

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2008-294619

(P2008-294619A)

(43) 公開日 平成20年12月4日(2008.12.4)

(51) Int.Cl. F I テーマコード(参考)
 HO 4 M 3/56 (2006.01) HO 4 M 3/56 Z 5 K 2 0 1

審査請求 未請求 請求項の数 5 O L (全 16 頁)

(21) 出願番号 特願2007-136427 (P2007-136427)
 (22) 出願日 平成19年5月23日(2007.5.23)

(71) 出願人 000004075
 ヤマハ株式会社
 静岡県浜松市中区中沢町10番1号
 (74) 代理人 100084548
 弁理士 小森 久夫
 (72) 発明者 畑 紀行
 静岡県浜松市中区中沢町10番1号
 ヤマハ株式会社内
 (72) 発明者 田丸 卓也
 静岡県浜松市中区中沢町10番1号
 ヤマハ株式会社内
 Fターム(参考) 5K201 AA05 BB09 BD06 CA01 CB10
 DA02 EC06 EF03 EF07

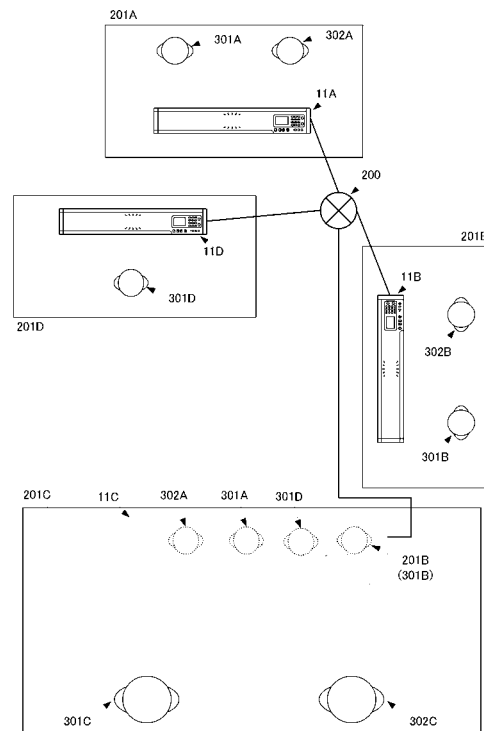
(54) 【発明の名称】 音声通信システム

(57) 【要約】

【課題】 会議に参加する各会議者が聞き取り易い多地点会議システムを提供する。

【解決手段】 ネットワーク200に接続された各收音側である放收音装置11A, 11B, 11Dは、自装置周りの発言を收音して收音データを形成するとともに、話者方位を検出する。放收音装置11A, 11B, 11Dは、話者方位情報に基づいて、話者識別データ、優先度データ、および装置識別IDを備えるメタデータを生成するか読み出し、收音データに基づく音声データに添付することで、フレームデータを形成して送信する。放音側である放收音装置11Cは、ネットワーク200経由で受信したフレームデータを解析して、メタデータに含まれる優先度データ、話者識別データ、装置識別IDを取得し、優先度が高い話者から優先して音像定位させるように放音環境を設定する。この際、音像定位数と話者識別データに基づく話者数の関係から適宜ミキシング等を行う。

【選択図】 図6



【特許請求の範囲】**【請求項 1】**

複数の音像定位を行う放音制御手段と、話者を検出するとともに話者方位からの音声を收音する收音制御手段との少なくともいずれか一方を備えた放收音装置と、

該放收音装置を含む複数の放收音装置を接続するネットワーク網と、
を備えた音声通信システムであって、

收音側の放收音装置は、前記收音制御手段を備え、收音音声に基づく音声データと、該音声データを識別させる識別データを生成して、前記收音音声に基づく音声データと前記識別データを含む通信用データを生成してネットワーク網へ送信し、

放音側の放收音装置は、前記放音制御手段を備え、前記ネットワーク網を介して受信した通信用データから前記識別データを取得して、該識別データに基づいて、関連する音声データの音像定位を決定する、

音声通信システム。

【請求項 2】

前記識別データは、話者識別データと、放收音装置に固有の装置識別 ID と、を、少なくとも備える、

請求項 1 に記載の音声通信システム。

【請求項 3】

前記識別データは、音像定位の優先性を示す優先度を、さらに備え、

放音側の放收音装置は、前記優先度が高いものを優先して音像定位させる、

請求項 2 に記載の音声通信システム。

【請求項 4】

前記收音側の放收音装置は、收音制御手段により複数の話者を並列して検出するとともに、それぞれの話者の音声を個別に收音し、收音音声毎に複数の音声データを形成するとともに音声データ毎にチャンネルを設定して、設定したチャンネル数を前記識別データに含ませる、

請求項 1 ~ 3 のいずれかに記載の音声通信システム。

【請求項 5】

前記音声通信システムは、複数の放收音装置間の前記通信用データの送受信を管理する管理サーバを備え、

該管理サーバは、複数の收音側の放收音装置から前記通信用データを受け付けると、複数の通信用データを含み、且つ拠点数を前記サーバ用識別データとして加えたサーバ通信用データを生成して、ネットワーク網へ送信し、

放音側の放收音装置は、前記サーバ用識別データおよび識別データに基づいて、自装置での收音音声を除く音声データの音像定位を決定する、

請求項 1 ~ 4 のいずれかに記載の音声通信システム。

【発明の詳細な説明】**【技術分野】****【0001】**

この発明は、互いに離れた場所に設置された複数の放收音装置をネットワーク接続して、互いに音声を通信する音声通信システムに関するものである。

【背景技術】**【0002】**

従来、それぞれに離れた複数の拠点に放收音装置を設置して、これら放收音装置をネットワークで接続することで、遠隔地間音声会議を行う音声通信システムが各種考案されている。そして、このようなシステムでは、音声データの使い勝手を良くするように、例えば特許文献 1 に示すように、音声データに話者情報等を添付して送信する收音装置が用いられている。

【特許文献 1】特開 2005 - 181391 公報

【発明の開示】

10

20

30

40

50

【発明が解決しようとする課題】

【0003】

上述のように話者位置データを添付した音声データを生成して送信するシステムの各音声データの話者同定が可能であるので、議事録の作成等に有用である。

【0004】

しかしながら、三台以上特に多数の放收音装置が接続される音声通信システムでは、放音装置（放收音装置）が複数の收音装置（放收音装置）からの話者情報付き音声データを取得しても、単に取得した順に放音したのでは、音声データを取得したタイミング毎に、放音態様が変化してしまう。例えば、同じ話者の声の放音位置が経時的に変化したり、複数の話者の音声を継続して放音している際に、一つの放音位置から経時的に異なる話者の音声が放音されてしまう。このような放音態様では、聞き取る側の会議者（聴者）にとって聞き難いものになってしまう。

10

【0005】

したがって、本発明の目的は、聴者が聴き取りやすいように放音を設定する音声通信システムを提供することにある。

【課題を解決するための手段】

【0006】

この発明は、複数の音像定位を行う放音制御手段と、話者を検出するとともに話者方位の音声を收音する收音制御手段との少なくともいずれか一方を備えた放收音装置と、該放收音装置を含む複数の放收音装置を接続するネットワーク網と、を備えた音声通信システムに関するものである。この発明では、次のような構成を特徴とする。收音側の放收音装置は、收音制御手段を備え、收音音声に基づく音声データと、該音声データを識別させる識別データとを生成する。收音側の放收音装置は、收音音声に基づく音声データと識別データとを含む通信用データを生成してネットワーク網へ送信する。放音側の放收音装置は、放音制御手段を備え、前記ネットワーク網を介して受信した通信用データから前記識別データを取得して、該識別データに基づいて、関連する音声データの音像定位を決定する。

