(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2012-105199
(P2012-105199A)
(43)公開日　平成24年5月31日(2012.5.31)

(51)Int.Cl.			F I			テーマコード（参考）
H04R	3/00	(2006.01)	H04R	3/00	320	5C164
H04R	1/40	(2006.01)	H04R	1/40	320A	5D018
H04N	7/15	(2006.01)	H04N	7/15	630Z	5D020

審査請求 有　請求項の数 10　ＯＬ　（全 21 頁）

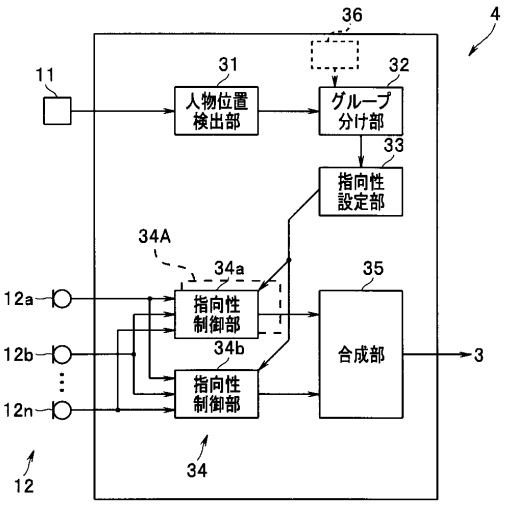
(21)出願番号　特願2010-253947 (P2010-253947)
(22)出願日　平成22年11月12日 (2010.11.12)

(71)出願人　000003078
株式会社東芝
東京都港区芝浦一丁目１番１号
(74)代理人　100076233
弁理士　伊藤　進
(72)発明者　天田　皇
東京都港区芝浦一丁目１番１号　株式会社東芝内

Ｆターム(参考)　5C164 FA10 UB92S VA04S VA06P VA07S VA52P
5D018 BB22
5D020 BB04

(54)【発明の名称】音響信号処理装置、テレビジョン装置及びプログラム

(57)【要約】
【課題】リアルタイム性を確保できる音響信号処理の可能な音響信号処理装置を提供すること。
【解決手段】ユニット４は、所定空間内に存在する人物の位置を検出する人物位置検出部３１と、人物位置検出部３１により検出された１又は２以上の人物を所定数以下の数のグループに割り振るグループ分けを行うグループ分け部３２と、設定されたそれぞれの指向性に基づいてマイクロホンアレー１２の指向性を制御する複数の指向性制御部３４と、グループ分け部３２でグループ分けがされた各グループの指向性を、対応する指向性制御部３４に設定する指向性設定部３３とを有する。
【選択図】図３



【特許請求の範囲】
【請求項１】
所定空間内に存在する人物の位置を検出する人物位置検出部と、
前記人物位置検出部により検出された１又は２以上の人物を所定数以下の数のグループに割り振るグループ分けを行うグループ分け部と、
設定されたそれぞれの指向性に基づいてマイクロホンアレーの指向性を制御する複数の指向性制御部と、
前記グループ分け部で前記グループ分けがされた各グループの指向性を、対応する指向性制御部に設定する指向性設定部と、
を有することを特徴とする音響信号処理装置。
【請求項２】
前記グループ分け部は、前記マイクロホンアレーの雑音抑圧性能の和が最大になるように、前記グループ分けを行うことを特徴とする請求項１に記載の音響信号処理装置。
【請求項３】
前記グループ分け部は、前記各グループに割り振られる人物の数が平均化するように、前記グループ分けを行うことを特徴とする請求項１に記載の音響信号処理装置。
【請求項４】
前記グループ分けの結果のグループ数あるいは前記所定数を、変更するためのグループ数変更部を有することを特徴とする請求項１から３のいずれか１つに記載の音響信号処理装置。
【請求項５】
発話者を検出する発話者検出部を有し、
前記グループ分け部は、前記１又は２以上の人物の中から、前記発話者だけを対象に、前記グループ分けを行うことを特徴とする請求項１から４のいずれか１つに記載の音響信号処理装置。
【請求項６】
前記人物位置検出部により検出された人物の増減を検出する人物増減検出部を有し、
前記グループ分け部は、前記人物増減検出部により検出された前記人物の増減に応じて、前記グループ分けを行うことを特徴とする請求項１から５のいずれか１つに記載の音響信号処理装置。
【請求項７】
前記指向性設定部によって指向性が設定された指向性制御部の出力監視あるいは前記人物位置検出部により検出された人物の発話監視を行う発話状態監視部を有し、
前記グループ分け部は、前記発話状態監視部により、前記指向性が設定された指向性制御部の出力あるいは前記検出された人物の発話が所定時間以上観測されなくなったときは、前記所定時間以上観測されなくなった指向性制御部あるいは人物を除いて、前記グループ分けを行うことを特徴とする請求項１から６のいずれか１つに記載の音響信号処理装置。
【請求項８】
前記人物位置検出部は、前記所定空間をカメラにより撮像して得られた画像により、あるいは前記所定空間からの音響信号により、前記人物の位置を検出することを特徴とする請求項１から７のいずれか１つに記載の音響信号処理装置。
【請求項９】
テレビジョン放送のコンテンツを表示する表示部と、
音響信号処理装置と、
通信回線を介して通信を行うための通信インターフェースと、
を有し、
前記音響信号処理装置は、
所定空間内に存在する人物の位置を検出する人物位置検出部と、
前記人物位置検出部により検出された１又は２以上の人物を所定数以下の数のグループ

に割り振るグループ分けを行うグループ分け部と、
設定されたそれぞれの指向性に基づいてマイクロホンアレーの指向性を制御する複数の指向性制御部と、
前記グループ分け部で前記グループ分けがされた各グループの指向性を、対応する指向性制御部に設定する指向性設定部と、
を有することを特徴とするテレビジョン装置。
【請求項１０】
音響信号を処理するプログラムであって、コンピュータに、
所定空間内に存在する人物の位置を検出する機能と、
検出された１又は２以上の人物を所定数以下の数のグループに割り振るグループ分けを行う機能と、
グループ分けがされた各グループの指向性を、マイクロホンアレーの指向性を制御する指向性制御部に設定する機能と、
を実現させるためのプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明の実施形態は、音響信号処理装置、テレビジョン装置及びプログラムに関する。
【背景技術】
【０００２】
従来より、マイクロホンアレーを用いて、目的とする方向からの音だけを強調する技術が知られている。
また、カメラを用いて人の位置を検出して、その検出した人の方向にマイクロホンアレーの指向性を向ける技術も提案されている。その提案に係る装置では、話者の位置を検出して、その検出された話者の位置関係に基づいて、話者毎に音声を抽出する方向と範囲が抽出され、音声抽出手段がその範囲内の音声を抽出する。
【先行技術文献】
【特許文献】
【０００３】
【特許文献１】特開２００５－２７４７０７号公報
【発明の概要】
【発明が解決しようとする課題】
【０００４】
しかし、そのような装置は、各話者に向かうようにマイクロホンアレーの指向性を制御して音声処理を行うようにするため、話者の数が増えると、抽出された各方向についての指向性制御部における音声信号の音声処理をするための演算量が増えるという問題がある。すなわち、話者数に応じて、指向性制御部の演算量が増える。
【０００５】
よって、このような装置は、話者の増えた分だけ計算量が増えるため、特に、リアルタイムで応答が要求されるシステムに適用した場合、そのシステムのリアルタイム性が確保できなくなってしまうという問題がある。例えば、遠隔地の者同士がテレビ会議等を行うような場合、音声信号がリアルタイムで音声処理できなくなるため、音声の途切れ、雑音の混入等が発生してしまう。
また、所謂計算コストの面から、余裕を持った数の指向性制御部を予め想定してハードウエアの性能を高めることは現実的でない。
【０００６】
そこで、本実施形態は、リアルタイム性を確保できる音響信号処理の可能な音響信号処理装置、テレビジョン装置及びプログラムを提供することを目的とする。
【課題を解決するための手段】
【０００７】

