

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6250297号
(P6250297)

(45) 発行日 平成29年12月20日(2017.12.20)

(24) 登録日 平成29年12月1日(2017.12.1)

(51) Int.Cl.

F I

H O 4 R 3/00 (2006.01)

H O 4 R 3/00 3 2 O

G 1 O L 15/20 (2006.01)

G 1 O L 15/20 3 7 O E

G 1 O L 21/0272 (2013.01)

G 1 O L 21/0272 1 O O A

H O 4 R 1/40 (2006.01)

H O 4 R 1/40 3 2 O A

H O 4 N 5/225 (2006.01)

H O 4 N 5/225

請求項の数 5 (全 19 頁) 最終頁に続く

(21) 出願番号 特願2013-84503 (P2013-84503)
 (22) 出願日 平成25年4月15日(2013.4.15)
 (65) 公開番号 特開2014-207589 (P2014-207589A)
 (43) 公開日 平成26年10月30日(2014.10.30)
 審査請求日 平成28年4月11日(2016.4.11)

(73) 特許権者 000005049
 シャープ株式会社
 大阪府堺市堺区匠町1番地
 (74) 代理人 110000338
 特許業務法人HARAKENZO WORLD PATENT & TRADEMARK
 (72) 発明者 徳井 圭
 大阪府大阪市阿倍野区長池町22番22号
 シャープ株式会社内

審査官 渡邊 正宏

最終頁に続く

(54) 【発明の名称】 音声入力装置、および画像表示装置

(57) 【特許請求の範囲】

【請求項1】

画像情報を取得する撮像素子と、音声情報を取得する複数のマイクロフォンとを備えた音声入力装置であって、

前記撮像素子が取得した画像情報からユーザを検出するユーザ検出部と、

ユーザの方向の音声情報を取得するユーザ音声取得部とを備え、

前記ユーザ音声取得部は、検出されたユーザの位置に基づいて音声情報を取得する際の基準となるマイクロフォンを設定し、

前記ユーザ音声取得部は、任意の、検出されたユーザの位置に対して、前記撮像素子を基準としたユーザの方向と、前記複数のマイクロフォンのうちの中央のマイクロフォンを基準としたユーザの方向とがなす角より、前記撮像素子を基準としたユーザの方向と、前記基準となるマイクロフォンを基準としたユーザの方向とがなす角が大きくなるように、前記基準となるマイクロフォンを設定することを特徴とする音声入力装置。

【請求項2】

画像情報を取得する撮像素子と、音声情報を取得する複数のマイクロフォンとを備えた音声入力装置であって、

前記撮像素子が取得した画像情報からユーザを検出するユーザ検出部と、

ユーザの方向の音声情報を取得するユーザ音声取得部とを備え、

前記ユーザ音声取得部は、検出されたユーザの位置に基づいて音声情報を取得する際の基準となるマイクロフォンを設定し、

10

20

前記ユーザ音声取得部は、任意の、検出されたユーザの位置に対して、前記撮像素子を基準としたユーザの方向と、前記複数のマイクロフォンのうちの中央のマイクロフォンを基準としたユーザの方向とがなす角より、前記撮像素子を基準としたユーザの方向と、前記基準となるマイクロフォンを基準としたユーザの方向とがなす角が小さくなるように、前記基準となるマイクロフォンを設定することを特徴とする音声入力装置。

【請求項 3】

画像情報を取得する撮像素子と、音声情報を取得する複数のマイクロフォンとを備えた音声入力装置であって、

前記撮像素子が取得した画像情報からユーザを検出するユーザ検出部と、

ユーザの方向の音声情報を取得するユーザ音声取得部とを備え、

前記ユーザ音声取得部は、検出されたユーザの位置に基づいて音声情報を取得する際の基準となるマイクロフォンを設定し、

前記ユーザ音声取得部は、

(i) 任意の、検出されたユーザの位置に対して、前記撮像素子を基準としたユーザの方向と、前記複数のマイクロフォンのうちの中央のマイクロフォンを基準としたユーザの方向とがなす角より、前記撮像素子を基準としたユーザの方向と、前記基準となるマイクロフォンを基準としたユーザの方向とがなす角が大きくなるように、前記基準となるマイクロフォンを設定する第一のモードと、

(i i) 任意の、検出されたユーザの位置に対して、前記撮像素子を基準としたユーザの方向と、前記複数のマイクロフォンのうちの中央のマイクロフォンを基準としたユーザの方向とがなす角より、前記撮像素子を基準としたユーザの方向と、前記基準となるマイクロフォンを基準としたユーザの方向とがなす角が小さくなるように、前記基準となるマイクロフォンを設定する第二のモードと、

を切り替える、ことを特徴とする音声入力装置。

【請求項 4】

前記ユーザ音声取得部は、前記第一のモードと前記第二のモードとを、前記ユーザ検出部で検出されたユーザの人数に基づいて切り換えることを特徴とする、請求項 3 に記載の音声入力装置。

【請求項 5】

請求項 1 ～ 4 のいずれか 1 項に記載の音声入力装置と、該音声入力装置が出力する音声情報を認識する音声認識部と、前記音声認識部で認識された結果に基づいて出力音声の制御を行う制御部と、前記音声情報に関連付けられた画像を表示する表示部と、を備えることを特徴とする画像表示装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は音声入力装置、および画像表示装置に関する技術であり、特に、音声情報を取得する複数のマイクロフォンと、画像情報を取得する撮像素子とを備え、取得される音声情報のノイズ低減技術に関する。

【背景技術】

【0002】

複数のマイクロフォンを備えるマイクアレイを使用して、特定方向からの音声情報を強調し、かつ、特定方向以外からの音声情報を抑制することにより、特定方向からの音声情報を取得する音声入力装置が開発されている。これは、複数のマイクロフォンから取得される音声情報から音源定位を行い、音源が存在する方向の音声情報を入力音声情報とすることで、特定方向の音声情報を入力音声情報として取得するものである。これにより、ユーザが発話した音声情報を取得することができる。

【0003】

しかし、音声情報を取得する方向を音声情報のみにより制御するため、ユーザ以外の音源が存在する場合、ユーザ以外の音源の方向を特定方向として制御してしまうと、ユーザ

10

20

30

40

50

が意図しない音声情報が入力音声情報として取得されてしまう。

【 0 0 0 4 】

そこで、撮像素子で撮影された画像からユーザの方向を検出し、音声情報を取得する特定方向を検出したユーザの方向になるように制御し、ユーザの音声情報を取得する方法が提案されている。

例えば、特許文献 1 には、カメラにより撮影された撮影画像を用いて、発話者の体格や着座位置などにより変化する話者方向を特定し、マイクロフォンの指向性方向を適切に制御することで音声認識の精度を向上させる音声処理装置が開示されている。

【 先行技術文献 】

【 特許文献 】

10

【 0 0 0 5 】

【 特許文献 1 】 特開 2 0 0 9 - 2 2 5 3 7 9 号公報

【 発明の概要 】

【 発明が解決しようとする課題 】

【 0 0 0 6 】

しかしながら特許文献 1 のように、カメラにより撮影された撮影画像に基づいてマイクロフォンの指向性方向を制御する方法は、以下のような課題を有する。

ユーザが存在する方向と同一の方向に他の音源が存在している場合、ユーザからの音声情報と他の音源からの音声情報を分離して取得することができない。例えば、ユーザの背面に他の人物が存在する場合、撮像素子で撮影された画像からユーザの方向を特定するが、他の人物も同一方向にいるため、他の人物が発話した音声情報として取得してしまう。

20

また、複数のユーザが存在する場合、音声情報を取得する方向を適切に設定しないと、複数のユーザの音声情報がお互いにノイズとして含まれてしまい、各ユーザからの音声情報を取得することができない。

【 0 0 0 7 】

そこで本発明は、上記課題に鑑みてなされたものであり、ノイズである他の音源から発せられる音声情報を低減し、ユーザから発話される音声情報を取得する音声入力装置、および画像表示装置を提供する。

【 課題を解決するための手段 】

【 0 0 0 8 】

30

本発明の音声入力装置は、画像情報を取得する撮像素子と、音声情報を取得する複数のマイクロフォンを備え、撮像素子が取得した画像情報からユーザを検出するユーザ検出部と、マイクロフォンが取得した音声情報のうち、特定方向の音声情報を出力音声情報とするユーザ音声取得部とを備え、ユーザ音声取得部が、特定方向の基準とするマイクロフォンの位置を、ユーザ検出部で検出されるユーザの位置に基づいて変化させることを特徴とする。