20

【0007】

この構成では、收音側の放收音装置は、自身での收音で得られる音声データに対して固有の識別データを添付して通信用データを生成し、ネットワーク網へ送信する。放音側の放收音装置は、ネットワークを介して受信した通信用データを解析し、識別データおよびこれに関連する音声データを取得する。放音側の放收音装置は、取得した識別データに基づいて関連する音声データの音像定位を決定する。すなわち、複数の通信用データを受信した場合に、それぞれ識別データと音声データとを取得し、識別データ毎に異なる位置に音像定位を設定する。そして、識別データ毎すなわち音源位置毎に、関連するそれぞれの音声データの放音を行う。これにより、音声データ毎に音像定位位置が決定されて、放音されるので、聞き取りやすくなる。

30

【0008】

また、この発明の音声通信システムでは、話者識別データと、放收音装置に固有の装置識別IDと、を少なくとも識別データに備えることを特徴としている。

40

【0009】

この構成では、具体的に、識別データとして、話者識別データ例えば話者方位データと、装置毎に固有に設定される装置識別IDとが設定される。これにより、装置毎に決まった音像定位が行われるので、装置毎の音声の聞き分けが可能になる。さらに、話者識別データに基づく音像定位を行うことで、話者単位での音声の聞き分けが可能になる。

【0010】

また、この発明の音声通信システムでは、音像定位の優先性を示す優先度を、さらに識別データに備える。そして、放音側の放收音装置は、優先度が高いものを優先して音像定位させることを特徴としている。

【0011】

50

この構成では、識別データに対して、優先度がさらに加えられる。放音側の放收音装置は、優先度が高いものから音像定位を決定するので、例えば、社長や専務、司会者等の收音音声に対して高い優先度を設定することで、これらの話者からの音声をより聞き分け易くすることができる。

【0012】

また、この発明の音声通信システムでは、收音側の放收音装置は、收音制御手段により複数の話者を並列して検出するとともに、それぞれの話者の音声を個別に收音し、收音音声毎に複数の音声データを形成するとともに音声データ毎にチャンネルを設定して、設定したチャンネル数を識別データに含ませることを特徴としている。

【0013】

この構成では、一つの装置に対して話者が複数存在するような場合に、話者毎の音声データを利用して話者毎の音像定位ができる。すなわち、一拠点内に複数の発言者（話者）がいても、これらの話者の音声を区別して聞き取ることができる。

【0014】

また、この発明の音声通信システムは、複数の放收音装置間の通信用データの送受信を管理する管理サーバを備え、次に示す特徴を備える。この発明の管理サーバは、複数の收音側の放收音装置から通信用データを受け付けると、複数の通信用データを含み、且つ拠点数をサーバ用識別データとして加えたサーバ通信用データを生成して、ネットワーク網へ送信する。放音側の放收音装置は、サーバ用識別データおよび識別データに基づいて、自装置での收音音声を除く音声データの音像定位を決定する。

【0015】

この構成では、音声通信システムに複数の放收音装置とは別の管理サーバを備える。管理サーバは、收音側の放收音装置のそれぞれからの通信用データを一括管理し、拠点数データをさらに添付してサーバ通信用データを生成し、各放收音装置へ送信する。放音を行う放收音装置は、受信したサーバ通信用データから拠点数データおよび他の識別データを取得し、複数の音声データをそれぞれ音像定位させる。これにより、管理サーバが存在するシステム環境であっても、前述の構成と同様に聞き取り易い放音が可能となる。さらに、拠点数データを付け加えることで、放音側の放收音装置が音像定位を設定する際に、自身の設定可能定位数との関係と比較しやすく、音像定位の設定判断が容易になる。

【発明の効果】

【0016】

この発明によれば、放收音装置毎や話者毎に異なる音源を定位させることができるので、聴者にとって非常に聞き取り易い多地点間音声通信を実現することができる。

【発明を実施するための最良の形態】

【0017】

本発明の実施形態に係る音声通信システムについて、図を参照して説明する。なお、以下の説明では、音声通信システムとして、多地点会議システムを例に説明する。

【0018】

図1は、本実施形態の多地点会議システムの構成を示す概念図である。なお、図1では、4拠点で会議を行う場合であり、拠点数、拠点に在席する会議者数は一例である。

【0019】

本実施形態の多地点会議システムは、各拠点201A～201Dに設置された放收音装置11A～11Dがネットワーク200で接続されてなる。拠点201Aには放收音装置11Aが設置され、会議者301A，302Aが在席している。拠点201Bには放收音装置11Bが配置され、会議者301B，302Bが在席している。拠点201Cには放收音装置11Cが配置され、会議者301C，302Cが在席している。拠点201Dには放收音装置11Dが配置され、会議者301Dが在席している

各拠点201A～201Dに設置されている放收音装置11A～11Dは、同じ構成からなり、話者方位を検出し、当該話者方位からの收音が可能であるとともに、複数位置に対して音像定位を行うことができる機能を有する。

10

20

30

40

50

図2は各放收音装置11A~11Dの主要構成を示すブロック図である。なお、放收音装置11A~11Dは、同じ構成、仕様であるので、放收音装置11Aを例に機能を説明する。

図3は多地点会議システムで用いられる通信用データすなわちフレームデータの構成を示す図であり、(A)は1台の装置に対してチャンネル数が一つだけ割り当てられる場合を示し、(B)は1台の装置に対して複数のチャンネル数が割り当てられる場合を示す。

【0020】

放收音装置11Aは、制御部111A、通信制御部112A、放音制御部113A、收音制御部114A、メモリ115A、スピーカアレイSPA、マイクアレイMCAを備える。なお、以下の説明では、説明を簡略化するために、まず話者方位情報が一つの場合、すなわち図3(A)に示すような場合について機能説明する。

10

【0021】

(收音時(話者方位情報が一つの場合))

收音制御部114Aは、マイクアレイMCAの各マイクMCでの收音で得られる個別收音信号に対して、所定パターンからなる遅延、振幅の制御を行うことで、それぞれに異なる複数の方位を主たる收音方位とする(言い換えれば異なる收音指向性パターンからなる)複数の收音ビーム信号を形成する。收音制御部114Aは、各收音ビーム信号のレベルを検出して、予め設定した話者検出閾値レベルを超える收音ビーム信号を検出すると、当該收音ビーム信号を收音データとして取得する。また、收音制御部114Aは、收音データとして取得した收音ビーム信号の方位を取得し、話者方位データとして制御部111Aへ与える。

20

【0022】

制御部111Aは、放收音装置全体の制御を行うとともに、收音時制御として、收音制御部114Aで生成される話者方位情報に基づいて、新たなメタデータ511を生成するか、メモリ115Aに予め記憶されているメタデータ511を読み出す。メタデータ511は、話者方位情報に基づく話者識別データ、装置ID、および優先度データを含む。例えば、制御部111Aは、話者方位情報が一つであれば、図3(A)に示すようなメタデータ511を読み出すか生成して通信制御部112Aに与える。ここで、メタデータ511が予め記憶されている場合とは、例えば、話者方位に高い優先度が設定されている場合や、装置自身に高い優先度が設定されている場合や、すでに一度メタデータを生成した方位である場合である。

30

通信制御部112Aは、收音時制御(送信制御)として、收音制御部114Aから与えられた所定の音声ファイル形式からなる收音データをネットワーク通信データ形式の音声データ510に変換し、当該音声データ510に、制御部111Aから与えられたメタデータ511をヘッダ等として添付することで、フレームデータ501を生成する。通信制御部112Aは、自装置の收音に基づくフレームデータ501を、ネットワーク200で接続された他の放收音装置11B~11Dの通信制御部112B~112Dへ送信する。

