

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7233035号

(P7233035)

(45)発行日 令和5年3月6日(2023.3.6)

(24)登録日 令和5年2月24日(2023.2.24)

(51)国際特許分類

F I

H 0 4 R 3/00 (2006.01)

H 0 4 R 3/00 3 2 0

G 1 0 K 11/34 (2006.01)

G 1 0 K 11/34 1 0 0

H 0 4 N 7/15 (2006.01)

H 0 4 N 7/15

H 0 4 R 1/40 (2006.01)

H 0 4 R 1/40 3 2 0 Z

請求項の数 11 (全25頁)

(21)出願番号 特願2019-539014(P2019-539014)

(86)(22)出願日 平成30年7月2日(2018.7.2)

(86)国際出願番号 PCT/JP2018/024996

(87)国際公開番号 WO2019/044157

(87)国際公開日 平成31年3月7日(2019.3.7)

審査請求日 令和3年6月8日(2021.6.8)

(31)優先権主張番号 特願2017-165717(P2017-165717)

(32)優先日 平成29年8月30日(2017.8.30)

(33)優先権主張国・地域又は機関

日本国(JP)

(73)特許権者 314012076

パナソニックIPマネジメント株式会社

大阪府大阪市中央区城見2丁目1番61

号

(74)代理人 100106518

弁理士 松谷 道子

(74)代理人 100132241

弁理士 岡部 博史

(72)発明者

廣瀬 良文

大阪府門真市大字門真1006番地 パ

ナソニック株式会社内

(72)発明者

中西 雅浩

大阪府門真市大字門真1006番地 パ

ナソニック株式会社内

(72)発明者

足立 祐介

最終頁に続く

(54)【発明の名称】 収音装置、収音方法、及びプログラム

(57)【特許請求の範囲】

【請求項1】

物体から出力される目的音を収音するための収音装置であって、
カメラによって生成された画像データの入力を受ける第1の入力部と、
マイクアレイから出力される前記目的音と非目的音を含む音響信号の入力を受ける第2
の入力部と、

前記目的音を収音する方向を決定する制御部と、

を有し、

前記制御部は、

前記画像データ内の部分領域が前記物体の特徴と一致する確率に基づいて、前記物体の
水平方向又は垂直方向の少なくともいずれか一方における位置を検出し、検出した前記物
体の位置を示す情報である物体位置情報を出力する検出部と、

前記音響信号内の所定の位置に前記目的音または非目的音の音源が存在する確率に基づ
いて、前記音源の水平方向又は垂直方向の少なくともいずれか一方における位置を推定し
、推定した前記音源の位置を示す情報である音源位置情報を出力する推定部と、

前記目的音の情報である第1情報と前記音響信号に含まれる音との第1尤度に応じて、
前記音源位置情報の重みを設定する重み付け部と、

前記物体位置情報と前記音源位置情報と前記音源位置情報の重みに基づいて、前記目
的音を収音する方向を決定する決定部と、

を含む、収音装置。

【請求項 2】

前記第 1 情報を収納する記憶部を含む、
請求項 1 に記載の收音装置。

【請求項 3】

前記記憶部は、前記非目的音の情報である第 2 情報を収納し、
前記重み付け部は、前記第 1 尤度と前記非目的音の前記第 2 情報による尤度との比に基づいて、前記音源位置情報の重みを設定する、
請求項 2 に記載の收音装置。

【請求項 4】

前記重み付け部は、前記画像データから前記物体の位置を検出する精度が変化する前記カメラの撮影状況に応じて、前記物体位置情報の重みを設定し、
前記決定部は、前記物体位置情報と前記物体位置情報の重みと前記音源位置情報と前記音源位置情報の重みに基づいて、前記目的音を收音する方向を決定する、
請求項 1 ～ 3 のいずれかに記載の收音装置。