【 0 0 0 9 】

さらに、本発明の音声入力装置は、ユーザ音声取得部が、ユーザ検出部で検出されたユーザの人数によって、特定方向の基準とするマイクロフォンの位置を変化させると好適である。

40

【 0 0 1 0 】

さらに、本発明の音声入力装置は、ユーザ音声取得部が、撮像素子を基準としたユーザの方向と、複数のマイクロフォンの中心を基準としたユーザの方向とがなす角より、撮像素子を基準としたユーザの方向と、特定方向とがなす角が大きくなるように、特定方向の基準を設定すると好適である。

【 0 0 1 1 】

さらに、本発明の音声入力装置は、ユーザ音声取得部が、撮像素子を基準としたユーザの方向と、複数のマイクロフォンの中心を基準としたユーザの方向とがなす角より、撮像素子を基準としたユーザの方向と、特定方向とがなす角が小さくなるように、特定方向の基準とするマイクロフォンの位置を設定すると好適である。

50

【 0 0 1 2 】

また、本発明の画像表示装置は、上記の音声入力装置と、音声入力装置が出力する音声情報を認識する音声認識部と、音声認識部で認識された結果に基づいて所定の制御を行う制御部を備えることを特徴とする。

【 発明の効果 】

【 0 0 1 3 】

本発明の音声入力装置によれば、マイクアレイで特定方向の音声情報を取得するときに、方向の基準となる基準マイクの位置を適切に設定することで、ノイズとなる他の音源から発せられる音声情報を低減することが可能となる。

また、本発明の画像表示装置では、ノイズとなる音声情報を低減できる音声入力装置を備えることにより、認識率の高い音声入力可能な画像表示装置を実現することができる。

10

【 図面の簡単な説明 】

【 0 0 1 4 】

【 図 1 】 本発明の音声入力装置の構成例を示す図である。

【 図 2 】 撮像素子により撮影された画像情報の例を示す図である。

【 図 3 】 カメラ基準ユーザ角度を示す図である。

【 図 4 】 カメラ基準ユーザ角度、カメラ基準ユーザ距離、およびマイクアレイ基準ユーザ角度の関係を示す図である。

【 図 5 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の例を示す図である。

【 図 6 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の他の例を示す図である。

20

【 図 7 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の更に他の例を示す図である。

【 図 8 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の更に他の例を示す図である。

【 図 9 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の更に他の例を示す図である。

【 図 1 0 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の更に他の例を示す図である。

【 図 1 1 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の更に他の例を示す図である。

30

【 図 1 2 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の更に他の例を示す図である。

【 図 1 3 】 本発明の音声入力装置の他の構成例を示す図である。

【 図 1 4 】 本発明の音声入力装置の更に他の構成例を示す図である。

【 図 1 5 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の更に他の例を示す図である。

【 図 1 6 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の更に他の例を示す図である。

【 図 1 7 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の更に他の例を示す図である。

40

【 図 1 8 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の更に他の例を示す図である。

【 図 1 9 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の更に他の例を示す図である。

【 図 2 0 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の更に他の例を示す図である。

【 図 2 1 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の更に他の例を示す図である。

【 図 2 2 】 カメラ基準ユーザ角度とマイクアレイ基準ユーザ角度の更に他の例を示す図である。

50

【図 2 3】本発明の画像表示装置の構成例を示す図である。

【図 2 4】本発明の音声入力装置を備えた音声情報記録装置の構成例を示す図である。

【発明を実施するための形態】

【0015】

以下、図面を使って本発明の実施の形態を詳細に説明する。なお、各図における表現は理解しやすいように誇張して記載しており、実際のものとは異なる場合がある。

【0016】

(実施形態 1)

図 1 は、本実施形態の構成を示す図である。本実施形態の音声入力装置 100 は、撮像素子 101、ユーザ検出部 102、カメラ基準ユーザ角度算出部 103、カメラ基準ユーザ距離算出部 104、マイクアレイ基準ユーザ角度算出部 105、マイクアレイ 106、ユーザ音声取得部 107 を備える。

10

撮像素子 101 は画像情報を取得するもので、CCD (Charge Coupled Device) や CMOS (Complementary Metal Oxide Semiconductor) センサといった固体撮像素子とレンズなどにより構成される。

【0017】

撮像素子 101 で取得した画像情報はユーザ検出部 102 に伝達され、画像情報内のユーザの情報が検出される。カメラ基準ユーザ角度算出部 103、およびカメラ基準ユーザ距離算出部 104 は、ユーザ検出部 102 で検出されたユーザ情報に基づいて、撮像素子 101 で撮影された画像情報を基準としたユーザがいる方向 (カメラ基準ユーザ方向)、および、ユーザまでの距離 (カメラ基準ユーザ距離) を各々算出する。算出された情報はマイクアレイ基準ユーザ角度算出部 105 に伝達する。

20

【0018】

マイクアレイ基準ユーザ角度算出部 105 では、伝達されたカメラ基準ユーザ方向とカメラ基準ユーザ距離とから、マイクアレイを基準としたユーザの方向 (マイクアレイ基準ユーザ方向) を算出する。算出した情報はユーザ音声取得部 107 に伝達する。

マイクアレイ 106 は複数のマイクロフォンを備え、所定の間隔で配置される。例えば、一定間隔で一列に配置する。マイクロフォンは周囲の音声情報を取得し、各マイクロフォンで取得された音声情報はユーザ音声取得部 107 に伝達する。

【0019】

ユーザ音声取得部 107 は、マイクアレイ基準ユーザ方向に基づいて、マイクアレイ 106 から伝達された入力音声情報から、ユーザ音声を取得して音声情報を出力する。

30

ここで、ユーザ検出部 102、カメラ基準ユーザ角度算出部 103、カメラ基準ユーザ距離算出部 104、マイクアレイ基準ユーザ角度算出部 105、およびユーザ音声取得部 107 は、CPU (Central Processing Unit) や GPU (Graphics Processing Unit) によるソフトウェア処理、ASIC (Application Specific Integrated Circuit) や FPGA (Field Programmable Gate Array) によるハードウェア処理によって実現することができる。

【0020】

図 2 は、撮像素子で取得した撮影画像の画像情報から顔領域を検出する例を説明するための図である。本実施例では、ユーザ検出部 102 におけるユーザ検出を顔検出によって行う。顔検出は一般に使用されている方法を使用することができる。例えば、多数の顔画像から算出した標準的な顔画像を参照データとして保有しておき、その参照データとの相関値から顔を検出する方法がある。顔検出により、入力された撮影画像 200 から、ユーザ 201 の顔検出領域 210 を検出する。これにより、撮影画像 200 における顔検出領域 210 の位置や大きさを検出することができる。

40

【0021】

カメラ基準ユーザ距離算出部 104 では、顔検出領域 210 の大きさに基づいてユーザ 201 までの距離を算出する。ユーザ 201 までの距離が遠ければ顔検出領域 210 は小さくなり、ユーザ 201 までの距離が近ければ顔検出領域 210 は大きくなる。ユーザ 2

50

01までの距離は、顔検出領域210の大きさと距離との関係をLUT (Look Up Table)などに記憶しておくことで算出することができる。

【0022】

ここで、ユーザ検出部102において顔検出を行うときに、年齢や性別といった情報も算出しておき、顔検出領域210の大きさから算出される距離を補正すると良い。例えば、大人よりも子供の方が顔は小さいので、大人と子供で顔検出領域210が同じ大きさであった場合、子供の方が撮像素子101に近い位置にいることになる。また、距離の補正は、共通のLUTを利用して、年齢などの情報により増減させる方法や、年齢などのグループごとにLUTを保有するようにしても良い。

【0023】

カメラ基準ユーザ角度算出部103では、顔検出領域210の位置に基づいてユーザ201の方向を示すカメラ基準ユーザ角度 θ を算出する。図3は、図2のシーンをユーザ上方から見たときを示す。ユーザ201の移動は、図3のような地面に対して水平方向に行われるため、本実施形態のカメラ基準ユーザ角度 θ は、撮像素子101の光軸であるカメラ基準軸10からの水平方向への角度とする。カメラ基準ユーザ角度 θ は、撮像素子101の焦点距離、解像度などの既知の値と、ユーザの顔検出領域の位置から算出することが可能である。