【0023】

この一連の收音時のフローを、図4を参照して説明する。

図4は、收音時の放收音装置の主要処理フローを示すフローチャートである。

40

放收音装置11Aは、電源がオン状態に制御されると初期状態となり、ネットワーク接続確認が行われるとともに、收音機能として発話検出待機状態となる。放收音装置11Aは、前述の話者検出閾値レベル以上の收音ビーム信号を検出すると、発話検出したと判断する(S101:Y)。放收音装置11Aは、この発話検出の判断が行われない間は、発話検出待機状態を維持する(S101:N)。放收音装置11Aは、発話検出を判断すると、今回検出した話者方位情報を含むメタデータが既に記憶されているかどうかを検出し、既にメタデータが存在すれば(S102:Y)、記憶されているメタデータを読み出す(S103)。これは、例えば、自装置の所定方位に司会者が在席し、当該司会者に高い優先度が与えられているような場合で、司会者の方位が予め記憶され、当該記憶された方位に対する話者識別データと、優先度データと、装置識別IDとからなるメタデータが予

50

め設定、記憶されているような場合に適用される。すなわち、放收音装置 1 1 A は、司会者の方位に対して話者検出を判断すると、当該司会者に応じて予め設定、記憶されたメタデータを読み出し、音声データに添付する。また、当該放收音装置 1 1 A が例えば本社の役員室等の拠点として優先度が高い場所であった場合でも、同様に、予め高い優先度と、装置識別 ID とからなるメタデータが設定されており、当該放收音装置 1 1 A は、常時高い優先度の優先度データを有するメタデータを読み出して音声データに添付する。なお、特に優先度が設定されていないような場合であっても、以前に話者方位情報に基づくメタデータが形成されていれば、同じ話者方位情報に対して同じメタデータが添付される。

【 0 0 2 4 】

一方、放收音装置 1 1 A は、取得した話者方位情報が、予め設定されたメタデータに含まれているものではなく、新たに検出されたものであれば (S 1 0 2 : N)、新規にメタデータを生成するとともに、メモリ 1 1 5 A に記憶する (S 1 0 4)。

10

【 0 0 2 5 】

放收音装置 1 1 A は、收音ビーム信号の取得を開始すると、順次音声データに変換し、音声データにメタデータを添付することでフレームデータを形成し、他の放收音装置へ送信する (S 1 0 5 S 1 0 6)。この処理は、発話検出が終了するまで継続して行われる (S 1 0 7 : N S 1 0 6)。放收音装置 1 1 A は、発話終了が検出されると、フレームデータの生成および送信を停止する (S 1 0 7 : Y S 1 0 8)。なお、ここで、発話終了の検出は、例えば、話者検出閾値レベルに達する收音ビーム信号が検出されない時間が予め設定した時間長に亘り継続した時点等で行う。

20

【 0 0 2 6 】

なお、前述の説明では、メタデータを生成する機能を有する放收音装置を例に説明したが、メタデータを生成する機能を有さない放收音装置の場合には、単に音声データのみからフレームデータを形成する。

【 0 0 2 7 】

(放音時 (受信したフレームデータ 5 0 2 の話者方位情報が一つの場合))

通信制御部 1 1 2 A は、放音時制御 (受信制御) として、他の放收音装置 1 1 B ~ 1 1 D から受信したフレームデータ 5 0 1 を解析して、メタデータ 5 1 1 と音声データ 5 1 0 とに分離し、メタデータ 5 1 1 を制御部 1 1 1 A へ与え、ネットワーク通信データ形式の音声データ 5 1 0 を所定の音声ファイル形式からなる放音データへ変換して放音制御部 1 1 3 A へ与える。この際、音声データ群 5 2 0 は、複数の音声データからなるので、通信制御部 1 1 2 A は、音声データ群 5 2 0 の各音声データを個別に放音データへ変換する。

30

【 0 0 2 8 】

制御部 1 1 1 A は、放收音装置全体の制御を行うとともに、放音時制御として、通信制御部 1 1 2 A を介して、ネットワーク 2 0 0 で接続された他の放收音装置 1 1 B ~ 1 1 D からメタデータ 5 1 1 を取得すると、当該メタデータ 5 1 1 を解析して、音像定位の設定等を含む放音制御データを生成して、放音制御部 1 1 3 A へ与える。なお、メタデータが存在しない場合には、送信元の IP アドレス等を判断材料として、新規に装置識別 ID を割り当てるとともに、高くない優先度を有するようにメタデータを新規設定する。制御部 1 1 1 A は、新規設定したメタデータをメモリ 1 1 5 A に記憶するとともに、放音制御データの生成要素として用いる。

40

【 0 0 2 9 】

放音制御部 1 1 3 A は、制御部 1 1 1 A からの放音制御データに基づいて、通信制御部 1 1 2 A からの放音データに対して、分離、遅延、振幅制御等の信号処理を行う。すなわち、放音制御部 1 1 3 A は、放音制御データで設定される複数の音像位置や放音ボリューム等を実現するように、スピーカアレイ S P A の各スピーカ S P へ与えるスピーカ駆動信号を生成する。スピーカアレイ S P A の各スピーカ S P は、個々に与えられたスピーカ駆動信号に基づいて放音する。

【 0 0 3 0 】

この一連の放音時のフローを、図 5 を参照して説明する。

50

図5は放音時の放收音装置の主要処理フローを示すフローチャートである。

放收音装置11Aは、電源がオン状態に制御されると初期状態となり、ネットワーク接続確認が行われるとともに、放音機能としてフレームデータの受信待機状態となる。放收音装置11Aは、ネットワークを介してフレームデータを受信すると、フレームデータの解析を行い、メタデータの検出、及び、音声データから放音データへの変換処理を行う(S201:Y)。一方、放收音装置11Aは、フレームデータが検出されなければ、受信待機状態を維持する(S201:N)。

【0031】

放收音装置11Aは、フレームデータの解析を行い、メタデータを検出すれば(S202:Y)、当該メタデータが、予め記憶されているかどうかを確認する(S203)。放收音装置11Aは、メタデータを検出できなければ(S202:N)、装置識別IDと高くない優先度とが設定された新規メタデータを生成して、記憶する(S211)。

10

【0032】

放收音装置11Aは、検出したメタデータが予め設定、記憶されているものであれば、放音環境設定の更新を行わない(S203:N)。一方、放收音装置11Aは、検出したメタデータが予め設定、記憶されたものでないか、放收音装置11Aで新規に作成したものであれば(S203:Y)、放音制御データに設定する放音環境設定の更新を行う(S212)。ここで、放音環境設定とは、例えば、異なる外部の放收音装置から受信して得られた複数の放音データが存在する場合に、各放音データの音像定位位置や音量を決定した内容を備えるものである。これにより、放收音装置11Aは、新規のメタデータが検出されれば、その時点までに設定されていた放音環境を、新規のメタデータの優先度や装置識別ID等を参照して、新たな放音環境へ再構築する。

20

【0033】

放收音装置11Aは、放音制御データ(放音環境設定)を決定すると、放音を開始する(S204)。放收音装置11Aは、放音開始とともに放音制御を行い、異なる複数位置に音像定位を行う等して、設定した放音環境を実現する(S205)。放收音装置11Aは、フレームデータの受信、放音データの生成が終了したことで放音の終了を検知するまで放音制御を継続し(S206:N S205)、放音終了を検知すれば(S206:Y)、放音制御等を終了する(S207)。

