

(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2016-201739

(P2016-201739A)

(43) 公開日 平成28年12月1日(2016.12.1)

(51) Int.Cl.	F I	テーマコード (参考)
<b>HO4M 1/738 (2006.01)</b>	HO4M 1/738	5K127
<b>HO4M 3/56 (2006.01)</b>	HO4M 3/56	B 5K201

審査請求 未請求 請求項の数 8 O L (全 24 頁)

(21) 出願番号	特願2015-81891 (P2015-81891)	(71) 出願人	000004226 日本電信電話株式会社 東京都千代田区大手町一丁目5番1号
(22) 出願日	平成27年4月13日 (2015.4.13)	(74) 代理人	100121706 弁理士 中尾 直樹
		(74) 代理人	100128705 弁理士 中村 幸雄
		(74) 代理人	100147773 弁理士 義村 宗洋
		(72) 発明者	齊藤 翔一郎 東京都千代田区大手町一丁目5番1号 日 本電信電話株式会社内
		(72) 発明者	加古 達也 東京都千代田区大手町一丁目5番1号 日 本電信電話株式会社内

最終頁に続く

(54) 【発明の名称】 音声会議システム、音声会議装置、その方法及びプログラム

(57) 【要約】

【課題】ハンズフリー通話において、1つの会議室などの1つの共通の音場において二つ以上の独立の会話を同時に行えるようにする技術を提供する。

【解決手段】音声分離通信装置は、 $K_s$ 個の音声收音再生装置のうちの一つである第一音声收音再生装置を介して指定された他地点に配置された音声分離通信装置に、 $K_s$ 個の音声信号のうちの一つの音声信号であって、第一音声收音再生装置の近傍に存在する話者の音声を強調した音声信号である第一音声信号を送信せず、残りの( $K_s - 1$ )個のうちの一つ以上の音声信号からなる第二音声信号を他地点に配置された全ての音声分離送信装置に送信する送信部と、第二音声信号の送信先から受信した再生信号を、第一音声收音再生装置以外の( $K_s - 1$ )個の音声收音再生装置のうちの一つで再生させる受信部とを含む。

【選択図】 図6

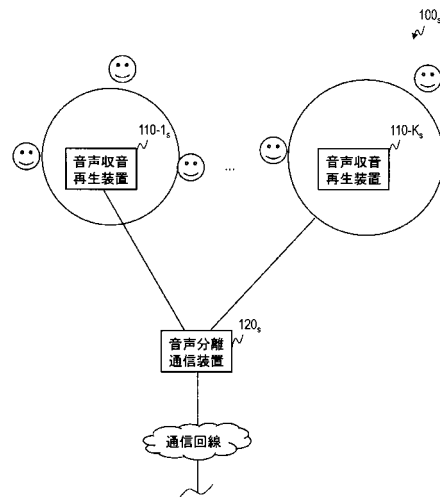


図6

## 【特許請求の範囲】

## 【請求項 1】

$K_s$  を 2 以上の整数の何れかとし、共通の音場に配置される  $K_s$  個の音声収録再生装置と、 $K$  個の前記音声収録再生装置に接続された音声分離通信装置とを含み、

前記音声収録再生装置は、音を収録し、収録信号を得、さらに、再生信号に基づき、音を再生し、

前記音声分離通信装置は、

$K_s$  個の音声収録再生装置からそれぞれ得られる  $K_s$  個の収録信号を用いて、各音声収録再生装置の近傍に存在する話者の音声を強調した  $K_s$  個の音声信号を得る音声分離部と、

$K_s$  個の前記音声収録再生装置のうちの一つである第一音声収録再生装置を介して指定された他地点に配置された音声分離通信装置に、 $K_s$  個の前記音声信号のうちの一つの音声信号であって、前記第一音声収録再生装置の近傍に存在する話者の音声を強調した音声信号である第一音声信号を送信せず、残りの  $(K_s - 1)$  個のうちの一つ以上の音声信号からなる第二音声信号を他地点に配置された全ての音声分離送信装置に送信する送信部と、

前記第二音声信号の送信先から受信した再生信号を、前記第一音声収録再生装置以外の  $(K_s - 1)$  個の音声収録再生装置のうちの一つ以上で再生させる受信部とを含む、

音声会議システム。

## 【請求項 2】

請求項 1 の音声会議システムであって、

前記他地点は 2 以上であり、

前記送信部は、前記第一音声収録再生装置を介して指定された他地点以外に配置された音声分離通信装置に、前記第一音声信号を送信し、

前記受信部は、前記第一音声信号の送信先から受信した再生信号を前記第一音声収録再生装置で再生させる、

音声会議システム。

## 【請求項 3】

請求項 1 または請求項 2 の音声会議システムであって、

前記音声分離部は、 $K_s$  個の前記収録信号の特徴量に基づいて、音声収録再生装置の近傍に存在する話者の音声を強調した音声信号として得る、

音声会議システム。

## 【請求項 4】

請求項 1 から請求項 3 の何れかの音声会議システムであって、

前記音声分離通信装置は、

$K_s$  個の収録信号に含まれる、再生信号に基づき  $K_s$  個の前記音声収録再生装置で再生した音に由来した成分を抑圧する

エコーキャンセル部を含む、

音声会議システム。

## 【請求項 5】

請求項 1 から請求項 4 の何れかの音声会議システムであって、

前記音声分離通信装置は、前記第二音声信号の送信中に、複数の前記音声収録再生装置の何れかを介してユーザにより、前記第一音声信号を送信しない音声分離通信装置を指定し、切り替え可能である、

音声会議システム。

## 【請求項 6】

$K_s$  を 2 以上の整数の何れかとし、共通の音場に配置される  $K_s$  個の音声収録再生部と、 $K_s$  個の前記音声収録再生部に接続された音声分離通信部とを含み、

前記音声収録再生部は、音を収録し、収録信号を得、さらに、再生信号に基づき、音を再生し、

前記音声分離通信部は、

$K_s$  個の音声収録再生部からそれぞれ得られる  $K_s$  個の収録信号を用いて、各音声収録再生

10

20

30

40

50

部の近傍に存在する話者の音声を強調した $K_s$ 個の音声信号を得る音声分離部と、

$K_s$ 個の前記音声收音再生部のうちの1つである第一音声收音再生部を介して指定された他地点に配置された音声分離通信部に、 $K_s$ 個の前記音声信号のうちの1個の音声信号であって、前記第一音声收音再生部の近傍に存在する話者の音声を強調した音声信号である第一音声信号を送信せず、残りの $(K_s - 1)$ 個のうちの1個以上の音声信号からなる第二音声信号を他地点に配置された全ての音声分離送信部に送信する送信部と、

前記第二音声信号の送信先から受信した再生信号を、前記第一音声收音再生部以外の $(K_s - 1)$ 個の音声收音再生部のうちの少なくとも1つで再生させる受信部とを含む、音声会議装置。

【請求項7】

$K_s$ を2以上の整数の何れかとし、共通の音場に配置される $K_s$ 個の音声收音再生部と、 $K_s$ 個の前記音声收音再生部に接続された音声分離通信部とを用いた音声会議方法であって、前記音声收音再生部は、音を收音し、收音信号を得、さらに、再生信号に基づき、音を再生するものとし、

前記音声分離通信部が、

$K_s$ 個の音声收音再生部からそれぞれ得られる $K_s$ 個の收音信号を用いて、各音声收音再生部の近傍に存在する話者の音声を強調した $K_s$ 個の音声信号を得る音声分離ステップと、

$K_s$ 個の前記音声收音再生部のうちの1つである第一音声收音再生部を介して指定された他地点に配置された音声分離通信部に、 $K_s$ 個の前記音声信号のうちの1個の音声信号であって、前記第一音声收音再生部の近傍に存在する話者の音声を強調した音声信号である第一音声信号を送信せず、残りの $(K_s - 1)$ 個のうちの1個以上の音声信号からなる第二音声信号を他地点に配置された全ての音声分離送信部に送信する送信ステップと、

前記第二音声信号の送信先から受信した再生信号を、前記第一音声收音再生部以外の $(K_s - 1)$ 個の音声收音再生部のうちの少なくとも1つで再生させる受信ステップとを含む、音声会議方法。

【請求項8】

請求項1から請求項5の何れかの音声分離通信装置としてコンピュータを機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、遠隔地を結んで行う音声会議で使用される音声会議システム、音声会議装置、その方法及びプログラムに関する。

【背景技術】

【0002】

遠隔地を結んで行う音声会議で使用される音声会議システムの従来技術として特許文献1が知られている。特許文献1では、会議室（自地点）の2つのエリアの音声を分けて收音し、相手側（他地点）で分けて再生するステレオ音声会議システムを示す。

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2010-288114号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

グループディスカッションを一つの会議室（拠点A）に集まって行うケースを考える（図1参照）。図1の場合、同じ空間（会議室内）にいる複数のグループ（例えば2～4）は、それぞれのグループ内での会話を行いながら、隣のグループともコミュニケーションを取ることができる。しかしながら、図2のようにメンバの一部が遠隔地（拠点B）にいる場合にハンズフリー通話（送受話器を手を持つことなく通話を行うことができるように

10

20

30

40

50

設計された通信機を利用して遠隔地を結んで行う音声通話)で同じことを実現しようとする、グループ内の会話とグループ間の会話とが、同じ音声信号内に混ざってしまい、遠隔地(拠点B)のメンバは聞き分けることができず、会話が破綻してしまう。特許文献1の音声会議システムを用いた場合であっても、この問題を解決することはできない。なお、拠点とは、音波の存在する空間(音場)を意味し、例えば、同じ建物内であっても、音波が届かない空間(例えば別の部屋)であれば、別の拠点とする。

【0005】

本発明は、ハンズフリー通話において、1つの会議室などの1つの共通の音場において二つ以上の独立の会話を同時に行えるようにする音声会議システム、音声会議装置、その方法及びプログラムを提供することを目的とする。

10

【課題を解決するための手段】

【0006】

上記の課題を解決するために、本発明の一態様によれば、音声会議システムは、 $K_s$ を2以上の整数の何れかとし、共通の音場に配置される $K_s$ 個の音声收音再生装置と、 $K_s$ 個の音声收音再生装置に接続された音声分離通信装置とを含む。音声收音再生装置は、音を收音し、收音信号を得、さらに、再生信号に基づき、音を再生する。音声分離通信装置は、 $K_s$ 個の音声收音再生装置からそれぞれ得られる $K_s$ 個の收音信号を用いて、各音声收音再生装置の近傍に存在する話者の音声を強調した $K_s$ 個の音声信号を得る音声分離部と、 $K_s$ 個の音声收音再生装置のうち1つである第一音声收音再生装置を介して指定された他地点に配置された音声分離通信装置に、 $K_s$ 個の音声信号のうち1個の音声信号であって、第一音声收音再生装置の近傍に存在する話者の音声を強調した音声信号である第一音声信号を送信せず、残りの $(K_s - 1)$ 個のうち1個以上の音声信号からなる第二音声信号を他地点に配置された全ての音声分離送信装置に送信する送信部と、第二音声信号の送信先から受信した再生信号を、第一音声收音再生装置以外の $(K_s - 1)$ 個の音声收音再生装置のうち少なくとも1つで再生させる受信部とを含む。