【請求項 5】

前記重み付け部は、
前記物体が前記画像データに含まれている確からしさを示す画像確度に基づいて前記物体位置情報の重みを設定する、
請求項 4 に記載の收音装置。

【請求項 6】

前記重み付け部は、前記画像データの平均輝度又は輝度の分散に基づいて、前記画像確度を算出する、
請求項 5 に記載の收音装置。

【請求項 7】

前記検出部は、前記画像データを複数の領域に分割し、
前記重み付け部は、前記複数の領域のうち、検出した前記物体の位置に対応する一の領域の平均輝度又は輝度の分散に基づいて、前記画像確度を算出する、
請求項 5 に記載の收音装置。

【請求項 8】

周囲の明るさを検知する輝度センサ、前記物体までの距離を検知する距離センサ、又は加速度センサのうちの少なくとも 1 つを含むセンサ部をさらに有し、
前記重み付け部は、前記センサ部の出力に基づいて、前記物体位置情報の重みを設定する、
請求項 1 に記載の收音装置。

【請求項 9】

前記カメラ又は前記マイクアレイのうちの少なくとも一方を備える、
請求項 1 に記載の收音装置。

【請求項 10】

物体から出力される目的音を收音するための收音方法であって、
カメラによって生成された画像データの入力を受けるステップと、
マイクアレイから出力される前記目的音と非目的音を含む音響信号の入力を受けるステップと、
前記画像データ内の部分領域が前記物体の特徴と一致する確率に基づいて、前記物体の水平方向又は垂直方向の少なくともいずれか一方における位置を検出し、検出した前記物体の位置を示す情報である物体位置情報を出力するステップと、
前記音響信号内の所定の位置に前記目的音または非目的音の音源が存在する確率に基づいて、前記音源の水平方向又は垂直方向の少なくともいずれか一方における位置を推定し、推定した前記音源の位置を示す情報である音源位置情報を出力するステップと、
前記目的音の情報である第 1 情報と前記音響信号に含まれる音との第 1 尤度に応じて、前記音源位置情報の重みを設定するステップと、

前記物体位置情報と前記音源位置情報と前記音源位置情報の重みとに基づいて、前記目的音を收音する方向を決定するステップと、

を含む、收音方法。

【請求項 11】

コンピュータに請求項 10 に記載の收音方法を実行させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、目的音を收音する收音装置、收音方法、及びプログラムに関する。

【背景技術】

10

【0002】

特許文献 1 は、一人以上の参加者による談話を記録したデータから、どの参加者がいつ発話したかを推定するための推定装置を開示している。この推定装置は、マイクロホンから出力される音響信号から得られる情報と、カメラから出力される映像信号から得られる情報とを統合することによって、談話参加者が発話した確率を算出している。これにより、談話参加者が発話のない状況で移動した場合であっても、談話参加者の位置を追跡することを可能にしている。

【先行技術文献】

【特許文献】

【0003】

20

【文献】特許第 5 2 1 5 8 2 6 号公報

【発明の概要】

【0004】

本開示は、目的音の收音の精度を向上させる收音装置、收音方法、及びプログラムを提供する。

【0005】

本開示の一態様の收音装置は、音源となる物体から出力される目的音を收音するための收音装置であって、第 1 の入力部と、第 2 の入力部と、制御部とを有する。第 1 の入力部は、カメラによって生成された画像データの入力を受ける。第 2 の入力部は、マイクアレイから出力される音響信号の入力を受ける。制御部は、目的音を收音する方向を決定する。制御部は、検出部と、推定部と、重み付け部と、決定部とを含む。検出部は、画像データに基づいて、物体の水平方向又は垂直方向の少なくともいずれか一方における位置を検出し、検出した物体の位置を示す情報である物体位置情報を出力する。推定部は、音響信号に基づいて、音源の水平方向又は垂直方向の少なくともいずれか一方における位置を推定し、推定した音源の位置を示す情報である音源位置情報を出力する。重み付け部は、カメラの撮影状況に応じて物体位置情報の重みを設定する。決定部は、物体位置情報と物体位置情報の重みと音源位置情報とに基づいて、目的音を收音する方向を決定する。