【0034】

なお、收音側の放收音装置で、複数の話者を識別した場合には、以下に示す処理を行う。

30

【0035】

(收音時(話者方位情報が複数の場合))

收音制御部114Aは、話者検出閾値レベルを超える收音ビーム信号を複数検出すると、各收音ビーム信号を個別の收音データとして取得し、対応する方位も個別に取得する。

【0036】

制御部111Aは、收音制御部114Aから話者方位情報を複数取得すると、それぞれの話者方位に対して話者識別データを生成し、話者識別データ数に対応するチャンネル数データを生成する。そして、制御部111Aは、図3(B)に示すようなメタデータ521を生成する。また、制御部111Aは、話者識別データ毎に優先度データを含むメタデータ5201A~5201Nを読み出すか生成して、通信制御部112Aに与える。

40

通信制御部112Aは、收音制御部114Aから与えられた複数の收音データをそれぞれ個別ファイルとしてネットワーク通信データ形式に変換する。通信制御部112Aは、変換された各音声データ5200A~5200Nを生成し、制御部111Aから与えられたメタデータ5201A~5201Nをヘッダ等として添付することで、要素データ520A~520Nを生成する。この際、各メタデータ5201A~5201Nと各音声データ5200A~5200Nとは関連付けされた状態で、要素データ520A~520Nは生成される。通信制御部112Aは、要素データ520A~520Nに、さらに装置識別IDおよびチャンネル数データを含むメタデータ521を添付することで、フレームデー

50

タ502を生成し、ネットワーク200で接続された他の放收音装置11B～11Dの通信制御部112B～112Dへ送信する。

【0037】

(放音時(受信した各フレームデータ502の話者方位情報が複数の場合))

通信制御部112Aは、放音時制御(受信制御)として、他の放收音装置11B～11Dから受信したフレームデータ502を解析して、メタデータ521と要素データ520A～520Nとを分離する。さらに、通信制御部112Aは、それぞれの要素データ520A～520Nをメタデータ5201A～5201Nと音声データ5200A～5200Nとに分離する。そして、通信制御部112Aは、メタデータ521、5201A～5201Nを制御部111Aへ与え、ネットワーク通信データ形式の音声データ5200A～5200Nを所定の音声ファイル形式からなる放音データへ変換して放音制御部113Aへ与える。

10

【0038】

制御部111Aは、放收音装置全体の制御を行うとともに、放音時制御として、通信制御部112Aを介して、各放收音装置11B～11Dからメタデータ521、5201A～5201Nを取得すると、当該メタデータ521、5201A～5201Nを解析して、音像定位の設定等を含む放音制御データを生成して、放音制御部113Aへ与える。なお、メタデータが存在しない場合には、送信元のIPアドレス等を判断材料として、新規に装置識別IDを割り当てるとともに、高くない優先度を有するようにメタデータを新規設定する。制御部111Aは、新規設定したメタデータをメモリ115Aに記憶するとともに、放音制御データの生成要素として用いる。

20

【0039】

放音制御部113Aは、制御部111Aからの放音制御データに基づいて、通信制御部112Aからの放音データに対して、分離、遅延、振幅制御等の信号処理を行う。すなわち、放音制御部113Aは、放音制御データで設定される複数の音像位置や放音ボリューム等を実現するように、スピーカアレイSPAの各スピーカSPへ与えるスピーカ駆動信号を生成する。スピーカアレイSPAの各スピーカSPは、個々に与えられたスピーカ駆動信号に基づいて放音する。

【0040】

次に、具体的な実施例に基づいて、放音環境の設定および実施状態を、図6～図9を参照して説明する。なお、図6～図9の図において、話者301Aが社長、話者302Aが専務、話者301Dが司会者、話者301B、302Bは一般の社員等であることを示し、話者301A、302A、301Dには高い優先度が設定されているものとする。そして、拠点201A、201B、201Dで收音した音声を拠点201Cで放音する場合を例に説明する。

30

【0041】

図6は、会議開始後で、話者301A、302A、301B、301Dが発言中の状態を示す図である。なお、発言中とは、一旦発言を行った後に一時的に発言をストップしている状態をも含む。

【0042】

(收音側の各放收音装置の処理)

拠点201Aの放收音装置11Aは、社長である話者301Aの発言と、専務である話者302Aの発言とを收音し、個別の收音データを生成する。放收音装置11Aは、話者301Aの收音データと話者302Aの收音データとを区別した状態で取得するとともに、前述の話者方位検出機能により、話者301Aと話者302Aとの話者方位を取得する。放收音装置11Aは、話者301Aに対応するメタデータと、話者302Aに対応するメタデータとを読み出す。ここで、話者301Aに対応するメタデータは、話者方位に基づいて得られる予め設定された話者識別データ(この場合、社長識別データ)と、当該話者方位に対して予め優先度高で設定された優先度データとを備える。話者302Aに対応するメタデータは、話者方位に基づいて得られる予め設定された話者識別データ(この場

40

50

合、専務識別データ)と、当該話者方位に対して予め優先度高で設定された優先度データとを備える。

放收音装置11Aは、さらに、装置識別IDと、2チャンネルを示すチャンネル数データとを生成する。放收音装置11Aは、図3(B)に示すような、話者301A, 302A毎のメタデータと音声データとからなる要素データと、装置識別IDとチャンネル数データとを含むメタデータと、からフレームデータを形成し、ネットワーク200に接続する各放收音装置へ送信する。なお、本例では、放收音装置11Aは、優先度の高い会議者のみが在席するので、装置識別IDに対応する優先度高の優先度データをメタデータに含むようにしても良い。

【0043】

拠点201Bの放收音装置11Bは、一般会議者である話者301Bの発言を收音し、收音データを生成するとともに、話者301Bの話者方位を取得する。放收音装置11Bは、話者301Bに対応するメタデータを読み出すか新規に作成する。ここで、話者301Bに対応するメタデータは、話者方位に基づいて得られる話者識別データと、当該話者方位に対して予め優先度中または優先度低で設定された優先度データとを備える。

放收音装置11Bは、さらに、装置識別IDを生成しメタデータに加えることで、図3(A)に示すような、装置識別ID、優先度中または優先度低の優先度データ、話者識別データからなるメタデータと音声データとからなるフレームデータを形成し、ネットワーク200に接続する各放收音装置へ送信する。

【0044】

拠点201Dの放收音装置11Dは、司会者である話者301Dの発言を收音して收音データを生成するとともに、話者301Dの話者方位を取得する。放收音装置11Dは、話者301Dに対応するメタデータを読み出す。ここで、話者301Dに対応するメタデータは、話者方位に基づいて得られる予め設定された話者識別データ(この場合、司会者識別データ)と、当該話者方位に対して予め優先度高で設定された優先度データとを備える。

放收音装置11Dは、さらに、装置識別IDを生成しメタデータに加えることで、図3(A)に示すような、装置識別ID、優先度高の優先度データ、話者識別データからなるメタデータと音声データとからなるフレームデータを形成し、ネットワーク200に接続する各放收音装置へ送信する。

【0045】

(放音側の放收音装置の処理)

拠点201Cの放收音装置11Cは、ネットワーク200を介して、放收音装置11A, 11B, 11Dからフレームデータを受信すると、それぞれのデータを解析して、装置識別ID、話者識別情報、各話者識別データに対応する優先度データを取得する。放收音装置11Cは、優先度高の優先度データを検出すると、対応する話者識別データを取得し、優先度データと話者識別データとに基づいて、予め設定した放音環境データから優先度高の特定話者に対する音像定位位置を設定する。例えば、図6であれば、放收音装置11Cは、優先度高である話者301A(社長)に対して、装置筐体の長尺方向の中心付近に音像を定位させる。また、放收音装置11Cは、優先度高である話者302A(専務)に対して、話者301Aから所定距離だけ操作パネルと反対側(図に向かって左側)に離れた位置に音像を定位させる。さらに、放收音装置11Cは、優先度高である話者301Dに対して、話者301Aから所定距離だけ操作パネル側に離れた位置に音像を定位させる。これらの音像定位は、話者301A, 302A, 301Dが会議から退席するまで維持される。なお、退席が確認できた場合には、設定した音像位置を開放し、他の話者に割り当てるようにしてもよい。