20

【0007】

上記の課題を解決するために、本発明の他の態様によれば、音声会議装置は、 $K_s$ を2以上の整数の何れかとし、共通の音場に配置される $K_s$ 個の音声收音再生部と、 $K_s$ 個の音声收音再生部に接続された音声分離通信部とを含む。音声收音再生部は、音を收音し、收音信号を得、さらに、再生信号に基づき、音を再生する。音声分離通信部は、 $K_s$ 個の音声收音再生部からそれぞれ得られる $K_s$ 個の收音信号を用いて、各音声收音再生部の近傍に存在する話者の音声を強調した $K_s$ 個の音声信号を得る音声分離部と、 $K_s$ 個の音声收音再生部のうち1つである第一音声收音再生部を介して指定された他地点に配置された音声分離通信部に、 $K_s$ 個の音声信号のうち1個の音声信号であって、第一音声收音再生部の近傍に存在する話者の音声を強調した音声信号である第一音声信号を送信せず、残りの $(K_s - 1)$ 個のうち1個以上の音声信号からなる第二音声信号を他地点に配置された全ての音声分離送信部に送信する送信部と、第二音声信号の送信先から受信した再生信号を、第一音声收音再生部以外の $(K_s - 1)$ 個の音声收音再生部のうち少なくとも1つで再生させる受信部とを含む。

30

【0008】

上記の課題を解決するために、本発明の他の態様によれば、音声会議方法は、 $K_s$ を2以上の整数の何れかとし、共通の音場に配置される $K_s$ 個の音声收音再生部と、 $K_s$ 個の音声收音再生部に接続された音声分離通信部とを用いる。音声会議方法は、音声收音再生部は、音を收音し、收音信号を得、さらに、再生信号に基づき、音を再生するものとし、音声分離通信部が、 $K_s$ 個の音声收音再生部からそれぞれ得られる $K_s$ 個の收音信号を用いて、各音声收音再生部の近傍に存在する話者の音声を強調した $K_s$ 個の音声信号を得る音声分離ステップと、 $K_s$ 個の音声收音再生部のうち1つである第一音声收音再生部を介して指定された他地点に配置された音声分離通信部に、 $K_s$ 個の音声信号のうち1個の音声信号であって、第一音声收音再生部の近傍に存在する話者の音声を強調した音声信号である第一音声信号を送信せず、残りの $(K_s - 1)$ 個のうち1個以上の音声信号からなる第二音声信号を他

40

50

地点に配置された全ての音声分離送信部に送信する送信ステップと、第二音声信号の送信先から受信した再生信号を、第一音声收音再生部以外の( $K_s - 1$ )個の音声收音再生部のうちの少なくとも1つで再生させる受信ステップとを含む。

【発明の効果】

【0009】

本発明によれば、ハンズフリー通話において、1つの会議室などの1つの共通の音場において二つ以上の独立の会話を同時に行うことができるという効果を奏する。

【図面の簡単な説明】

【0010】

【図1】発明が解決しようとする課題を説明するための図。 10

【図2】発明が解決しようとする課題を説明するための図。

【図3】第一実施形態のポイントを説明するための図。

【図4】第一実施形態のポイントを説明するための図。

【図5】第一実施形態のポイントを説明するための図。

【図6】第一実施形態に係る音声会議システムの機能ブロック図。

【図7】図7Aは音声会議システムの收音及び送信時の処理フローを示す図、図7Bは音声会議システムの受信及び再生の処理フローを示す図。

【図8】音声收音再生装置の機能ブロック図。

【図9】音声分離通信装置の機能ブロック図。

【図10】アドレステーブルの例を示す図。 20

【図11】セッションテーブルの例を示す図。

【図12】音声分離部の機能ブロック図。

【図13】入出力先選択部124Aの機能ブロック図。

【図14】タッチパネルに表示される画面の例を示す図。

【図15】音声收音再生装置の配置例を示す図。

【図16】拠点Bの音声收音再生装置の通話先の設定変更の画面遷移例を示す図。

【図17】図16中の部分での拠点Cの端末状態を示す図。

【図18】拠点Bの様子をさらに詳細に記載した図。

【図19】音声收音再生装置の配置例を示す図。

【図20】音声收音再生装置の配置例を示す図。 30

【図21】音声收音再生装置の配置例を示す図。

【発明を実施するための形態】

【0011】

以下、本発明の実施形態について、説明する。なお、以下の説明に用いる図面では、同じ機能を持つ構成部や同じ処理を行うステップには同一の符号を記し、重複説明を省略する。また、ベクトルや行列の各要素単位で行われる処理は、特に断りが無い限り、そのベクトルやその行列の全ての要素に対して適用されるものとする。

【0012】

<第一実施形態のポイント>

まず、ハンズフリー通話において、1つの共通の音場において二つ以上の独立の会話を同時に行えるようにすることの目的について説明する。例えば、本実施形態の音声会議システムは、以下のような場面で利用される。 40

(1)複数の遠隔地でのグループディスカッションを、グループごとに会議室を設ける(図3)ことなく、同一の会議室で行いたい(図4、図中、太線の矢印間、及び、一点鎖線の矢印間でそれぞれ独立して会話を行いたい)。例えば、図3では、拠点B、拠点Cのメンバーとそれぞれグループディスカッションを行うために、2つの会議室(拠点A及びD)を設けている。

(2)3地点以上の拠点間会議を行っていて、一部の拠点間会議を他の拠点に対して秘匿したい。例えば、拠点Aが発話を行っているが、それに対する意見を拠点Bと拠点Cで事前相談してから拠点Aへ発言したい。つまり、図5に示すように拠点A、拠点B、拠点Cで 50

1つのグループ3（図中、太線の矢印間で会話を行いたい）を作り、さらに、拠点B，拠点Cで1つのグループ4（図中、一点鎖線の矢印間で会話を行いたい）を作り、グループ4の発話内容を拠点Aに対して秘匿したい。

【0013】

通常の電話回線によるハンズフリー通話では、上述の(2)の場面で目的を達成しようとした場合、ハンズフリー通話の1つの回線に宛先の違う音声信号を混在させることはできないため、拠点A，拠点B，拠点C間でハンズフリー通話を行う回線とは別に、拠点B，拠点C間で別途通話回線を設定する必要がある。しかし、ハンズフリー通話を開催した当人でない場合は各拠点の電話番号を把握していない場合が多く、また通話回線の別設定自体が手間のかかる作業である。

10

【0014】

また、近年主流になりつつある、IP網での通話を利用すると、通話先の制御は技術的には音声パケットの宛先の変更により可能となる。しかし、実際に上述の(2)の場面で行おうとした場合、発話ごとにソフトウェアを操作して手動で宛先を変更するのは煩雑な作業であり、また手動での宛先変更では異なる宛先への通話を同時に行うことはできず、どちらかを選択的にしか実行できない。

【0015】

上述の(1),(2)の場面で、発話音声と宛先を自動的に判定する方法として、音声会議装置などに付属する拡張マイクを、それぞれの話者グループの近くへ持ってくる方法が考えられる。例えば、図4の場合、拠点Aにおいて、グループ4のメンバがいるテーブルに拡張マイク5を設置し、拡張マイク5で収録した音声信号を拠点Bに送信する。また、拠点Aにおいて、グループ3のメンバがいるテーブルに拡張マイク6を設置し、拡張マイク6で収録した音声信号を拠点Cに送信する。この方法の場合、拡張マイク5、6で収録した収録信号には意図した発話者以外の声も混入する（つまり、グループ4のメンバが発した声が共通の音場に存在する拡張マイク6で収録され、また、グループ3のメンバが発した声が共通の音場に存在する拡張マイク5で収録される）ため通話が分離されない。特に、一方のグループの通話内容が他方のグループのメンバに聞かれない内容の通話だった場合（例えば、グループ4の通話内容が、拠点Cにいるグループ3のメンバに聞かれない内容だった場合）に問題となる。

20

【0016】

拠点Cにいるグループ3のメンバに、グループ4の通話内容が漏れることを完全に遮断したい場合は、図3のように、会議に使う部屋（拠点）を2つ用意してそれぞれの部屋（拠点A，D）で、それぞれ拠点B，Cにいるメンバと通話を行うことが考えられる。しかし、会議室（拠点）を2つ用意することがそれだけ手間である。

30

【0017】

特許文献1のステレオシステムを使えば、エリアごとに話者の音声を分けて収録することはできるが、そもそも両方の音声を同時に受聴するのが目的なので両方の音声は再生されるため、両方の音声は完全に分離されず、二つ以上の独立の会話を同時に行えるようにするという目的では利用できない。またステレオシステムのため同一の拠点に3つ以上のグループが存在する場合には対応できない。また、特許文献1の音声会議システムは、一体型であることから、二つのグループの物理的な距離を離すことが出来ないため、(1)の場面には不向きである。

40

【0018】

本実施形態では、複数のマイクロホンから、特定の発話者の音声のみを強調する技術とマルチチャンネルエコーキャンセラ技術とを組み合わせ、共通の音場における複数の発話を分離して送信できるようにすることにより、発話者に応じて音声の送受信先を制御し、複数のハンズフリー通話ないし会議を共通の音場で行えるようにする。

【0019】

< 第一実施形態に係る音声会議システム100<sub>s</sub> >

図6は第一実施形態に係る音声会議システム100<sub>s</sub>の機能ブロック図を、図7Aは音

50

声会議システム 100<sub>s</sub> の收音及び送信時の処理フローを、図 7 B は音声会議システム 100<sub>s</sub> の受信及び再生の処理フローを示す。

【0020】

音声会議システム 100<sub>s</sub> は、 $K_s$  個 ( $K_s$  は 2 以上) の音声收音再生装置 110 -  $k_s$  (以下、「端末」ともいう) と、少なくとも 1 つの音声分離通信装置 120<sub>s</sub> (以下、「本体」ともいう) とを含む。ただし、 $s$  は拠点を表すインデックスであり、 $s=1, 2, \dots, S$  であり、 $S$  は拠点の総数を表し、 $K_s$  は拠点  $s$  毎の音声收音再生装置の個数を表し、 $k_s=1, 2, \dots, K_s$  である。ただし、音声收音再生装置 110 - 1<sub>s</sub>、音声收音再生装置 110 - 2<sub>s</sub>、...、音声收音再生装置 110 -  $K_s$  と示した場合、拠点  $s$  に配置された音声收音再生装置であることを示す。音声收音再生装置 110 -  $k_s$  は各拠点  $s$  の各グループに対して少なくとも 2 つ配置される。

10

【0021】

$K_s$  個の音声收音再生装置 110 -  $k_s$  は、共通の音場に配置される。

【0022】

音声分離通信装置 120<sub>s</sub> と  $K_s$  個の音声收音再生装置 110 -  $k_s$  とは、通信可能に接続される。例えば、無線通信(例えば Bluetooth(登録商標))により接続されてもよい。無線通信とすることで、配線の手間を省略することができ、配線による制限を受けずに自由に配置することができ、配線により美観を損なわないというメリットがある。

【0023】

図 8 は、音声收音再生装置 110 -  $k_s$  の機能ブロック図を示す。音声收音再生装置 110 -  $k_s$  は、送受信情報設定部 111 と、收音部 112 と、音声送信部 113 と音声受信部 114 と再生部 115 とを含む。

20

【0024】

音声收音再生装置 110 -  $k_s$  は、音を收音し、收音信号を得、さらに、音声信号に基づき、音を再生する。例えば、收音部 112 は音を收音し、收音信号  $x_{k,s}$  を得、音声送信部 113 を介して音声分離通信装置 120<sub>s</sub> に出力する。例えば、收音部 112 は 1 個以上のマイクロホンからなる。また、音声收音再生装置 110 -  $k_s$  は、音声受信部 114 を介して、再生信号  $z_{k,s}$  を受け取り、再生部 115 は受け取った再生信号  $z_{k,s}$  に基づき、音を再生する。例えば、再生部 115 は 1 個以上のスピーカからなる。音声收音再生装置 110 -  $k_s$  として、例えば、スマートフォン等を利用してよい。

30

【0025】

図 9 は、音声分離通信装置 120<sub>s</sub> の機能ブロック図を示す。音声分離通信装置 120<sub>s</sub> は、 $K$  個の音声受信部 121 -  $k$  と、音声分離部 123 と、送受信部 124 と、 $K$  個の音声送信部 126 -  $k$  とを含む。送受信部 124 は、入出力先選択部 124 A と、通話制御部 124 B とを含む。各部の処理内容については、処理の流れに沿って説明する。

【0026】

会議開始時の設定について

音声分離通信装置 120<sub>s</sub> の通話制御部 124 B を操作することにより、拠点間の通話を開始する。通話に関しては、相手先の音声收音再生装置と音声の RTP(Real-time Transport Protocol) パケット(参考文献 1 参照)を送受信するなどにより実現する。

40

(参考文献 1): "RTP: A Transport Protocol for Real-Time Applications," IETF Network Working Group Request for Comments: 3550, July 2003.