30

【0006】

これらの概括的かつ特定の態様は、システム、方法、及びコンピュータプログラム、並びに、それらの組み合わせにより、実現されてもよい。

40

【0007】

本開示の收音装置、收音方法、及びプログラムによれば、カメラの撮影状況及びマイクアレイの收音状況に応じて收音方向が決定されるため、目的音の收音の精度が向上する。

【図面の簡単な説明】

【0008】

【図 1】図 1 は、第 1 実施形態の收音装置の構成を示すブロック図である。

【図 2】図 2 は、第 1 実施形態の制御部の機能を示すブロック図である。

【図 3 A】図 3 A は、マイクアレイのみを使用した場合の收音方向を説明するための図である。

【図 3 B】図 3 B は、マイクアレイとカメラを使用した場合の收音方向を説明するための

50

図である。

【図 4】図 4 は、第 1 実施形態における收音方法を示すフローチャートである。

【図 5】図 5 は、顔位置の検出単位の領域を説明するための図である。

【図 6】図 6 は、顔の識別を説明するための図である。

【図 7】図 7 は、顔の確率を説明するための図である。

【図 8】図 8 は、画像確度の判定を示すフローチャートである。

【図 9】図 9 は、音源位置の推定を説明するための図である。

【図 10】図 10 は、音源の確率を説明するための図である。

【図 11】図 11 は、音響確度の判定を示すフローチャートである。

【図 12】図 12 は、画像確度と音響確度の重みを示す図である。

10

【図 13 A】図 13 A は、顔の確率の一例を示す図である。

【図 13 B】図 13 B は、音源の確率の一例を示す図である。

【図 13 C】図 13 C は、図 13 A と図 13 B とにより算出される收音方向を示す図である。

【図 14 A】図 14 A は、顔の確率の他の例を示す図である。

【図 14 B】図 14 B は、音源の確率の他の例を示す図である。

【図 14 C】図 14 C は、図 14 A と図 14 B とにより算出される收音方向を示す図である。

【図 15】図 15 は、第 2 実施形態の收音装置の構成を示すブロック図である。

【図 16】図 16 は、第 3 実施形態における制御部の機能を示すブロック図である。

20

【図 17】図 17 は、第 4 実施形態における顔位置の検出単位の領域を説明するための図である。

【図 18】図 18 は、他の実施形態の制御部の機能を示すブロック図である。

【図 19】図 19 は、他の実施形態の制御部の機能を示すブロック図である。

【発明を実施するための形態】

【0009】

(本開示の基礎となった知見)

特許文献 1 の推定装置において、談話参加者が発話した確率を精度良く算出するためには、カメラから出力される映像信号から得られる情報とマイクロホンから出力される音響信号から得られる情報とが正確である必要がある。よって、映像信号から得られる情報と音響信号から得られる情報とが正確でない場合には、談話参加者が発話した確率を精度良く算出することができない。

30

【0010】

本開示は、カメラから出力される映像信号から得られる情報及びマイクアレイから出力される音響信号から得られる情報の精度がよくない場合であっても、目的音の收音の精度を向上させる收音装置を提供する。具体的には、本開示の收音装置は、映像信号から得られる情報に対してカメラの撮影状況に応じた重みを設定し、且つ音響信号から得られる情報に対してマイクアレイの受音状況に応じた重みを設定する。これによって、本開示の收音装置は、精度良く目的音の收音方向を決定する。

【0011】

40

(第 1 実施形態)

以下、第 1 実施形態について、図面を参照しながら説明する。本実施形態では、人物の音声を目的音として收音する例について説明する。本実施形態では、カメラの撮影状況に応じた重みは、人物の顔が画像データに含まれている確からしさを示す確度に基づいて、設定される。また、マイクアレイの受音状況に応じた重みは、目的音が音響信号に含まれている確からしさを示す確度に基づいて、設定される。「確度」とは、確実さの度合いのことである。