放收音装置11Cは優先度高の話者301A, 302A, 301Dに対する音像定位を設定すると、次に、優先度中または優先度低の話者301B(拠点201B)に対する音像定位を行う。例えば、図6の例であれば、優先度高の話者に対して設定された音像定位以外の音像定位位置から選択して、筐体の操作パネル側の端部付近に音像を定位させる。

10

20

30

40

50

【 0 0 4 6 】

放收音装置 1 1 C は、このように話者毎または拠点毎に音像定位を設定してなる放音環境を設定する。そして、放收音装置 1 1 C は、当該放音環境に準じて放音制御を行う。これにより、拠点 2 0 1 C に在席する会議者 3 0 1 C , 3 0 2 C には、話者 3 0 1 A , 3 0 2 A , 3 0 1 D、および話者 3 0 1 B (拠点 2 0 1 B) の発言がそれぞれ異なる位置から聞こえるようにすることができる。この結果、会議者にとって話者を識別しやすく聞き取り易い会議を提供することができる。

【 0 0 4 7 】

なお、上述の説明では、優先度高の話者が最初から発言をしている場合を例に示したが、優先度高の話者が途中から加わることもある。このような場合、放收音装置 1 1 C は、優先度高の話者が加わった時点で放音環境の設定更新を行う。具体的には、放收音装置 1 1 C は、優先度高の話者の音源定位を優先的に行い、当該優先度高の話者の位置で放音していた優先度中または低の話者の放音位置を移動させる。これにより、優先度高の話者の発言は、最初からであれ途中からであれ、決まった位置から放音されるので、優先度高の話者の発言を聞き取り易くなる。

【 0 0 4 8 】

次に、図 6 に示す状態から、図 7 に示すような拠点 2 0 1 E が増加した場合の処理について説明する。

図 7 は、図 6 の状態に加えて、拠点 2 0 1 E の 3 0 1 E が発言した状態を示す図である。

(追加の收音側の放收音装置の処理)

拠点 2 0 1 E の放收音装置 1 1 E は、一般会議者である話者 3 0 1 E の発言を收音し、收音データを生成するとともに、話者 3 0 1 E の話者方位を取得する。放收音装置 1 1 E は、話者 3 0 1 E に対応するメタデータを読み出すか新規に作成する。ここで、話者 3 0 1 E に対応するメタデータは、話者方位に基づいて得られる話者識別データと、当該話者方位に対して予め優先度中または優先度低で設定された優先度データとを備える。

放收音装置 1 1 E は、さらに、装置識別 ID を生成しメタデータに加えることで、図 3 (A) に示すような、装置識別 ID、優先度中または優先度低の優先度データ、話者識別データからなるメタデータと音声データとからなるフレームデータを形成し、ネットワーク 2 0 0 に接続する各放收音装置へ送信する。

【 0 0 4 9 】

(放音側の放收音装置の処理)

拠点 2 0 1 C の放收音装置 1 1 C は、すでに、話者 3 0 1 A , 3 0 2 A , 3 0 1 D、および話者 3 0 1 B (拠点 2 0 1 B) に対する音像定位が行われた状態で、新たに放收音装置 1 1 E からフレームデータを受信すると、放音環境の設定を更新する。具体的には、放收音装置 1 1 C は、フレームデータを解析し、話者 3 0 1 E (拠点 2 0 1 E) の優先度が中または低であることを検出すると音像定位の設定位置の空き状況を確認する。放收音装置 1 1 C は、空きを検出すると、当該空きである位置に音像を定位させる。例えば、図 7 の例であれば、空きの位置である筐体の操作パネルと反対側の端部付近に、話者 3 0 1 E (拠点 2 0 1 E) の音像を定位させる。

【 0 0 5 0 】

放收音装置 1 1 C は、このように話者毎または拠点毎に音像定位を設定してなる放音環境を更新設定する。そして、放收音装置 1 1 C は、当該更新設定した放音環境に準じて放音制御を行う。これにより、拠点 2 0 1 C に在席する会議者 3 0 1 C , 3 0 2 C には、話者 3 0 1 A , 3 0 2 A , 3 0 1 D や、話者 3 0 1 B (拠点 2 0 1 B)、話者 3 0 1 E (拠点 2 0 1 E) の発言がそれぞれ異なる位置から聞こえるようにすることができる。

【 0 0 5 1 】

なお、空きの音像定位位置が無い場合には、放收音装置 1 1 C は、同じ優先度中または優先度低からなる話者 3 0 1 B (拠点 2 0 1 B) と同じ位置に音像を定位させる。すなわち、放收音装置 1 1 C は、話者 3 0 1 B (拠点 2 0 1 B) の放音データと話者 3 0 1 E (

拠点 201E) とをミキシングして、元々話者 301B (拠点 201B) の音源定位位置であった筐体の操作パネル側の端部付近に音像定位を設定する。これにより、取得した放音データ数が音像定位数よりも多くても、放音を行うことができるとともに、優先度の高い話者の音声については、優先的に独立した放音を行うことができる。この結果、司会者や社長、専務のような優先的に発言すべき人の声を、他の会議者の声よりも明確且つ簡単に聞き分けることができ、より聞き取り易い会議を提供することができる。なお、図 7 に対する上述の説明では、図 6 の状態から図 7 の状態に遷移した場合を例にしたが、会議開始当初から図 7 のような状態になっても、図 6 の場合と同様に放音環境を設定することができる。

【0052】

次に、図 7 の状態から、さらに、図 8 に示すように拠点 201B の話者 302B が発言した場合の処理について説明する。

図 8 は、図 7 の状態に加えて拠点 201B の話者 302B が発言した状態を示す図である。

(収音側の放收音装置の処理)

拠点 201B の放收音装置 11B は、新たに話者 302B が発言すると、当該話者 302B の音声を収音するとともに、話者 302B に対する話者方位を検出する。

【0053】

放收音装置 11B は、話者 302B に対応するメタデータを読み出すか生成する。ここで、話者 302B に対応するメタデータは、話者方位に基づいて得られる予め設定された話者識別データと、当該話者方位に対して予め優先度中または低で設定された優先度データとを備える。

放收音装置 11B は、さらに、装置識別 ID と、2 チャンネルを示すチャンネル数データとを生成する。放收音装置 11B は、図 3 (B) に示すような、話者 301B, 302B 毎のメタデータと音声データとからなる要素データと、装置識別 ID とチャンネル数データとを含むメタデータと、からフレームデータを形成し、ネットワーク 200 に接続する各放收音装置へ送信する。

【0054】

(放音側の放收音装置の処理)

拠点 201C の放收音装置 11C は、すでに、話者 301A, 302A, 301D、話者 301B (拠点 201B)、301E (拠点 201E) に対する音像定位が行われた状態で、放收音装置 11B から複数チャンネルを有するフレームデータを受信すると、放音環境の設定を更新する。具体的には、放收音装置 11C は、更新された放收音装置 11B からのフレームデータを解析し、話者 302B の優先度が中または低であることを検出すると音像定位の設定位置の空き状況を確認する。放收音装置 11C は、空きが無いことを検出すると、同じ装置識別 ID である話者 301B と同じ位置に音像定位させる。例えば、図 8 の例であれば、話者 301B の音像定位位置である筐体の操作パネル側の端部付近に、拠点 201B の音声として音像を定位させる。