【0027】

音声分離通信装置 120<sub>s</sub> は、通信中の他地点(他の拠点)のアドレス・ポート番号と通話単位となるセッション(以下「通話セッション」ともいう)の ID(以下「セッション ID」ともいう)の組み合わせが格納されたアドレステーブルを記憶する。図 10 は、アドレステーブルの例を示す。なお、IP アドレスが他地点の音声分離通信装置 120<sub>s</sub> のアドレスを表し(ただし、 $s'=1, 2, \dots, S$  であり、 $s \neq s'$  とする)、同一拠点に対し複数の通話セッションを設定する場合はポート番号により区別する。他地点の音声分離通信装置 120<sub>s</sub> では、ポート番号で通話セッションを区別し、通話セッションに括り付けられた端末を

50

判定する。

【0028】

まず、通話を開始する際に、アドレステーブルにおける通話セッション(図中、セッションIDとする)＝「1」に対して、全拠点(全他地点)のアドレスが登録される。また、セッションテーブルにおける自地点のすべての音声収音再生装置110-k(図中、端末番号情報とする)に対して通話セッション＝「1」が登録される。図11は、通話セッション＝「1」に登録されている音声収音再生装置を表すセッションテーブルの例を示す。言い換えると、通話を開始する際に、新たにセッションを作成し、全拠点の全音声収音再生装置を登録する。

【0029】

同じセッションIDが付与された自地点の音声収音再生装置と他地点の音声収音再生装置の間で通話でき、会議開始時には、すべての音声収音再生装置110- $k_s$ はすべての拠点と通話できる状態、もしくは音声分離通信装置120 $_s$ のアドレステーブルに残っている拠点と通話できる状態になっている。

【0030】

会議中の動作について

システム全体の動作を示す。 $K_s$ 個の音声収音再生装置110- $k_s$ の収音部112が、音を収音し(図7AのS11)、収音信号 $x_{k,s}$ を得、出力する。音声収音再生装置110- $k_s$ の収音部112で収音された収音信号 $x_{k,s}$ は、音声送信部113を経て音声分離通信装置120 $_s$ の音声受信部121-kへ伝達される。 $K_s$ 個の音声収音再生装置110- $k_s$ からそれぞれ収音された $K_s$ 個の収音信号 $x_{k,s}$ を、音声分離部123で特定の発話者(特定の音声収音再生装置の近傍に存在する発話者)ごとの音声信号 $y_{k,s}$ に分離する(S12)。その音声信号 $y_{k,s}$ と特定の音声収音再生装置を示す情報である端末番号情報 $k_{c,s}$ とを入出力先選択部124Aへ入力し、入出力先選択部124Aは端末番号情報 $k_{c,s}$ を用いて、上述の設定の値(セッションテーブル及びアドレステーブル)に従い音声信号からなるパケットの送信先を設定する。通話制御部124Bは入出力先選択部124Aから渡されたパケットを宛先へ送信する(S13)。

【0031】

また、通話制御部124Bは、音声パケットを受信すると(図7BのS21)、入出力先選択部124Aに渡す。入出力先選択部124Aは再生信号と宛先情報(送信先アドレスとポート番号)とをパケットから取り出す。宛先情報から、再生対象の音声収音再生装置を選択し、選択された音声収音再生装置に対する再生信号を出力して音声分離部123へ渡す。なお、音声分離部123内のエコーキャンセラ部123Cで再生信号を利用する。音声分離部123は、音声送信部126-kを介して再生対象の音声収音再生装置110-kに再生信号 $z_{k,s}$ を出力する。 $K_s$ 個の音声収音再生装置110- $k_s$ の再生部115は、音声受信部114を介してそれぞれ再生信号 $z_{k,s}$ を受け取り、再生する(S22)。

【0032】

<音声分離部123>

音声分離部123は、 $K_s$ 個の音声収音再生装置からそれぞれ得られる $K_s$ 個の収音信号 $x_{k,s}$ を受け取り、これらの収音信号 $x_{k,s}$ を用いて、各音声収音再生装置の近傍に存在する話者の音声を強調した $K_s$ 個の音声信号 $y_{k,s}$ を得(S123)、送受信部124に出力する。

【0033】

音声分離部123の動作を説明する。音声分離部123ではたとえば参考文献2,3の技術を用いる(図12に構成を示す)。

(参考文献2)特開2014-92705号

(参考文献3)特開2014-112190号

なお、本実施形態では音声分離部123の概要を説明する。詳細については参考文献2,3に記載の技術を用いればよい。なお、以下の音声分離部123内の処理において、下付き添え字sを省略する。

【0034】

10

20

30

40

50

## &lt; サンプル周波数変換部 1 2 3 A &gt;

サンプル周波数変換部 1 2 3 A は、K個の收音信号 $x_k(i_k)$ を受け取り、サンプル周波数変換し、特定のサンプル周波数のK個の変換後の收音信号 $cx_k(i_k)$ を得、出力する。ただし、 $i_k$ は時間領域のサンプル点を表す整数のインデックスである。すなわち、 $x_k(i_k)$ は、インデックス $i_k$ で表されるサンプル点の收音信号を表す。

## 【 0 0 3 5 】

## &lt; 信号同期部 1 2 3 B &gt;

信号同期部 1 2 3 B は、K個の変換後の收音信号 $cx_k(i_k)$ を受け取り、K個の変換後の收音信号 $cx_k(i_k)$ をチャンネル間で同期させ、K個の收音信号 $tx_k(i_k)$ を得て出力する。

## 【 0 0 3 6 】

## &lt; エコーキャンセラ部 1 2 3 C &gt;

エコーキャンセラ部 1 2 3 C は、K個の收音信号 $tx_k(i_k)$ とK個の再生信号 $z_k$ を受け取り、收音信号 $tx_k(i_k)$ に含まれる、K個の再生信号 $z_k$ に基づき再生した音に由来する成分を抑制し、抑制後の收音信号 $sx_k(i_k)$ を得て出力する。この部分に関しては、既存のマルチチャンネルエコーキャンセラ等を利用する（例えば参考文献 4 参照）。

（参考文献 4）特開 2 0 0 5 - 3 4 7 9 5 7 号公報

## 【 0 0 3 7 】

このとき、エコーキャンセラの参照信号としては、後述する入出力先選択部 1 2 4 A から送られてくるK個の再生信号 $z_k$ を利用する。K個の音声收音再生装置 1 1 0 - k で再生された音に由来する成分も收音信号 $x_k$ に含まれているため、その再生音の影響を除去するためにエコーキャンセラ部 1 2 3 C を設ける。

## 【 0 0 3 8 】

なお、参照信号として利用したK個の再生信号 $z_k$ は、それぞれK個の音声送信部 1 2 6 - k に出力する。

## 【 0 0 3 9 】

## &lt; フレーム分割部 1 2 3 D &gt;

フレーム分割部 1 2 3 D は、エコー抑制後のK個の收音信号 $sx_k(i_k)$ を受け取り、所定の時間区間であるフレームに分割する。以下では、チャンネルkの r 番目のフレームrに属する收音信号を $sx_k(i_k, r, 0), \dots, sx_k(i_k, r, L-1)$ と表現する。ただし、Lはフレーム長を表す。

## 【 0 0 4 0 】

## &lt; VAD 判定部 1 2 3 E &gt;

VAD 判定部 1 2 3 E は、各チャンネルk ( $k=1, 2, \dots, K$ ) の各フレームrに属する收音信号 $sx_k(i_k, r, 0), \dots, sx_k(i_k, r, L-1)$ を受け取り、各フレームrが音声区間であるか非音声区間であるかを判定する。各フレームrに判定結果を表すラベル $r$ を付与し、出力する。例えば、また、音声区間であることを表すラベルの例は $r=1$ であり、非音声区間であることを表すラベルの例は $r=0$ である。

## 【 0 0 4 1 】

## &lt; S/Nベクトル生成部 1 2 3 G &gt;

S/Nベクトル生成部 1 2 3 G は、各チャンネルk ( $k=1, 2, \dots, K$ ) の各フレームrに属する收音信号 $sx_k(i_k, r, 0), \dots, sx_k(i_k, r, L-1)$ とラベル $r$ を受け取り、チャンネルkごとに音声区間の收音信号の大きさを非音声区間の收音信号の大きさで正規化した特徴量を得、チャンネル $k=1, \dots, K$ に対して得られた特徴量を要素とするS/Nベクトル（特徴量列）を得て出力する。「特徴量」の例は、非音声区間の收音信号の大きさに対する音声区間の收音信号の大きさの比を表す値である。本実施形態では、收音信号のパワーの平均値を「收音信号の大きさ」とする。

## 【 0 0 4 2 】

フレームrに属する收音信号 $sx_k(i_k, r, 0), \dots, sx_k(i_k, r, L-1)$ の平均パワー $P_N(k, r)$ を計算し、平均パワー $P_N(k, r)$ をk番目の要素とする平均パワーベクトル $P_N(r)=(P_N(1, r), \dots, P_N(K, r))$ を非音声パワー記憶部 1 2 3 F に格納する。

10

20

30

40

50

## 【 0 0 4 3 】

ラベル  $r$  が音声区間を表す場合、S / Nベクトル生成部 1 2 3 Gは、非音声パワー記憶部 1 2 3 Fに格納されている非音声区間のフレーム  $r'$  の平均パワーベクトル  $P_N(r') = (P_N(1, r'), \dots, P_N(K, r'))$  を取り出す。

## 【 0 0 4 4 】

さらにS / Nベクトル生成部 1 2 3 Gは、すべてのチャンネル  $k$  について、音声区間のフレーム  $r$  に属する収音信号  $s_{x_k}(i_k, r, 0), \dots, s_{x_k}(i_k, r, L-1)$  の平均パワーを  $P_N(k, r')$  で除算し、正規化平均パワー  $P_V(k, r)$  を得る。S / Nベクトル生成部 1 2 3 Gは、得られた正規化平均パワー  $P_V(k, r)$  を  $k$  番目の要素とするS / Nベクトル  $P_V(r) = (P_V(1, r), \dots, P_V(K, r))$  を出力する。上述の処理を全てのフレーム  $r$  に対して行う。S / Nベクトル  $P_V(r) = (P_V(1, r), \dots, P_V(K, r))$  が「音声区間の収音信号の大きさを非音声区間の収音信号の大きさを正規化した特徴量」に相当する。