実施形態の音響信号処理装置は、所定空間内に存在する人物の位置を検出する人物位置検出部と、前記人物位置検出部により検出された１又は２以上の人物を所定数以下の数のグループに割り振るグループ分けを行うグループ分け部と、設定されたそれぞれの指向性に基づいてマイクロホンアレーの指向性を制御する複数の指向性制御部と、前記グループ分け部で前記グループ分けがされた各グループの指向性を、対応する指向性制御部に設定する指向性設定部と、を有する。

【図面の簡単な説明】

【０００８】

【図１】図１は、第１の実施の形態に係る音響信号処理装置が適用されるテレビ電話システムの例を説明するための図である。

【図２】第１の実施形態に係わる、各テレビ２と各ユニット４の構成を示すブロック図である。

【図３】第１の実施形態に係わる、ユニット４のソフトウエア構成を示すブロック図である。

【図４】第１の実施形態に係わる、人物位置検出部３１によって、マイクロホンアレー１２あるいはテレビ画面２ａに対して、所定の位置を基準として、３人の人物が検出されたときに、２グループにグループ分けされた一の状態を説明するための図である。

【図５】第１の実施形態に係わる、マイクロホンアレー１２の雑音抑圧性能の和が最大になるようにグループ分けを行う方法を説明するための図である。

【図６】第１の実施形態に係わる、ユニット４における指向性設定処理の流れの例を示すフローチャートである。

【図７】第１の実施形態に係わる、グループ数すなわちビーム数の変更を説明するための図である。

【図８】第２の実施の形態に係るユニット４のソフトウエア構成を示すブロック図である。

【図９】第２の実施の形態に係るユニット４の変形例に係るソフトウエア構成を示すブロック図である。

【図１０】第２の実施の形態に係るユニット４における指向性設定処理の流れの例を示すフローチャートである。

【図１１】第３の実施の形態に係るユニット４のソフトウエア構成を示すブロック図である。

【図１２】第３の実施の形態に係るユニット４における指向性設定処理の流れの例を示すフローチャートである。

【図１３】第４の実施の形態に係るユニット４のソフトウエア構成を示すブロック図である。

【図１４】図１３のユニット４の変形例に係るソフトウエア構成を示すブロック図である。

【図１５】第４の実施の形態に係るユニット４における指向性設定処理の流れの例を示すフローチャートである。

【図１６】第５の実施の形態に係るユニット４のソフトウエア構成を示すブロック図である。

【図１７】マイクロホンアレー１２の各マイクロホンからの音響信号のみから、人物の位置を検出するユニットのソフトウエア構成を示すブロック図である。

【発明を実施するための形態】

【０００９】

以下、図面を参照して実施形態を説明する。

【００１０】

（第１の実施形態）

（構成）

図１は、第１の実施の形態に係る音響信号処理装置が適用されるテレビ電話システムの

例を説明するための図である。

図１に示すように、テレビ電話システム１は、テレビジョン放送を受信するテレビジョン装置（以下、テレビという）を利用して、遠隔地の者同士が、テレビ画面に映し出された相手の画像を見ながら会話をすることができるシステムである。テレビ電話システム１は、互いに離れた場所に設置された２台のテレビ２A、２Bを含み、テレビ２A、２Bは、互いに通信回線としてのインターネット３を介して通信可能に接続されている。

【００１１】

そして、各テレビ２A、２B（以下、２台のテレビの両方あるいは一方を指すとき、テレビ２ともいう）は、テレビジョン放送を受信可能であると共に、インターネット接続下で、インターネット３を介して互いに画像信号及び音声信号の通信が可能である。そのために、各テレビ２A、２Bには、それぞれカメラとマイクロホンアレーを内蔵するユニット４A、４B（以下、２つのユニットの両方あるいは一方を指すとき、ユニット４ともいう）が取り付けられている。

各テレビ２のテレビ画面２ａには、テレビ２が放送受信モードのときは、テレビジョン放送のコンテンツが表示され、かつ後述するテレビ電話モードのときは、相手方の画像が表示される。

【００１２】

テレビ電話ユニットとしての各ユニット４は、対応するテレビ２の画面の前を撮像するカメラと、テレビ２の前の音声を取り込むマイクロホンアレーを有する（図２）。ユーザは、リモコン５A、５B（以下、２つのリモコンの両方あるいは一方を指すとき、リモコン５ともいう）を操作することによって、テレビジョン放送をテレビ画面上に表示させたり、ユニット４のテレビ電話機能を動作させたりすることができる。ユニット４は、音響信号処理装置を構成する。

【００１３】

図２は、各テレビ２と各ユニット４の構成を示すブロック図である。

ユニット４は、カメラ１１、マイクロホンアレー１２、中央処理装置（CPU）１３、ROM１４、RAM１５、及びインターフェース（以下。I/Fと略す）１６，１７，１８を含む。CPU１３、ROM１４及びRAM１５は、バス１９を介して互いに接続されている。テレビ２とユニット４は、インターネット３に接続するための通信ライン２０に接続されている。

【００１４】

カメラ１１は、CCDなどの撮像素子を有して、デジタルの画像信号を出力する。マイクロホンアレー１２は、複数の（ここでは、ｎ個の）マイクロホンを有し、各マイクロホンの音響信号を出力する。テレビ２に載置されたユニット４内において、カメラ１１は、テレビ２のテレビ画面を視るユーザ側を撮像するように配置され、マイクロホンアレー１２は、テレビ画面を視るユーザの音声を取り込むように配置されている。

【００１５】

I/F１６は、カメラ１１とバス１９を接続するインターフェースである。I/F１７は、マイクロホンアレー１２とバス１９を接続するインターフェースである。I/F１８は、インターネット３に接続された通信ライン２０とバス１９とを接続するためのインターフェースである。通信ライン２０には、テレビ２も接続されているので、テレビ２とユニット４は、互いに通信可能となっている。ユニット４と合わせてテレビ２は、テレビ電話可能なテレビジョン装置を構成する。

【００１６】

CPU１３は、ROM１４に格納された、後述する各種ソフトウエアプログラム（以下、単にプログラムという）を実行する処理部である。ROM１４は、後述する各種プログラムを格納する不揮発性の記憶部である。RAM１５は、CPU１３が各種プログラムを実行するときに作業領域として利用する記憶部である。

【００１７】

ROM１４には、テレビ電話システム１が機能するときに、ユニット４が音響信号処理装置としての機能を実行するための各種プログラムが含まれる。

なお、図１と図２に示すテレビ電話システム１では、テレビ２とユニット４が別体のものであり、ユニット４がテレビ２に載置されるようになっているが、ユニット４がテレビ２の本体内に内蔵されるような構成でもよい。

【００１８】

図３は、ユニット４のソフトウエア構成を示すブロック図である。ユニット４内の各ブロックは、プログラムにより構成される。ユニット４は、人物位置検出部３１、グループ分け部３２，指向性設定部３３，複数の（ここでは２つの）指向性制御部３４ａ、３４ｂ、及び合成部３５を有して構成される。

【００１９】

カメラ１１からのデジタルの画像信号が、人物位置検出部３１に入力される。

人物位置検出部３１は、入力された画像信号に基づいて、撮像された人物を判別し、各人物の位置を検出して、その位置情報をグループ分け部３２に出力する。カメラ１１は、テレビ２の前の所定空間を撮像するので、人物位置検出部３１は、その所定空間内に存在する人物の位置を検出する処理部を構成する。人物位置検出部３１は、画像の中から人の顔を認識する顔画像認識処理により、複数の人物のそれぞれの位置を検出する。検出された各顔の位置と、所定の基準位置との位置関係から、カメラ１１が撮像した所定空間内における各人物の位置が算出される。すなわち、人物位置検出部３１は、複数の顔が検出されれば、各顔の位置に対応する各人物の位置を算出する。

【００２０】

グループ分け部３２は、入力された位置情報に基づいて、検出された人物のグループ分けを行い、複数の人物が検出されれば、所定のグループ数のグループに纏められる。判別された１又は２以上の人物は、予め設定された上限のグループ数までグループ数にグループ分けされる。よって、グループ分け部３２は、人物位置検出部３１により検出された１又は２以上の人物を所定数以下の数のグループに割り振るグループ分けを行う処理部を構成する。