【0024】

図4は、カメラ基準ユーザ角度 θ 、およびカメラ基準ユーザ距離Lとマイクアレイ基準ユーザ角度 ϕ との関係を示す図である。本実施形態のマイクアレイ106は、撮像素子101から水平方向にマイクロフォンが離間するように配置され、カメラ基準軸10とマイクアレイ基準軸20とが間隔Wだけ離間して配置している。カメラ基準軸10は、撮像素子101の光軸に一致する軸であり、マイクアレイ基準軸20は、マイクアレイ106のうちの特定のマイクロフォンを通りカメラ基準軸10に平行な軸である。

【0025】

カメラ基準ユーザ角度 θ は、カメラ基準ユーザ角度算出部103により算出されている。また、カメラ基準ユーザ距離Lは、カメラ基準ユーザ距離算出部104により算出される。カメラ基準ユーザ距離Lは、カメラ基準軸10方向における撮像素子101からユーザ201までの距離であり、カメラ基準ユーザ距離算出部104が算出した撮像素子101からユーザ201までの距離と、カメラ基準ユーザ角度算出部103が算出したカメラ基準ユーザ角度 θ から算出することができる。

【0026】

撮像素子101（カメラ基準軸10）とマイクアレイ106（マイクアレイ基準軸20）との距離である間隔Wは、音声入力装置が備えるものであり既知である。したがって、カメラ基準ユーザ角度 θ 、カメラ基準ユーザ距離L、間隔Wからマイクアレイ基準ユーザ角度 ϕ を算出することができる。

ここではカメラ基準軸10からユーザ方向にカメラ基準ユーザ角度 θ だけ傾いた方向をカメラ基準ユーザ方向11とし、マイクアレイ基準軸20からユーザ方向にマイクアレイ基準ユーザ角度 ϕ だけ傾いた方向をマイクアレイ基準ユーザ方向21とする。また、カメラ基準ユーザ方向11とマイクアレイ基準ユーザ方向21とのなす角度を α で表す。このときの角度 α は、カメラ基準ユーザ方向11とマイクアレイ基準ユーザ方向21とのなす角度のうち、カメラ基準ユーザ方向11とマイクアレイ基準ユーザ方向21との交点、撮像素子101、及びマイクアレイ基準軸20を設定したマイクロフォンの3点を頂点とする3角形の内角側の角度を示すものとする。

【0027】

ユーザ音声取得部107は、マイクアレイ106から取得された入力音声情報から、マイクアレイ基準ユーザ角度算出部105で算出されたマイクアレイ基準ユーザ角度 ϕ に基づき、マイクアレイ基準ユーザ方向21からの音声情報を取得する。このとき音声情報を取得する基準となるマイクロフォンは、マイクアレイ基準軸20を設定したマイクロフォンとする。そしてマイクアレイ基準軸20を設定したマイクロフォンを基準として、マイ

10

20

30

40

50

クアレイ基準ユーザ方向 2 1 を特定方向とし、その特定方向からの音声情報を取得する。特定方向から音声情報を取得する角度範囲は、マイクアレイ基準ユーザ方向 2 1 を含む一定の範囲を設定することができる。

特定の角度の音声取得は一般的な方法を使用することができ、例えば、マイクアレイ 1 0 6 の各マイクロフォンに到達する音声の時間差や音量差と、各マイクロフォンの音声取得特性と、各マイクロフォンの位置関係とから計算する。

【 0 0 2 8 】

時間差を使用する場合には、音源から遠いマイクロフォンは他のマイクロフォンに比べ遅延が大きくなる。また、音量を使用する場合には、音源から遠いマイクロフォンは他のマイクロフォンに比べ音量が小さくなる。このようにして、特定方向からの音声情報がどのようにマイクロフォンに到達するかを推定して音声情報を取得する。取得された音声情報は出力音声情報として出力する。

【 0 0 2 9 】

ここで、図 4 では、マイクアレイ基準軸 2 0 をマイクアレイ 1 0 6 の一番右側のマイクロフォンに設定しているが、適切に設定することで、ユーザ以外が発した音声情報であるノイズを低減することができる。ノイズとなる音声情報は、画像情報から検出されたユーザの背後に存在するかもしれない他の人物から発生してしまう。画像情報は透視投影などで取得した 2 次元の画像情報であり、ユーザの背後の音源についての情報までは取得することができない。したがって、他の人物が発話した音声情報はノイズとなり、音声入力装置の出力音声情報で音声認識を行った場合誤動作をする可能性がある。

【 0 0 3 0 】

図 5 は、撮像素子 1 0 1 がマイクアレイ 1 0 6 の中央のマイクロフォン位置に配置され、マイクロフォンの配列方向において、マイクアレイ 1 0 6 の右端のマイクロフォンに近い位置にユーザが検出された場合を示す図である。マイクアレイ基準軸 2 0 は、撮像素子 1 0 1 に対してユーザが検出された側と反対側で撮像素子 1 0 1 から最も遠いマイクロフォンに設定されている。一方、図 6 は、図 5 と同じ環境でマイクアレイ基準軸 2 0 が、撮像素子 1 0 1 に対してユーザが検出された側で撮像素子 1 0 1 から最も遠いマイクロフォンに設定されている。

【 0 0 3 1 】

マイクアレイ 1 0 6 から特定方向の音声情報を取得する場合、マイクアレイ基準ユーザ方向 2 1 を含む所定の角度幅の音声情報を取得する。例えば、マイクアレイ基準ユーザ角度 $\varphi \pm 10$ 度の角度幅とする。このとき、撮像素子 1 0 1 で撮影された画像情報の周辺で歪が生じている場合などは、画像情報周辺ではユーザ検出部 1 0 2 でのユーザの位置検出精度が低下する可能性があるため、角度 θ の絶対値が大きくなるほど音声情報を取得する角度幅を大きくすると、ユーザの音声情報が取得できないなどの不具合を低減することができ好適である。

また、ユーザ検出部 1 0 2 でのユーザの顔領域が小さい場合、ユーザ位置の検出精度が低下する可能性があるため、カメラ基準ユーザ距離 L が大きくなるほど音声情報を取得する角度幅を大きくすると、ユーザの音声情報が取得できないなどの不具合を低減することができ好適である。

【 0 0 3 2 】

図 5 および図 6 において、撮像素子 1 0 1 によりユーザ 2 0 1 が撮影される範囲をユーザ撮影範囲 1 2 とし、マイクアレイ 1 0 6 の特定のマイクロフォンを基準としてユーザ 2 0 1 による音声を取得できる範囲をユーザ音声取得範囲 2 2 とする。ユーザ音声取得範囲 2 2 は、マイクアレイ 1 0 6 による所定の角度幅の音声情報の取得範囲ではなく、ユーザ 2 0 1 による音声が取得できる範囲を示している。

撮像素子 1 0 1 からユーザ 2 0 1 の方向を見たとき、ユーザ 2 0 1 の背後領域にユーザ以外の人物が存在するかなどの情報は画像情報からは取得できない。つまり、ユーザ撮影範囲 1 2 内でユーザ 2 0 1 の背後領域にいる他の人物は、画像情報からは認識することができない。

10

20

30

40

50

【 0 0 3 3 】

一方、マイクアレイ 1 0 6 のマイクアレイ基準軸 2 0 を設定したマイクロフォンでは、そのマイクロフォンを基準としてユーザ 2 0 1 の方向からの音声を取得するが、このとき、ユーザ音声取得範囲 2 2 内でユーザ 2 0 1 の背後に別の人物がいた場合、その別の人物からの音声も取得するため、ユーザ 2 0 1 からの音声情報と他の人物による音源からの音声情報を分離して取得することができずにノイズとなってしまう。

【 0 0 3 4 】

マイクアレイ 1 0 6 が音声情報を取得する特定方向の領域は、カメラ基準軸 1 0 とマイクアレイ基準軸 2 0 が異なると、撮像素子 1 0 1 からユーザ 2 0 1 の方向を見たときのユーザ領域とは異なる。つまり、マイクアレイ基準軸 2 0 を設定する位置により、カメラから見たユーザ背後領域とマイクアレイ 1 0 6 による音声情報取得領域との重なりが変化する。

10

このときに、ユーザ撮影範囲 1 2 内でユーザ 2 0 1 の背後の領域は、他のユーザが存在しているかもしれないため、この領域からは音声を取得しないようにすることが好ましい。このためには、ユーザ撮影範囲 1 2 内のユーザ 2 0 1 の背後の領域と、ユーザ音声取得範囲 2 2 とが重なる領域を最も小さくすることが好ましい。これにより、撮影画像からは認識できないユーザ 2 0 1 の背後のユーザからの音声取得をできる限り回避し、ノイズを低減させることができる。