## 【 0 0 4 5 】

< 非音声パワー記憶部 1 2 3 F >

前述のように、非音声パワー記憶部 1 2 3 Fは、S / Nベクトル生成部 1 2 3 Gで得られた平均パワーベクトル  $P_N(r)$  を格納する。

## 【 0 0 4 6 】

< ベクトル分類部 1 2 3 H >

ベクトル分類部 1 2 3 Hは、複数個のS / Nベクトル  $P_V(r)$  (K個のチャンネルに対して得られた特徴量からなる特徴量列) を受け取り、複数個のS / Nベクトル  $P_V(r)$  をクラスタリングし、各S / Nベクトル  $P_V(r)$  が属する信号区間分類 (クラスタ) を決定する。ここでは、各クラスタのラベルを  $CL$  とし、ラベル  $CL$  は非音声区間を表すラベル  $r$  (本実施形態では0) 以外の値 (本実施形態では1以上の整数) をとる。例えば、コサイン類似度を距離関数とするクラスタリングによって得られたラベル  $CL$  が、入力されたS / Nベクトル  $P_V(r)$  が属する信号区間分類を表す。ベクトル分類部 1 2 3 Hは、入力されたS / Nベクトル  $P_V(r)$  に対して得られたラベル  $CL$  をラベル  $r$  に代入してラベル  $r$  を更新する。これにより、音声区間のフレーム  $r$  のラベル  $r$  はラベル  $CL$  の値となり、非音声区間のフレーム  $r$  のラベル  $r$  は非音声区間を表す値となる。ベクトル分類部 1 2 3 Hは各フレーム  $r$  のラベル  $r$  を出力する。

## 【 0 0 4 7 】

< スペクトル算出部 1 2 3 K >

スペクトル算出部 1 2 3 Kは、フレーム分割部 1 2 3 Dで分割された、各チャンネル  $k$  の各フレーム  $r$  に属する収音信号  $s_{x_k}(i_k, r, 0), \dots, s_{x_k}(i_k, r, L-1)$  を受け取る。ここで、フレーム  $r$  での各チャンネル  $k$  の収音信号  $s_{x_k}(i_k, r, j)$  を要素とする  $K$  次元の縦ベクトルを  $x(j, r) = [s_{x_1}(i_1, r, j), \dots, s_{x_K}(i_K, r, j)]^T$  と記述する。ただし、 $[\ ]^T$  は  $[\ ]$  の転置を表す。また、フレーム  $r$  に属する  $K$  次元ベクトル  $x(0, r), \dots, x(L-1, r)$  の要素を周波数領域に変換して得られる値を要素とする  $K$  次元の縦ベクトルを  $X(f, r)$  と記述する。すなわち、フレーム  $r$  に属する  $s_{x_k}(i_k, r, 0), \dots, s_{x_k}(i_k, r, L-1)$  を周波数領域に変換して得られる値  $X(k, f, r)$  を  $k$  番目の要素とする  $K$  次元の縦ベクトルをスペクトルベクトル  $X(f, r) = [X(1, f, r), \dots, X(K, f, r)]^T$  と記述する。ただし、 $f$  は離散周波数を表すインデックスである。周波数領域への変換方法の例は、FFT (Fast Fourier Transform) などの離散フーリエ変換である。また、 $X(k, f, r)$  の振幅スペクトル  $A(k, f, r)$  を  $k$  番目の要素とする  $K$  次元の縦ベクトルを振幅スペクトルベクトル  $A(f, r) = [A(1, f, r), \dots, A(K, f, r)]^T$  と記述する。さらに、 $X(k, f, r)$  の位相スペクトル  $\phi(k, f, r)$  を  $k$  番目の要素とする  $K$  次元の縦ベクトルを位相スペクトルベクトル  $\phi(f, r) = [\phi(1, f, r), \dots, \phi(K, f, r)]^T$  と記述する。スペクトル算出部 1 2 3 Kは、 $x(j, r) = [s_{x_1}(i_1, r, j), \dots, s_{x_K}(i_K, r, j)]^T$  を周波数領域に変換し、フレーム  $r$  ごとに、 $k$  個の振幅スペクトル  $A(k, f, r)$  からなる振幅スペクトルベクトル  $A(f, r)$  と、 $k$  個の位相スペクトル  $\phi(k, f, r)$  からなる位相スペクトルベクトル  $\phi(f, r)$  を得て出力する。

## 【 0 0 4 8 】

10

20

30

40

50

< 振幅スペクトル記憶部 1 2 3 L 及び位相スペクトル記憶部 1 2 3 M >

振幅スペクトルベクトル  $A(f, r)$  は振幅スペクトル記憶部 1 2 3 L に格納され、位相スペクトルベクトル  $(f, r)$  は位相スペクトル記憶部 1 2 3 M に格納される。

【 0 0 4 9 】

< フィルタ係数算出部 1 2 3 I 及びフィルタ係数記憶部 1 2 3 J >

フィルタ係数算出部 1 2 3 I は、ベクトル分類部 1 2 3 H から出力された各フレーム  $r$  のラベル  $r$ 、および振幅スペクトル記憶部 1 2 3 L から読み出した振幅スペクトルベクトル  $A(f, r)$  を受け取る。ここでラベル  $r$  がとり得る値 (分類ラベル番号) のうち、音を強調する信号区間分類 (強調信号区間分類) を表す分類ラベル番号を  $c$  とする。1 個の分類ラベル番号  $c$  のみが設定されてもよいし、複数個の分類ラベル番号  $c$  が設定されてもよい。よって、 $r=c$  は、フレーム  $r$  が強調信号区間分類に分類されていることを表す。

10

【 0 0 5 0 】

フィルタ係数算出部 1 2 3 I は、強調信号区間分類に属する  $S/N$  ベクトル  $P_V(r)$  に対応する振幅スペクトル  $A(k, f, r)$  を強調するフィルタリングのためのフィルタ係数  $w_c(f)$  を算出し、出力する。なお、このフィルタ係数  $w_c(f)$  は、収音信号に含まれる各分類ラベル番号  $c$  の音声を強調するフィルタ係数である。またフィルタ係数  $w_c(f)$  は、チャンネル  $k$  に対応する係数  $w_c(f, k)$  を  $k$  番目の要素とする  $K$  次元の横ベクトル  $[w_c(f, 1), \dots, w_c(f, K)]$  である。フィルタ係数算出部 1 2 3 I は、各インデックス  $f$  および各分類ラベル番号  $c$  についてフィルタ係数  $w_c(f)$  を得て出力する。さらにフィルタ係数算出部 1 2 3 I は、 $r=c$  である各フレーム  $r$  の  $S/N$  ベクトル  $P_V(r)$  の要素のうち最大の要素に対応するチャンネルを、最大チャンネル番号  $k_{c, r}$  として得る。なお、以下、「最大チャンネル番号」を「端末番号情報」ともいう。フィルタ係数算出部 1 2 3 I は、フィルタ係数  $w_c(f)$  と端末番号情報  $k_{c, r}$  とを各分類ラベル番号  $c$  に対応付け、フィルタ係数記憶部 1 2 3 J に格納する。 $k_{c, r}$  はフレーム  $r$  において音声を最もよく収音している音声収音再生装置を表す。

20

【 0 0 5 1 】

< フィルタリング部 1 2 3 N >

フィルタリング部 1 2 3 N は、フィルタ係数記憶部 1 2 3 J から読み出したフィルタ係数  $w_c(f)$ 、および振幅スペクトル記憶部 1 2 3 L から読み出した振幅スペクトルベクトル  $A(f, r)$  を入力として受け取る。フィルタリング部 1 2 3 N は、振幅スペクトルベクトル  $A(f, r)$  を構成する複数個の振幅スペクトル  $A(1, f, r), \dots, A(K, f, r)$  に対し、フィルタ係数  $w_c(f) = [w_c(f, 1), \dots, w_c(f, K)]$  によるフィルタリングを行い、処理後振幅スペクトル  $A_c'(f, r)$  を得て出力する。例えばフィルタリング部 1 2 3 N は、次式のように、フィルタ係数  $w_c(f)$  と振幅スペクトルベクトル  $A(f, r)$  との内積を処理後振幅スペクトル  $A_c'(f, r)$  として得る。

30

$$A_c'(f, r) = w_c(f)A(f, r)$$

【 0 0 5 2 】

上述の処理により複数個の振幅スペクトル  $A(1, f, r), \dots, A(K, f, r)$  に対し、強調信号区間分類に属する  $S/N$  ベクトル  $P_V(r)$  に対応する振幅スペクトルを強調する処理が行われ、複数個の処理後振幅スペクトル  $A_c'(f, r)$  が得られる。

【 0 0 5 3 】

< 位相付与部 1 2 3 O >

位相付与部 1 2 3 O は、処理後振幅スペクトル  $A_c'(f, r)$  に、それに対応する位相スペクトルを付与して複素スペクトル  $Y_c(f, r)$  を得て出力する。本実施形態では、位相付与部 1 2 3 O は、フィルタ係数記憶部 1 2 3 J から各フレーム  $r$  および各分類ラベル番号  $c$  に対応する端末番号情報  $k_{c, r}$  を読み出す。位相付与部 1 2 3 O は、位相スペクトル記憶部 1 2 3 M から全チャンネル  $k$  に対応する位相スペクトル  $(k, f, r)$  を読み出し、それらから端末番号情報  $k_{c, r}$  に対応する位相スペクトル  $(k_{c, r}, f, r)$  を選択する。さらに位相付与部 1 2 3 O は、フィルタリング部 1 2 3 N から出力された処理後振幅スペクトル  $A_c'(f, r)$  を入力として受け取る。位相付与部 1 2 3 O は、以下の次式のように処理後振幅スペクトル  $A_c'(f, r)$  に位相スペクトル  $(k_{c, r}, f, r)$  を付与し、複素スペクトル  $Y_c(f, r)$  を得て出力する。

40

50

$$Y_c(f, r) = A_c'(f, r) \exp(i(k_{c,r}, f, r))$$

ただし、 $i$ は虚数単位であり、 $\exp$ は指数関数である。

【0054】

<時間領域変換部123P及び音声信号記憶部123Q>

時間領域変換部123Pは、複素スペクトル $Y_c(f, r)$ を入力として受け取り、複素スペクトル $Y_c(f, r)$ を時間領域に変換して強調音響信号 $y_c(n, r)$  ( $n=0, \dots, L-1$ )を得る。ただし、 $n$ はサンプル点を表すインデックスである。時間領域に変換する方法としては、スペクトル算出部123Kにおいて用いた周波数領域に変換する方法に対応する方法を用いればよい。さらに時間領域変換部123Pは、オーバーラップ法を用いて強調音響信号 $y_c(n, r)$  ( $n=0, \dots, L-1$ )を合成して時間領域の音声信号を得る。さらに、時間領域変換部123Pは、端末番号情報 $k_{c,r}$ が一致する音声信号を加算し、音声信号 $y_k$ を取得する。

10

【0055】

このような構成により、音声分離部123は、複数のマイクロホンの收音信号 $x_1, x_2, \dots, x_k$ から特定の発話者(特定の音声收音再生装置の近傍に存在する発話者)の音声を強調した音声信号 $y_k$ と、その過程で推定される発話者のチャネル情報(発話者の音声がどの音声收音再生装置から最も得られるかを示す情報であり、本実施形態では端末番号情報 $k_c$ である)を取り出す。参考文献2で分類ラベル番号 $c$ が複数出ることがあるため、音声信号の出力が複数になる場合を考慮し、音声信号 $y_k$ と端末番号情報 $k_c$ とを一度、音声信号記憶部123Qに記憶させ、音声信号 $y_k$ と端末番号情報 $k_c$ を1セットずつ入出力先選択部124Aに渡す。音声分離部123では、 $K$ 個の收音信号 $x_k$ の特徴量である $S/N$ ベクトル $P_V(r) = (P_V(1, r), \dots, P_V(K, r))$ に基づいて、ベクトル分類部123Hでクラスタリングし、各 $S/N$ ベクトル $P_V(r)$ が属する信号区間分類(クラスタ)を決定し、クラスタを利用してフィルタ係数を算出する。さらに、算出したフィルタ係数を用いて、音声收音再生装置の近傍に存在する話者の音声を強調した音声信号として得る。