【0012】

1. 收音装置の構成

図 1 は、本開示の第 1 実施形態の收音装置の構成を示している。收音装置 1 は、カメラ

50

１０、マイクアレイ２０、制御部３０、記憶部４０、入出力インタフェース部５０、及びバス６０を備える。收音装置１は、例えば、会議中の人の音声を收音する。收音装置１は、一例では、カメラ１０、マイクアレイ２０、制御部３０、記憶部４０、入出力インタフェース部５０、及びバス６０が一体化された専用の收音機である。

【００１３】

なお、別の例では、收音装置１は、カメラ１０とマイクアレイ２０のいずれか一方又は両方を内蔵していなくてもよい。この場合、收音装置１は、外付けのカメラ１０又はマイクアレイ２０と電氣的に接続される。例えば、收音装置１は、カメラ１０を備えたスマートフォンなどの電子機器であって、マイクアレイ２０を備えた外部機器と電氣的及び機械的に接続されてもよい。

【００１４】

カメラ１０は、ＣＣＤイメージセンサ、ＣＭＯＳイメージセンサ、又はＮＭＯＳイメージセンサなどを備える。カメラ１０は、映像信号である画像データを生成して、出力する。

【００１５】

マイクアレイ２０は、複数のマイクロホンを備える。マイクアレイ２０は、音波を受信して電気信号である音響信号に変換して出力する。

【００１６】

制御部３０は、カメラ１０から得られた画像データと、マイクアレイ２０から得られた音響信号に基づいて、收音方向を決定する。制御部３０は、決定した收音方向に基づいて、音響信号から目的音を取り出す。制御部３０は、半導体素子などで実現可能である。制御部３０は、例えば、マイコン、ＣＰＵ、ＭＰＵ、ＤＳＰ、ＦＰＧＡ、ＡＳＩＣで構成することができる。

【００１７】

記憶部４０は、カメラ１０から得られた画像データ及びマイクアレイ２０から得られた音響信号を格納する。記憶部４０は、例えば、ハードディスク（ＨＤＤ）、ＳＳＤ、ＲＡＭ、ＤＲＡＭ、強誘電体メモリ、フラッシュメモリ、磁気ディスク、又はこれらの組み合わせによって実現できる。

【００１８】

入出力インタフェース部５０は、所定の通信規格（例えばＬＡＮ、ＷｉＦｉ、Ｂｌｕｅｔｏｏｔｈ（登録商標）、ＵＳＢ、ＨＤＭＩ（登録商標））に準拠して外部機器との通信を行う回路を含む。

【００１９】

バス６０は、カメラ１０、マイクアレイ２０、制御部３０、記憶部４０、及び入出力インタフェース部５０を電氣的に接続する信号線である。

【００２０】

制御部３０が画像データを記憶部４０から取り出すときは、制御部３０が画像データの入力部に相当する。制御部３０が音響信号を記憶部４０から取り出すときは、制御部３０が音響信号の入力部に相当する。入出力インタフェース部５０が收音装置１に外付けされたカメラ１０から画像データを入力するときは、入出力インタフェース部５０が画像データの入力部に相当する。入出力インタフェース部５０が收音装置１に外付けされたマイクアレイ２０から音響信号を入力するときは、入出力インタフェース部５０が音響信号の入力部に相当する。

【００２１】

図２は、制御部３０の機能を示している。制御部３０の機能は、ハードウェアのみで構成してもよいし、ハードウェアとソフトウェアとを組み合わせることにより実現してもよい。