【００２１】

例えば、所定数としての上限が２の場合、検出された人物が一人であれば、グループ数は１であり、検出された人物が二人であれば、グループ数は２となる。さらに、上限が２の場合、判別された人物が三人以上であっても、グループ分けされたグループ数は２となる。

【００２２】

グループ数の上限は、CPU１３の処理能力によって決定される。CPU１３の処理能力に応じて、プログラムの処理時間は異なる。CPU１３の処理能力が高ければ、CPU１３における指向性制御部３４における１つの指向性制御部の処理時間が短くなり、CPU１３の処理能力が低ければ、１つの指向性制御部の処理時間が長くなる。

【００２３】

特に、指向性制御部３４において、指定された方向にマイクロホンアレー１２のビームを形成して音声処理するための計算量が、テレビ電話システム１に要求されるリアルタイム性が維持できない計算量にならないように、上限が決定される。

【００２４】

例えば、CPU１３の音声処理のための処理能力が、100MIPS（百万命令毎秒）であるときに、１つの指向性制御部の処理時間が50MIPSであれば、２つの指向性制御部の処理がCPU１３の処理能力の限界である。その場合、グループ数の上限は２となり、ユニット４は、指向性制御部を２つ有することができる。検出された人物についてのグループ分けの方法については、後述する。

【００２５】

グループ分け部３２は、グループ分けされた人物の位置情報に基づいて、グループ毎の指向性の情報（すなわちビームの情報）を算出して決定し、指向性設定部３３に出力する。ビームの情報は、指向性の設定方向dsと設定範囲θsの情報を含む。設定方向dsは、設定範囲θsの中心方向である。例えば、設定範囲θsは、設定方向dsを中心とする角度幅で

ある。

【００２６】

なお、グループ分け部３２は、所定のルールに従って検出された人物の位置情報に基づいてグループ分けを行うが、雑音抑圧能力すなわち性能が最も高くなるように、すなわち最適なグループ分けを行うようにしてもよい。最適なグループ分けの方法については後述する。

グループ分け部３２は、グループ分けされたグループ毎のビームの情報を、指向性設定部３３に供給する。

【００２７】

指向性設定部３３は、グループ毎のビームの情報に基づいて、各グループの話者の音声を強調するように、各ビームを形成するための設定情報をグループ毎に生成して、対応する指向性制御部３４ａ、３４ｂに各設定情報を供給する。例えば、検出された人物が一人の場合、指向性制御部３４ａのみに設定情報が供給されて設定され、指向性制御部３４ａのみがその設定情報に基づくビームを形成する。検出された人物が二人以上の場合、指向性制御部３４ａ、３４ｂのそれぞれに設定情報が供給されて設定され、指向性制御部３４ａ、３４ｂは、それぞれ設定情報に基づくビームを形成する。

すなわち、グループ分け部３２は、人物のグルーピングを行い、指向性設定部３３は、指向性をグループ単位で形成するように、各指向性制御部３４を制御する。よって、指向性設定部３３は、グループ分け部３２でグループ分けがされた各グループの指向性を、対応する指向性制御部３４に設定する処理部を構成する。

【００２８】

複数の指向性制御部３４ａ、３４ｂは、設定されたそれぞれの指向性に基づいてマイクロホンアレーの指向性を制御する処理部を構成する。すなわち、各指向性制御部３４は、設定された方向から到来する音声を強調する。指向性制御部３４ａ、３４ｂにおいて音声強調処理されたデジタルの音声信号は、合成部３５で加算されて合成信号として、インターネット３へ送信される。指向性制御部３４は、それぞれに設定された指向性を実現すべくｎ個の音声入力信号に対してアレー処理を行う。アレー処理の例は、例えば、特許第３７９５６１０号公報、特開２００７－１０８９７号公報に開示されている。

なお、指向性制御部３４ａ、３４ｂにおいて音声強調処理されて得られたデジタルの複数の音声信号は、合成部３５で合成しないで、それぞれ個別に、インターネット３へ送信するようにしてもよい。

【００２９】

上述したように、グループ分けにより生成された各グループの人物の音声を強調するのに適した指向性を形成するための設定が決定される。指向性制御部３４がその設定に基づいて複数のマイクロホン１２ａ～１２ｎからの音声信号に対して所定のフィルタ演算を行い、その演算結果を加算する処理を行うことによって、マイクロホンアレー１２の指向性すなわちビームは形成される。ここでは、上限数である２つの指向性制御部３４ａ、３４ｂは、それぞれ、互いに異なる特性のフィルターセットを動作させるようにしてビームを形成する。

【００３０】

なお、本実施の形態及びこれに続く他の実施の形態においても、話者の位置は、水平方向における角度で特定する場合で説明するが、これに限られず、複数のマイクロホンを２次元配置等することによって、奥行き方向も含めた複数の話者のグループ化を行って、話者の存在する空間の奥行き方向の制御を行うようにしてもよい。

【００３１】

（グループ分けの方法）

ここで、グループ分け部３２におけるグループ分けの方法について説明する。図４は、人物位置検出部３１によって、マイクロホンアレー１２あるいはテレビ画面２ａに対して、所定の位置を基準として、３人の人物が検出されたときに、２グループにグループ分けされた一の状態を説明するための図である。

【００３２】
図４は、３人の人物P1,P2,P3が、それぞれ、マイクロホンアレー１２の所定の中心位置P0から、方向d1,d2,d3に存在することが検出された場合であって、人物P1とP2が第１のグループを構成し、人物P3が第２のグループを構成する例を示す。
【００３３】
例えば、検出された各人物の顔の中心の方向が、人物の方向として決定される。図４の場合、人物P1は、方向d1に存在する。同様に、人物P2とP3は、それぞれ方向d2とd3に存在する。
【００３４】
（ルール）
検出された人物に対するグループ化は、所定のルールで行われるが、所定のルールは、種々のルールが適用可能である。まず、検出された人物の数がグループ数の上限に達するまでは、検出された人毎にグループ分けが行われ、その人物の数が上限を超えると、所定のルールによるグループ化が行われる。所定のルールの中で簡単なルールは、例えば、マイクロホンアレー１２の所定の基準方向（０度）から所定の方向（１８０度の方向）に向かって、検出された人物を、各グループが所定の人数になるようにグループ化する、というようなルールである。
【００３５】
図４の場合、検出された三人の人物P1,P2,P3が、０度方向から１８０度方向に向かって、二人のグループと一人のグループにグループ分けされている。人物P1とP2が第１のグループとなり、人物P3がもう一つの第２のグループとなっている。第１のグループでは、方向d1とd2の真ん中の方向D1が、目的音源方向としての設定方向dsとなる。第２のグループでは、方向d3の方向D2が、目的音源方向としての設定方向dsとなる。さらに、第１のグループでは、目的音源方向D1の前後（θ１）／２の範囲θ１が、設定範囲θsとなる。第２のグループでは、目的音源方向D2の前後θp／２の範囲θpが、設定範囲θsとなる。
【００３６】
ここでは、図４の第２のグループのように、一つのグループに人物が一人だけいる場合、設定範囲θsは、所定の範囲θpを有するものとする。図４の第１のグループのように、一つのグループに人物が二人以上存在する場合、それぞれの所定の前後（θp）／２を加味した人物P1,P2間の角度θｄを含むように、設定範囲θsは、角度θ１となる。
【００３７】
他にも、グループ分け部３２は、各グループに割り振られる人物の数が平均化するというルールに基づいて、グループ分けを行うようにしてもよい。
以上のように、グループ分けは、所定のルールに従って行われる。その所定のルールは、グループ数の上限に達するまでは、検出された人毎にグループ分けが行われ、検出された人物の数がグループ数の上限を超えると、所定のルールで、グループ数が上限を超えないようにグループ分けするルールである。
【００３８】
（最適なグループ分け）
グループ分け部３２は、マイクロホンアレー１２の雑音抑圧性能の和が最大になるように最適なグループ分けを行うようにしてもよい。
図５は、マイクロホンアレー１２の雑音抑圧性能の和が最大になるようにグループ分けを行う方法を説明するための図である。人物の配置は、図４と同じであるが、３人の人物P1,P2,P3は、図４とは異なるグループにグループ分けされている。
【００３９】
図５の場合、人物P1が第１のグループであり、人物P2,P3が第２のグループとなっている。第１のグループ（人物P1のグループ）は、図４の第２のグループ（人物P3のグループ）と同じ設定範囲θsを有する。第２のグループ（人物P2,P3のグループ）は、図４の第１のグループ（人物P1,P2のグループ）よりも、設定範囲θsが狭くなっている。
【００４０】

