成されていてもよい。メタデータは別個のストリームとして、たとえば下流オーディオ・ストリームとの同期のためのタイムスタンプとともに、端末120に送られてもよいことは注意しておくべきである。これは、メタデータを必要としないまたはどのようにメタデータをデコードすべきかを知らない端末120でも基本的な下流オーディオ・ストリームを処理できる（そしてオーディオ信号を端末120のところにいる聴取者に対してレンダリングできる）ということを意味する。換言すれば、前記一組の下流オーディオ信号124および前記メタデータは、2Dまたは3D会議シーンのレンダリングに対応していない端末120でも（たとえば混合形式の）下流オーディオ信号をレンダリングするために使用されよう、後方互換な仕方でもエンコードされてもよい。

20

30

40

【0040】

よって、オーディオ・サーバー112は、複数の下流オーディオ・ストリームおよび／または複数の下流オーディオ信号ならびに会議シーンを記述するメタデータを生成するた

50

めに、複数の上流オーディオ・ストリームおよび／または複数の上流オーディオ信号の（たとえばデジタル信号プロセッサを使った）実際の信号処理を実行するよう構成されていてもよい。オーディオ・サーバー 112 は、（図 1 の a に示されるように）対応する端末 120 の専用であってもよい。あるいはまた、オーディオ・サーバー 112 は、複数の端末 120 のために、たとえば全部の端末 120 のために信号処理を実行するよう構成されていてもよい。

【0041】

端末 120 の上流オーディオ信号 123 は、端末 120 のところで話している、たとえば端末 120 のマイクロホンに話している会議参加者によって生成されるオーディオ信号を含むので、語り手オーディオ信号 123 と称されてもよいことを注意しておくべきである。同様に、端末 120 に送られる一組の下流のオーディオ信号 124 は、一組の聞き手オーディオ信号 124 と称されてもよい。該組 124 は、端末 120 のところにいる参加者がたとえばヘッドホンまたはラウドスピーカーを使って聞く複数のオーディオ信号を含むからである。

【0042】

ある特定の端末のための一組の下流オーディオ信号 124 は、前記複数の上流オーディオ信号 123 から中央会議コントローラ 111 およびオーディオ・サーバー 112、たとえばその特定の端末 120 のためのオーディオ・サーバー 112（または処理モジュールまたはソフトウェア・スレッド）を使って生成される。中央会議コントローラ 111 およびオーディオ・サーバー 112 は、その特定の端末 120 のところにいる会議参加者によって知覚されるべきように 2D または 3D 会議シーンの像を生成する。オーディオ会議の M 人の参加者がいれば、すなわち会議サーバー 110 に接続された M 個の端末 120 があれば、会議サーバー 110 は、(M-1) 個の上流オーディオ信号 123 のグループ M 個を、M 個の 2D または 3D 会議シーン内に配列するよう構成されていてもよい（M は $M > 2$ である整数であり、たとえば $M > 3, 4, 5, 6, 7, 8, 9, 10$ ）。より精密には、会議サーバー 110 は、M 個の端末 120 について M 個の会議シーンを生成するよう構成されていてもよい。ここで、各端末 120 について、残りの (M-1) 個の他の上流オーディオ信号 123 が 2D または 3D 会議シーン内に配列される。

【0043】

例として、会議サーバー 110 は、2D または 3D の空間的配列内での M 人の会議参加者の配列を記述するマスター会議シーンを利用してもよい。会議サーバー 110 は、M 人の会議参加者について（すなわち、M 人の会議参加者それぞれの M 個の端末 120 について）マスター会議シーンの異なる観点を生成するよう構成されていてもよい。そうすることによって、会議参加者全員が、他の会議参加者がどこに配置されているかの同じ相対的なビューをもつことが保証できる。これは、M 人の会議参加者が「テーブルのまわりに」（たとえば円）位置されている場合および M 個の「個々の」会議シーンにおける聴取者が「テーブル」の端に（たとえばその円上に）位置されている場合に特にそうである。

【0044】

もう一つの例として、会議サーバー 110 は、会議シーンへの会議参加者の逐次的な到着に従って会議シーンの語り手位置を割り当ててもよい。語り手位置は、本稿で記述されるように、最も内側の語り手位置から最も外側の語り手位置へと割り当てられてもよい。そのような逐次順の結果、会議参加者は、異なる聴取者のために予定されている諸会議シーン内で、同じ語り手位置に配置されうる。会議サーバーは、計算量を節約するために、諸会議シーンのこの類似性（または同一性）を利用してもよい。

【0045】

さらなる例では、会議サーバー 110 は、M 人の会議参加者全員について単一の会議シーンを利用し、M 人の会議参加者全員を単一の会議シーン内のあらかじめ決定された語り手位置に配置してもよい。そのような場合、M 人の会議参加者全員は同じ 2D または 3D 会議シーンを知覚することになる。特定の端末 120 にその単一の会議シーンを提供するとき（特定の会議参加者が聴取者となる）、その特定の会議参加者の語り手位置は空のままに

10

20

30

40

50

されてもよい。この例は、会議サーバー 110 による単一の会議シーンの管理を要求するだけなので、実装するのが効率的でありうる。

【0046】

M 個の会議シーンは典型的には、M 人の参加者が会議シーンの中心内に配置されるという点で異なっている。例として、第一の端末 120 のための会議シーンは典型的には、第一の端末 120 がその会議シーンの中心にあり、他の (M-1) 個の端末が第一の端末 120 のまわりに配置されると想定する。よって、第一の端末 120 についてのオーディオ・サーバー 112 は、第一の端末 120 からの上流オーディオ信号 123 以外の (M-1) 個の上流オーディオ信号 123 から、(M-1) 個までの下流オーディオ信号 124 の組（および対応するメタデータ）を生成する。下流オーディオ信号 124 のこの端末固有の組は、第一の

10

【0047】

ある実施形態では、中央会議コントローラ 111 はオーディオ会議の制御を受け持つ。ここで、諸オーディオ・サーバー 112 が上流オーディオ信号 123 を操作し、それらの対応する端末 120 のための下流オーディオ信号 124 の諸組を、中央会議コントローラ 111 の制御のもとで、生成する。例として、中央会議コントローラ 111 は、ペアラール情報（すなわち、上流オーディオ信号 123 内の実際のオーディオ・データ）を処理しなくてもよく、信号伝達情報（たとえば、呼び出される当事者および呼び出し側当事者のアドレッシング情報、端末 120 の機能など）を処理してもよい。中央会議コントローラ 111 は、オーディオ会議をセットアップするためにその信号伝達情報を使ってもよい。上流オーディオ信号 123 の実際の混合、下流オーディオ信号 124 の組の生成、端末固有の会議シーンを定義する適切なメタデータの生成およびオーディオ・ビットストリームからの／へのオーディオ信号のデコード／エンコードは、オーディオ・サーバー 112 によって、たとえばデジタル信号プロセッサを使って処理されてもよい。

20

【0048】

端末 120 は、下流オーディオ信号 124 の端末固有の組（および対応するメタデータ）を受領し、該下流オーディオ信号 124 の組をオーディオ・トランシーバ 122（たとえば、ヘッドホンまたはラウドスピーカー）を介してレンダリングする。この目的のために、端末 120（たとえば、端末 120 内に含まれるオーディオ処理ユニット 121）が、下流オーディオ信号および対応するメタデータを抽出するために、下流オーディオ・ビットストリームの組をデコードするよう構成されていてもよい。さらに、オーディオ処理ユニット 121 は、オーディオ・トランシーバ 122 によってレンダリングするための混合バイノーラル・オーディオ信号を生成するよう構成されていてもよい。ここで、混合バイノーラル・オーディオ信号は、この端末 120 について会議サーバー 110 においてデザインされた端末固有の会議シーンを反映する。換言すれば、オーディオ処理ユニット 121 は、受領されたメタデータを解析し、下流オーディオ信号 124 の受領された組を端末固有の会議シーン中に配置するよう構成されていてもよい。結果として、会議参加者はバイノーラル・オーディオ信号を知覚し、該バイノーラル・オーディオ信号は端末 120

30

40

【0049】

下流オーディオ信号 124 のそれぞれについてのバイノーラル・オーディオ信号の生成は、立体音響化アルゴリズムを通じて（モノ）下流オーディオ信号を処理することによって実行されてもよい。そのようなアルゴリズムは、左および右耳信号を与えるために、一対の頭部伝達関数（HRTF）を使ってその下流オーディオ信号のサンプルをフィルタリングすることであることができる。HRTF は、空間中の特定の位置に位置された（下流オーディオ信号の）音源と聴取者の耳との間で自然に生じたであろうフィルタリングを記述する。HRTF は、両耳間時間差、両耳間レベル差およびスペクトル手がかりといった、音のバイ