【 0 0 3 5 】

図 5 と図 6 において、ユーザ撮影範囲 1 2 内のユーザ 2 0 1 の背後の領域と、マイクアレイ基準軸 2 0 を設定したマイクロフォンによるユーザ音声取得範囲 2 2 との重なり領域（図中斜線で示す領域）S の大きさは、図 5 よりも図 6 の場合の方が狭くなっている。これは、マイクアレイ基準軸 2 0 の設定により、カメラ基準ユーザ方向 1 1 とマイクアレイ基準ユーザ方向 2 1 とがなす角度 α が変化するためである。図 5 に比べ図 6 では、カメラ基準ユーザ方向 1 1 とマイクアレイ基準ユーザ方向 2 1 とがなす角度 α が大きくなっており、図 6 の方が、撮影画像におけるユーザ背後領域からの音声取得領域（重なり領域 S）が狭くなっている。これにより、ユーザ背後領域に他の人物が存在しても、音声情報として取得してしまう可能性を低減することができる。

20

【 0 0 3 6 】

撮像素子 1 0 1 とマイクアレイ 1 0 6 との配置が変化しても同様に効果を得ることができる。図 7 から図 1 2 は、撮像素子 1 0 1 の片側にマイクロフォンが配置されている場合を示した図である。

30

図 7 および図 8 は、マイクロフォンの配列方向において撮像素子 1 0 1 とマイクアレイ 1 0 6 との間にユーザ 2 0 1 が検出された場合で、撮像素子 1 0 1、検出されたユーザ 2 0 1、マイクアレイ 1 0 6 の順で位置している。

【 0 0 3 7 】

図 7 では、マイクアレイ基準軸 2 0 が撮像素子 1 0 1 に最も近いマイクロフォンに設定されており、図 8 では、マイクアレイ基準軸 2 0 が撮像素子 1 0 1 から最も遠いマイクロフォンに設定されている。図 7 におけるカメラ基準ユーザ方向 1 1 とマイクアレイ基準ユーザ方向 2 1 とがなす角 α よりも、図 8 におけるカメラ基準ユーザ方向 1 1 とマイクアレイ基準ユーザ方向 2 1 とがなす角 α の方が大きくなっており、ユーザ撮影範囲 1 2 内のユーザ 2 0 1 の背後の領域と、マイクアレイ基準軸 2 0 を設定したマイクロフォンによるユーザ音声取得範囲 2 2 との重なり領域（図中斜線で示す領域）S は、図 8 の方が小さくなっている。これにより、ユーザの背後領域に存在するかもしれない他の人物からの音声情報を、ノイズとして取得することを低減できる。

40

【 0 0 3 8 】

撮像素子 1 0 1 とマイクアレイ 1 0 6 の配置が同様の状態で、ユーザの位置が変化しても同様の効果が得られる。図 9 および図 1 0 は、マイクロフォンの配列方向において、検出されたユーザ 2 0 1、撮像素子 1 0 1、マイクアレイ 1 0 6 の順で位置している。図 9 ではマイクアレイ基準軸 2 0 が撮像素子 1 0 1 に最も近いマイクロフォンに設定されてお

50

り、図10ではマイクアレイ基準軸20が撮像素子101から最も遠いマイクロフォンに設定されている。

【0039】

図9におけるカメラ基準ユーザ方向11とマイクアレイ基準ユーザ方向21とがなす角 α よりも、図10におけるカメラ基準ユーザ方向11とマイクアレイ基準ユーザ方向21とがなす角 α の方が大きくなっており、ユーザ撮影範囲12内のユーザ201の背後の領域と、マイクアレイ基準軸20を設定したマイクロフォンによるユーザ音声取得範囲22との重なり領域(図中斜線で示す領域)Sは、図10の方が小さくなっている。これにより、ユーザの背後領域に存在するかもしれない他の人物からの音声情報を、ノイズとして取得することを低減できる。

10

【0040】

また、図11および図12は、マイクロフォンの配列方向において、撮像素子101、マイクアレイ106、検出されたユーザ201の順で位置している。図11ではマイクアレイ基準軸20が撮像素子101に最も近いマイクロフォンに設定されており、図12ではマイクアレイ基準軸20が撮像素子101から最も遠いマイクロフォンに設定されている。図11におけるカメラ基準ユーザ方向11とマイクアレイ基準ユーザ方向21とがなす角 α よりも、図12におけるカメラ基準ユーザ方向11とマイクアレイ基準ユーザ方向21とがなす角 α の方が大きくなっており、ユーザ撮影範囲12内のユーザ201の背後の領域と、マイクアレイ基準軸20を設定したマイクロフォンによるユーザ音声取得範囲22との重なり領域(図中斜線で示す領域)Sは、図12の方が小さくなっている。これにより、ユーザの背後領域に存在するかもしれない他の人物からの音声情報を、ノイズとして取得することを低減できる。

20

【0041】

以上で説明したように、カメラ基準ユーザ方向11とマイクアレイ基準ユーザ方向21とがなす角 α が、大きくなるようにマイクアレイ基準軸20を設定することで、ユーザの背後領域に存在するかもしれない他の人物からの音声情報を、ノイズとして取得することを低減できる。

ここでは初期設定としてマイクアレイ106の中央のマイクロフォンにマイクアレイ基準軸20を設定しておき、初期設定よりカメラ基準ユーザ方向11とマイクアレイ基準ユーザ方向21とがなす角 α が大きくなるようにマイクアレイ基準軸20を設定することで

30

【0042】

また、本実施形態では実在するマイクロフォンにマイクアレイ基準軸20を設定するように説明をしたが、仮想的なマイクロフォンにマイクアレイ基準軸20を設定しても良い。例えば、マイクロフォンAとマイクロフォンBの間に仮想マイクロフォンが存在すると仮定し、各マイクロフォンで取得された複数の音声情報から、仮想マイクロフォンにおける特定方向からの音声情報算出結果を推定する。これは、各マイクロフォンからの相対的な距離が分かっているため、音声情報の遅延量や音量などを推定することで実現できる。

【0043】

ここで、ユーザ201の位置のみでマイクアレイ基準軸20を再設定すると、ユーザ201が動きながら発話する場合、ユーザ201が発話した単語の途中で音声情報の取得結果が変化してしまい、音声認識などが正常に動作しない可能性がある。そこで、取得される音声情報が0に近づいたとき、すなわち、ユーザ201の発話が無い状態のときに、マイクアレイ基準軸20を再設定することで、発話した音声情報に切れ目が生じていない音声情報を取得することが可能になり好適である。

40

【0044】

上述した実施形態では、検出された顔の大きさからユーザまでの距離を算出する方法を説明したが、他の方法で距離を算出しても同様の効果を得ることができる。例えば、図13のように、撮像素子108と距離算出部109とを備える方法が有る。撮像素子108は撮像素子101と同様に、固体撮像素子とレンズなどを備える。距離算出部109は、

50

撮像素子 101 と撮像素子 108 とで取得される 2 つの画像情報の視差を算出し、ユーザ検出部と同様に各種のハードウェアやソフトウェアによって処理が実現される。

【0045】

視差は一般に使用される方法が適用でき、例えば、2 つの画像のブロックマッチングにより実現でき、撮像素子 101 で取得される画像情報に基準探索窓を設定し、撮像素子 108 で取得される画像情報に参照探索窓を設定し、参照探索窓を移動していく。ブロックマッチングは、SAD (Sum of Absolute Difference) や SSD (Sum of Squared Difference) などにより類似度または相違度を評価することで行われる。算出された視差 D から距離を算出することができ、距離 Z は $Z = B \times f / D$ により算出される。ここで、 B は 2 つの撮像素子間の距離である基線長であり、 f は撮像素子の焦点距離である。

10

【0046】

距離算出部 109 で算出された視差情報は、カメラ基準ユーザ距離算出部 104 に伝達され、ユーザ検出部 102 で検出されたユーザの画像内の位置に対応した距離情報を取得する。このようにして、2 つの撮像素子から距離情報を算出しても良い。また、上記では画像情報全体の視差情報を算出する方法を説明したが、ユーザ検出部 102 で検出されたユーザの位置に基づいて、ユーザの顔領域に対してブロックマッチングを行うと、処理量を削減でき好適である。