図５において、第２のグループの設定範囲θsは、角度θ３であり、図４の第１のグループの設定範囲θsであるθ１よりも狭い。

設定範囲θsが狭い方が、目的とする方向からの音だけを強調する性能が高い。図４の場合、人物P2とP3の間には隙間がなく、仮にこの隙間から雑音が到来しても、抑圧することはできない。これに対して、図５の場合は、人物P1とP2の間には隙間があり、この隙間から雑音が到来しても抑圧することができる。

【００４１】

よって、図５に示すグループ分けは、図４のグループ分けよりも、よりよいグループ化ということが言える。これは、互いに異なる設定範囲θsを有する２つのグループ分けがあったときに、評価値EVとしての（１／θs）の和が大きいグループ分けの方が、他のグループ分けよりも、システム全体としての強調性能が高いので、図５の方が図４よりも、よりよいグループ分けということができるからである。

【００４２】

従って、最適なグループ分けの方法としては、検出された人物の数が、グループ数の上限を超えた場合、検出された全ての人物について取り得るグループ分けの組み合わせを仮定し、取り得る組み合わせの中で、各グループの評価値EV（＝１／θs）の総和EVｓが最も大きい組み合わせのグループ分けが、最適なグループ分けであるというルールを用いることができる。すなわち、グループ数の上限の範囲内で、全ての組み合わせについて所定の評価値の総和EVsを算出し、その総和EVｓが最も大きな値のグループ分けを選択するというルールを、評価基準として、グループ分け部３２が採用することができる。

【００４３】

例えば、図４において、第１のグループの設定範囲θｓ（＝θ１）が６５度、第２のグループの設定範囲θｓ（＝θp）が２０度とすると、上記の評価値の総和EVs1は、（（１／θs）+（１／θp）＝（（１／６５）＋（１／２０））である。これに対して、図５において、第１のグループの設定範囲θｓ（＝θp）が２０度、第２のグループの設定範囲θ３が３８度とすると、上記の評価値の総和EVs2は、（（１／θp）+（１／θ３）＝（（１／２０）＋（１／３８））である。EVs2＞EVs1となるので、図５の組み合わせの方が、図４の組み合わせよりも、評価値EVが高い。

【００４４】

よって、グループ分け部３２は、このような評価値の総和EVsの比較を、全てのグループ分け可能な組み合わせ同士で行い、最も評価値の総和EVsが高い組み合わせのグループ分けを、最適なグループ分けとして決定する。

【００４５】

（処理）

図６は、ユニット４における指向性設定処理の流れの例を示すフローチャートである。図６の処理は、テレビ電話システム１におけるテレビ電話機能がオンになると、CPU１３によって実行される。ユーザは、リモコン５を操作して、テレビ電話機能をオンにすることができる。

【００４６】

まず、CPU１３は、人物位置検出部３１により人物位置の検出を行い（S1）、続いて、グループ分け部３２による、上述したようなグループ分けを行う（S2）。CPU１３は、指向性設定部３３による、グループ分けの結果に基づいて各指向性制御部３４に対する指向性の設定を行う（S3）。

【００４７】

指向性制御部３４は、設定された指向性の情報に基づいて、ビームを制御して音声処理を行う。指向性制御部３４で処理された音声信号は、合成部３５で合成されて、通信回線であるインターネット３を介して、相手方のテレビ電話システムへ伝送される。

【００４８】

図６の処理が実行された後、において、人物位置検出部３１により検出された人物の位置に変化がなければ、図６の処理は実行されないが、人物の位置に変化があると、図６の

グループ分け及び指向性設定の処理が実行される。

【００４９】

その結果、遠隔に離れた者同士が、テレビ２を利用して、リアルタイムで音声が途切れたり等なく、テレビ電話を行うことができる。

【００５０】

なお、上述した例では、上限数まではグループ数が増えていくが、ユーザによって、予め設定された上限を変更できるようにしてもよい。

図７は、グループ数すなわちビーム数の変更を説明するための図である。ユーザは、自己が見ているテレビ２のリモコン５に対して所定の操作を行うと、図７に示すようなビーム数上限の設定変更画面２１を表示させることができる。設定変更画面２１は、テレビ２の画面２ａ上に、サブウインドウとして表示される。設定変更画面２１を表示させるプログラムは、ROM１４中に予め格納されている。

【００５１】

設定変更画面２１は、変更可能なビーム数を表示する表示部２１Aを含み、ユーザは、リモコン５を操作して、カーソル（図７では斜線で示されている）を所望のビーム数の表示部２１Aに移動させて選択することができる。例えば、図７では、上限数の「２」よりも少ない「１」の表示部が選択されている。選択の確定は、ユーザが、リモコン５の確定ボタンなどを操作することによって行うことができる。

なお、上限が「３」であれば、上限より少ない「２」と「１」が選択可能となるように、設定画面変更画面２１には、上限の数と、上限よりも少ない数の表示部２１Aが、選択可能に表示される。

【００５２】

例えば、ユーザが上限「２」よりも少ないビーム数である「１」を選択すると、ビーム形成に必要な指向性制御部３４の数が少なくなるので（１つの指向性制御部でよいことになるので）、上述したCPU１３の演算量が少なくなる。その場合、CPU１３の処理時間に余裕ができるので、より処理時間のかかる別の指向性制御部３４A（図３では点線で示す）を利用して、例えば、より高音質な音声信号が得られる指向性制御を実行させるようにすることも可能である。

【００５３】

例えば、一人しかいないような状況では、ユーザは、設定変更画面２１を用いて、ビーズ数の上限を１に設定して、より高性能な指向性制御部３４Aによるビーム形成を行わせることができる。その結果、相手方には、より高音質で音声信号を送信することができる。

【００５４】

ROM１４に格納された図７の画面に関する処理を実行するプログラムは、上述したビーム数言い換えるグループ数を変更するためのグループ数変更部３６（図３で点線で示す）を構成する。

なお、上述した例では、グループ数変更部３６は、グループ数を上限数よりも少ない数に変更できる例であるが、予め設定された上限を、より少ない数に変更するようにしてもよい。よって、グループ数変更部３６は、グループ分けの結果のグループ数あるいは所定の上限を変更するための処理部を構成する。

【００５５】

以上のように、上述した本実施の形態によれば、CPUの処理能力などに応じたグループ数の上限内でグループ化が行われるので、リアルタイム性を確保できる音響信号処理が可能なテレビ電話システムを提供することができる。

【００５６】

（第２の実施の形態）

第１の実施の形態では、顔検出された人物の位置を検出して、グループ分けを行っていたが、第２の実施の形態は、検出された人物の中で発話した人物のみを、グループ分けするようにした点が、第１の実施の形態と異なる。

以下、第２の実施の形態を説明するが、第１の実施の形態と同様の構成については、同じ符号を付し説明は省略し、異なる点を主として説明する。

【００５７】

図８は、第２の実施の形態に係るユニット４のソフトウエア構成を示すブロック図である。図８は、発話者検出部４１を含む点で、図３と異なる。発話者検出部４１は、人物位置検出部３１により検出された人物の中から、発話者を検出する処理部である。

【００５８】

よって、本実施の形態は、発話者検出部４１によって実際に発話したかどうかを検出し、発話が検出された人物の位置情報のみが、後段のグループ分け部３２に出力される。グループ分け部３２は、人物位置と検出された発話者との対応関係の情報を有している。

【００５９】

なお、発話者検出部４１にマイクロホンアレー１２からの音声信号を入力させ、口の動きと共に、音声信号が入力された場合に、発話者であるという判定をするようにしてもよい。図９は、本実施の形態に係るユニット４の変形例に係るソフトウエア構成を示すブロック図である。図９では、口の動きと音声信号による判定のために、マイクロホンアレー１２の各マイクロホンからの音声信号が、発話者検出部４１に入力されている。