20

【0056】

<送受信部124>

送受信部124は、送信時には、 $K_s$ 個の音声信号 $y_{k,s}$ と端末番号情報 $k_{c,s}$ とを受け取る。また、送受信部124は、受信時には、通話相手の個数分の再生信号と宛先アドレスとからなる音声パケットを受け取る。前述の通り、会議開始時には、すべての音声收音再生装置110- $k_s$ はすべての拠点と通話できる状態、もしくは音声分離通信装置120 $_s$ のアドレステーブルに残っている拠点と通話できる状態になっているため、送受信部124は、送受信設定情報 $p_{k,s}$ を受け取るまでは、 $K_s$ 個の音声信号 $y_{k,s}$ を全ての他地点に配置された音声分離通信装置120 $_s$ に送信する。また、入出力先選択部124Aの音声信号送出部124Ai(図13参照)は、セッションテーブルを参照して、通話相手の個数分の再生信号を、各再生信号を再生する音声收音再生装置110- $k_s$ の個数に応じてコピーして、 $K_s$ 個の再生信号 $z_{k,s}$ を生成する。送受信部124は、音声信号送出部124Aiで生成された $K_s$ 個の再生信号 $z_{k,s}$ をそれぞれ $K_s$ 個の音声收音再生装置110- $k_s$ で再生させる。

30

【0057】

送受信部124は、送受信設定情報 $p_{k,s}$ を受け取ると以下の処理を行う。なお、送受信設定情報 $p_{k,s}$ は、 $K_s$ 個の音声收音再生装置110- $k_s$ のうちの1つである音声收音再生装置110- $k'_s$ を介してユーザの操作により入力される情報であって、音声收音再生装置110- $k'_s$ の近傍に存在する話者の音声を強調した音声信号 $y_{k,s}$ を送信しない他地点を指定する情報である。例えば、図14は、ユーザの操作により、二つの拠点AとCのうち、拠点Aが送信しない他地点として指定された画面状態を表している。

40

【0058】

送受信部124は、音声收音再生装置110- $k'_s$ を介して指定された他地点に配置された音声分離通信装置120 $_d$  ( $d$ は1, 2, ...,  $S$ の何れかであって、 $d \neq s$ )に、 $K_s$ 個の音声信号 $y_{k,s}$ のうちの1個の音声信号 $y_{k',s}$ を送信せずに、指定された他地点以外に配置された音声分離通信装置120 $_d$  ( $d'=1, 2, \dots, S$ 、ただし $d' \neq d, d' \neq s$ )に、音声信号 $y_{k',s}$

50

を送信する。

【0059】

また、送受信部124は、音声収音再生装置110 -  $k'_s$  以外の音声収音再生装置110 -  $k_s$  の近傍に存在する話者の音声を強調した( $K_s - 1$ )個の音声信号 $y_{k,s}$ のうち1個以上の音声信号 $y_{m,s}$  ( $m \in \{1, 2, \dots, K_s\}$ ,  $m \neq k'_s$ )を他地点に配置された全ての音声分離送信装置120 $_{s'}$ に送信する。

【0060】

送受信部124は、音声信号 $y_{k',s}$ の送信先(指定された他地点以外に配置された音声分離通信装置120 $_d$ )から受信した再生信号 $z_{k',s}$ を音声収音再生装置110 -  $k'_s$ で再生させる。

10

【0061】

また、送受信部124は、音声信号 $y_{m,s}$ の送信先から受信した再生信号 $z_{m,s}$ を、音声収音再生装置110 -  $k'_s$ 以外の( $K_s - 1$ )個の音声収音再生装置110 -  $k_s$ のうち少なくとも1つで再生させる。

【0062】

例えば、送受信部124は入出力先選択部124A及び通話制御部124Bを用いて以下のように実現する。

【0063】

<入出力先選択部124A>

図13は、入出力先選択部124Aの機能ブロック図を示す。

20

【0064】

(送信時)

通話セッション変換部124Aaは、端末番号情報 $k_{c,r}$ を受け取り、セッションテーブル124Abを参照して、端末番号情報 $k_{c,r}$ に対応するセッションIDを取得し、出力する。例えば、音声分離部123から端末番号情報 $k_{c,r}$ を取得すると、通話セッション変換部124Aaは、セッションテーブル124Abを用いてセッション名に変換する。例えば、

(1)セッションID=「1」：端末番号情報=「1」

(2)セッションID=「2」：端末番号情報=「2」

のようなセッションテーブルがあるとき、 $k_{c,r}$ =「2」の場合、セッションID=「2」という情報を出力する。つまり、セッションテーブル124Abには、セッションIDとそのセッションに接続される音声収音再生装置とが対応付けられている。

30

【0065】

アドレス変換部124Acは、セッションIDを受け取り、アドレステーブル124Adを参照して、セッションIDに対応する送信先を取得し、出力する。例えば、アドレス変換部124Acでは、セッションIDを受信して送信先アドレスへ変換する。通話開始時に入力された「セッションID=「2」」という情報を受け取ると、たとえば

(1)セッションID=「1」：192.168.1.1:5004,192.168.1.3:5004

(2)セッションID=「2」：129.168.1.3:5005

のようなアドレステーブル124Adを参照し、送信先、つまり、宛先アドレスとポート番号「129.168.1.3:5005」を得る。つまり、アドレステーブル124Adには、セッションIDと送受信先のアドレスとが対応付けられている。

40

【0066】

音声パケット送信部124Aeは、送信先アドレスと音声信号 $y_{k,s}$ を受け取り、送信先アドレスに対して、音声信号をペイロードに持つ音声パケット(たとえばRTPパケット)を生成し、送信する。

(受信時)

音声パケット受信部124Afは、後述する通話制御部124Bから音声パケットを受信すると、宛先アドレスとポートの情報をアドレス変換部124Agへ渡し、再生信号 $z_{k,s}$ をペイロードに持つ音声パケットを音声信号送出部124Aiに渡す。

50

## 【0067】

アドレス変換部124Agは、宛先アドレスとポートの情報を受け取り、アドレステーブル124Adを参照して、宛先アドレスとポートの情報に対応するセッションIDを取得し、出力する。例えば、宛先アドレスとポートの情報をセッションIDに変換して、セッションIDを通話セッション変換部124Ahに渡す。

## 【0068】

通話セッション変換部124Ahは、セッションIDを受け取り、セッションテーブル124Abを参照して、セッションIDに対応する端末番号情報を受得し、音声信号送出处124Aiに出力する。

## 【0069】

音声信号送出处124Aiは、端末番号情報と音声パケットとを受け取り、端末番号情報に対応する音声收音再生装置へのみ受信した音声パケットを再生信号 $z_{k,s}$ に変換して送信する。

## 【0070】

要は、入出力先選択部124Aは、端末番号情報から送信アドレスを特定し、音声信号を送信し、宛先アドレスから自地点の音声收音再生装置を特定し、その音声收音再生装置で再生信号が再生されるように再生信号を送信する。

## 【0071】

通話先変更の動作について

会議中に音声收音再生装置110- $k'_s$ の送受信情報設定部111により、音声送受信先の設定を行う動作を示す。送受信情報設定部111は音声收音再生装置に設置されたタッチパネルなどで操作することを想定する。タッチパネルに表示される画面の例を図14に示す。ここで行う設定は音声分離通信装置120<sub>s</sub>の入出力先選択部124Aに結果が反映される。より詳しくは、セッションテーブル124Ab及びアドレステーブル124Adの内容を変更する。

## 【0072】

拠点A、B、C間で通話していると想定し、各地点の音声收音再生装置110-kは2台とする。図15は音声收音再生装置110-kの配置例を示し、拠点Aには音声分離通信装置120<sub>1</sub>、音声收音再生装置110-1<sub>1</sub>、110-2<sub>1</sub>、拠点Bには音声分離通信装置120<sub>2</sub>、音声收音再生装置110-1<sub>2</sub>、110-2<sub>2</sub>、拠点Cには音声分離通信装置120<sub>3</sub>、音声收音再生装置110-1<sub>3</sub>、110-2<sub>3</sub>が配置される。図15中、太線の矢印で接続された音声收音再生装置間で通話可能である。拠点Bの音声收音再生装置110-1<sub>2</sub>、110-2<sub>2</sub>の通話先の設定変更の画面遷移例を図16に示す。拠点Bの音声收音再生装置110-1<sub>2</sub>、110-2<sub>2</sub>には拠点A、Cのボタンが表示されている。ここで音声收音再生装置110-2<sub>2</sub>で拠点Aを選択し、拠点Aのハイライト(太線の囲い)を解除すると、「通話先設定を変更しますか?」というメッセージが表示される。OKを押すと、通話先との通信を行い、設定が完了するとメッセージが再度表示され、音声收音再生装置110-1<sub>2</sub>付近では拠点AとC、音声收音再生装置110-2<sub>2</sub>付近では拠点Cのみと通話をする設定が完了する。

## 【0073】

図16中の 部分での拠点Cの端末状態を図17に示す。

## 【0074】

拠点Bが通話先の設定変更を行うと、新しい通話先である拠点Cの音声收音再生装置110-1<sub>3</sub>、110-2<sub>3</sub>にその旨のメッセージが表示される。拠点Bとの新たな通話グループを作成する場合は、通話先設定を変更したい方の音声收音再生装置110-k<sub>3</sub>で設定変更の依頼を承認する。2拠点間であればただちに、他に通話グループに拠点がある場合は全拠点の承認が出た時点で通話先設定が変更される。通話グループを別に作らない、もともと拠点Cに音声收音再生装置110-k<sub>3</sub>が1台しかなく承認すると全拠点との通話ができなくなる場合は、承認依頼を拒否することも出来る。

## 【0075】

10

20

30

40

50

この場合の入出力先選択部 1 2 4 A での動作を図 9 に従って説明する。音声収録再生装置 1 1 0 -  $k_s$  から送受信設定情報  $p_{k, s}$  を受信した通話制御部 1 2 4 B は、対象の拠点へ新たな通話セッションの確立を要求する。要求が承認されるとその旨を音声収録再生装置 1 1 0 -  $k_s$  へ知らせるとともに、対象の拠点での承認を受けて通話セッションの確立を行う。その際に確立された新しいセッションと宛先アドレスのセットを、入出力先選択部 1 2 4 A のアドレステーブルへ登録する。また、入出力先選択部 1 2 4 A のセッションテーブルに登録されている音声収録再生装置のうち、新しいセッションへ参加した音声収録再生装置の情報を更新する。

#### 【 0 0 7 6 】

この例で通話状態がどのように変更されるかの簡略図を図 1 5 に示す。初期状態の設定変更前ではすべての音声収録再生装置同士が音声通話できるようになっていたが、設定変更により拠点 B と C の音声収録再生装置 1 1 0 -  $2_2$  , 1 1 0 -  $2_3$  は全体通話（全拠点においてなされる通話）から切り離され、音声収録再生装置 1 1 0 -  $2_2$  , 1 1 0 -  $2_3$  間での通話に切り替わっている。拠点 B の様子をさらに詳細に記載したのが図 1 8 である。音声収録再生装置 1 1 0 -  $1_2$  の近く（実線の丸）にいる会議参加者（メンバ）a ~ c は、全拠点との会話を行うことができる。一方、音声収録再生装置 1 1 0 -  $2_2$  の近く（点線の丸）にいる会議参加者 d は、拠点 C の音声収録再生装置 1 1 0 -  $2_3$  の近くにいる参加者とのみ通話でき、その会話内容は拠点 A には聞こえない。ただし、拠点 A からの音声は拠点 B の音声収録再生装置 1 1 0 -  $1_2$  から再生されるため、拠点 A の会話内容は参加者 d は多少離れているものの把握することができ、また拠点 C の音声収録再生装置 1 1 0 -  $2_3$  の近くにいる参加者の声も拠点 B の参加者 a ~ c は聞くことができる。そのため、参加者 a ~ d は両方の会話の内容を把握しながら、部分的に拠点 C とだけ会話を行うことができる。よって、図 2 のような場合に、拠点 A において、独立して二つ以上のグループディスカッション（会話）を行うことができる。