【0055】

放收音装置 11C は、このように話者毎または拠点毎に音像定位を設定してなる放音環境を更新設定する。そして、放收音装置 11C は、当該更新設定した放音環境に準じて放音制御を行う。これにより、拠点 201C に在席する会議者 301C, 302C には、話者 301A, 302A, 301D や、拠点 201B (話者 301B + 話者 302B)、話者 301E (拠点 201E) の発言がそれぞれ異なる位置から聞こえるようにすることができる。これにより、取得した放音データ数 (発言中の話者数) が音像定位数よりも多くても、放音を行うことができるとともに、優先度の高い話者の音声については、優先的に独立した放音を行うことができる。さらに、優先度が低い話者に対しては、同じ拠点に在席する話者の音声をミキシングで聞くことができる。これにより、拠点毎等でも識別でき、聞き取り易い会議を提供することができる。

【0056】

10

20

30

40

50

なお、図 8 に対する上述の説明では、図 7 の状態から図 8 の状態に遷移した場合を例にしたが、会議開始当初から図 8 のような状態になっても、拠点単位での音源定位等を用いることで、図 6 の場合と同様に放音環境を設定することができる。

【 0 0 5 7 】

ところで、前述の各説明では、すべて同じ放收音装置を用いた場合を例に説明したが、図 9 に示すように、拠点毎に放收音装置が異なるような場合でも、上述の構成を利用することができる。この場合、話者検出ができない放收音装置や、音源定位ができない放收音装置を用いてもよい。

【 0 0 5 8 】

図 9 は本実施形態の他の多地点会議システムの構成を示す概念図である。

10

図 9 に示す会議では、図 1 に示した会議に対して、さらに、拠点 2 0 1 F、2 0 1 G が加わったものである。

拠点 2 0 1 F には、放收音装置 1 2 F が設置されるとともに、放收音装置 1 2 F の周囲に話者 3 0 1 F、3 0 2 F が在席している。放收音装置 1 2 F は、円周上に複数のスピーカが配置されたスピーカアレイと、円周上に複数のマイクが配置されたマイクアレイとを備え、音像定位数が少ないものの、複数の音像定位および話者検出が可能な構成からなる。このような放收音装置 1 2 F を用いる場合、上述のように音像定位数に達するまでは優先度の高い順に音像定位を設定し、話者数が音像定位数を超える場合には優先度の低いものから順にミキシングして放音する。これにより、上述の構成を利用することができる。

【 0 0 5 9 】

20

拠点 2 0 1 G には、ネットワーク通信用のパソコン 1 3 G、ヘッドホン音声変換器 1 4 G が設置され、話者 3 0 1 G はヘッドホン 1 5 G を装着している。このような場合、ヘッドホン音声変換器 1 4 G の仕様により、ヘッドホン 1 5 G から疑似サラウンドを放音することが可能であり、疑似サラウンドを用いる場合には、上述の音像定位を疑似サラウンドに置き換えて放音制御を行えばよい。また、ステレオ放音を行う場合には、優先度を参照しながら取得した放音データを適宜ミキシングし、2 チャンネルからなるステレオ放音を実現すればよい。一方、收音は話者がヘッドホン装着者と決まっているので話者検出を行わず話者識別データを生成しなくてもよい。この場合、パソコン 1 3 G から送信されるフレームデータは、図 1 0 に示すようになる。

図 1 0 は話者識別データの無いフレームデータ 5 0 3 の構成を示す図である。

30

このように、話者識別データが無い場合、装置識別 ID と優先度データとからなるメタデータ 5 3 1 と、音声データ 5 3 0 とからフレームデータ 5 0 3 が構成される。

【 0 0 6 0 】

ネットワーク 2 0 0 にて、パソコン 1 3 G に接続する各放收音装置 1 1 A ~ 1 1 D、1 2 F は、フレームデータを解析する際に、メタデータに話者識別データが無いことを確認すると、装置識別 ID と一意的に対応すると判断し、上述の放音環境の設定に利用する。

【 0 0 6 1 】

このように、各種の放收音装置が混在するような構成であっても、各話者（会議者）にとって聞き取り易い会議を提供することができる。

【 0 0 6 2 】

40

また、前述の説明では、サーバレスのシステムを例に説明したが、図 1 1 に示すようなサーバ 2 1 0 を有するシステムであっても、前述の構成を利用することができる。

図 1 1 はサーバ 2 1 0 を有する多地点会議システムの構成を示す概念図である。また、図 1 2 はサーバ用フレームデータ 5 0 5 の構成を示す図である。

図 1 1 に示す例は、図 1 に示した会議システムにおいて、サーバ 2 1 0 が存在する場合を示す。このような会議システムでは、各放收音装置 1 1 A ~ 1 1 D のフレームデータは、サーバ 2 1 0 へ送信される。サーバ 2 1 0 は、受信した各フレームデータを一つのサーバ用フレームデータ 5 0 5 にまとめ、当該サーバ用フレームデータ 5 0 5 のメタデータとして拠点数データを添付する。拠点数データは、受信したフレームデータ数を計測することで得られる。サーバ 2 1 0 は、サーバ用フレームデータ 5 0 5 を生成すると、各放收音

50

装置 11A ~ 11D へ送信する。

【0063】

各放收音装置 11A ~ 11D は、サーバ用フレームデータ 505 を受信すると、解析を行い、自装置からのフレームデータを除外して音像定位の設定を行う。この際、拠点数データは、少なくとも拠点単位での音像定位が可能かどうか等に用いることができる。このような構成であっても、上述の構成および処理を利用して各話者（会議者）が聞き取り易い会議を提供することができる。

【図面の簡単な説明】

【0064】

【図 1】本発明の実施形態の多地点会議システムの構成を示す概念図である。

10

【図 2】各放收音装置 11A ~ 11D の主要構成を示すブロック図である。

【図 3】多地点会議システムで用いられる通信用データすなわちフレームデータの構成を示す図である。

【図 4】收音時の放收音装置の主要処理フローを示すフローチャートである。

【図 5】放音時の放收音装置の主要処理フローを示すフローチャートである。

【図 6】会議開始後で、話者 301A, 302A, 301B, 301D が発言中の状態を示す図である。

【図 7】図 6 の状態に加えて拠点 201E の 301E が発言した状態を示す図である。

【図 8】図 7 の状態に加えて拠点 201B の話者 302B が発言した状態を示す図である。

20

【図 9】本発明の実施形態の他の多地点会議システムの構成を示す概念図である。

【図 10】話者識別データの無いフレームデータ 503 の構成を示す図である。

【図 11】サーバ 210 を有する多地点会議システムの構成を示す概念図である。

【図 12】サーバ用フレームデータ 505 の構成を示す図である。

【符号の説明】

【0065】

200 - ネットワーク、201A ~ 201G - 拠点、301A, 302A, 301B, 302B, 301C, 302C, 301D, 301E, 301F, 302F, 301G - 話者（会議者）、