【００６０】

図１０は、本実施の形態に係るユニット４における指向性設定処理の流れの例を示すフローチャートである。図１０の処理は、人物位置の検出の後に、発話者検出部４１による発話者の検出処理を行い（S11）、発話者が検出されたか否かを判定し（S12）、発話者が検出された場合に（S12:YES）、グループ分けが行われるようになっている。そして、グループ分け部３２は、人物位置検出部３１により検出された１又は２以上の人物の中から、発話者だけを対象に、グループ分けを行う。

【００６１】

よって、存在はしているが、発話しない人物にマイクロホンアレー１２のビームを向けておくような無駄をなくすことができ、さらに人物の誤検出による人でない対象に指向性が向けられるということがないようにすることができる。また、発話者が検出されると、グルーピングを再度行うので、常に最適なグループ分けの状態で、テレビ電話システム１は動作することができる。

【００６２】

また、発話者検出部４１を追加することにより、人物位置検出部３１の検出感度を、人物を検出し易いように高く設定できるので、人物位置検出部３１において、誤検出はあっても、検出漏れのないようにして、検出漏れによる指向性の設定がされないという問題を回避することもできる。

【００６３】

（第３の実施の形態）

第２の実施の形態では、顔検出された人物の中から発話した人物のみを、グループ分けするようにしているが、第３の実施の形態は、人物の増減を監視し、増減に応じてグループ分けを行うようにした点が、第２の実施の形態と異なる。

【００６４】

以下、第３の実施の形態を説明するが、第１及び第２の実施の形態と同様の構成については、同じ符号を付し説明は省略し、異なる点を主として説明する。

【００６５】

図１１は、第３の実施の形態に係るユニット４のソフトウエア構成を示すブロック図である。図１１は、人物増減検出部４２を含む点で、図３と異なる。人物増減検出部４２は、人物位置検出部３１により検出された人物の増減を検出する処理部である。

【００６６】

図１２は、本実施の形態に係るユニット４における指向性設定処理の流れの例を示すフローチャートである。図１２の処理は、人物位置の検出の後に、人物増減検出部４２による人物の増減の検出処理を行い（S21）、検出された人物の数に増減があったか否かを判

定し（S22）、その増減があった場合に（S22:YES）、グループ分けが行われるようになっている。そして、グループ分け部３２は、人物増減検出部４２により検出された人物の増減に応じて、グループ分けを行う。

【００６７】

よって、存在はしているが、途中から新たな人物がテレビ電話に参加したり、逆にそれまで居た人物が途中で退席してテレビ電話に参加しなくなったような場合にも、適切な指向性の制御が可能となる。また、人物の増減が検出されると、グルーピングを再度行うので、常に最適なグループ分けの状態で、テレビ電話システムは動作することができる。

【００６８】

（第４の実施の形態）

第３の実施の形態では、人物の増減に応じてグループ分けするようにしているが、第４の実施の形態は、人物に増減はないが、途中から発話がなくなった人物がいる場合には、そのような人物の位置情報を一旦削除して、再度グループ分けを行うようにした点が、第１，第２及び第３の実施の形態と異なる。

【００６９】

以下、第４の実施の形態を説明するが、第１、第２及び第３の実施の形態と同様の構成については、同じ符号を付し説明は省略し、異なる点を主として説明する。

【００７０】

図１３は、第４の実施の形態に係るユニット４のソフトウエア構成を示すブロック図である。ユニット４内の各ブロックは、プログラムにより構成される。図１３は、不活性ビーム検出部４３を含む点で、図３と異なる。不活性ビーム検出部４３は各指向性制御部３４の出力に基づいて、不活性ビームの検出を行う。

【００７１】

不活性ビーム検出部４３は、そのビーム方向からの発話が所定の時間以上に亘って観測されなくなったか否かを検出する。不活性ビーム検出部４３は、指向性設定部３３によって指向性が設定された指向性制御部３４の出力監視を行い、発話状態を監視する発話状態監視部を構成する。不活性ビーム検出部４３は、発話が観測されなくなったビームを検出すると、そのビームに対応する位置の人物の位置情報を削除した人物位置情報をグループ分け部３２に出力する。グループ分け部３２は、その人物位置情報に基づいて、再度グループ分けする。

【００７２】

このような構成によれば、例えば途中で寝てしまった人物がいたような場合に、発話が所定の時間観測されなくなると、無駄なビームを削除することができる。

【００７３】

なお、上述した図１３の構成では、１つのグループの中に複数人いる場合は、全員が発話を止めた場合にのみ、再度グループ分けされるが、不活性ビーム検出部５１の構成を変更することによって、グループの中の一人でも発話を止めた場合に、再グループ分けを行うようにしてもよい。

【００７４】

図１４は、図１３のユニット４の変形例に係るソフトウエア構成を示すブロック図である。図１４では、マイクロホンアレー１２の各マイクロホンからの音声信号が、不活性ビーム検出部４３ａに入力されている。不活性ビーム検出部４３ａは、検出された人物の数ｋだけ、指向性制御部５１-a～５１-k（以下、複数の指向性制御部５１-a等を指すとき、あるいは一つの指向性制御部を指すとき、指向性制御部５１という）を有している。

【００７５】

検出された人物毎に指向性制御部５１が生成される。指向性制御部５１は、人物位置検出部３１からの人物位置の人物の発話が所定時間観測されなかったか否かを検出する。従って、不活性ビーム検出部４３ａは、人物位置検出部３１により検出された人物の発話監視を行う発話状態監視部を構成する。

【００７６】

なお、指向性制御部５１のプログラムは、テレビ電話によるリアルタイムな通話目的のものではないので、より高性能で処理時間のかかるプログラムであってもよい。

【００７７】

不活性ビーム検出部４３ａは、人物位置検出部３１で検出された人物の発話が検出されなくなると、その人物の位置情報を削除した人物位置情報をグループ分け部３２に出力する。グループ分け部３２は、その人物位置情報に基づいて、再度グループ分けする。

【００７８】

図１５は、本実施の形態に係るユニット４における指向性設定処理の流れの例を示すフローチャートである。図１５の処理は、人物位置の検出の後に、不活性ビーム検出部４３による不活性ビームの検出処理（あるいは図１４の不活性ビーム検出部４３ａによる発話を止めた人物の検出処理）を行い（S31）、不活性ビーム（あるいは発話を止めた人物）が検出されたか否かを判定し（S32）、その不活性ビーム（あるいは図１４の発話を止めた人物）があった場合に（S32:YES）、グループ分けが再度行われるようになっている。

【００７９】

よって、グループ分け部３２は、発話状態監視部である不活性ビーム検出部４３，４３ａにより、指向性が設定された指向性制御部３４の出力あるいは検出された人物の発話が所定時間以上観測されなくなったときは、所定時間以上観測されなくなった指向性制御部あるいは人物を除いて、グループ分けを行う。

【００８０】

なお、上述した不活性ビームの検出などにより、グループ分けの対象から外された人物も、上述した第２の実施の形態の処理を利用することによって、グループ分けの対象人物として、再度加入することができる。

【００８１】

さらになお、発話をしなくなったグループあるいは人物を画像データからのみに基づいて検出する場合は、不活性ビーム検出部４３，４３ａへの、図１３及び図１４に示すような、指向性制御部３４及びマイクロホンアレー１２からの入力は不要である。

【００８２】

よって、本実施の形態によれば、発話がなくなったビームあるいは人物を検出して、無駄なビームを削除することができる。また、発話の有無が検出されると、グルーピングを再度行うので、常に最適なグループ分けの状態で、テレビ電話システムは動作することができる。

【００８３】

（第５の実施の形態）

第５の実施の形態に係るユニット４は、上述した第２から第４の実施の形態において説明した発話者検出部４１，４１ａ、人物増減検出部４２，及び不活性ビーム検出部４３を含む検出統合部４４を有する。