【0047】

また、図 14 のように測距素子 110 を備え、測距素子 110 から距離情報を取得しても同様の効果を得ることができる。測距素子 110 には一般に使用される測距素子が適用でき、例えば、赤外線を照射して反射されてくるまでの時間により距離を計測する TOF (Time Of Flight) センサなどがある。また、照射する赤外線を 2 次元のパターンとし、その形状の変化から距離を取得するセンサもある。

20

【0048】

測距素子 110 で取得された距離情報はカメラ基準ユーザ距離算出部 104 に伝達され、ユーザ検出部 102 で検出されたユーザの位置に対応した距離情報を取得してユーザまでの距離とする。以上の方法により、測距素子 110 から距離情報を取得しても同様の効果を得ることができる。

【0049】

(実施形態 2)

30

実施形態 1 では、ユーザ検出部 102 で検出されるユーザ 201 が 1 名である場合であったが、本実施形態 2 では複数人の人物が検出された場合の実施形態を説明する。本実施形態における音声入力装置 100 の構成は、実施形態 1 と同様で図 1、図 13、図 14 に示す構成を有するものであり、共通する各部の詳細な説明は省略する。

【0050】

図 15 は、ユーザが第 1 のユーザ 201 および第 2 のユーザ 202 の 2 名であった場合で、ユーザ検出部 102 で検出されるユーザは第 1 のユーザ 201 および第 2 のユーザ 202 の 2 名となり、検出されるユーザの情報も第 1 のユーザ 201 および第 2 のユーザ 202 の 2 名分となる。

図 15 では、検出された第 1 のユーザ 201 に対してマイクアレイ基準軸 20 を撮像素子 101 から最も遠いマイクロフォンに設定している。このとき、カメラ基準ユーザ方向 11 とマイクアレイ基準ユーザ方向 21 とがなす角 α が大きくなるようにマイクアレイ基準軸 20 を設定しているが、音声情報を取得する領域に第 2 のユーザ 202 が存在するため、第 1 のユーザ 201 から取得したい音声情報に、第 2 のユーザ 202 から発せられた音声情報が含まれてしまう可能性がある。

40

【0051】

この場合、第 2 のユーザ 202 から発せられた音声情報はノイズとなる。同様に、カメラ基準ユーザ方向 11 とマイクアレイ基準ユーザ方向 21 とがなす角 α が大きくなるようにマイクアレイ基準軸 20 を設定して、第 2 のユーザ 202 の音声情報を取得しようすると、第 1 のユーザ 201 から発せられた音声情報が含まれてしまう可能性がある。

50

【 0 0 5 2 】

図 1 6 では、ユーザ検出部 1 0 2 で検出された第 2 のユーザ 2 0 2 の位置が図 1 5 と異なっている。同様に、カメラ基準ユーザ方向 1 1 とマイクアレイ基準ユーザ方向 2 1 とがなす角 α が大きくなるようにマイクアレイ基準軸 2 0 を設定すると、第 1 のユーザ 2 0 1 から取得する音声情報には第 2 のユーザ 2 0 2 が発した音声情報が含まれる可能性があり、第 2 のユーザ 2 0 2 から取得する音声情報には第 1 のユーザ 2 0 1 が発した音声情報が含まれる可能性がある。したがって、所望とするユーザからの音声情報にノイズが含まれる可能性がある。

【 0 0 5 3 】

そこで、カメラ基準ユーザ方向 1 1 とマイクアレイ基準ユーザ方向 2 1 とがなす角 α が小さくなるように、マイクアレイ基準軸 2 0 を設定して取得する音声情報のノイズを低減する。図 1 7 は図 1 5 と同じようにユーザ検出部 1 0 2 で第 1 のユーザ 2 0 1 および第 2 のユーザ 2 0 2 が検出された場合を示す。マイクアレイ基準軸 2 0 は、カメラ基準ユーザ方向 1 1 とマイクアレイ基準ユーザ方向 2 1 とがなす角 α が小さくなるように、撮像素子 1 0 1 に近いマイクロフォンに設定されている。カメラ基準ユーザ方向 1 1 とマイクアレイ基準ユーザ方向 2 1 とがなす角 α が小さくなるように、マイクアレイ基準軸 2 0 を設定することにより、第 1 のユーザ 2 0 1 の音声情報を取得する領域から第 2 のユーザ 2 0 2 が離れるため、第 1 のユーザ 2 0 1 から取得される音声情報から第 2 のユーザ 2 0 2 で発せられる音声情報を低減することができる。すなわち、第 1 のユーザ 2 0 1 から取得される音声情報のノイズを低減することが可能となる。

【 0 0 5 4 】

図 1 8 は図 1 6 と同じようにユーザ検出部 1 0 2 で第 1 のユーザ 2 0 1 および第 2 のユーザ 2 0 2 が検出された場合を示す。マイクアレイ基準軸 2 0 は、カメラ基準ユーザ方向 1 1 とマイクアレイ基準ユーザ方向 2 1 とがなす角 α が小さくなるように、撮像素子 1 0 1 に近いマイクロフォンに設定されている。このときも図 1 7 の場合と同様に、カメラ基準ユーザ方向 1 1 とマイクアレイ基準ユーザ方向 2 1 とがなす角 α が小さくなるように、マイクアレイ基準軸 2 0 を設定することにより、第 1 のユーザ 2 0 1 の音声情報を取得する領域から第 2 のユーザ 2 0 2 が離れるため、第 1 のユーザ 2 0 1 から取得される音声情報における、第 2 のユーザ 2 0 2 で発せられる音声情報を低減することができる。すなわち、第 1 のユーザ 2 0 1 から取得される音声情報のノイズを低減することが可能となる。

【 0 0 5 5 】

また、図 1 7 および図 1 8 において、第 2 のユーザ 2 0 2 に対する音声情報の取得も可能で、各々図 1 9 と図 2 0 に対応する。第 2 のユーザ 2 0 2 の音声情報を取得するときのマイクアレイ基準軸 2 0 は、第 1 のユーザ 2 0 1 の音声情報を取得するときのマイクアレイ基準軸 2 0 と同じにしている。このとき、カメラ基準ユーザ方向 1 1 とマイクアレイ基準ユーザ方向 2 1 とがなす角 α が小さくなるように、マイクアレイ基準軸 2 0 は設定されており、第 2 のユーザ 2 0 2 から取得される音声情報における、第 1 のユーザ 2 0 1 で発せられる音声情報を低減することができる。すなわち、第 2 のユーザ 2 0 2 から取得される音声情報のノイズを低減することが可能となる。

【 0 0 5 6 】

ここで、音声情報を取得する方向は、マイクアレイ基準ユーザ方向 2 1 に対して一定の角度幅を設定することができるため、図 2 1 のように、マイクロフォンにより音声情報を取得する音声取得範囲 2 3 が重複する可能性がある。図 2 1 に示す音声取得角度範囲 2 3 は、これまで示したようなマイクロフォンによりユーザの音声情報が取得されるユーザ音声取得範囲 2 2 ではなく、マイクアレイ 1 0 6 により取得する特定方向の音声情報の取得角度範囲を示すものであり、ユーザの大きさに関わりなく、例えばマイクアレイ基準ユーザ角度 $\varphi \pm 10$ 度の範囲で設定されるものである。

音声取得角度範囲 2 3 が重複すると、第 1 のユーザ 2 0 1 から取得される音声情報には第 2 のユーザ 2 0 2 が発した音声情報が含まれてしまい、第 2 のユーザ 2 0 2 から取得される音声情報には第 1 のユーザ 2 0 1 が発した音声情報が含まれてしまう可能性がある。

【 0 0 5 7 】

そこで、第 1 のユーザ 2 0 1 の音声取得角度範囲 2 3 と第 2 のユーザ 2 0 2 の音声取得角度範囲 2 3 とが重複しないように、音声情報を取得する範囲を変化させる。図 2 2 に示すように、第 1 のユーザ 2 0 1 と第 2 のユーザ 2 0 2 の間に共通の角度範囲の境界 B を設定する。これにより、第 1 のユーザ 2 0 1 および第 2 のユーザ 2 0 2 から各々ノイズを低減した音声情報を取得できるため好適である。