【００２２】

制御部３０は、物体位置検出部３１、音源位置推定部３２、重み付け部３００、收音方向決定部３６、及びビームフォーム部３７を含む。

【００２３】

10

20

30

40

50

物体位置検出部 31 は、カメラ 10 によって生成された画像データ v から物体の位置を検出する。本実施形態では、検出対象となる物体は人の顔である。物体位置検出部 31 は、検出した物体の位置を示す情報である物体位置情報を出力する。具体的には、物体位置情報は、動画 1 フレーム分又は静止画 1 枚分に相当する画像データ v 内の複数の領域のそれぞれの画像が顔である確率 $P(\theta, \phi | v)$ を示す情報である。画像データ v 内の各領域の座標系の位置は、カメラ 10 の設計情報に基づいて、カメラ 10 の画角の水平角 θ 及び垂直角 ϕ と対応付けられる。

【0024】

音源位置推定部 32 は、マイクアレイ 20 から得られる音響信号 s から、音源の位置を推定する。本実施形態では、音源は、音声を発する人である。音源位置推定部 32 は、推定した音源の位置を示す情報である音源位置情報を出力する。具体的には、音源位置情報は、水平角 θ 及び垂直角 ϕ によって特定される位置に音源が存在する確率 $P(\theta, \phi | s)$ を示す情報である。

10

【0025】

重み付け部 300 は、画像データ v に基づいて、物体位置情報である確率 $P(\theta, \phi | v)$ に重み付けを行う。また、重み付け部 300 は、音響信号 s に基づいて、音源位置情報である確率 $P(\theta, \phi | s)$ に重み付けを行う。

【0026】

重み付け部 300 は、画像確度判定部 33、音響確度判定部 34、及び重み設定部 35 を含む。画像確度判定部 33 は、画像データ v に基づいて画像確度 CM_v を判定する。画像確度 CM_v は、顔が画像データ v に含まれている確からしさを示す確度である。すなわち、画像確度 CM_v は、物体位置情報である確率 $P(\theta, \phi | v)$ の信頼性を表す。音響確度判定部 34 は、音響信号 s に基づいて音響確度 CM_s を判定する。音響確度 CM_s は、目的音である音声が音響信号に含まれている確からしさを示す確度である。すなわち、音響確度 CM_s は、音源位置情報である確率 $P(\theta, \phi | s)$ の信頼性を表す。

20

【0027】

重み設定部 35 は、画像確度 CM_v に基づいて、物体位置情報の重み W_v を設定する。画像確度 CM_v に基づいて設定された物体位置情報の重み W_v は、カメラ 10 の撮影状況に応じた値になる。重み設定部 35 は、音響確度 CM_s に基づいて、音源位置情報の重み W_s を設定する。音響確度 CM_s に基づいて設定された音源位置情報の重み W_s は、マイクアレイ 20 の受信状況に応じた値になる。

30

【0028】

收音方向決定部 36 は、物体位置情報である確率 $P(\theta, \phi | v)$ 及びその重み W_v と、音源位置情報である確率 $P(\theta, \phi | s)$ 及びその重み W_s とに基づいて、收音方向を決定する。

【0029】

ビームフォーム部 37 は、決定された收音方向に基づいて、音響信号 s から目的音を取り出す。これにより、雑音が低減された、クリアな音声を收音できる。

【0030】

図 3A は、マイクアレイ 20 のみを使用した場合の收音方向を模式的に示している。図 3B は、マイクアレイ 20 とカメラ 10 を使用した場合の收音方向を模式的に示している。図 3A に示すように、マイクアレイ 20 から出力される音響信号のみによって收音方向を決定した場合、周囲の雑音（スピーカの音）が人の音声よりも大きいときに、音源の方向として雑音の方向が検出される。この場合、大きな雑音を收音してしまい、人の音声をクリアに收音することができない。しかし、図 3B に示すように、カメラ 10 を併用した場合、カメラ 10 の画像データに含まれる人の顔の位置を検出することによって、人がいる方向を音源の方向として特定できる。

40

【0031】

一方、カメラ 10 の撮影状況に応じて、画像データ v から得られる顔位置の検出の精度は変化する。例えば、暗い環境で撮影されて生成された画像データ v では、顔の位置を誤