【００８４】

図１６は、本実施の形態に係るユニット４のソフトウエア構成を示すブロック図である。図１６に示すように、ユニット４は、発話者検出部４１，４１ａ、人物増減検出部４２，及び不活性ビーム検出部４３を含む検出統合部４４を有する。なお、これらの３つの検出部４１，４２，４３の全てを含まなくてもよい。なお、図１６は、ｍ個の指向性制御部３４を有する例である。

【００８５】

その結果、ユニット４は、第２から第４の実施の形態で説明した利点を含むユニットとなるので、テレビ電話システム１は、ユーザにとって、より使いやすいものとなる。また、発話者の検出、人物の増減、あるいは発話の有無が検出されると、グルーピングを再度行うので、常に最適なグループ分けの状態で、テレビ電話システムは動作することができる。

【００８６】

以上のように、上述した第１から第５の実施の形態によれば、リアルタイム性を確保で

きる音響信号処理が可能な音響信号処理装置、テレビジョン装置、テレビ電話システムを提供することができる。特に、指向性がグループ単位で形成されるため、人物が多い場合、計算リソースが限られている場合等においても、少ない計算量で話者全員をカバーし、かつ雑音抑圧性能の高い音響信号処理装置、及びその応用装置を実現することができる。

【００８７】

なお、上述した各実施の形態では、カメラ１１を用いて人物の位置を検出しているが、カメラを用いないで、人物を検出するようにしてもよい。

【００８８】

図１７は、マイクロホンアレー１２の各マイクロホンからの音響信号のみから、人物の位置を検出するユニットのソフトウエア構成を示すブロック図である。

【００８９】

人物位置検出部３１Aは、複数のマイクロホンの音響信号から音の到来方向を推定する処理部である。推定方法としては、MUSIC法、ESPRIT法などを用いたDOA（Direction Of Arraival）推定が利用可能である。

DOA推定については、例えば、菊間信良著、「アレーアンテナによる適応信号処理」（科学技術出版、２００４年）の第１０章などを参照されたし。

【００９０】

図１７の構成によれば、カメラが不要となるので、コストの低減を図ることができる。さらに、人物以外の音が少ない環境では、雑音方向に誤って人物を検出してしまう等の誤動作も発生しにくい。

【００９１】

なお、以上説明した動作を実行するプログラムは、コンピュータプログラム製品として、フレキシブルディスク、CD-ROM等の可搬媒体や、ハードディスク等の記憶媒体に、その全体あるいは一部を記録し、あるいは記憶するようにしてもよい。そのプログラムは、コンピュータにより読み取られて、動作の全部あるいは一部が実行される。あるいは、そのプログラムの全体あるいは一部を通信ネットワークを介して流通または提供することができる。利用者は、通信ネットワークを介してそのプログラムをダウンロードしてコンピュータにインストールしたり、あるいは記録媒体からコンピュータにインストールすることで、容易に上述した実施形態の音響信号処理装置、テレビジョン装置、テレビ電話システムを実現することができる。

【００９２】

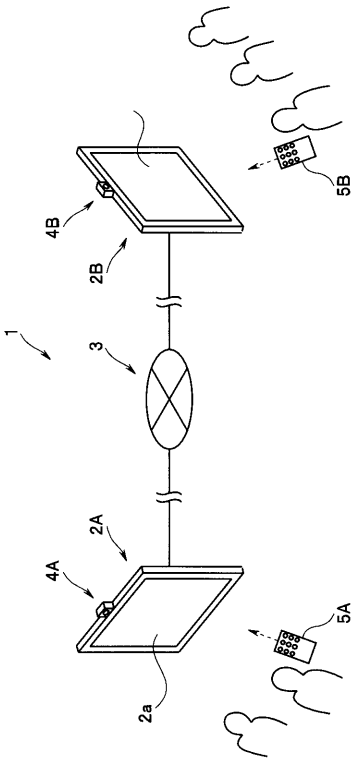
本発明は、上述した実施の形態に限定されるものではなく、本発明の要旨を変えない範囲において、種々の変更、改変等が可能である。

【符号の説明】

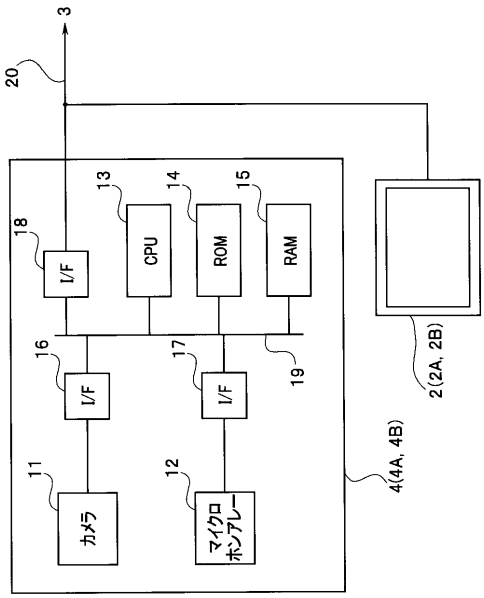
【００９３】

１　テレビ電話システム、２、２A、２B　テレビ、２ａ　テレビ画面、３　インターネット、４、４A、４B　ユニット、５、５A、５B　リモコン、１１　カメラ、１２　マイクロホンアレー、１３　CPU、１４　ROM、１５　RAM、１６，１７，１８　I/F、１９　バス、２０　通信ライン、３１　人物位置検出部、３２　グループ分け部、３３　指向性設定部、３４　指向性制御部、３５　合成部、３６　グループ数変更部、４１、４１ａ　発話者検出部、４２　人物増減検出部、４３，４３ａ　不活性ビーム検出部

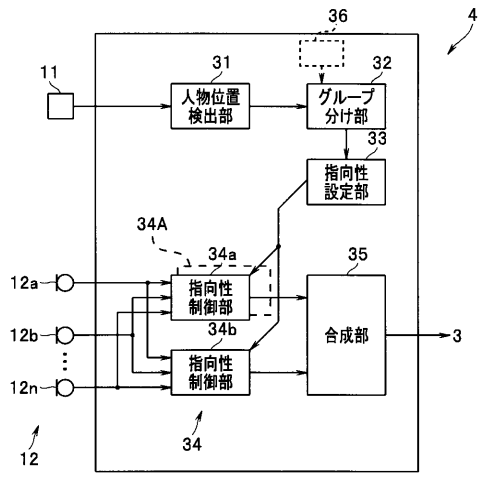
【図１】


1
3
4B
2B
5B
2A
4A
5A
2a

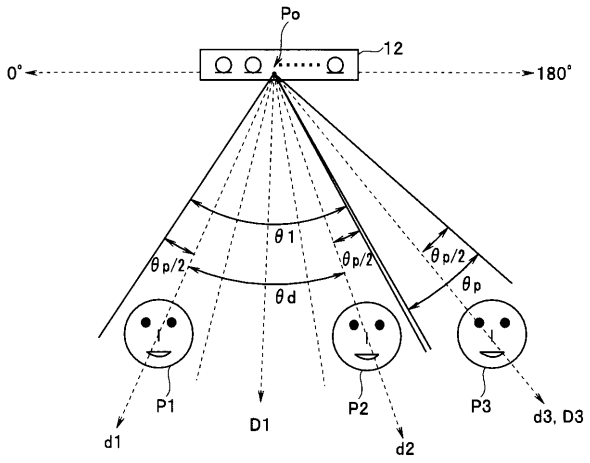
【図２】


3
20
18
I/F
13
CPU
14
ROM
15
RAM
16
I/F
17
I/F
19
11
カメラ
12
マイクロホンアレー
4(4A, 4B)
2(2A, 2B)

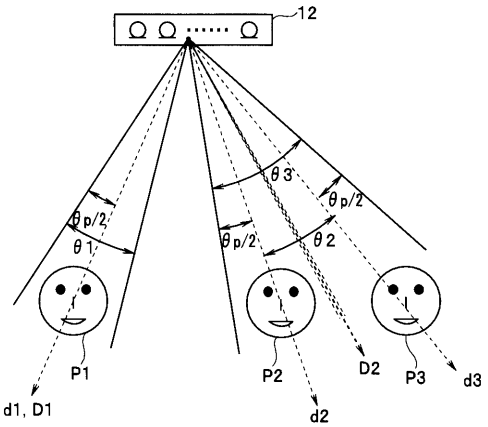
【図３】


4
36
11
31
人物位置検出部
32
グループ分け部
33
指向性設定部
34A
34a
35
12a
12b
12n
指向性制御部
34b
指向性制御部
合成部
3
12
34