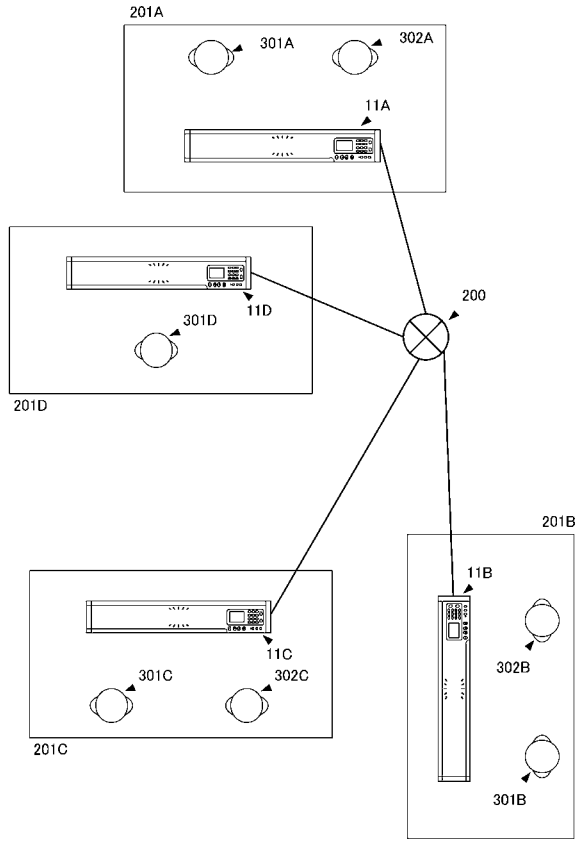
11A ~ 11E - 放收音装置、111A ~ 111D - 制御部、112A ~ 112D - 通信制御部、113A ~ 113D - 放音制御部、114A ~ 114D - 收音制御部、115A ~ 115D - メモリ、SPA - スピーカアレイ、SP - スピーカ、MCA - マイクアレイ、MC - マイク、

30

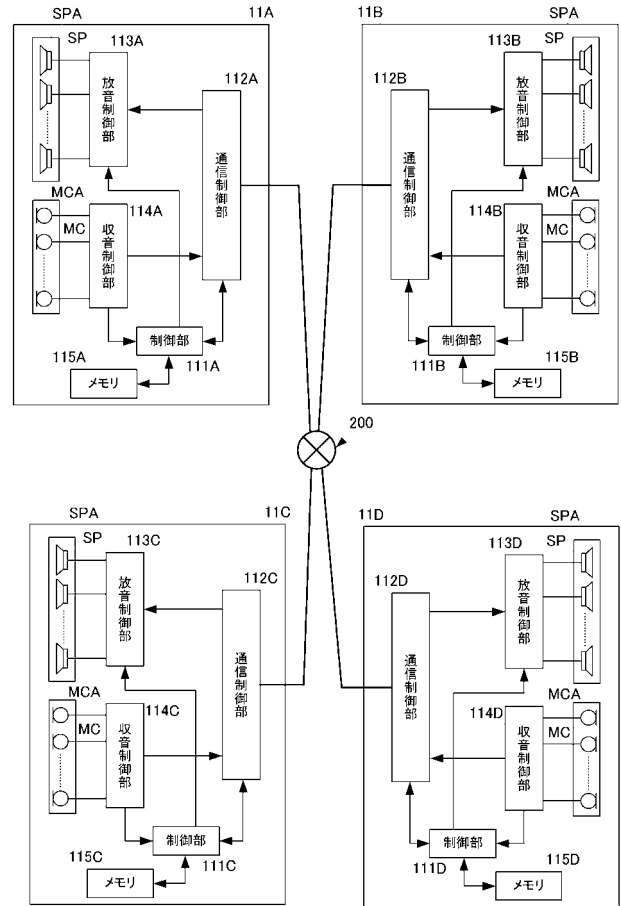
12F - 放收音装置、13G - パソコン、14G - ヘッドホン音声変換器、15G - ヘッドホン、

501, 502, 503, 504A ~ 504N - フレームデータ、505 - サーバ用フレームデータ、511, 521, 5201A ~ 5201N, 531 - メタデータ、510, 5200A ~ 5200N, 530 - 音声データ