50

ノーラル・レンダリングのためのあらゆる手がかりを含む。HRTFは、音源の位置に（すなわち、下流オーディオ信号の語り手位置に）依存する。会議シーン内の各特定の位置について、HRTFの異なる固有の対が使われてもよい。あるいはまた、特定の位置についてのフィルタリング特性は、HRTFが利用可能な隣り合う位置の間の補間によって生成されることができる。よって、端末120は、下流オーディオ信号の語り手位置を、付随するメタデータから識別するよう構成されていてもよい。さらに、端末120は、識別された語り手位置について適切なHRTFの対を決定するよう構成されていてもよい。さらに、端末120は、下流オーディオ信号にHRTFの対を適用し、それにより識別された語り手位置から到来するものとして知覚されるバイノーラル・オーディオ信号を与えるよう構成されていてもよい。端末120が下流オーディオ信号123の組内の二つ以上の下流オーディオ信号を受領する場合、上記の処理は、それらの下流オーディオ信号のそれぞれについて実行されてもよく、結果として得られるバイノーラル信号が重畳されて組み合わせられたバイノーラル信号を与えてもよい。

10

【0050】

混合バイノーラル・オーディオ信号の生成の代わりにまたはそれに加えて、端末120（たとえばオーディオ処理ユニット121）は、適切に配置されたラウドスピーカー122を使って端末120においてレンダリングされうるサラウンドサウンド（たとえば5.1または7.1サラウンドサウンド）信号を生成するよう構成されていてもよい。さらに、端末120は、モノ・ラウドスピーカー122を使ってレンダリングするために、下流オーディオ信号124の組から混合オーディオ信号を生成するよう構成されていてもよい。

20

【0051】

単一の混合されたオーディオ信号が会議サーバーから端末に伝送される通常のモノラル・オーディオ会議システムとは対照的に、図1のaのオーディオ会議システム100では、(M-1)個までの下流オーディオ信号124の組および対応するメタデータが、会議サーバー110から各端末120に（たとえばビットストリームとして）伝送される。基礎になる通信ネットワークの帯域幅制限に鑑み、下流オーディオ信号124の組内で伝送されるオーディオ信号（たとえばビットストリーム）の数を制限することが有益であることがある。以下では、Nは特定の端末120についての下流オーディオ信号124の組内の下流オーディオ信号124の数であるとする。ここで、Nは整数で、たとえば $N < M$ である。Nは端末120および／または端末120と会議サーバー110との間の通信ネットワークに依存してもよいことを注意しておくべきである。すなわち、Nは、異なる端末120については異なってもよい。例として、端末120は会議サーバー110に無線ネットワークを介して接続される携帯電話であってもよい。そのような場合、その携帯電話に送信するための比較的少数、たとえば $N=1$ の下流オーディオ信号を選択する、あるいは下流のオーディオ・ストリームを生成するために使われるコーデックのパラメータを変更することが有益であることがある。

30

【0052】

上記で概説したように、会議サーバー110は、2Dまたは3D会議シーン内に配置されるM個までの上流オーディオ信号123を受領する。会議サーバー110は、複数の一連の時点において（および／または複数の一連のフレームにおいて）M個の上流オーディオ信号123の語り手活動の度合いを決定および解析してもよい。上流オーディオ信号の語り手活動の度合いは、上流オーディオ信号のエネルギー（たとえば平均平方エネルギー）に基づいていてもよい。会議参加者（および対応する上流オーディオ信号）は、（ある特定の時点において）「アクティブな」語り手として（該特定の時点において）「非アクティブな」語り手として分類されうる。この分類は、時点によって変わりうる。会議サーバー110は次いで、語り手活動度を考慮に入れることによって、特定の会議参加者についての（すなわち特定の端末120についての）下流オーディオ信号124の組（および付随するメタデータ）を決定してもよい。下流オーディオ信号124の組に挿入される上流オーディオ信号123の選択は、時点によって（語り手活動度の関数として）変わりうる。例として、会議サーバー110は、下流オーディオ信号124の組のために、特定の時点

40

50

における「アクティブな」語り手の上流オーディオ信号123のみを考慮するよう構成されていてもよい。

【0053】

語り手活動度を考慮に入れる結果として、会議サーバー110は、下流オーディオ信号124の諸組を種々の端末120に送信するための必要とされる帯域幅を軽減するよう構成されていてもよい。単一のアクティブな語り手の場合、下流オーディオ信号124の組は、単一のオーディオ信号（すなわち、アクティブな語り手の上流オーディオ信号）を含むだけでもよく、それにより会議サーバー110と端末120の間の通信リンクに対する帯域幅を著しく軽減しうる。下流オーディオ信号124の組はそれでも、アクティブな語り手（単数または複数）の空間位置を示すメタデータを含んでいてもよい（あるいはかかるメタデータに関連付けられていてもよい）。よって、端末120はそれでも、アクティブな語り手のオーディオ信号を立体音響化された仕方でレンダリングできるようにされてもよい。メタデータは、語り手活動の変化に伴って、時点によって変わりうる。よって、メタデータは、各時点において、下流オーディオ信号124の組内に反映されるアクティブな語り手（単数または複数）の空間配列を示しうる。

10

【0054】

帯域幅を軽減するさらなる方策として、源装置から（すなわち、端末120から）会議サーバー110への不連続的な伝送が使用されてもよい。例として、端末120は、端末120において記録されたオーディオ信号に基づいて、語り手活動度を決定するよう構成されていてもよい。語り手活動度が低い（たとえば、あらかじめ決定されたエネルギー閾値未満）場合、端末120は、端末120からサーバー110への上流オーディオ信号123の伝送を中断し、それにより必要とされる帯域幅を軽減するよう構成されていてもよい。よって、複数の語り手が、会議シーン内の同じ空間的位置に割り当てられてもよいが、それでも、衝突を引き起こすのはそれらの語り手が同時に話す場合のみであろう。

20

【0055】

図1のaは、中央集中式のアーキテクチャをもつ2Dまたは3D会議システム110を示している。2Dまたは3Dオーディオ会議は、図1のbの会議システム150によって示されるように、分散式のアーキテクチャを使って提供されてもよい。図示した例では、端末170は、会議参加者のオーディオ信号を混合する、および／またはオーディオ信号を会議シーン中に配置するよう構成されたローカル会議コントローラ175を有する。中央集中式の会議サーバー110の中央会議コントローラ111と同様に、ローカル会議コントローラ175は、会議シーンを生成するために受領されたオーディオ信号の信号伝達情報を解析することに限定されてもよい。オーディオ信号の実際の操作は、別個のオーディオ処理ユニット171によって実行されてもよい。

30

【0056】

分散式のアーキテクチャでは、端末170は、その上流オーディオ信号173を（たとえばビットストリームとして）他の参加端末170に通信ネットワーク160を介して送るよう構成されている。この目的のために、端末170は、マルチキャスト方式および／または他の参加端末170の直接アドレッシング方式を使ってもよい。よって、M個の参加端末170の場合、各端末170は、(M-1)個の他の端末170の上流オーディオ信号173に対応する(M-1)個までの下流オーディオ信号174を（たとえばビットストリームとして）受領する。受信端末170のローカル会議コントローラ175は、受領された下流オーディオ信号174を2Dまたは3D会議シーン中に配置するよう構成されている。ここで、受信端末170は典型的にはその会議シーンの中央に配置される。受信端末170のオーディオ処理ユニット171は、受領された下流オーディオ信号174から混合バイノーラル信号を生成するよう構成されている。ここで、混合バイノーラル信号は、ローカル会議コントローラ175によってデザインされた2Dまたは3D会議シーンを反映する。混合バイノーラル信号は、その後、オーディオ・トランシーバ122によってレンダリングされる。

40

【0057】

50

中央集中式の会議システム１００および分散式の会議システム１５０は組み合わせられてハイブリッド・アーキテクチャを形成してもよいことを注意しておくべきである。例として、（他のユーザーが端末１２０を使用しうる一方で）端末１７０が会議サーバー１１０との関連で使用されてもよい。ある例示的な実施形態では、端末１７０は下流オーディオ信号１２４の組（および対応するメタデータ）を会議サーバー１１０から受領する。端末１７０内のローカル会議コントローラ１７５は、会議サーバー１１０によって提供される会議シーンをデフォルト・シーンとしてセットアップしてもよい。加えて、端末１７０のユーザーは、会議サーバー１１０によって提供されるデフォルト・シーンを修正することができるようにされてもよい。

【００５８】

代替的または追加的に、会議サーバー１１０のコンポーネントは、たとえばオーディオ会議によって必要とされる帯域幅を軽減するために、ネットワーク内で分散されていてもよい。例として、中央会議コントローラ１１１が第一の位置（たとえば中央位置）に位置されてもよく、オーディオ・サーバー１１２はネットワーク内の一つまたは複数の他の位置に位置されてもよい。これは、オーディオ会議を扱うために必要とされる全体的なネットワーク容量を軽減する諸オーディオ・サーバー１１２の位置を選択するために有益であることがある。オーディオ会議の参加端末１２０の地域的分布に基づいてオーディオ・サーバー１１２を配置することがたとえば有益であることもある。オーディオ・サーバー１１２と中央会議コントローラ１１１との間の通信は、（実際のオーディオ・データを交換する必要なしに）信号伝達情報に限定されてもよい。