【 0 0 5 8 】

音声情報を取得する音声取得角度範囲 2 3 の制御を行う場合、マイクアレイ基準ユーザ方向 2 1 に対する角度幅は変化させないで、マイクアレイ基準ユーザ方向 2 1 を音声取得角度範囲 2 3 が重複しないように補正する方法がある。また、マイクアレイ基準ユーザ方向 2 1 を変化させないで、マイクアレイ基準ユーザ方向 2 1 に対しての角度幅を音声取得角度範囲 2 3 が重複しないように補正する方法でも良い。このとき、角度幅の補正は片方でも両方でも良いが、音声情報の取得角度範囲が極端に小さくならないようにするため、重複する側の片方を補正すると好適である。

【 0 0 5 9 】

以上により、カメラ基準ユーザ方向 1 1 とマイクアレイ基準ユーザ方向 2 1 とがなす角が小さくなるように、マイクアレイ基準軸 2 0 を設定することにより、第 1 のユーザ 2 0 1 の音声情報を取得する領域から第 2 のユーザ 2 0 2 が離れるため、第 1 のユーザ 2 0 1 から取得される音声情報における、第 2 のユーザ 2 0 2 で発せられる音声情報を低減することが可能となる。

【 0 0 6 0 】

上記の実施形態では、ユーザ検出部 1 0 2 で検出されるユーザの人数が 2 名の場合で説明したが、本発明に係る実施形態は、3 名以上の場合にも適用することが可能である。また、ユーザ検出部 1 0 2 で検出されるユーザの人数が変化して、検出されたユーザが 2 名から 1 名になった場合には、実施形態 1 で説明した制御にすることで、シーンに合わせてノイズを低減した音声情報を取得できる。すなわち、ユーザ検出部 1 0 2 で検出されるユーザに基づいて、マイクアレイ基準軸 2 0 の設定方法を切り換えると、シーンに合わせてノイズを低減した音声情報を取得できるため好適である。

【 0 0 6 1 】

以上で説明したように、複数のユーザが検出されたときに、カメラ基準ユーザ方向 1 1 とマイクアレイ基準ユーザ方向 2 1 とがなす角 α が、小さくなるようにマイクアレイ基準軸 2 0 を設定することで、複数のユーザの音声情報を取得することによるノイズを低減できる。

ここでは初期設定としてマイクアレイ 1 0 6 の中央のマイクロフォンにマイクアレイ基準軸 2 0 を設定しておき、初期設定よりカメラ基準ユーザ方向 1 1 とマイクアレイ基準ユーザ方向 2 1 とがなす角 α が小さくなるようにマイクアレイ基準軸 2 0 を設定することでノイズの低減を実現できる。

【 0 0 6 2 】

(実施形態 3)

実施形態 3 は、実施形態 1 および実施形態 2 で説明した音声入力装置を備える画像表示装置である。本実施形態における音声入力装置の構成は、実施形態 1 および実施形態 2 と同様であるため、共通する各部の詳細な説明は省略する。

【 0 0 6 3 】

図 2 3 は本実施形態の構成を示す図である。画像表示装置 3 0 0 は、音声入力装置 1 0 0、音声認識部 3 0 1、制御部 3 0 2、画像表示部 3 0 3、音声出力部 3 0 4 を備える。音声認識部 3 0 1 は、音声入力装置 1 0 0 から出力された音声情報を認識する。音声情報を認識する方法は一般的な方法を使用することができる。例えば、単語の音声データを保有しておき、そのデータと入力された音声情報を比較して、その類似度により認識を行う方法がある。

【 0 0 6 4 】

音声認識部 301 で認識された音声認識結果は制御部 302 に伝達される。制御部 302 では音声認識結果に基づいて画像表示部 303 や音声出力部 304などを制御する。ここで、音声認識部 301 や制御部 302 は、CPUでのソフトウェア処理、ASICでのハードウェア処理により実現できる。また、画像表示部 303 は、画像情報が表示可能な表示デバイスなどから構成され、例えば、液晶パネルとバックライト、有機EL (Electro Luminescence) パネルなどである。音声出力部 304 はスピーカなどにより構成されている。

【0065】

制御部 302 による画像表示部 303 の制御としては、例えば、テレビ放送のチャンネル変更、画面の明るさ増減、などがある。音声入力装置 100 からの音声認識結果が「明るく」であれば、画像表示部 303 の画面の明るさを高くし、音声認識結果が「暗く」であれば、画像表示部 303 の画面の明るさを低くする。また、音声出力部 304 の制御としては、例えば、テレビ放送の音量の増減がある。音声入力装置 100 からの音声認識結果が「大きく」であれば、音声出力部 304 の音量を大きくし、音声認識結果が「低く」であれば、音声出力部 304 の音量を小さくする。

【0066】

ここで、音声入力装置 100 から出力される音声情報にノイズが多く含まれていると、音声認識部 301 での認識率の低下や誤認識が発生してしまう。つまり、制御部 302 が動作しなかったり、意図しない動作したりすることになる。したがって、実施形態 1 や実施形態 2 で説明したノイズを低減した音声情報が取得できる音声入力装置 100 を備えることにより、音声入力操作による認識率を高め、誤動作を低減することが可能な画像表示装置 200 を実現することが可能になる。

【0067】

以上の説明では、制御部 302 による制御は画像表示部 303 と音声出力部 304 とで説明したが、他の機能を制御することも可能である。例えば、電源のOFF、インターネットへの接続、選択肢の選択および決定などがある。さらに、画像表示装置 300 の制御だけでなく、画像表示装置 300 に接続される機器などの制御も可能で、例えば、録画機への録画、エアコンの温度設定、照明器具のON/OFFなどがある。

【0068】

ここで、上記で説明した音声入力装置 100 を使用した音声情報による制御は、複数のユーザにより同時に操作しようとする、と相反する制御が同時に発生する場合がある。例えば、異なる放送番組への切り替え、音量の増加と減少などである。そこで、音声入力装置 100 のユーザ検出部 102 において、操作の権利を特定のユーザのみに与えるようにすると好適である。例えば、画像表示装置 300 からの距離、撮像素子で撮影されている時間などの評価基準を設定して、その基準に合わせて操作可能なユーザを決定する。

【0069】

また、ジェスチャなどと組み合わせると、操作可能なユーザの切り換えが容易となり好適である。例えば、ユーザの顔の下領域に手の平を配置するジェスチャを最も先に行ったユーザを音声による操作が可能なユーザとする。さらに、上記ジェスチャを行っている間だけ音声入力が可能としておくことで、意図しない音声情報が入力されて誤動作が発生しないようにできるため好適である。このとき、音声による操作が可能であることをユーザに通知するように、LED (Light Emitting Diode) を点灯したり、画像表示部に音声入力可能であることを表示したりすると、ユーザのジェスチャが認識されていることが確認できるため好適である。

【0070】

(実施形態 4)

実施形態 4 は、実施形態 1 および実施形態 2 で説明した音声入力装置を備える音声情報記録装置である。本実施形態における音声入力装置の構成は、実施形態 1 および実施形態 2 と同様であるため、共通する各部の詳細な説明は省略する。

図 24 は本実施形態の構成を示す図である。音声情報記録装置 400 は、音声入力装置

10

20

30

40

50

１００、音声認識部３０１、記録部４０１を備える。本実施形態の音声入力装置１００は、ユーザ検出部１０２において、検出されたユーザが誰であるかの認識を行う。これは、人物画像と人物名とをあらかじめ登録しておくことで実現できる。音声入力装置１００は、ユーザ情報と音声情報とを関連付けて音声認識部３０１に伝達する。音声認識結果はテキストデータとしてユーザ情報と関連付けて記録部４０１に伝達される。

【００７１】

記録部４０１では、音声認識結果とユーザ情報とをデータとして記録する。これにより、発話したユーザの情報と、発話内容とを関連付けて記録できるようになる。ここで、音声入力装置１００は、ノイズを低減した音声情報を取得できるため、音声認識部３０１での認識率が向上し、ユーザの発話内容を精度良く記録することが可能になる。例えば、ユーザ情報と音声認識結果とを合わせて記録することにより、会議で誰がどんな発言をしたかなど、自動で会議の議事録を作成することも可能となる。

10

【００７２】

以上説明したように、本発明の音声入力装置は、画像情報を取得する撮像素子と、音声情報を取得する複数のマイクロフォンとを備えた音声入力装置であって、撮像素子が取得した画像情報からユーザを検出するユーザ検出部と、マイクロフォンが取得した音声情報から音源定位を行い、音源が存在する特定方向から取得した音声情報を出力音声情報とするユーザ音声取得部とを備え、ユーザ音声取得部は、特定方向の基準とするマイクロフォンの位置を、ユーザ検出部で検出されるユーザの位置に基づいて変化させる。これにより、ノイズである他の音源から発せられる音声情報を低減し、ユーザから発話される音声情報を取得することができる。