#### 【 0 0 7 7 】

この際に、参加者 d は音声収録再生装置 1 1 0 -  $1_2$  から極端に離れているわけではないため、参加者 d の声は音声収録再生装置 1 1 0 -  $1_2$  にも収録される。このままでは拠点 A に参加者 d の会話内容が伝わってしまうが、前述の音声分離部 1 2 3 及び送受信部 1 2 4 により、拠点 A へ送信される音声に参加者 d の音声は含まれないように処理を行うことが可能である。また、拠点 C の音声収録再生装置 1 1 0 -  $2_3$  の音声についても、拠点 B の音声収録再生装置 1 1 0 -  $2_2$  の再生部 1 1 5 を通して音声収録再生装置 1 1 0 -  $1_2$  に収録されるが、こちらも前述のエコーキャンセル機能により相手に伝わることはない。よって、図 1 5 に示すように拠点 A , 拠点 B , 拠点 C で 1 つのグループを作り、さらに、拠点 B , 拠点 C で 1 つのグループを作り、拠点 B , 拠点 C で作られたグループの発話内容を拠点 A に対して秘匿することができる。

#### 【 0 0 7 8 】

このような通話先の設定変更は、一度変更を行った後も再度行うことができる。例えば拠点 B の音声収録再生装置 1 1 0 -  $2_2$  が拠点 C の音声収録再生装置 1 1 0 -  $2_3$  とのみ会話していた状態から、拠点 A の音声収録再生装置 1 1 0 -  $1_1$  とのみ会話をする状態へセッション状態を変更することが可能である。この場合、拠点 C のセッションからの離脱承認と、拠点 A のセッション参加承認を受けてセッションの再作成を行う。また、作成したセッションを削除し、全拠点通話のみに戻すことも出来る。この場合は拠点 B のセッション終了承認を受けてセッションを削除する。

#### 【 0 0 7 9 】

< 効果 >

以上の構成により、ハンズフリー通話において、1 つの会議室などの 1 つの共通の音場において二つ以上の独立の会話を同時に行うことができる。ハンズフリー通話装置において、発話者に応じて音声の送信先を変更することができるようになる。また、送信音声はグループごとに分離されるが、再生音声はある程度拠点内でグループをまたいで聞こえるようにすることも出来るため、3 地点以上の拠点間会議を行っていて、拠点 A が発話を行っ

10

20

30

40

50

ているが、それに対する意見を拠点Bと拠点Cで事前相談してから拠点Aへ発言したいようなケースで全体の会話を聞きながら部分的な通話を並行して行うということも出来るようになる。

#### 【0080】

また、このような構成により、発話者に応じて自動で通話相手先を変えることができる装置を実現した。さらに、発話者分離とエコーキャンセラを組み合わせることにより、ハンズフリーで同一空間にいながら2つの通話を独立に行うことを可能にした。

エコーキャンセラ部123Cでは、收音信号 $x_{k,s}$ に含まれる $K_s$ 個の音声收音再生装置110-k<sub>s</sub>で再生された音に由来する成分を削除する。これにより、遠隔地同士でも会議室と同様のグループディスカッションが行えるようになる。また、音声が必要な他の地点に漏れないことを利用し、多地点遠隔会議中の秘匿通話にも利用することができる。

10

#### 【0081】

<変形例>

本実施形態では、拠点が3つのときについて説明したが、2つでもよいし、4つ以上であってもよい。例えば、図19を用いて、拠点が2つの場合について説明する。図19の場合、会議開始時には拠点間の全ての音声收音再生装置間で会話が可能となっている。例えば、拠点Aの音声收音再生装置110-2<sub>1</sub>を介して、拠点Bに配置された音声分離通信装置120<sub>2</sub>に音声信号 $y_{2,1}$ を送信しないと指定され、かつ、拠点Bの音声收音再生装置110-2<sub>2</sub>を介して、拠点Aに配置された音声分離通信装置120<sub>1</sub>に音声信号 $y_{2,2}$ を送信しないと指定された場合、拠点Aの音声收音再生装置110-1<sub>1</sub>と拠点Bの音声收音再生装置110-1<sub>2</sub>との間で音声パケットの送受信が行われる。この場合には、音声信号 $y_{2,1}$ 、 $y_{2,2}$ は何れの拠点にも送信しない構成となり、音声收音再生装置110-2<sub>1</sub>、110-2<sub>2</sub>は再生信号を受信せず、再生しない構成となる。この場合であっても、ハンズフリー通話において、1つの会議室などの1つの共通の音場において2つ以上の独立の会話を同時に行うことができる。つまり、拠点A及び拠点Bでは、各拠点における会話と、拠点Aと拠点Bとの間の会話とを同時に行うことができる。なお、この場合には、設定変更の依頼を承認するステップを省略してもよい。

20

#### 【0082】

3つ以上の拠点の場合、各拠点には、全拠点と通話するための音声收音再生装置と、一部の拠点に対して通話を秘匿するための音声收音再生装置とを備えればよい。例えば、図20の場合、拠点Aでは、音声收音再生装置110-1<sub>1</sub>と音声收音再生装置110-1<sub>2</sub>との間で拠点A及び拠点B間の秘匿通話が可能となり、音声收音再生装置110-3<sub>1</sub>と音声收音再生装置110-1<sub>3</sub>との間で拠点A及び拠点C間の秘匿通話が可能となり、音声收音再生装置110-2<sub>1</sub>と音声收音再生装置110-2<sub>2</sub>と音声收音再生装置110-2<sub>3</sub>との間で全拠点間の通話である全体通話が可能となる。このように、1つの拠点において2つ以上の秘匿通話を行ってもよい。

30

#### 【0083】

例えば、Nを1以上( $K_s-1$ )以下の整数の何れかとし、 $n=1,2,\dots,N$ とすると、N個の送受信設定情報 $p_{n,s}$ は、 $K_s$ 個の音声收音再生装置110-k<sub>s</sub>のうちのN個の音声收音再生装置110-n<sub>s</sub>を介してユーザの操作によりそれぞれ入力される情報であって、音声收音再生装置110-n<sub>s</sub>の近傍に存在する話者の音声を強調した音声信号 $y_{n,s}$ を送信しない他地点d( $d \in \{1,2,\dots,S\}$ 、 $d \neq s$ )を指定する情報である。なお、音声收音再生装置110-n<sub>s</sub>毎に他地点dを選択することができる。また、1つの音声收音再生装置110-n<sub>s</sub>を介して2つ以上の他地点dを指定してもよい。送受信部124は、音声收音再生装置110-n<sub>s</sub>を介して指定された他地点に配置された音声分離通信装置120<sub>d</sub>に、 $K_s$ 個の音声信号 $y_{k,s}$ のうちの1個の音声信号 $y_{n,s}$ を送信せずに、指定された他地点以外に配置された音声分離通信装置120<sub>d'</sub>( $d' \in \{1,2,\dots,S\}$ 、ただし $d' \neq d, d' \neq s$ )に、音声信号 $y_{n,s}$ を送信する。また、送受信部124は、音声收音再生装置110-n<sub>s</sub>以外の音声收音再生装置110-k<sub>s</sub>の近傍に存在する話者の音声を強調した( $K_s-N$ )個の音声信号 $y_{k,s}$ のうちの1個以上の音声信号 $y_{m,s}$ ( $m \in \{1,2,\dots,K_s\}$ 、 $m \neq n_s$ )を他地点に配置された全て

40

50

の音声分離送信装置  $120_s$  に送信する。送受信部  $124$  は、音声信号  $y_{n,s}$  の送信先（指定された他地点以外に配置された音声分離通信装置  $120_d$ ）から受信した再生信号  $z_{n,s}$  を音声収音再生装置  $110 - n_s$  で再生させる。また、送受信部  $124$  は、音声信号  $y_{m,s}$  の送信先から受信した再生信号  $z_{m,s}$  を、音声収音再生装置  $110 - n_s$  以外の  $(K_s - N)$  個の音声収音再生装置  $110 - k_s$  のうちの少なくとも1つで再生させる。この場合であっても、 $N$  は1以上の整数であり、第一実施形態と同様の処理を含むことになる。

【0084】

また、必ずしも全ての拠点で秘匿通話が必要となるわけではない。例えば、拠点Aでは秘匿通話を必要としない場合には、図21のような構成としてもよい。

【0085】

つまり、ハンズフリー通話において、1つの共通の音場において2つ以上の独立の会話を同時に行うためには、少なくとも、2つの音声収音再生装置を含む音声会議システムが、2拠点以上に設置されればよく、全体通話と秘匿通話を実現するためには、3つ以上の拠点を含み、3つ以上の拠点のうち2拠点以上に2つの音声収音再生装置を含む音声会議システムが設置されればよい。

【0086】

$K_s$  個の音声収音再生装置  $110 - k_s$  のうちの1つが音声分離通信装置  $120_s$  の機能を備える構成としてもよい。また、 $K_s$  個の音声収音再生装置  $110 - k_s$  と音声分離通信装置  $120_s$  とを一つの装置により構成してもよい。その場合、音声収音再生装置及び音声分離通信装置をそれぞれ音声収音再生部及び音声分離通信部ともいう。また、送受信部  $124$  は送信機能を持つ送信部と、受信機能を持つ受信部とに分けて構成してもよいし、本実施形態のように1つの送受信部として構成してもよい。

【0087】

なお、図16の説明において、拠点Aを選択し、拠点Aのハイライト（太線の囲い）を解除すると説明したが、拠点Bを選択し、拠点Aのハイライト（太線の囲い）を解除してもよい。この場合、通話する拠点として拠点Bを選択することは、結果的に、通話しない拠点として拠点Aを選択することに相当する。要は、インターフェースの違いであって、本実施形態のように、直接、音声信号を送信しない拠点を選択してもよいし、音声信号を送信する拠点を選択することで間接的に音声信号を送信しない拠点を選択してもよい。

【0088】

なお、本実施形態では、会議開始時には、すべての音声収音再生装置  $110 - k_s$  はすべての拠点と通話できる状態、もしくは音声分離通信装置  $120_s$  のアドレステーブルに残っている拠点と通話できる状態になっているが、アドレステーブルとセッションテーブルの設定を予め変更しておくことで、会議開始時に、秘匿通話が行われる構成としてもよい。

【0089】

<その他の変形例>

本発明は上記の実施形態及び変形例に限定されるものではない。例えば、上述の各種の処理は、記載に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的あるいは個別に実行されてもよい。その他、本発明の趣旨を逸脱しない範囲で適宜変更が可能である。

【0090】

<プログラム及び記録媒体>

また、上記の実施形態及び変形例で説明した各装置における各種の処理機能をコンピュータによって実現してもよい。その場合、各装置が有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、上記各装置における各種の処理機能がコンピュータ上で実現される。

【0091】

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気

10

20

30

40

50

記録装置、光ディスク、光磁気記録媒体、半導体メモリ等のようなものでもよい。

【0092】

また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させてもよい。

【0093】

このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶部に格納する。そして、処理の実行時、このコンピュータは、自己の記憶部に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。また、このプログラムの別の実施形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよい。さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆるASP（Application Service Provider）型のサービスによって、上述の処理を実行する構成としてもよい。なお、プログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの（コンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等）を含むものとする。