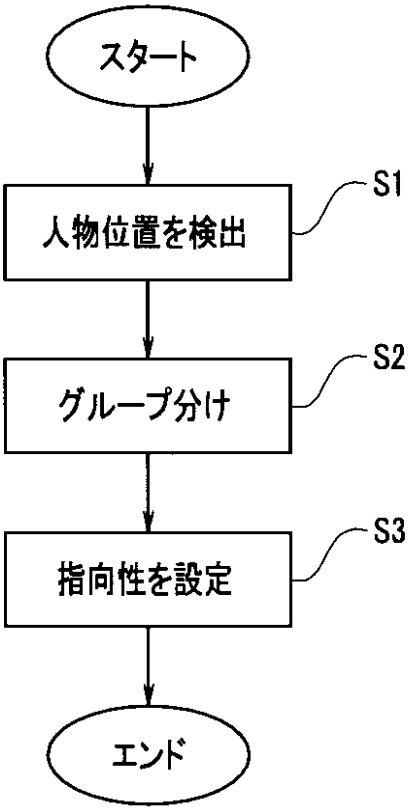
【図４】


Po
12
0°
180°
θp/2
θ1
θp/2
θp/2
θp
θd
P1
P2
P3
D1
d1
d2
d3, D3

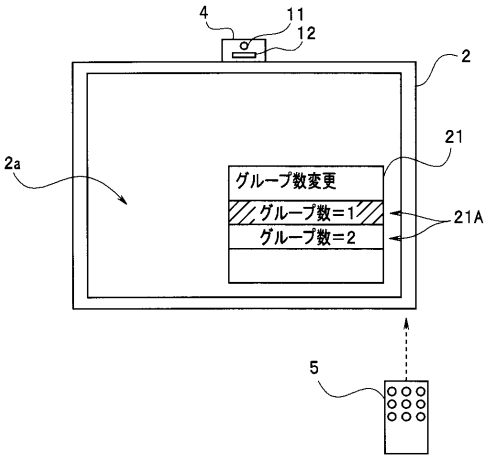
【図５】


12
θp/2
θ1
θp/2
θ3
θp/2
θp/2
θ2
P1
P2
D2
P3
d3
d1, D1
d2

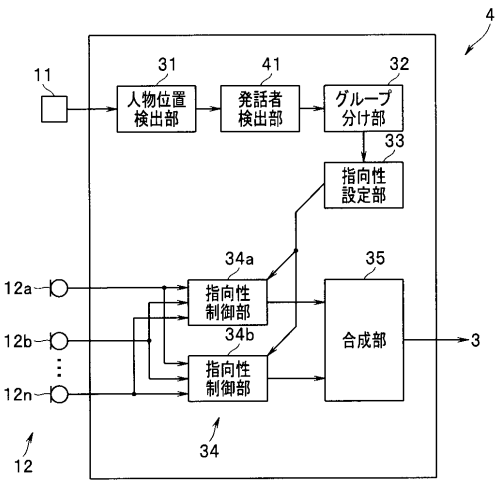
【図６】


スタート
人物位置を検出
S1
グループ分け
S2
指向性を設定
S3
エンド

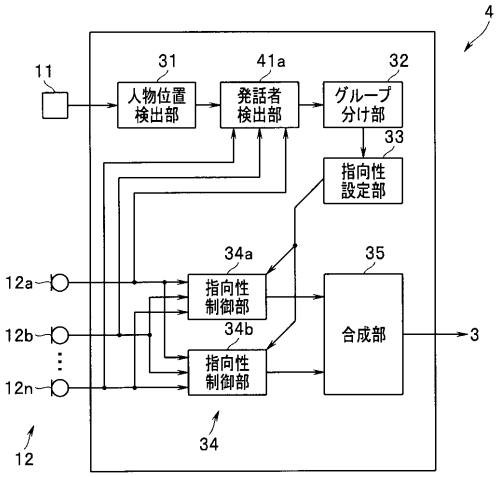
【図７】


4
11
12
2
21
2a
グループ数変更
グループ数＝1
グループ数＝2
21A
5

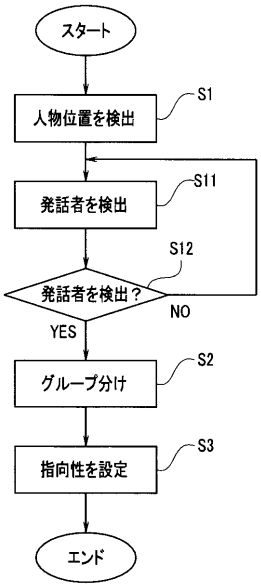
【図８】


4
11
31
人物位置検出部
41
発話者検出部
32
グループ分け部
33
指向性設定部
34a
指向性制御部
34b
指向性制御部
35
合成部
3
12a
12b
12n
12
34