【００５９】

以下では、中央集中式の会議システム１００を参照する。しかしながら、本開示は分散式のアーキテクチャ１５０および任意のハイブリッド形式の会議システムにも適用可能であることを注意しておくべきである。

【００６０】

図２は、会議サーバー１１０および／または端末１２０において提供されてもよい会議シーンのグラフィカル・ユーザー・インターフェース（GUI）２００を示している。端末１２０において提供される場合、GUI ２００は、参加者が、端末１２０において知覚される会議シーンを修正できるようにしてもよい。ある実施形態では、GUI ２００は、オーディオ会議の司会者が、会議参加者を会議シーン内に配置できるようにする。GUI ２００は、オーディオ会議の参加者２０１を示してもよい。参加者２０１は、下流オーディオ信号の組内に含まれるオーディオ信号の発生元に対応してもよい。よって、GUI ２００は、（M-1）人までの参加者２０１を示してもよい。さらに、GUI ２００は、会議シーン２１０を示してもよい。図２では、2D（二次元）会議シーンが、たとえば円として示されている。GUI ２００は3D（三次元）会議シーンをたとえば球として示すよう適応されてもよいことを注意しておくべきである。聴取者２１１（すなわち、下流オーディオ信号１２４の端末固有の組を受領する端末１２０）は、デフォルトでシーン２１０の中心に配置されてもよい。GUI ２００は、聴取者２１１の位置の修正を許容するよう構成されていてもよい。さらに、GUI ２００は（図２において白丸２１２として示されている）複数の語り手位置２１２を提供する。複数の語り手位置２１２はあらかじめ決定されていてもよく、あるいはGUI ２００のユーザーによって選択されてもよい。語り手位置２１２は、会議参加者２０１の一つまたは複数に割り当てられてもよい。これは、（たとえば下流オーディオ信号１２４の組と一緒に受領されるメタデータに基づいて）自動的になされてもよい。代替的または追加的に、GUI ２００は、（たとえば、指示される参加者２０１の語り手位置２１２への「ドラッグアンドドロップ」操作を使って）ユーザー固有の割り当てを許容してもよい。参加者２０１および語り手位置２１２の割り当ては、たとえば、ポップアップ・ウィンドー２１３を使って指示されてもよい。さらに、GUI ２００は、（図２では黒丸２１４で示されている）追加的な音位置２１４を指示および／または修正することを許容してもよい。そのような音位置２１４は、参加者のオーディオ信号（典型的には声の信号）以外の音、たとえば会議通知および／または音楽をレンダリングする

10

20

30

40

50

ために使われてもよい。

【0061】

会議システム100は、複数の配置ルールの一つまたは複数に基づいて会議シーン210の語り手位置212を自動的に決定するよう構成されていてもよい。これらの配置ルールは、語り手201が会議シーン210内の種々の位置212に配置され、聴取者211についての好まれるレンダリング位置212が決定された知覚試験に基づく。これらの知覚上の実験は以下のことを示した。

【0062】

・聴取者211は典型的には、会議における語り手が聴取者211の頭部の正面に、好ましくは聴取者の頭部の背後でないところに、立体音響化されることを好む。

10

【0063】

・より精密には、聴取者211は典型的には、語り手201が聴取者の頭部の正面の、聴取者211の正面の中心線215から約 -30° ないし 30° 以内の円錐内に、好ましくはより狭い円錐内に、すなわち中心線215から $\pm 30^{\circ}$ より小さな角度によって定義される円錐内に配置されることを好む。聴取者211にとって、語り手201に長い間傾聴することは、該語り手が大きく偏ったところに、たとえば中心線215から 20° よりも著しく大きな角度のところに配置される場合には、快適ではないことがあることが観察された。よって、語り手位置212が聴取者211の頭部の前方の円錐内に位置されるよう語り手位置212を選択することが有益でありうる。円錐は、該円錐の中心軸215と該円錐の母線216との間の角度があらかじめ決定された最大円錐角、たとえば 15° 、 20° または 30° より小さいようなものであってもよい。本稿で言及される角度は、聴取者211の頭部の前方における中心線215に対する角度をいう。負の角度は、反時計方向の角度を指し、正の角度は中心線215から時計方向の角度を指す。

20

【0064】

・語り手201を互いから分離する能力は、典型的にはいくらかの角度分離を必要とし、語り手識別および了解性を補助するためには約 5° 以上の角度の角度分離を必要とする。よって、二つの隣り合う語り手位置212の間の角度分離がたとえば 5° の最小角距離より大きいように語り手位置212を選択することが有益であることがある。

【0065】

・中央線215（中心線とも称される）のまわりの完全に対称的なレンダリングは好まれない。その理由は、対称的なレンダリングは、時に、たとえば中央線215に対して対称的な点212に配置された二人の語り手201の間で会話が起これば、聴取者211の真正面での像形成効果につながる可能性があるということである。よって、語り手位置212を中心線215に対して非対称な仕方で配列することが有益であることがある。

30

【0066】

・非対称なレンダリングは、オーディオ会議に加わる追加的な参加者201のためにシーン210に語り手201が追加されるときに比較的「均衡した」シーン210を与えるというさらなる利点をもつ。例として、中央線215のまわりに対称的な仕方で配列された最大6個のあらかじめ定義された語り手位置212を含むデフォルト会議シーン210は、1、3または5の語り手201だけがシーン内に配置される場合（すなわち、6個のあらかじめ定義された語り手位置がまだ実際の語り手201で満たされていないとき）には、中央線215を横断して著しく不均衡である。

40

【0067】

上述したルールの一部または全部が、デフォルト語り手位置212をもつデフォルト・シーン210を定義するために使われてもよい。さらに、これらのルールの一部または全部が、自動シーン・マネージャ（たとえば、中央会議コントローラ111および／またはローカル会議コントローラ175によって具現される）の決定論的な振る舞いを指定するために使用されてもよい。シーン・マネージャは、参加者201が会議シーン210にはいるにつれて、参加者201を会議シーン210中に自動的に配置するよう構成されていてもよい。換言すれば、シーン・マネージャ（中央会議コントローラ111および／また

50

はローカル会議コントローラ175上で実行される)は、新しい会議参加者201の語り手位置212を、デフォルト会議シーン210に基づいてかつ会議シーン210内にすでに配置されている参加者201に基づいて、自動的に決定するよう構成されていてもよい。

【0068】

以下では、例示的な三点シーン・マネージャ(デフォルトの三点会議シーンに分布させる)および例示的な六点シーン・マネージャ(デフォルトの六点会議シーンに分布させる)が記述される。本稿で記述される配置ルールを使って、一般的なX点シーンおよび対応するX点シーン・マネージャが規定されうることを注意しておくべきである(Xは整数で、 $X=1,2,3,4,5,6,7,8,9,10$ などであり、異なる空間位置に配置されるM人の会議参加者がいる会議についてはたとえば $X=M-1$)。

10

【0069】

図3aは、例示的な三点シーン300を示しており、中央語り手位置303が中央線301から 2° オフセットされ、両側の対である語り手位置302、304は中心語り手位置303からそれぞれ $\pm 8^\circ$ オフセットされている。図示した例では、各音源、すなわち各語り手位置302、303、304は、聴取者211から同じ知覚される動径方向距離をもつ。

【0070】

より一般的な表現では、三点シーンは以下の特性をもつ。

【0071】

20

・三点シーンのある好ましい実装では、会議参加者201は三つの固定した語り手位置302、303、304の一つに割り当てられる。会議参加者201が割り当てられる実際の語り手位置は、会議参加者201がオーディオ会議に加わる逐次順に依存してもよい。

【0072】

・中心語り手位置303(シーンの中心点)は、シーン300の中央線301のまわりの -5° ないし 5° の中心角のところに配置される。好ましい実装は、中心角 0° のところではなく、中心角 1° ないし 5° または -1° ないし -5° のところである。中心語り手位置300を中央線301から外れた位置にする結果として、全体的な会議シーンが中央線301に対して非対称的になりうる。

30

【0073】

・シーンの他の二つの点(すなわち、他の語り手位置302、303)は、 -30° から 30° の間に任意のところで当人の正面の円錐内に配置されることができる。他の語り手位置302、303の好ましい実装は -15° から 15° の間の円錐内である。

【0074】

・他の二つの語り手位置302、303は、中心語り手位置303のそれぞれの側に配置され、語り手位置302、303の中心点から少なくとも 5° の角度だけ離間されるべきである。

【0075】

・語り手位置の好ましい間隔は、中央線301に対して非対称であるべきである。これは、他の語り手位置302、304を中心点303に対して対称的な仕方で配置することによって達成できる(中心点303が 0° に配置されているのではないとして)。

40

【0076】

図3bは、例示的な六点シーン310を示しており、各語り手201は互いに 5° 離間され、シーン310全体は中央線301に対して固定角 2° だけ回転されている。換言すれば、六点シーン310の語り手位置311、312、313、314、315、316は 2° の角度だけ回転された中央線に対して対称的である。各音源、すなわち各語り手位置311、312、313、314、315、316は、聴取者211から同じ知覚される動径方向距離をもつ。六点シーン310は語り手が、六つの異なる固定点311、312、313、314、315、316に割り当てられることを許容する。本稿で与えられ