50

検出する場合がある。よって、本実施形態においては、顔の確率 $P(\theta, \phi | v)$ に対して、カメラ 10 の撮影状況に応じた重み付けを行う。また、マイクアレイ 20 の受音状況に応じて、音響信号 s から得られる音源位置の推定の精度も変わる。例えば、雑音が大きすぎると、音響信号 s から目的音の音源の位置を正確に推定できない場合がある。よって、本実施形態においては、音源の確率 $P(\theta, \phi | s)$ に対して、マイクアレイ 20 の受音状況に応じた重み付けを行う。そして、重み付けされた確率 $P(\theta, \phi | v)$ 、 $P(\theta, \phi | s)$ に基づいて、收音方向を決定する。

【0032】

2. 收音装置の動作

図 4 は、制御部 30 による收音動作を示している。

【0033】

物体位置検出部 31 は、カメラ 10 によって生成された画像データ v に基づいて、顔の位置を検出する (S1)。具体的には、物体位置検出部 31 は、画像データ v 内において水平角 θ 及び垂直角 ϕ で特定される位置の画像が顔である確率 $P(\theta, \phi | v)$ を算出する。顔の位置の検出方法は、任意である。一例として、顔の位置の検出は、動画 1 フレーム分又は静止画 1 枚分に相当する画像データ v を複数の領域に分割して、各領域が顔の特徴と一致しているか否かを判定することによって行う (「Rapid Object Detection using a Boosted Cascade of Simple Features」ACCEPTED CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION 2001 を参照)。以下、この顔検出の方法について説明する。

【0034】

図 5 は、画像データ v における顔位置の検出単位の領域 $r(\theta, \phi)$ を示している。物体位置検出部 31 は、画像データ v を複数の領域 $r(\theta, \phi)$ に分割し、各領域 $r(\theta, \phi)$ の画像が顔であるか否かを判定する。なお、図 5 では、画像データ v を格子状に分割し、それぞれの軸を θ 、 ϕ に対応付けたが、全周カメラを使用した場合は、円周状に θ 、 ϕ を対応付けるようにしても良い。

【0035】

図 6 は、顔の識別の一例を示している。物体位置検出部 31 は、例えば、 N 個の弱識別器 310 (弱識別器 310 (1) ~ 310 (N)) を備える。弱識別器 310 (1) ~ 310 (N) は、それぞれ、顔の特徴を示す情報を有する。顔の特徴を示す情報は、 N 個の弱識別器 310 のそれぞれにおいて、異なる。物体位置検出部 31 は、領域 $r(\theta, \phi)$ が顔であると判定した回数 $C(r(\theta, \phi))$ を計算する。具体的には、物体位置検出部 31 は、最初に、一つ目の弱識別器 310 (1) によって、領域 $r(\theta, \phi)$ が顔であるか否かを判定する。弱識別器 310 (1) が、領域 $r(\theta, \phi)$ は顔でないと判定した場合、「 $C(r(\theta, \phi)) = 0$ 」となる。一つ目の弱識別器 310 (1) が、領域 $r(\theta, \phi)$ は顔であると判定した場合は、二つ目の弱識別器 310 (2) が、一つ目の弱識別器 310 (1) とは異なる顔の特徴の情報を用いて、領域 $r(\theta, \phi)$ が顔であるか否かを判定する。二つ目の弱識別器 310 (2) が、領域 $r(\theta, \phi)$ が顔であると判定した場合、三つ目の弱識別器 310 (3) が、領域 $r(\theta, \phi)$ が顔であるか否かを判定する。このように、動画 1 フレーム分又は静止画 1 枚分に相当する画像データ v に対して、領域 $r(\theta, \phi)$ 毎に、 N 個の弱識別器 310 を用いて顔であるか否かを判定する。例えば、 N 個の弱識別器 310 の全てが領域 $r(\theta, \phi)$ は顔であると判定した場合、顔であると判定された回数は「 $C(r(\theta, \phi)) = N$ 」となる。