20

【００７３】

また、本発明の音声入力装置は、ユーザ音声取得部が、撮像素子を基準としたユーザの方向と、複数のマイクロフォンの中心を基準としたユーザの方向とがなす角より、撮像素子を基準としたユーザの方向と、特定方向とがなす角が大きくなるように、特定方向の基準とするマイクロフォンの位置を設定する。これにより、ユーザ背後領域に他の人物が存在しても、音声情報として取得してしまう可能性を低減することができる。

【００７４】

また、本発明の音声入力装置は、ユーザ音声取得部が、ユーザ検出部で検出されたユーザの人数によって、特定方向の基準とするマイクロフォンの位置を変化させる。これによりユーザの人数に応じて最適な音声情報取得制御を行うことができる。

30

【００７５】

また、本発明の音声入力装置は、ユーザ音声取得部が、ユーザ検出部で複数のユーザが検出された場合、撮像素子を基準としたユーザの方向と、複数のマイクロフォンの中心を基準としたユーザの方向とがなす角より、撮像素子を基準としたユーザの方向と、特定方向とがなす角が小さくなるように、特定方向の基準とするマイクロフォンの位置を設定する。これにより、複数のユーザが検出されたときにユーザ背後領域に他の人物が存在しても、音声情報として取得してしまう可能性を低減することができる。

【００７６】

また、本発明の画像表示装置は、上記の音声入力装置と、音声入力装置が出力する音声情報を認識する音声認識部と、音声認識部で認識された結果に基づいて所定の制御を行う制御部と、を備える。これにより、ノイズである他の音源からの発せられる音声情報を低減し、ユーザから発話される音声情報を取得し、これに基づいて制御を行う画像表示装置が得られる。

40

【００７７】

また、本発明の音声入力装置を備える音声情報記録装置は、上記の音声入力装置と、音声入力装置が出力する音声情報を認識する音声認識部とを備え、音声入力装置のユーザ検出部で検出されたユーザ情報と、音声入力装置の音声認識部で認識された結果を関連付けて記録する記録部を備える。これにより、ノイズである他の音源からの発せられる音声情報を低減し、ユーザから発話される音声情報を取得し、音声情報により認識された結果と

50

ユーザ情報とを関連付けて記憶できる音声情報記録装置が得られる。

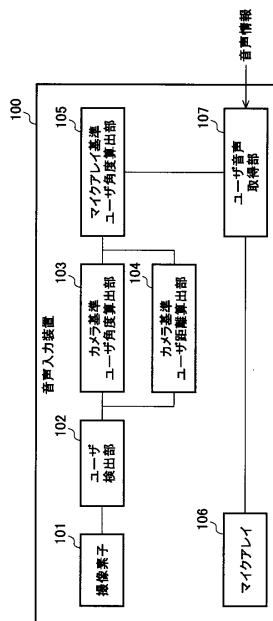
【符号の説明】

【 0 0 7 8 】

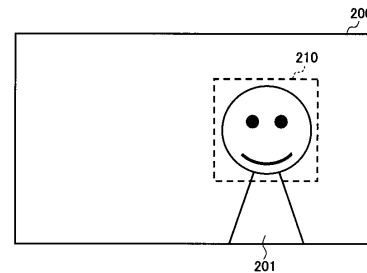
1 0 ... カメラ基準軸、1 1 ... カメラ基準ユーザ方向、1 2 ... ユーザ撮影範囲、2 0 ... マイクアレイ基準軸、2 1 ... マイクアレイ基準ユーザ方向、2 2 ... ユーザ音声取得範囲、2 3 ... 音声取得角度範囲、1 0 0 ... 音声入力装置、1 0 1 ... 撮像素子、1 0 2 ... ユーザ検出部、1 0 3 ... カメラ基準ユーザ角度算出部、1 0 4 ... カメラ基準ユーザ距離算出部、1 0 5 ... マイクアレイ基準ユーザ角度算出部、1 0 6 ... マイクアレイ、1 0 7 ... ユーザ音声取得部、1 0 8 ... 撮像素子、1 0 9 ... 距離算出部、1 1 0 ... 測距素子、2 0 0 ... 撮影画像、2 0 1 ... ユーザ、2 0 2 ... ユーザ、2 1 0 ... 顔検出領域、3 0 0 ... 画像表示装置、3 0 1 ... 音声認識部、3 0 2 ... 制御部、3 0 3 ... 画像表示部、3 0 4 ... 音声出力部、4 0 1 ... 記録部。