50

る配置ルールを使って六点シーン 3 1 0 の他の構成が指定されてもよいことを注意しておくべきである。

【0077】

図 3 a および図 3 b は、語り手位置が互いから固定角だけ離間して配置され、語り手位置の配列が中央線 3 0 1 から固定角だけ回転されているシーン 3 0 0、3 1 0 を示している。しかしながら、最小角がある最小の好ましい角またはある最小の角距離、たとえば 5° より大きい限り、語り手位置は互いから固定角のところに配置される必要はないことを注意しておくべきである。また、隣り合う語り手位置の間の動径方向距離は、追加的な距離手がかりを与えるために変化してもよい。

【0078】

シーン・マネージャ（たとえば、中央またはローカル会議コントローラ）は、各語り手が会議にはいる際に語り手を会議シーンに配置するために、あらかじめ決定された X 点シーン（たとえば、それぞれ図 3 a および図 3 b に示した三点シーン 3 0 0 および／または六点シーン 3 1 0）を使ってもよい。合計 M 人の会議参加者をもつ会議について、M 人の会議参加者のそれぞれが異なる語り手位置に割り当てられうるよう、 $X=(M-1)$ の X 点シーンが使われてもよい。

【0079】

典型的には、オーディオ会議における語り手の実際の数は、会議が始まる時には既知ではない。よって、シーン・マネージャは、会議参加者がコールしてきたときにあらかじめ決定された X 点シーンに会議参加者を追加するよう構成されていてもよい。特に、シーン・マネージャは、あらかじめ決定された X 点シーン内の特定の位置を加わってくる参加者に割り当てよう構成されていてもよい。この目的のために、シーン・マネージャは、会議参加者をあらかじめ決定された X 点シーン中に追加する（または除去する）ための一組のルールを使ってもよい。例示的な配置ルールは次のようなものであってもよい：

- ・新たな会議参加者を、X 点シーンの中央線 3 0 1 にできるだけ近い利用可能な語り手位置に配置する；
- ・X 点シーンの中央線 3 0 1 に対するおよび／または X 点シーンの中心位置 3 0 3 に対する、割り当てられた語り手位置の最大限の均衡を保証する；
- ・X 点シーンを去った会議参加者によって空にされた空の語り手位置を充填する。

【0080】

上述した配置ルールは、新たな参加者を X 点シーンに配置するために、単独でまたは組み合わせて使用されうる。よって、新たな参加者は X 点シーンの内側の諸点から外向きに、および／または X 点シーンの中心の語り手位置 3 0 3 またはシーンの中央線 3 0 1 のまわりの均衡を最大にするような仕方で、会議シーンに追加されてもよい。会議シーン内の語り手の数 $(M-1)$ が X 点シーンの語り手位置の数 X を超える場合は、シーン・マネージャは複数の語り手を同じ語り手位置に割り当てよう構成されていてもよい。図 3 a に示される三点シーン 3 0 0 については、上流の参加者はシーン・マネージャによって次のように配置されることができる：

- ・参加者 1： -2° のところに（すなわち語り手位置 3 0 3 に）配置、
- ・参加者 2： 6° のところに（すなわち語り手位置 3 0 4 に）配置、
- ・参加者 3： -10° のところに（すなわち語り手位置 3 0 2 に）配置、
- ・参加者 4： -2° のところに（すなわち語り手位置 3 0 3 に）配置、
- ・参加者 5： 6° のところに（すなわち語り手位置 3 0 4 に）配置、
- ・以下同様。

【0081】

本稿では、角度の値は記号「 $^\circ$ 」、用語「度」または可能性としては両方によって表わされることを注意しておく。六点シーン 3 1 0 については、新たな会議参加者は次のようにシーンに加わることができる（図 3 b を参照として使う）。

- ・参加者 1： -2° のところに（すなわち語り手位置 3 1 3 に）配置、
- ・参加者 2： 3° のところに（すなわち語り手位置 3 1 4 に）配置、

- ・参加者3： -7° のところに（すなわち語り手位置312に）配置、
- ・参加者4： 8° のところに（すなわち語り手位置315に）配置、
- ・参加者5： -12° のところに（すなわち語り手位置311に）配置、
- ・参加者6： 13° のところに（すなわち語り手位置316に）配置、
- ・参加者7： -2° のところに（すなわち語り手位置313に）配置、
- ・以下同様。

【0082】

聴取者211が異なる語り手位置311、312、313、314、315、316に配置された異なる参加者を区別する能力に関して特に良好な性質をもつことが示された特定の六点シーン310は、語り手位置311、312、313、314、315、316

10

について以下の角度を利用する。この特定の六点シーンは、隣り合う語り手位置の間の最小離間の制約条件を満たし、 $\pm 20^{\circ}$ の円錐内に留まり、中央線301に対してわずかに非対称である：

- ・語り手位置314（たとえば第一の参加者について）は中央線301から 2° のところ；
- ・語り手位置313（たとえば第二の参加者について）は中央線301から -5° のところ；
- ・語り手位置315（たとえば第三の参加者について）は中央線301から 9° のところ；
- ・語り手位置312（たとえば第四の参加者について）は中央線301から -12° のところ；
- ・語り手位置316（たとえば第五の参加者について）は中央線301から 16° のところ；
- ・語り手位置311（たとえば第六の参加者について）は中央線301から -19° のところ；

20

デフォルト・シーンの上述した記述は、例示的な三点および例示的な六点シーン300、310に限られていた。二点シーン・マネージャから(M-1)点シーンまで(M人の参加者の会議の場合)、シーン内の点の他の数も可能であることを注意しておくべきである。シーン内の点の数は典型的には、本稿に記述されるデザインおよび配置ルールによって制限されるのみである。さらに、角度の指示された値は単に例であることを注意しておくべきである。選択される角度値は ± 1 度または ± 2 度だけ変化してもよい。よって、本稿に記述される角度値は近似的な指示として理解されるべきである。

30

【0083】

（語り手の数(M-1)が語り手位置の数Xを超えるために）同じ語り手位置に複数の語り手を割り当てることの代わりにまたはそれに加えて、シーン・マネージャは、会議シーンをより多数の語り手位置をもつ会議シーンに（たとえば三点シーンから六点シーンに）アップグレードするよう構成されていてもよい。例として、シーン・マネージャ（たとえば会議サーバー110）は、（端末120の一つに位置している）オーディオ会議のまとめ役に、会議シーンがアップグレードされるべきかどうかを促してもよい。承認されたら、シーン・マネージャは会議参加者をアップグレードされた会議シーンに移してもよい。こうすることにより、会議シーンのサイズが会議参加者の実際の数に合わせて柔軟に適応できる。さらに、種々のサイズをもつ会議シーンが会議サービス・プロバイダーによって、付加価値サービスとして提供されてもよい。

40

【0084】

会議シーン内での語り手の水平方向の分布の代わりにまたはそれに加えて、会議シーンは、特にエンドポイントが3Dレンダリングの機能をもつ場合、垂直方向に拡張されてもよいことを注意しておくべきである。たとえば、異なる語り手位置の間で同じ方位角の角度離間が、ただしたとえば10度の仰角離間をもって使われてもよい。このようにして、語り手の複数の層を作り出すことができ、それによりさらに会議内の異なる語り手の空間的な分離の可能性を高める。より一般的な表現では、会議シーン内の複数の語り手位置は方位

50

角 ϕ （聴取者211の頭部の前方の、中央線215を含む水平面での）および傾斜角 θ （聴取者の頭部の前方の、中央線215を含む鉛直面内での）によって記述されてもよい。会議シーンは複数列の語り手位置を有していてもよく（列内の各語り手位置は異なる方位角 ϕ および同じ傾斜角 θ によって記述される）、各列は異なる傾斜角 θ に位置される。

【0085】

以下では、オーディオ会議のための必要とされるネットワーク資源を軽減するためのさまざまな方式が概説される。上記で論じたように、本稿で記述されるオーディオ会議システムは、オーディオ会議の端末120における会議シーンのバイノーラル・レンダリング（またはマルチチャンネル・レンダリング）を許容することに向けられる。バイノーラル・レンダリングは、2Dまたは3D空間内の会議シーンにおける語り手の配置を許容するべきである。これは、二つの（モノ）オーディオ信号を単一の（モノ）信号に混合する（すなわち加算する）こと（これは二つのオーディオ信号の空間的分離は許容しない）とは対照的である。会議シーンにおける語り手のバイノーラル・レンダリングは、会議システム内のさまざまな位置において実装されることができる。図1のaの例示的な会議システム100は、中央集中式の会議サーバー110を利用し、該サーバー110は、特定の会議シーンを形成するために、下流オーディオ信号124の対応する組がどのように組み合わせられるべきかを指定するメタデータを生成する。その特定の会議シーンを反映するバイノーラル信号がそれぞれの端末120において決定され、それによりバイノーラル・レンダリングが端末120（エンドポイントとも称される）におけるオーディオ・トランシーバ122のレンダリング特性に柔軟に適応することを許容する。典型的には、バイノーラル信号の生成は、下流オーディオ信号124の組に基づきかつメタデータ内に含まれる配置情報に基づく。さらに、バイノーラル信号の生成は、オーディオ・トランシーバ122（たとえばラウドスピーカーまたはヘッドホン）の型に依存してもよい。中央集中式の会議サーバー110は、端末120において使用されるオーディオ・トランシーバ122の型を知らなくてもよく、したがって、バイノーラル信号の生成を端末120において実行することが有益であることがある。