10

20

【0094】

また、コンピュータ上で所定のプログラムを実行させることにより、各装置を構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

【 図 1 】

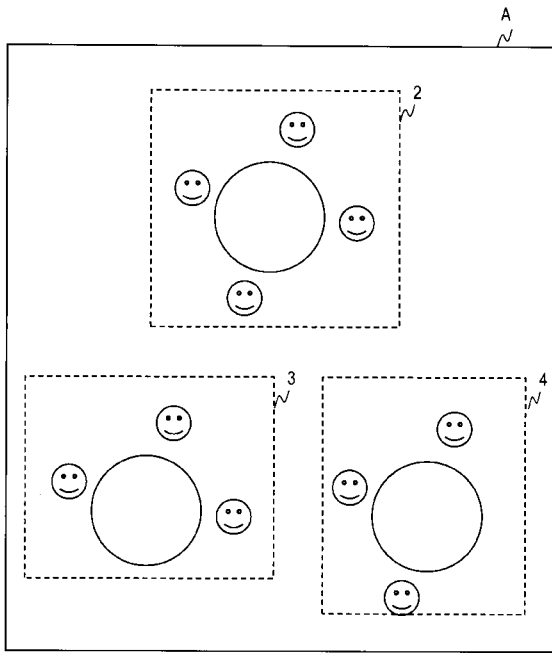


図1

【 図 2 】

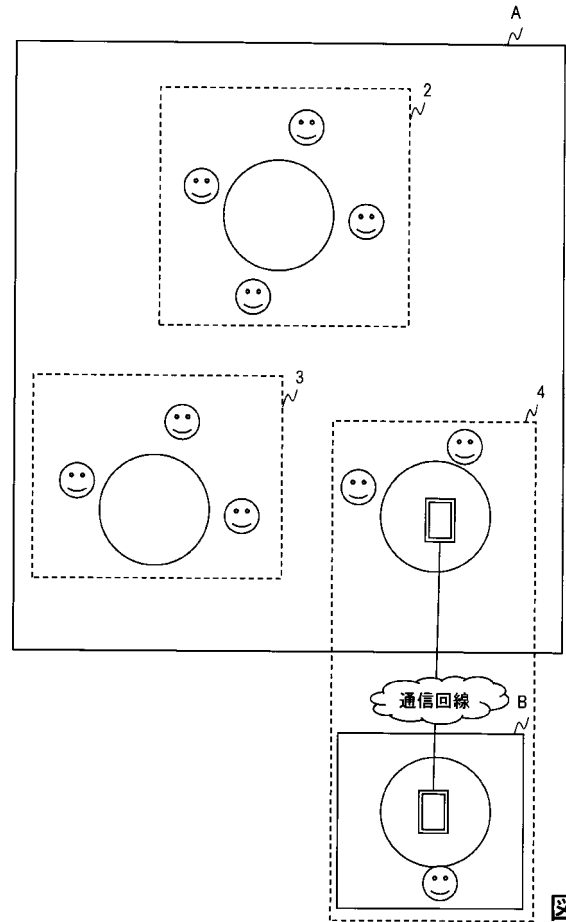


図2

【 図 3 】

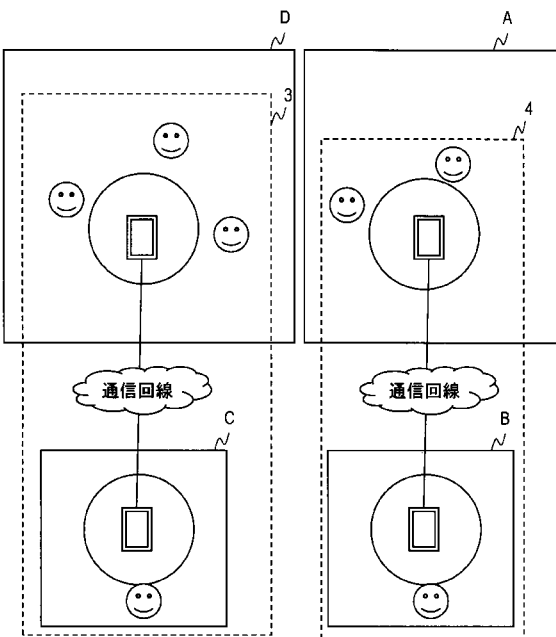


図3

【 図 4 】

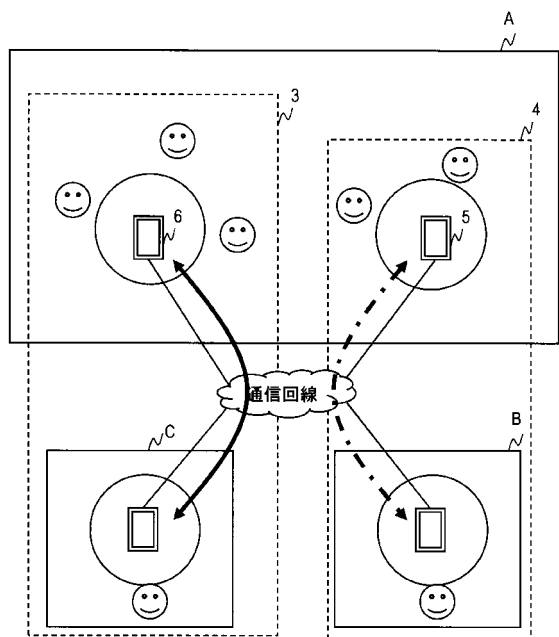


図4

【 図 5 】

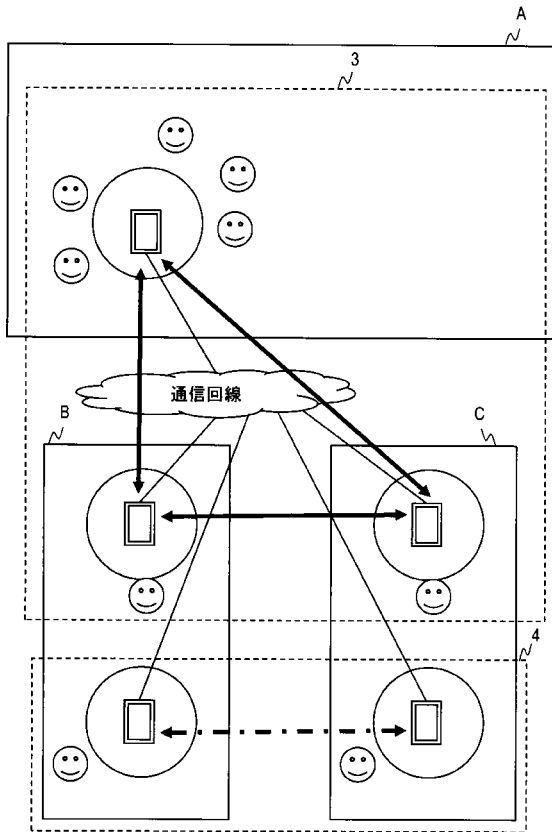


図5

【 図 6 】

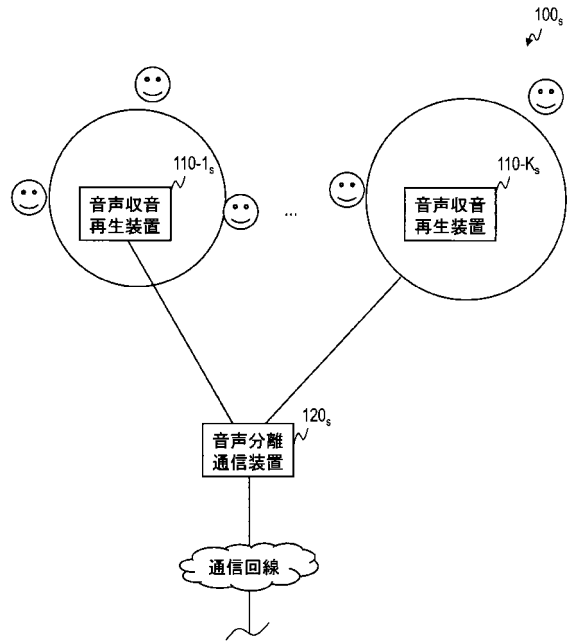


図6

【 図 7 】

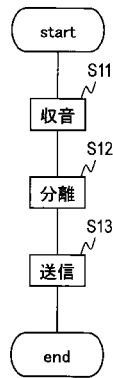


図7A

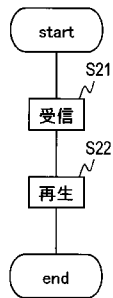


図7B

【 図 8 】

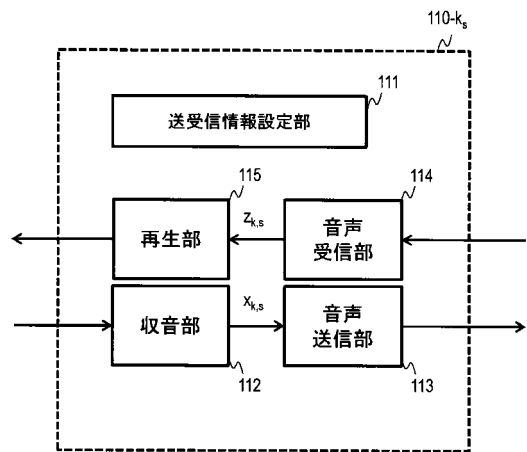


図8

【 図 9 】

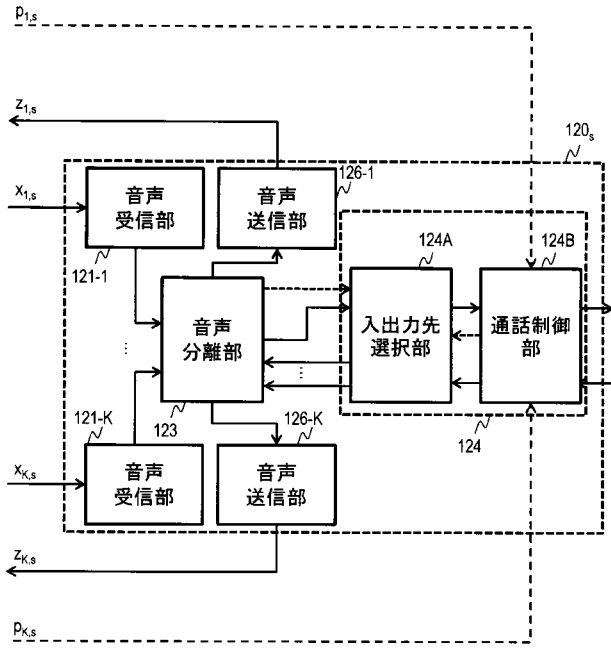


図9

【 図 1 0 】

セッションID	IPアドレス	ポート番号
1	192.168.1.1	5004
1	192.168.1.3	5004
1	192.168.1.3	5005
...	...	...

図10

【 図 1 1 】

セッションID	端末番号情報
1	1
1	2
1	3
...	...