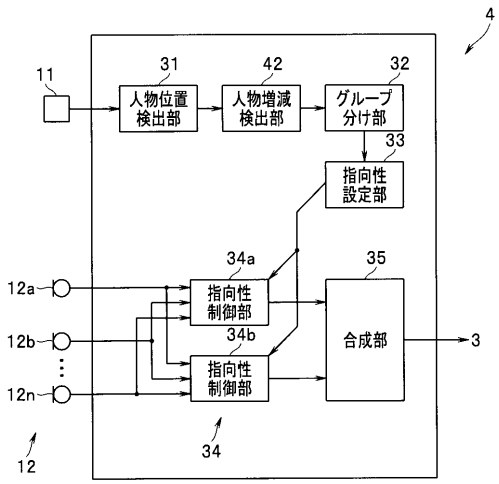
【図９】



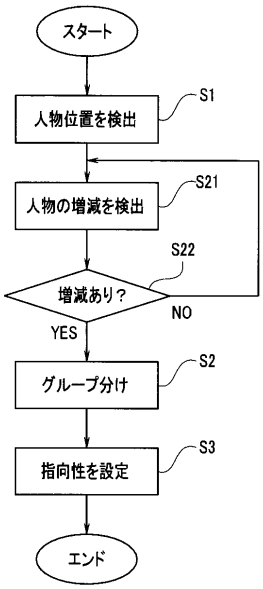
【図１０】



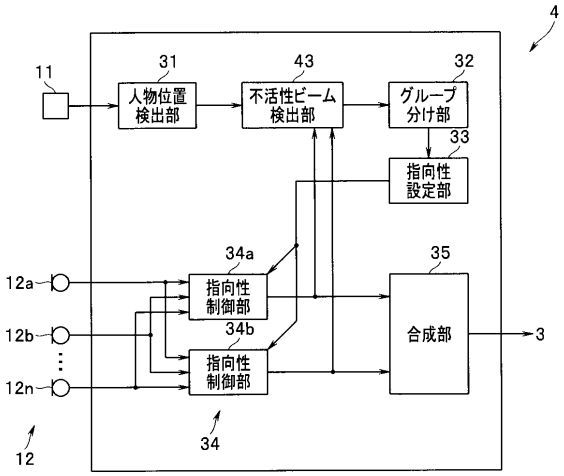
【図１１】



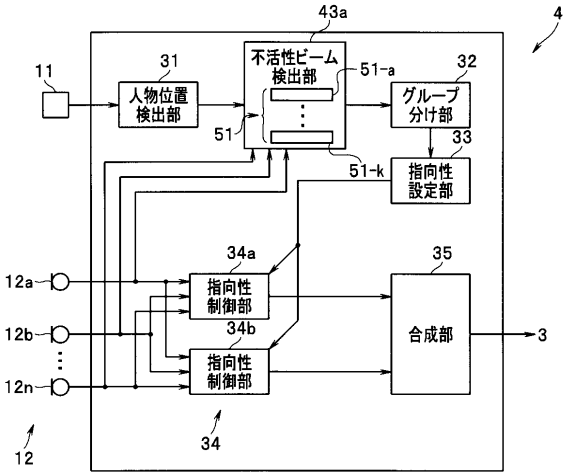
【図１２】



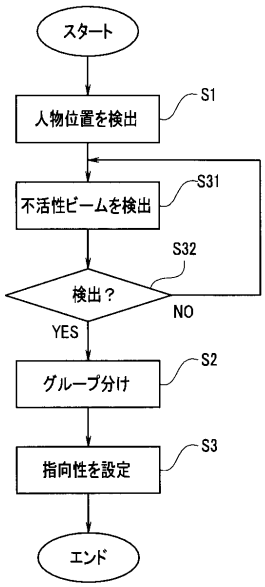
【図１３】



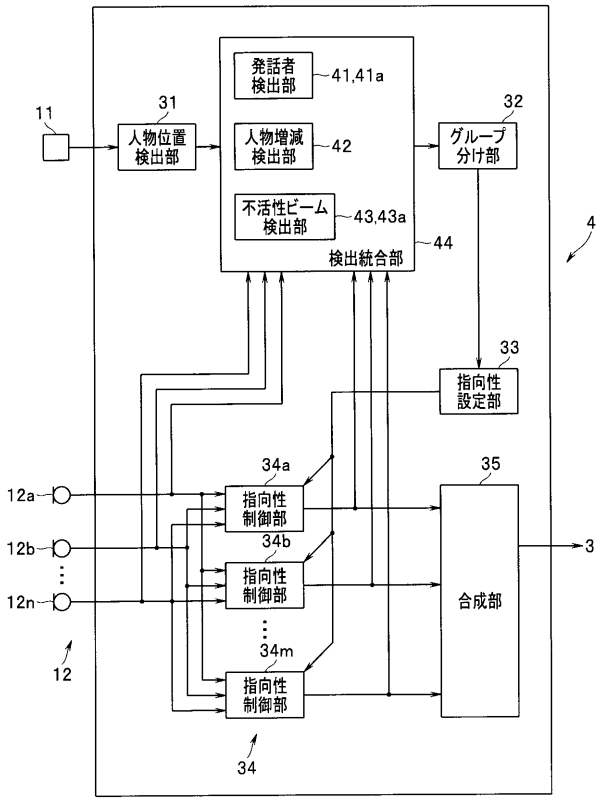
【図１４】



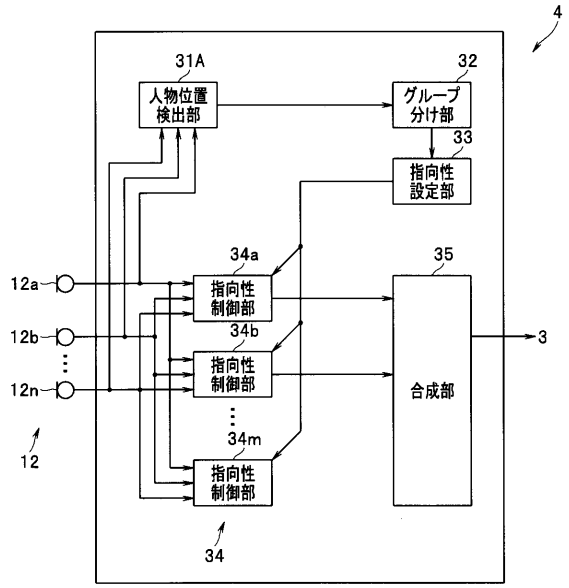
【図１５】



【図１６】



【図１７】



【手続補正書】
【提出日】平成23年11月22日(2011.11.22)
【手続補正１】
【補正対象書類名】明細書
【補正対象項目名】０００７
【補正方法】変更
【補正の内容】
【０００７】
実施形態の音響信号処理装置は、所定空間内に存在する人物の位置を検出する人物位置検出部と、前記人物位置検出部により検出された２以上の人物の数が２以上の所定のグループ数を超えると、前記検出された２以上の人物を、前記所定のグループ数以下の数のグループに割り振るグループ分けを行うグループ分け部と、前記グループ分け部で前記グループ分けがされた各グループの指向性を、マイクロホンアレーの指向性を制御する複数の指向性制御部に設定する指向性設定部と、を有する。
【手続補正２】
【補正対象書類名】特許請求の範囲
【補正対象項目名】全文
【補正方法】変更
【補正の内容】
【特許請求の範囲】
【請求項１】
所定空間内に存在する人物の位置を検出する人物位置検出部と、
前記人物位置検出部により検出された２以上の人物の数が２以上の所定のグループ数を超えると、前記検出された２以上の人物を、前記所定のグループ数以下の数のグループに割り振るグループ分けを行うグループ分け部と、

前記グループ分け部で前記グループ分けがされた各グループの指向性を、マイクロホンアレーの指向性を制御する複数の指向性制御部に設定する指向性設定部と、
を有することを特徴とする音響信号処理装置。
【請求項２】
前記グループ分け部は、前記マイクロホンアレーの雑音抑圧性能の和が最大になるように、前記グループ分けを行うことを特徴とする請求項１に記載の音響信号処理装置。
【請求項３】
前記グループ分け部は、前記各グループに割り振られる人物の数が平均化するように、前記グループ分けを行うことを特徴とする請求項１に記載の音響信号処理装置。
【請求項４】
前記グループ分けの結果のグループ数あるいは前記所定数を、変更するためのグループ数変更部を有することを特徴とする請求項１から３のいずれか１つに記載の音響信号処理装置。
【請求項５】
発話者を検出する発話者検出部を有し、
前記グループ分け部は、前記２以上の人物の中から、前記発話者だけを対象に、前記グループ分けを行うことを特徴とする請求項１から４のいずれか１つに記載の音響信号処理装置。
【請求項６】
前記人物位置検出部により検出された人物の増減を検出する人物増減検出部を有し、
前記グループ分け部は、前記人物増減検出部により検出された前記人物の増減に応じて、前記グループ分けを行うことを特徴とする請求項１から５のいずれか１つに記載の音響信号処理装置。
【請求項７】
前記指向性設定部によって指向性が設定された指向性制御部の出力監視あるいは前記人物位置検出部により検出された人物の発話監視を行う発話状態監視部を有し、
前記グループ分け部は、前記発話状態監視部により、前記指向性が設定された指向性制御部の出力あるいは前記検出された人物の発話が所定時間以上観測されなくなったときは、前記所定時間以上観測されなくなった指向性制御部あるいは人物を除いて、前記グループ分けを行うことを特徴とする請求項１から６のいずれか１つに記載の音響信号処理装置。
【請求項８】
前記人物位置検出部は、前記所定空間をカメラにより撮像して得られた画像により、あるいは前記所定空間からの音響信号により、前記人物の位置を検出することを特徴とする請求項１から７のいずれか１つに記載の音響信号処理装置。
【請求項９】
テレビジョン放送のコンテンツを表示する表示部と、
音響信号処理装置と、
通信回線を介して通信を行うための通信インターフェースと、
を有し、
前記音響信号処理装置は、
所定空間内に存在する人物の位置を検出する人物位置検出部と、
前記人物位置検出部により検出された２以上の人物の数が２以上の所定のグループ数を超えると、前記検出された２以上の人物を、前記所定のグループ数以下の数のグループに割り振るグループ分けを行うグループ分け部と、
前記グループ分け部で前記グループ分けがされた各グループの指向性を、マイクロホンアレーの指向性を制御する複数の指向性制御部に設定する指向性設定部と、
を有することを特徴とするテレビジョン装置。
【請求項１０】
音響信号を処理するプログラムであって、コンピュータに、

所定空間内に存在する人物の位置を検出する機能と、
検出された２以上の人物の数が２以上の所定のグループ数を超えると、前記検出された２以上の人物を、前記所定のグループ数以下の数のグループに割り振るグループ分けを行う機能と、
グループ分けがされた各グループの指向性を、マイクロホンアレーの指向性を制御する複数の指向性制御部に設定する機能と、
を実現させるためのプログラム。