【0086】

例として、エンドポイント120はオーディオ会議の間に動的に適応する必要があることがある。たとえば、エンドポイント120における聴取者211は、バイノーラル・ヘッドセットを使うことによってオーディオ会議を開始してもよい。のちの段階になって、その部屋において聴取者211に第二の会議参加者が加わってもよく、よって二人とも参加できるよう彼らはバイノーラル・ヘッドセットを切断してエンドポイント・ラウドスピーカーおよびマイクロホンを使う。結果として、会議シーンのレンダリングは、ヘッドホンからラウドスピーカーに切り換えるために、適応される必要がある。よって、エンドポイント120は、2Dまたは3D会議シーンのレンダリングを、エンドポイント120において使用されるオーディオ・トランシーバ122に適応させるよう構成されていてもよい。

【0087】

よって、 $(M-1)$ 個までの個々の下流オーディオ信号（オーディオ会議内の $(M-1)$ 人の語り手に対応）および関連付けられたメタデータを端末120に送信することが有益であることがある。会議シーンが X 個の語り手位置に制限されている場合、複数の語り手が同じ語り手位置に割り当てられたことがありうる。同じ語り手位置に割り当てられた語り手のオーディオ信号は、それぞれの語り手位置についての下流オーディオ信号を形成するよう、混合されてもよい。よって、 X 個までの下流オーディオ信号（ X 点会議シーンの X 個の語り手位置に対応）の組および関連付けられたメタデータが端末120に送られてもよい。端末120は、下流オーディオ信号の組および関連付けられたメタデータを使って、2Dまたは3Dの X 点会議シーンをレンダリングするよう構成されていてもよい。さらに、端末120（たとえば端末170）は、ローカル会議コントローラ175を使って会議シーンを修正する（たとえば、語り手および語り手位置を入れ替える、会議シーンをシフトさせるなど）よう構成されていてもよい。しかしながら、端末120が X 点会議シーンのバイノーラルまたは多チャンネル・レンダリングを実行できるようにするために、 X 個までの個々

の下流オーディオ信号の組および関連付けられたメタデータが端末120に伝送される必要がある。

【0088】

あるいはまた、端末120のためのバイノーラル信号は、会議サーバー110において生成されてもよい。これは、会議サーバー110と端末120との間のリンク上での必要とされる帯域幅に関して有益であることがある。バイノーラル信号（すなわちステレオ信号）の送信は、 $(M-1)$ 個までの下流オーディオ信号の組および対応するメタデータ（これは典型的にはビットストリーム、たとえばG.711、G722.2（AMR-WB: Adaptive Multi-Rate-Wide Band [適応マルチレート・ワイドバンド]）、MPEG2またはMPEG4ビットストリームの形で送信される）の送信よりも、必要とされる帯域幅が少ないことがあるからである。他方、会議サーバー110においてバイノーラル信号を生成すると、宛先端末120において使用されるオーディオ・トランシーバ122に関しておよび／または宛先端末120（聴取者端末120とも称される）における会議シーンの操作に関して許容される柔軟性が低くなる。

【0089】

（図1のaおよびbのコンテキストにおいて概説したように）端末120においてバイノーラル信号の生成を実行するとき、下流オーディオ信号124の組のオーディオ信号の数は、最大 N 個の同時アクティブ・オーディオ信号に制限されてもよい（ここで、 N は参加者の数 M より小さく、たとえば $N < M-1$ であるおよび／または N は X 点シーン内の語り手位置の数 X より小さく、すなわち $N < X$ である）。これは、会議サーバー110と端末120との間のリンクに対する帯域幅制限のためであってもよい。換言すれば、サーバー110とエンドポイント120との間の帯域幅を制限するために、サーバー110からエンドポイント120に送られる同時のアクティブなストリーム（すなわちオーディオ信号）の最大数を制限する必要があることがある。このように、たとえば会議参加者201が会議シーン210内で $(M-1)$ 個の離散的な点121に配置されたとしても（たとえば $M=7$ ）、たとえば $(M-1)$ 点シーン310を実装しているときであっても、エンドポイント120に同時に送達されるストリームの数は N 、たとえば $N=1, 2$ または3個の同時のアクティブなストリームに制限されてもよい。たとえば、同時のアクティブなストリームの最大数 N は、多者会議の知覚的に快いレンダリングを提供しながら、サーバー110とエンドポイント120との間の必要とされる帯域幅を制限するために選択されてもよい。 $N=1$ の場合、一つの下流オーディオ信号124だけが（たとえばオーディオ・ストリームとして）サーバー110からエンドポイント120に送られ、レンダリングまたは混合はサーバー110において実行されてもよい。この場合、端末120におけるレンダリングはモノ出力に制限されてもよい。 $N=2$ については、最大二つの同時のオーディオ信号124が（たとえばオーディオ・ストリームとして）サーバー110からエンドポイント120にレンダリングのために送られてもよい。 $N=3$ については、最大三個の同時のオーディオ信号124が（たとえばオーディオ・ストリームとして）サーバー110からエンドポイント120に送られてもよい。上記のそれぞれの場合において、サーバー110は、会議内の同時の語り手の数 $(M-1)$ があらかじめ定義された最大 N より大きい場合に、いくつかのストリームを混合してもよい。三点または六点シーン300、310についてのある実施形態では、シーン・マネージャは、エンドポイント120に送られるストリームの数を $N=3$ 個のストリームに制限するよう構成されていてもよい。換言すれば、下流オーディオ信号の組内のオーディオ信号の数は $N=3$ に制限されてもよい。

【0090】

N 個の下流オーディオ信号が対応するメタデータとともに提供されてもよいことを注意しておくべきである。よって、端末120は、立体音響化された仕方で N 個の下流オーディオ信号をレンダリングできるようにされてもよい。例として、たとえば $N=1$ であっても、単一の下流オーディオ信号は、その単一の下流オーディオ信号を2Dまたは3D空間的会議シーンにおいてどこに配置するかを示すメタデータとともに送信されてもよい。単一の語り手のみがアクティブである場合、下流オーディオ信号（単一のアクティブな語り手の上流

オーディオ信号に対応する)は、その単一の語り手の語り手位置に配置されることができる。これは、通常のモノ・レンダリング(立体音響化なし)とは異なる。複数の語り手(かつ $N=1$)の場合にのみ、複数の語り手の単一の下流オーディオ信号への混合のために、複数の語り手の空間的な曖昧さ解消が失われる。

【0091】

上記で概説したように、会議サーバー110は、中央会議コントローラ111および複数のオーディオ・サーバー112を有していてもよい。会議コントローラ111は、会議シーン内の会議参加者の配置を定義するよう構成されていてもよい。さらに、会議コントローラ111は、一または複数の会議参加者のオーディオ信号が混合される必要があるかどうか、どのオーディオ信号が混合されるべきかおよび混合操作の優先度を決定するよう構成されていてもよい。換言すれば、会議コントローラ111は以下の処理を実行するよう構成されていてもよい。

【0092】

・一または複数の会議参加者のオーディオ信号を混合する必要性を決定する。この目的のために、会議参加者の数 M と下流オーディオ信号124の組内のオーディオ信号の最大数 N が比較されてもよい。

【0093】

・どのオーディオ信号が混合されるべきかを決定する。このコンテキストにおいて、会議コントローラ111は、一つまたは複数の混合ルールを利用してもよい。たとえば、会議シーンの中央線301のより近くにレンダリングされている語り手よりも、会議シーン内でより大きな角度に配置されている語り手をより高い優先度をもって混合されるようにすることが好ましいことがありうる。換言すれば、聴取者211の前方に配置されている語り手の混合を回避することが有益であることがある。これは、聴取者211は典型的には、会議シーン内の動きを、より大きな角度のところで生起する動きに比べて動きが聴取者211の真正面で生起する場合に、より観察するという事実のためである。さらに、会議に加わる最初の人々が会議のまとめ役である可能性が高いことが想定されてもよい。上記で概説したように、シーン・マネージャは、会議シーン210内の語り手位置212を、会議に加わる順に従って、中心位置から外側位置に向かって分布させていくよう構成されていてもよい。よって、会議のまとめ役が中心位置に位置されると想定されてもよく、従って、会議のまとめ役に(他の会議参加者と混合されるリスクがより低い)優先的な分離を提供することが望ましいことがありうる。