10

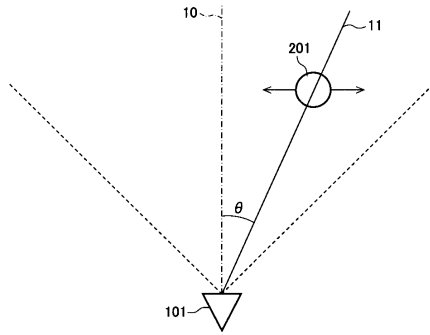
【図 1】



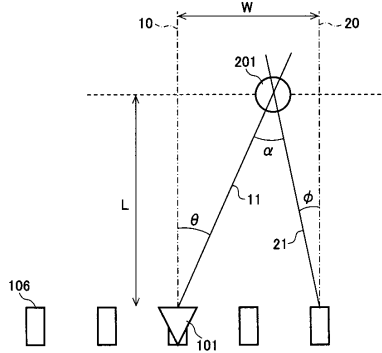
【図 2】



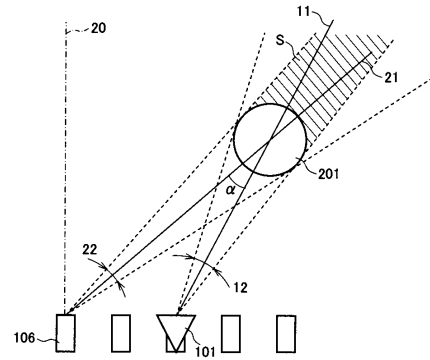
【図 3】



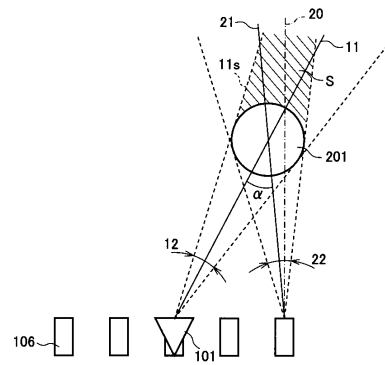
【図 4】



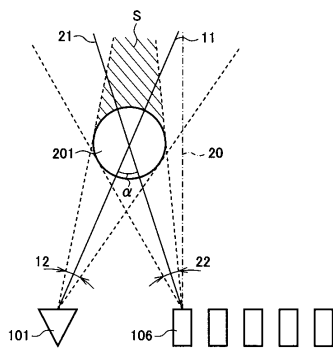
【図 5】



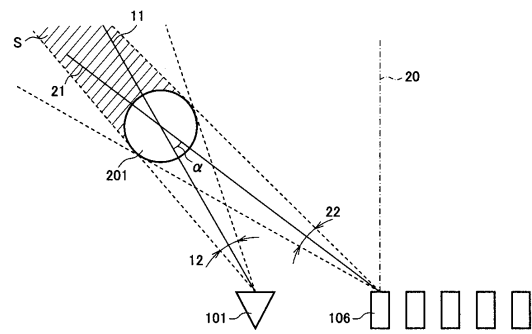
【図 6】



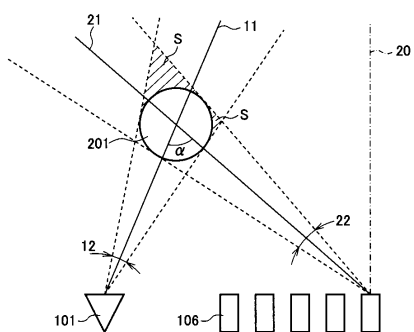
【図 7】



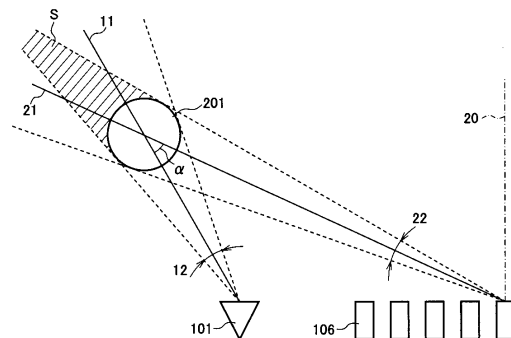
【図 9】



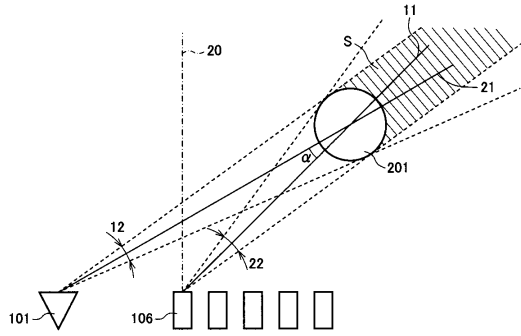
【図 8】



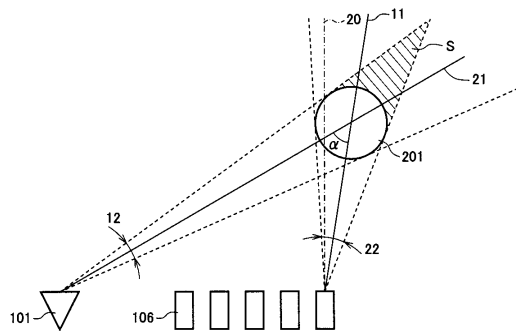
【図 10】



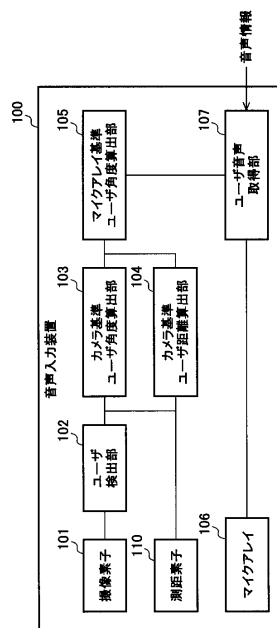
【 図 1 1 】



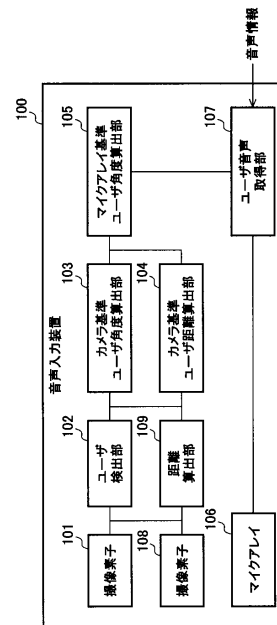
【 図 1 2 】



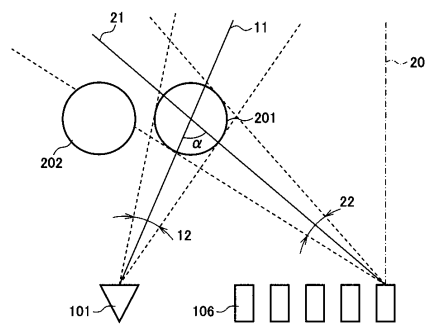
【 図 1 4 】



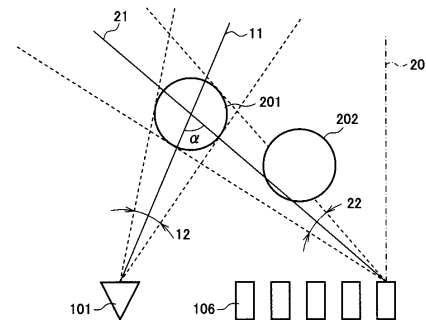
【 図 1 3 】



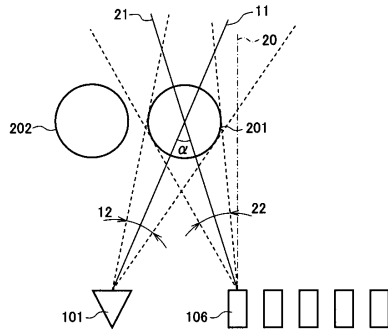
【 図 1 5 】



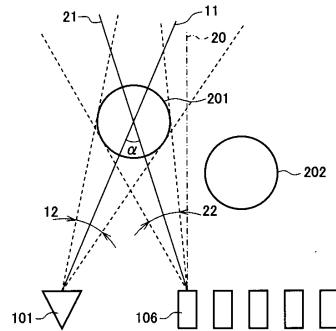
【 図 1 6 】



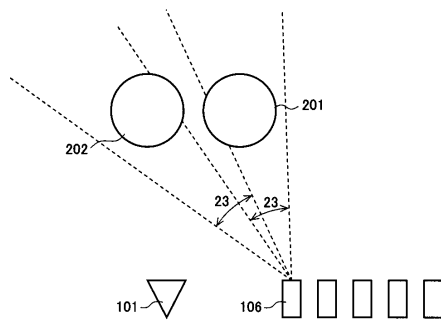
【図 17】



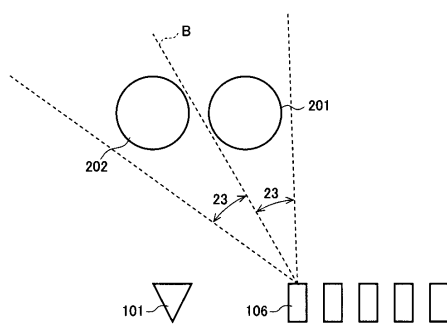
【図 18】



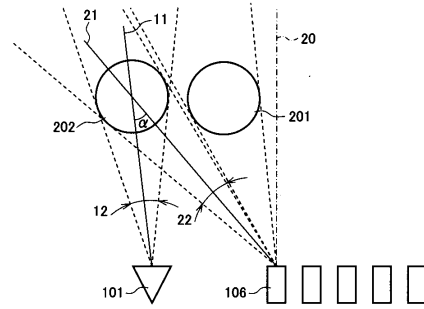
【図 21】



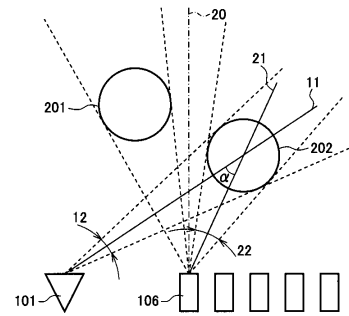
【図 22】



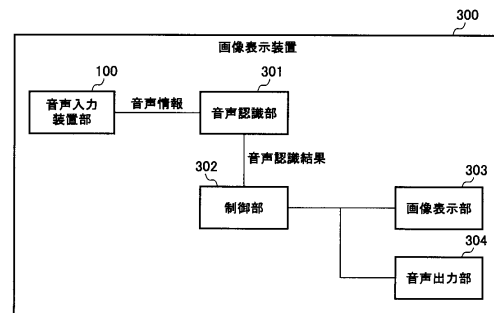
【図 19】



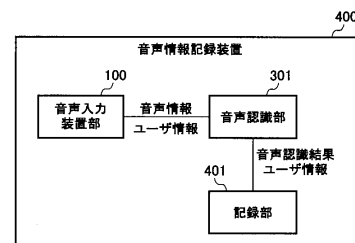
【図 20】



【図 23】



【図 24】



フロントページの続き

(51)Int.Cl. F I
H 0 4 N 5/232 (2006.01) H 0 4 N 5/232

(56)参考文献 特開 2 0 1 2 - 1 8 6 5 5 1 (J P , A)
特開 2 0 0 7 - 3 0 2 1 5 5 (J P , A)
特開 2 0 1 1 - 0 7 1 7 0 2 (J P , A)
特開 2 0 0 5 - 1 5 9 7 3 1 (J P , A)
特開 2 0 1 1 - 1 9 3 1 7 6 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)

G 1 0 L 1 3 / 0 0 - 1 3 / 1 0
G 1 0 L 1 5 / 0 0 - 1 7 / 2 6
G 1 0 L 1 9 / 0 0 - 1 9 / 2 6
G 1 0 L 2 1 / 0 0 - 2 1 / 1 8
G 1 0 L 2 5 / 0 0 - 2 5 / 9 3
G 1 0 L 9 9 / 0 0
H 0 4 N 5 / 2 2 2 - 5 / 2 5 7
H 0 4 R 1 / 2 0 - 1 / 4 0
H 0 4 R 3 / 0 0 - 3 / 1 4