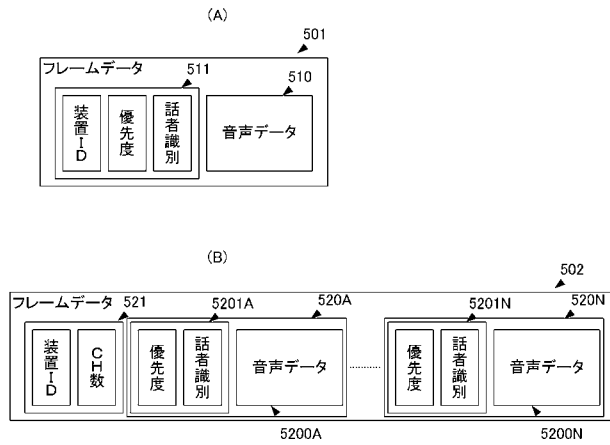
【図 1】



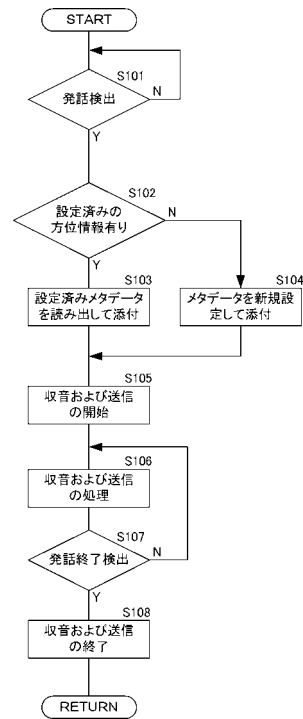
【図 2】



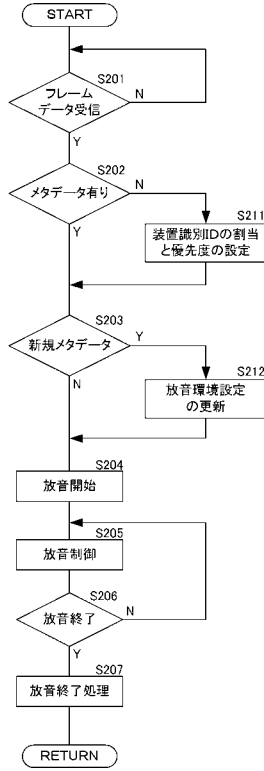
【図 3】



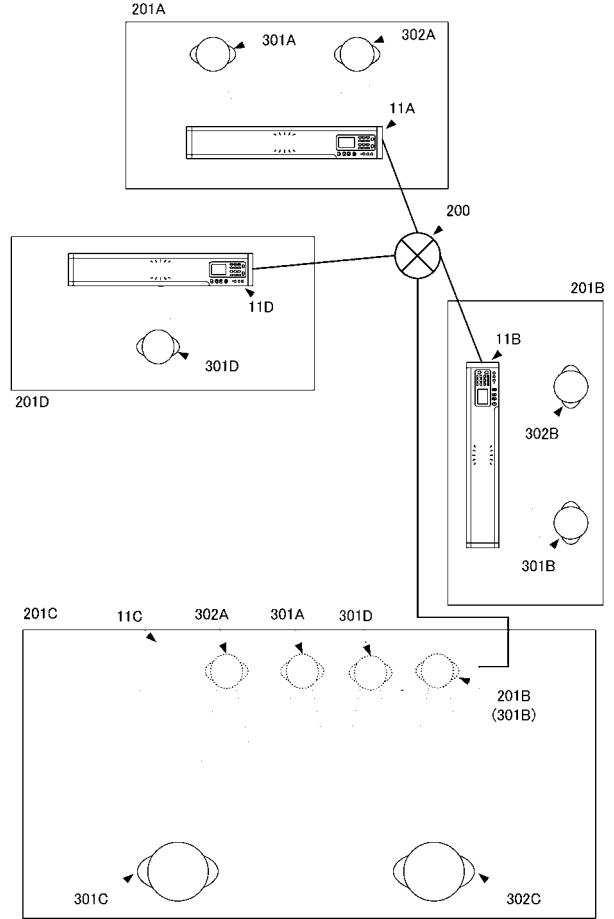
【図 4】



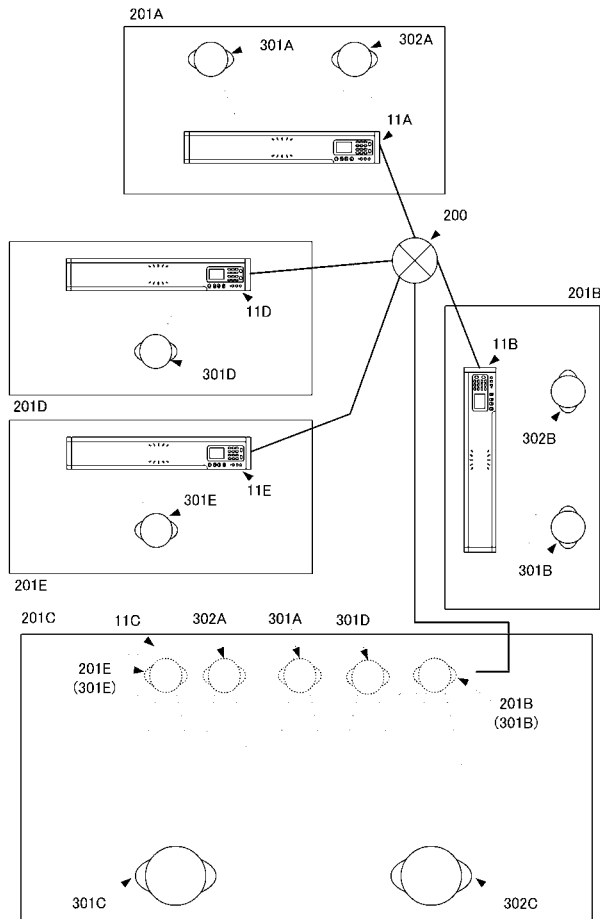
【 図 5 】



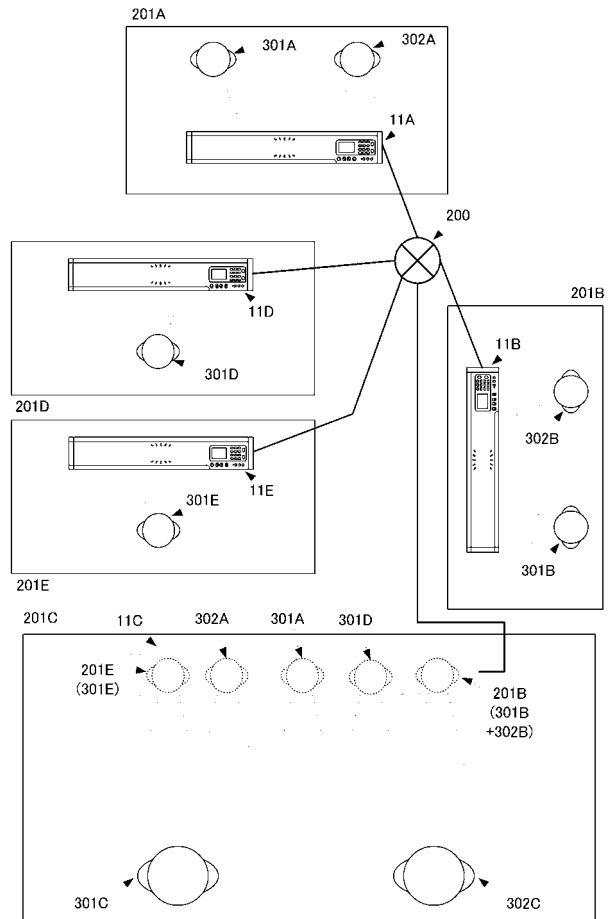
【 図 6 】



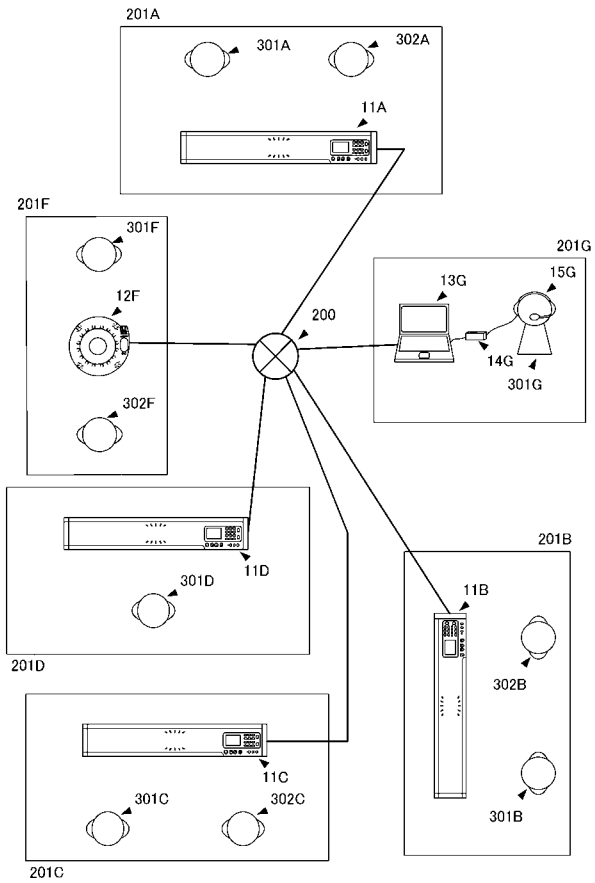
【 図 7 】



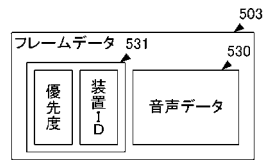
【 図 8 】



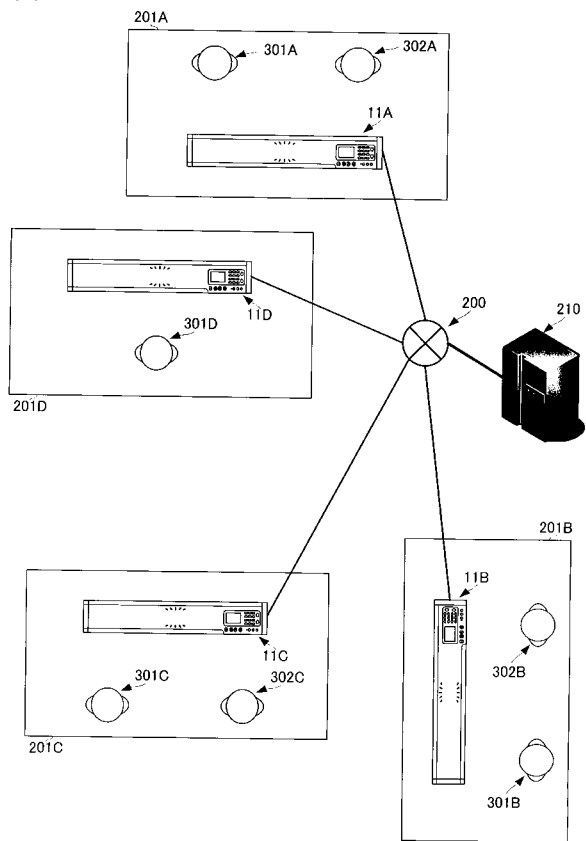
【図 9】



【図 10】



【図 11】



【図 12】