【0094】

・混合された語り手の配置を決定する。この目的のために、会議コントローラ111は、一つまたは複数の配置ルールを適用してもよい(たとえば、本稿で記述した一つまたは複数の配置ルール)。換言すれば、会議コントローラ111は、混合された語り手が会議シーンのどこに配置されるかのあらかじめ定義された方針を利用してもよい。例として、会議コントローラ111は、 X の異なる値をもつ複数のあらかじめ定義された X 点会議シーンを有していてもよい。許容されるオーディオ信号の数 N が、 M を会議参加者の数として、必要とされるオーディオ信号の数 $(M-1)$ より小さいことが判別されると、会議コントローラ111は、あらかじめ定義された N 点会議シーンに従って混合されたオーディオ信号を配置するよう構成されていてもよい。換言すれば、会議コントローラ111は、会議シーンを選択するよう構成されていてもよい。ここで、会議シーン内の語り手位置の数は、端末120に個々に送信されることのできるオーディオ信号の数 N に適応されてもよい。

【0095】

このように、会議コントローラ111によって使用される混合方針に対して少なくとも二つの要素がある。これらの要素は、どの語り手が一緒に混合されるかを決定し、混合される語り手についての最終的な空間的位置が会議シーン内のどこにあるかを決定するものである。例として、六点シーンについて、会議コントローラ111は、混合のために、隣接して配置された角度(すなわち語り手位置)を同定するよう構成されていてもよい。これは、六点シーンから三点シーンへの低下を許容する($N=3$ の場合)。これは、図4にお

10

20

30

40

50

いて、例示的な六点会議シーン400において示されている。シーン400において話している四人の人々しかいない場合、好ましい混合方針は、これらの位置410における語り手がアクティブである場合には、オーディオ信号の数を最大数 $N=3$ に減らすために、クラスター1 403および／またはクラスター2 401を混合することでありうる。これが十分でない場合にのみ、クラスター3 402内の語り手がサーバー110において混合されうる。下記でさらに詳細に概説するように、混合は典型的には、語り手活動の解析に基づいて実行される。これは、複数の時点のそれぞれにおいて、アクティブな上流オーディオ信号の数が決定されてもよいことを意味している。ある特定の時点におけるアクティブな上流オーディオ信号が N より大きい場合、アクティブな上流オーディオ信号の一部または全部が（本稿で記述する混合ルールに従って）混合されてもよい。

10

【0096】

さらに別の言い方をすれば、会議コントローラ111は、会議内に存在しているストリームの数に基づいて、かつ許容可能なストリームの最大数に基づいて、オーディオ・ストリームを混合する（すなわち、オーディオ信号を混合する）よう構成されていてもよい。ストリームの数が N ストリームを超える場合には、エンドポイント120に送信されるストリーム124の数を制限するために混合方針が適用される。混合方針は、偏りの大きなものを常に先に混合するという混合ルールを含んでいてもよい。さらに、会議コントローラ111は、混合ストリームを、混合されたストリームがもともと配置されていた二つ（以上）のあらかじめ定義された語り手位置の一つに配置するよう構成されていてもよい。あるいはまた、混合ストリームは、二つ（以上）のあらかじめ定義された語り手位置の中間のどこかに配置されてもよい。ある好ましい実装では、混合ストリームは、混合された諸ストリームの語り手位置の間の中間に配置される。会議コントローラ111は、最後の手段としてのみ（すなわち、低下した優先度で）、会議シーンの中央線301の近くに配置される語り手（たとえば、図4のクラスター3 402）の混合を実行してもよい。

20

【0097】

上記で論じたように、会議参加者のオーディオ信号の混合は典型的には、アクティブな語り手の数（すなわち、 M 人の参加者がいる会議については $M-1$ ）がオーディオ信号124の組内の許容されるオーディオ信号の最大数 N を超える場合および／またはアクティブな語り手の数（ $M-1$ ）が X 点シーン内の語り手位置212の数をを超える場合（すなわち、 $M-1 > X$ ）にのみ必要とされる。例として、六点シーン内での混合は、四人以上の語り手がいるときにのみ必要とされる。この例では、シーンは「混雑」しており、したがって音の小さな動きに気づくことは難しくなる。換言すれば、会議における参加者の数が増すにつれて、すなわち会議が「混雑」してくるにつれて、オーディオ信号の混合から帰結する音の空間的な移動は聴取者211によって知覚されにくくなる傾向がある。

30

【0098】

オーディオ・サーバー112は、会議コントローラ111によって定義されるオーディオ信号の混合を実装するよう構成されていてもよい。換言すれば、オーディオ・サーバー112は、オーディオ信号を処理し、オーディオ信号のマージを実行してもよい。さらに別の言い方をすれば、パケット・ベースの通信ネットワークにおいて、オーディオ・サーバー112は、オーディオ信号のパケット毎に、ストリームの総数を減らすために混合を実装するかどうかを決定してもよい。例として、オーディオ・サーバー112は、複数の相続く時点のそれぞれにおいて M 個の上流オーディオ信号の活動度を決定するよう構成されていてもよい（ここで、たとえば、それらの時点はオーディオ・ビットストリームのパケットと同時であってもよい）。会議コントローラは、活動度を解析し、下流オーディオ信号の総数を許容される最大 N まで低下させるための上流オーディオ信号の選択および／または混合について決定してもよい。さらに、会議コントローラは、 N 個の下流オーディオ信号に関する配置情報を提供してもよい。その際、実際の混合およびメタデータの生成は、オーディオ・サーバー112において、会議コントローラによって与えられる決定および配置情報に基づいて、実行されてもよい。

40

【0099】

50

端末120に向かうオーディオ・ストリームの数を最大N個のオーディオ・ストリームに制限する上記の例は、端末120に伝送される（可能性としては混合された）オーディオ・ストリームの固定した選択に基づいている。典型的には、会議内でのアクティブな語り手の数は限られている。理想的できわめて整然としたオーディオ会議では、一人のアクティブな語り手のみがあって、その間、他の会議参加者は傾聴しているであろう。よって、そのようにきわめて整然としたオーディオ会議では、単一のオーディオ・ストリーム（すなわち、アクティブな語り手のオーディオ・ストリーム）を会議シーン内でのそのアクティブな語り手の配置を指示するメタデータとともに送信するだけで十分であることがある。異なる語り手位置に位置される別の参加者がアクティブな語り手になると、単一の送信されたオーディオ・ストリームは、新しい語り手位置を指示するメタデータを伴った、新しいアクティブな語り手に対応するオーディオ・ストリームに変更されうる。よって、単一のオーディオ・ストリーム（および関係したメタデータ）のみを送信しつつ、すべての異なる語り手が端末120においてそれぞれの語り手位置においてレンダリングされうる。

10

【0100】

より一般的な表現では、会議参加者111は、X点会議シーン内に配置されるX個の（混合されたまたは混合されていない）オーディオ信号の語り手活動度に基づいて、N個の下流オーディオ信号124の組を動的に選択するよう構成されていてもよい。ある特定の時点において、会議コントローラ111は、端末120に向けた伝送のためのX個の（混合されたまたは混合されていない）オーディオ信号のうちN個の最もアクティブなものを（選択されたオーディオ信号を会議シーン内で配置するための関連するメタデータとともに）選択してもよい。端末120への送信のためのオーディオ信号の選択は、相続く時点について（たとえば1秒毎または100ms毎に）繰り返されてもよい。よって、端末120に送信されるオーディオ・ストリームの低下した数Nを提供しつつ、端末120においてレンダリングできる空間的位置の数Xは維持されうる。ある実施形態では、X個の（混合されたまたは混合されていない）オーディオ信号のうちの(N-1)個の最もアクティブなものが端末120への送信のために選択される。よって、(N-1)個の最もアクティブな語り手が端末120において立体音響化された仕方でレンダリングされうる。

20

【0101】

本稿では、オーディオ会議の2Dまたは3Dシーンを管理するためのさまざまな側面が記述された。これらの側面は、音声会議システムの開発者または音声会議システムのユーザーが（種々の会議参加者から発する）音声信号および／または音信号（たとえば通知、音声プロンプト、音楽）の会議シーンへの配置を管理することを許容するために、API（アプリケーション・プログラミング・インターフェース）またはGUI（グラフィカル・ユーザー・インターフェース）のコンテキストにおいて与えられることがある。本稿は、あらかじめ決定されるX点シーンを定義するためおよび／または語り手を会議シーン内で自動的に位置決めするためにシーン・マネージャによって使用されるルールおよび論理を提供する。シーン・マネージャは、シーンがヘッドホンによってレンダリングされているかラウドスピーカーによってレンダリングされているかに依存して異なりうる、装置固有の空間位置（たとえば、シーン内の端末固有の位置）および／または装置固有のX点シーンを

30

40

【0102】

本稿で記述された方法およびシステムは、ソフトウェア、ファームウェアおよび／またはハードウェアとして実装されうる。ある種のコンポーネントはたとえば、デジタル信号プロセッサまたはマイクロプロセッサ上で実行されるソフトウェアとして実装されてもよい。他のコンポーネントはたとえば、ハードウェアとしておよび／または特定用途向け集積回路として実装されてもよい。記載される方法およびシステムにおいて出てくる信号は、ランダム・アクセス・メモリまたは光学式記憶媒体といった媒体に記憶されてもよい。そうした信号は、電波ネットワーク、衛星ネットワーク、無線ネットワークまたは有線ネットワーク、たとえばインターネットのような、ネットワークを介して転送されてもよい