図11

【 図 1 2 】

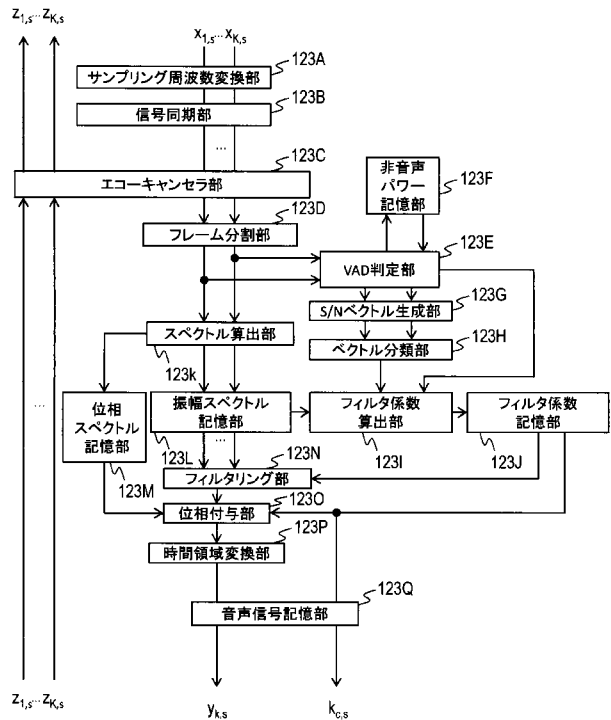


図12

【 図 1 3 】

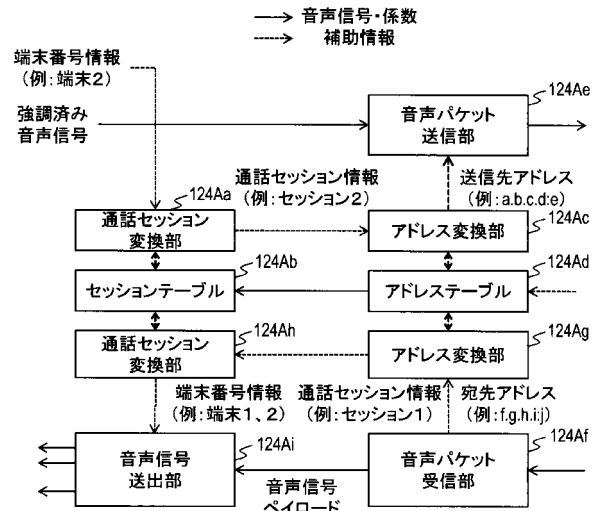


図13

【 図 1 4 】

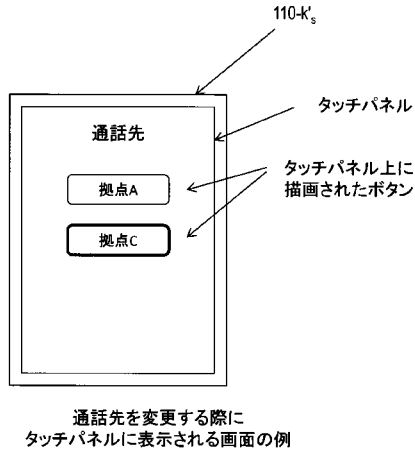


図 14

【 図 1 5 】

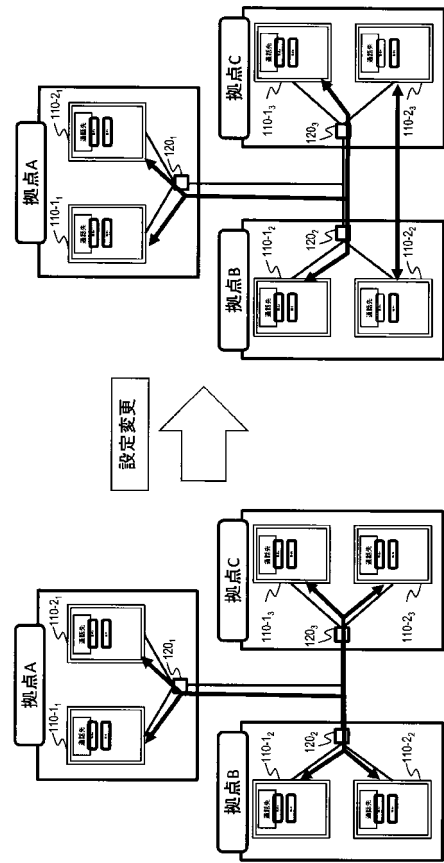


図15

【 図 1 6 】

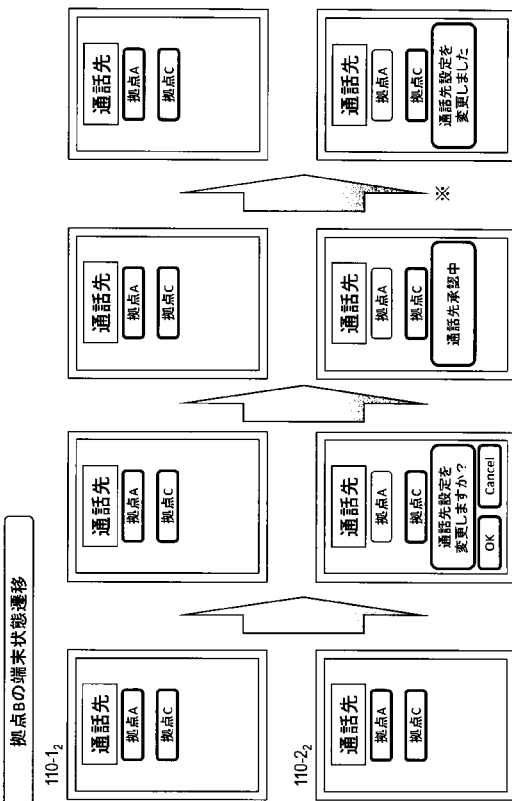


図16

【 図 1 7 】

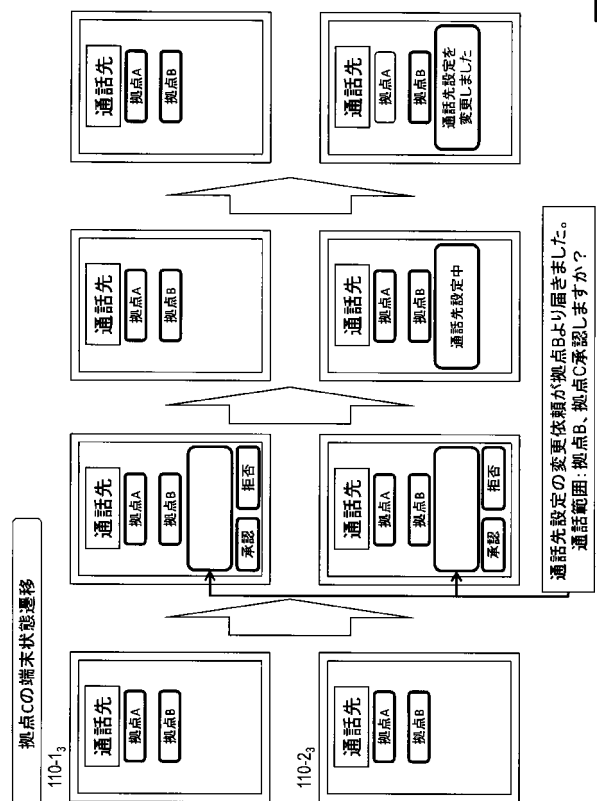


図17

【 図 18 】

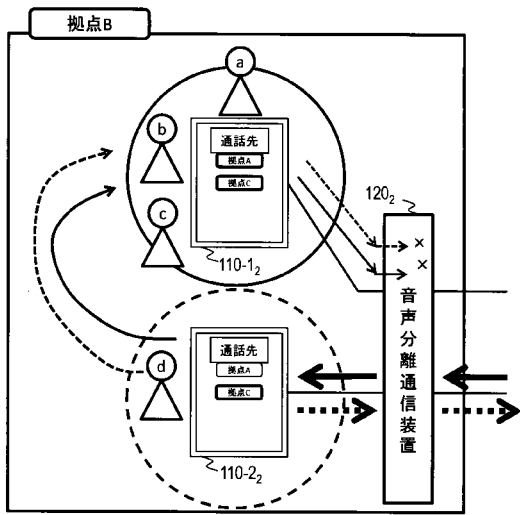


図 18

【 図 19 】

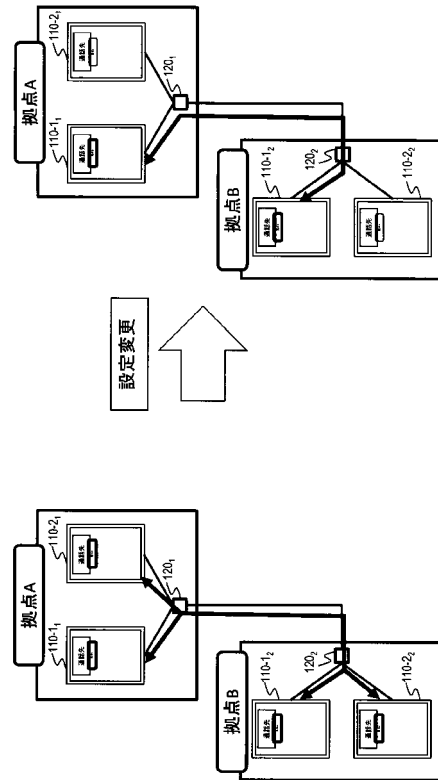


図19

【 図 20 】

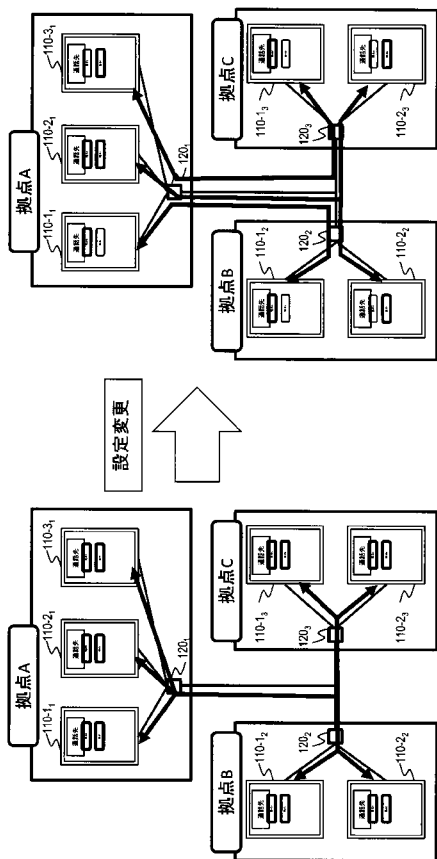


図20

【 図 21 】

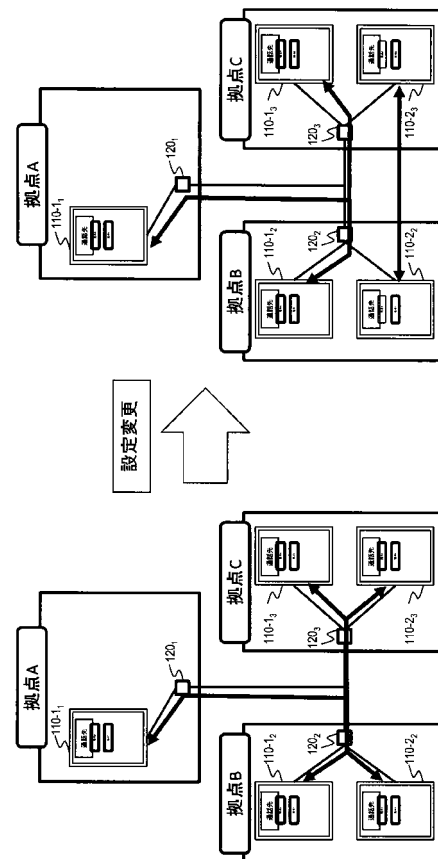


図21

フロントページの続き

(72)発明者 植松 尚

東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内

Fターム(参考) 5K127 AA36 GB47 GB72 JA14 KA16 MA02

5K201 BB09 CA01 CC04 DC06