50

。本稿に記載される方法およびシステムを利用する典型的な装置は、ポータブル電子装置またはオーディオ信号を記憶および／またはレンダリングするために使用される他の消費者設備である。

いくつかの態様を記載しておく。

〔態様 1〕

会議参加者に関連付けられた上流オーディオ信号（1 2 3、1 7 3）を、聴取者（2 1 1）に対してレンダリングされる2Dまたは3D会議シーン内で配置するよう構成された会議コントローラ（1 1 1、1 7 5）であって：

・前記会議シーン内のX個の異なる空間的語り手位置（2 1 2）をもつX点会議シーンをセットアップする段階であって、Xは整数であり、 $X > 0$ であり、前記X個の語り手位置（2 1 2）は、聴取者の頭部の前方の中央線（2 1 5）のまわりの円錐内に位置され、前記円錐の母線（2 1 6）と前記中央線（2 1 5）は、あらかじめ決定された最大円錐角以下の角度をなす、段階と；

・前記上流オーディオ信号を前記X個の語り手位置の一つに割り当てる段階と；

・割り当てられた語り手位置を同定するメタデータを生成し、オーディオ処理ユニット（1 2 1、1 7 1）が前記上流オーディオ信号に基づいて立体音響化されたオーディオ信号を生成できるようにする段階であって、立体音響化されたオーディオ信号を聴取者に対してレンダリングするとき、聴取者は前記上流オーディオ信号を、割り当てられた語り手位置から到来するものとして知覚する、段階とを実行するよう構成されている、
会議コントローラ。

〔態様 2〕

前記X個の語り手位置は円または球上に配置され、聴取者はその円または球の中心に配置される、態様 1 記載の会議コントローラ。

〔態様 3〕

・前記X個の語り手位置のうち二つの隣接する語り手位置は、少なくともある最小の角距離だけ離間しており、

・前記最小の角距離は 5° またはそれ以上である、
態様 1 または 2 記載の会議コントローラ。

〔態様 4〕

前記X個の語り手位置のうちの隣り合う語り手位置の間の角距離が異なる語り手位置については異なる、態様 1 ないし 3 のうちいずれか一項記載の会議コントローラ。

〔態様 5〕

前記X個の語り手位置のうちの隣り合う語り手位置の間の角距離が、該隣り合う語り手位置の前記中央線からの距離が増すにつれて増大する、態様 4 記載の会議コントローラ。

〔態様 6〕

前記最大円錐角が30度である、態様 1 ないし 5 のうちいずれか一項記載の会議コントローラ。

〔態様 7〕

前記最大円錐角が20度である、態様 1 ないし 5 のうちいずれか一項記載の会議コントローラ。

〔態様 8〕

前記最大円錐角が15度である、態様 1 ないし 5 のうちいずれか一項記載の会議コントローラ。

〔態様 9〕

前記X個の語り手位置が前記中央線に対して非対称である、態様 1 ないし 8 のうちいずれか一項記載の会議コントローラ。

〔態様 1 0〕

Xが奇数の場合、前記X個の語り手位置は、該X個の語り手位置のうちの中央語り手位置に対して対称的である、態様 1 ないし 9 のうちいずれか一項記載の会議コントローラ。

〔態様 1 1〕

・Xが偶数の場合、前記X個の語り手位置の半分が前記中央線の一方の側または前記中央線を通る鉛直面の一方の側にあり、前記X個の語り手位置の残りの半分が前記中央線の反対側または前記中央線を通る前記鉛直面の反対側にある、および／または

・Xが奇数の場合、前記X個の語り手位置のうち $(X-1)/2$ 個が一方の側にあり、 $(X+1)/2$ 個が前記中央線または前記鉛直面の反対側にある、

態様1ないし10のうちいずれか一項記載の会議コントローラ。

〔態様12〕

・前記X個の語り手位置は、聴取者の頭部からのX本の射線上に位置され、前記中央線とX個の語り手角をなし；

・前記中央線からある方向の語り手角は正であり、前記中央線から反対方向の語り手角は負であり；

・前記X個の語り手角の和は0とは異なる、

態様1ないし11のうちいずれか一項記載の会議コントローラ。

10

〔態様13〕

語り手角の前記和は2度以下である、態様12記載の会議コントローラ。

〔態様14〕

X=3であり、前記X個の語り手位置は、前記中央線から-2度、+6度および-10度の語り手角のところに位置される、態様12または13記載の会議コントローラ。

〔態様15〕

X=6であり、前記X個の語り手位置は、前記中央線から-2度、+3度、-7度、+8度、-12度および+13度の語り手角のところに位置される、態様12または13記載の会議コントローラ。

20

〔態様16〕

X=6であり、前記X個の語り手位置は、前記中央線から+2度、-5度、+9度、-12度、+16度および-19度の語り手角のところに位置される、態様12または13記載の会議コントローラ。

〔態様17〕

複数の会議参加者に関連付けられた複数の上流オーディオ信号を前記X点会議シーン内に配置するよう構成されている、態様1ないし16のうちいずれか一項記載の会議コントローラであって、当該会議コントローラは、前記複数の上流オーディオ信号を、前記複数の上流オーディオ信号のある逐次順に従って、前記X個の語り手位置に割り当てるよう構成されている、会議コントローラ。

30

〔態様18〕

前記複数の上流オーディオ信号の前記逐次順は、当該会議コントローラによる前記複数の上流オーディオ信号の検出の順序に基づいている、態様17記載の会議コントローラ。

〔態様19〕

当該会議コントローラは、前記中央線からの絶対的な角距離が増す順に前記X個の語り手位置を割り当てるよう構成されている、態様17または18記載の会議コントローラ。

〔態様20〕

前記複数の上流オーディオ信号からの上流オーディオ信号を、前記X個の語り手位置のうちからの最も外側の語り手位置に割り当てることを受けて、当該会議コントローラは、前記複数の上流オーディオ信号のうちからの次の上流オーディオ信号を、前記X個の語り手位置のうちからの最も内側の語り手位置に割り当てるよう構成されている、態様19記載の会議コントローラ。

40

〔態様21〕

当該会議コントローラは、上流オーディオ信号の数が語り手位置の数Xより多い場合、前記複数の上流オーディオ信号のうちからのいくつかの上流オーディオ信号を、前記X個の語り手位置のうちの少なくとも一つに割り当てるよう構成されている、態様17ないし20のうちいずれか一項記載の会議コントローラ。

〔態様22〕

50

同じ語り手位置に割り当てられた前記いくつかの上流オーディオ信号の混合を開始するよう構成されている、態様 2 1 記載の会議コントローラ。

〔態様 2 3〕

前記複数の上流オーディオ信号のそれぞれが、前記X個の語り手位置のうちの単一の位置のみに割り当てられる、態様 1 7 ないし 2 2 のうちいずれか一項記載の会議コントローラ。

〔態様 2 4〕

・前記複数の上流オーディオ信号からの次の上流オーディオ信号を、前記X個の語り手位置のうちの、まだ割り当てられておらず、かつ、前記中央線にできるだけ近い語り手位置に割り当てる；

・割り当てられた諸語り手位置の、前記中央線に対するおおよび／または前記X個の語り手位置のうち中心の語り手位置に対する最大の均衡を保証する；

・前記X個の語り手位置のうちからの、前記複数の上流オーディオ信号のうちからのすでに割り当てられた上流オーディオ信号の中断によって空にされた空の語り手位置を割り当て直す、

という配置ルールの一つまたは複数に従って前記複数の上流オーディオ信号を配置するよう構成されている、

態様 1 7 ないし 2 3 のうちいずれか一項記載の会議コントローラ。

〔態様 2 5〕

・前記立体音響化されたオーディオ信号をレンダリングするオーディオ・トランシーバ (1 2 2) のレンダリング特性を判別し；

・前記オーディオ・トランシーバの前記レンダリング特性に依存して前記複数の上流オーディオ信号の一つまたは複数の混合を開始するよう構成されている、

態様 1 7 ないし 2 4 のうちいずれか一項記載の会議コントローラ。

〔態様 2 6〕

前記レンダリング特性が前記オーディオ・トランシーバがモノ・オーディオ信号のレンダリングに限定されていることを示す場合には、前記会議シーンに配置されるべき前記複数の上流オーディオ信号の全部の混合を開始するよう構成されている、態様 2 5 記載の会議コントローラ。

〔態様 2 7〕

前記上流オーディオ信号が発話信号を含むことを判別するよう構成されている、態様 1 ないし 2 6 のうちいずれか一項記載の会議コントローラ。

〔態様 2 8〕

・聴取者の好まれる側を判別する段階であって、前記好まれる側は、前記中央線から左側または右側である、段階と；

・到来するオーディオ信号を、聴取者の前記好まれる側にある語り手位置に割り当てる段階とを実行するよう構成されている、

態様 1 ないし 2 7 のうちいずれか一項記載の会議コントローラ。

〔態様 2 9〕

当該会議コントローラは、前記好まれる側に関する指示を聴取者の端末から、たとえば聴取者の前記端末における選好設定に基づいて受領するよう構成されている、態様 2 8 記載の会議コントローラ。

〔態様 3 0〕

・前記会議シーンが3D会議シーンであり；

・前記X個の異なる空間的語り手位置は、前記中央線に対して異なる方位角および／または傾斜角に位置される、

態様 1 ないし 2 9 のうちいずれか一項記載の会議コントローラ。

〔態様 3 1〕

・前記X個の異なる空間的語り手位置は、異なる傾斜角にある複数の層に配列される；および／または

10

20

30

40

50

・各鉛直方向層は、異なる方位角にある複数の異なる空間的語り手位置を含む；および／または

・第一の鉛直方向層における異なる空間的語り手位置と第二の鉛直方向層における異なる空間的語り手位置は対応する方位角を有する、

態様 3 0 記載の会議コントローラ。

〔態様 3 2〕

・前記上流オーディオ信号の地位に関する情報を受領する段階と；

・前記上流オーディオ信号を、該上流オーディオ信号の地位に基づいて、空間的語り手位置に割り当てる段階とを実行するよう構成されている、

態様 1 ないし 3 1 のうちいずれか一項記載の会議コントローラ。

〔態様 3 3〕

・前記地位が、前記上流オーディオ信号が会議の司会者またはまとめ役に対応することを示し；

・前記上流オーディオ信号が割り当てられる空間的語り手位置は、最も中央寄りの空間的語り手位置である、

態様 3 2 記載の会議コントローラ。

〔態様 3 4〕

あらかじめ決定された地位をもつ上流オーディオ信号のために前記X個の空間的語り手位置のうちの一つまたは複数を予約するよう構成されている、態様 1 ないし 3 3 のうちいずれか一項記載の会議コントローラ。

〔態様 3 5〕

前記X個の語り手位置が聴取者の頭部の前方の前記中央線のまわりの前記円錐内に位置されるよう、X個の異なる空間的語り手位置をもつ前記X点会議シーンを計算するよう構成されている、態様 1 ないし 3 4 のうちいずれか一項記載の会議コントローラ。

〔態様 3 6〕

X個の異なる空間的語り手位置をもつ前記X点会議シーンを、X個の異なるあらかじめ決定された空間的語り手位置をもつあらかじめ決定されたX点会議シーンの集合から選択するよう構成されている、態様 1 ないし 3 5 のうちいずれか一項記載の会議コントローラ。

〔態様 3 7〕

オーディオ会議システム（1 0 0、1 5 0）であって、

・会議参加者に関連付けられた上流オーディオ信号を生成するよう構成された語り手端末と；

・態様 1 ないし 3 6 のうちいずれか一項記載の会議コントローラであって、2Dまたは3D会議シーン内の語り手位置に上流オーディオ信号を割り当てるよう構成されており、割り当てられた語り手位置を同定するメタデータを生成するよう構成されている会議コントローラと；

・聴取者が上流オーディオ信号を、割り当てられた語り手位置から到来するものとして知覚するよう、前記メタデータを使って聴取者に対して上流オーディオ信号をレンダリングするよう構成された聴取者端末を有する、

オーディオ会議システム。

〔態様 3 8〕

会議参加者に関連する上流オーディオ信号を、聴取者に対してレンダリングされるべき2Dまたは3D会議シーン内で配置する方法であって、

・会議シーン内でX個の異なる空間的語り手位置をもつX点会議シーンをセットアップする段階であって、Xは整数であり、 $X > 0$ であり、前記X個の語り手位置は、聴取者の頭部の前方の中央線のまわりの円錐内に位置され、前記円錐の母線と前記中央線はあらかじめ決定された最大円錐角以下の角度をなす、段階と；

・前記上流オーディオ信号を前記語り手位置の一つに割り当てる段階と；

・割り当てられた語り手位置を同定するメタデータを生成し、オーディオ処理ユニット（1 2 1、1 7 1）が前記上流オーディオ信号に基づいて立体音響化されたオーディオ信号

10

20

30

40

50

を生成できるようにする段階とを含んでおり、立体音響化されたオーディオ信号を聴取者に対してレンダリングするとき、聴取者は、前記上流オーディオ信号を、割り当てられた語り手位置から到来するものとして知覚する、方法。

〔態様 39〕

プロセッサ上での実行のために、該プロセッサ上で実行されたときに態様 38 記載の方法段階を実行するために適応されている、ソフトウェア・プログラム。

〔態様 40〕

プロセッサ上での実行のために、コンピューティング装置上で実行されたときに態様 38 記載の方法段階を実行するために適応されたソフトウェア・プログラムを有している記憶媒体。

〔態様 41〕

コンピュータ上で実行されたときに態様 38 記載の方法段階を実行するための実行可能命令を有するコンピュータ・プログラム・プロダクト。

10

【図 1 a】

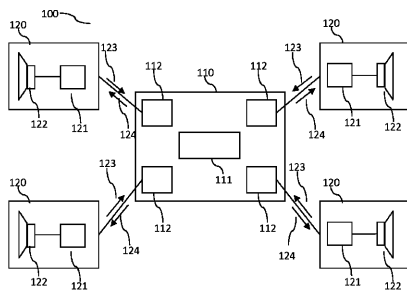


Fig. 1a

【図 2】

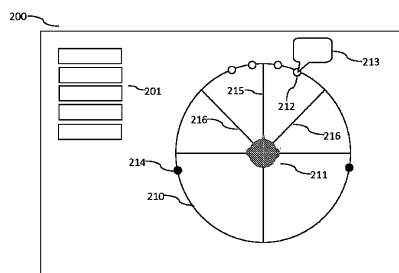


Fig. 2

【図 1 b】

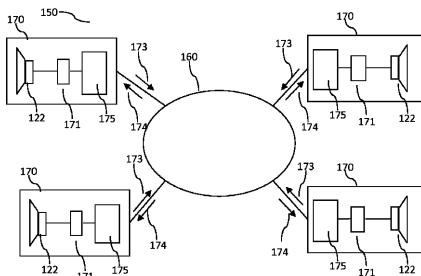
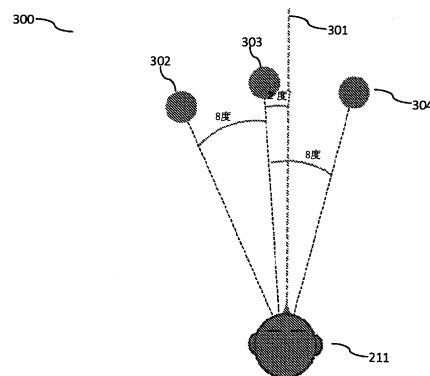
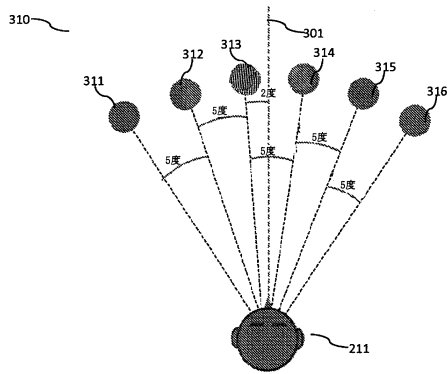


Fig. 1b

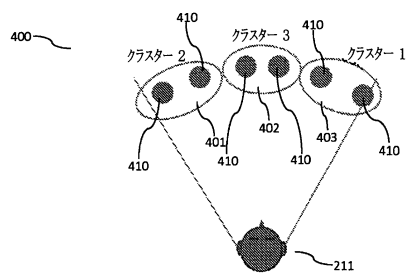
【図 3 a】



【図 3 b】



【図 4】



フロントページの続き

- (72)発明者 エッカート, マイケル
オーストラリア国, ニュー・サウス・ウェールズ 2060, マクマホーンズ・ポイント, ミッチェル・ストリート 35 レベル 3 ドルビー・オーストラリア・プロプライエタリー・リミテッド内
- (72)発明者 スピットル, デイリー
アメリカ合衆国, カリフォルニア州 94103, サンフランシスコ, ポットレロ アヴェニュー 100, ドルビー ラボラトリーズ インコーポレイテッド内
- (72)発明者 ホリアー, マイケル ピー
アメリカ合衆国, カリフォルニア州 94103, サンフランシスコ, ポットレロ アヴェニュー 100, ドルビー ラボラトリーズ インコーポレイテッド内

合議体

審判長 國分 直樹

審判官 関谷 隆一

審判官 酒井 朋広

- (56)参考文献 特開平9-90963 (JP, A)
特開2006-279492 (JP, A)
米国特許出願公開第2010/0215164 (US, A1)
特開2006-180251 (JP, A)

(58)調査した分野(Int.Cl., DB名)

H04M3/00, 3/16-3/20, 3/38-3/58, 7/00-7/16, 11/00-11/10, H04S1/00-7/00