【0036】

顔を検出するときの領域 $r(\theta, \phi)$ の大きさは、一定であってもよいし、可変であってもよい。例えば、顔を検出するときの領域 $r(\theta, \phi)$ の大きさは、動画の 1 フレーム分又は静止画 1 枚分の画像データ v 毎に変わってもよい。

【0037】

物体位置検出部 31 は、画像データ v 内の領域 $r(\theta, \phi)$ の全てについて、顔であるか否かの判定を行う。そして、物体位置検出部 31 は、下記式 (1) によって、画像デー

10

20

30

40

50

タ v 内において水平角 θ 及び垂直角 ϕ で特定される位置の画像が顔である確率 $P(\theta, \phi | v)$ を算出する。

【0038】

【数1】

$$P(\theta, \phi | v) = \frac{1}{N} C(r(\theta, \phi)) \quad \cdots(1)$$

【0039】

なお、水平角 θ のみで顔の位置を検出できる場合は、下記式(2)によって、顔の確率 $P(\theta | v)$ を算出してもよい。

【0040】

【数2】

$$P(\theta | v) = \sum_{\phi} \frac{1}{N} C(r(\theta, \phi)) \quad \cdots(2)$$

【0041】

図7は、画像データ v 内における水平角 θ で特定される位置の画像が顔である確率 $P(\theta | v)$ を例示している。確率 $P(\theta | v)$ が高いほど、その水平角 θ に顔がある可能性が高いことを表している。

【0042】

図4に示すように、物体位置検出部31によって顔の位置が検出されると、画像確度判定部33は、顔が画像データ v に含まれている確からしさを示す画像確度 CM_v を判定する(S2)。具体的には、画像確度判定部33は、カメラ10によって生成された画像データ v に基づいて、画像確度 CM_v を設定する。なお、画像確度 CM_v の判定(S2)は、顔の位置を検出する(S1)よりも前に行ってもよい。

【0043】

図8は、画像確度 CM_v の判定方法(S2の詳細)の一例を示している。図8の例では、画像確度判定部33は、画像データ v の平均輝度 Y_{ave} に基づいて、画像確度 CM_v を判定する。画像確度判定部33は、まず、画像データ v の平均輝度 Y_{ave} を算出する(S201)。それから、画像確度判定部33は、平均輝度 Y_{ave} を推奨輝度($Y_{min_base} \sim Y_{max_base}$)と比較する(S202)。推奨輝度は、最小推奨輝度(Y_{min_base})から最大推奨輝度(Y_{max_base})までの範囲を有する。推奨輝度を示す情報は、予め記憶部40に格納されている。平均輝度 Y_{ave} が最小推奨輝度よりも低ければ(S203でYes)、画像確度判定部33は、画像確度 CM_v を「 Y_{ave} / Y_{min_base} 」に設定する(S204)。平均輝度 Y_{ave} が最大推奨輝度よりも高ければ(S205でYes)、画像確度判定部33は、画像確度 CM_v を「 Y_{max_base} / Y_{ave} 」に設定する(S206)。平均輝度 Y_{ave} が推奨輝度の範囲内であれば(S205でNo)、画像確度判定部33は、画像確度 CM_v を「1」に設定する(S207)。平均輝度 Y_{ave} が最小推奨輝度 Y_{min_base} よりも低かったり、最大推奨輝度 Y_{max_base} よりも高かったりすれば、顔を誤検出する場合がある。よって、平均輝度 Y_{ave} が推奨輝度の範囲内のときは画像確度 CM_v を最大値「1」に設定し、平均輝度 Y_{ave} が推奨輝度より高い、或いは、低いほど画像確度 CM_v を低くする。

【0044】

図4に示すように、音源位置推定部32は、マイクアレイ20から出力された音響信号 s に基づいて、音源の位置を推定する(S3)。具体的には、物体位置検出部31は、水

10

20

30

40

50

平角 θ 及び垂直角 ϕ で特定される位置に音源が存在する確率 $P(\theta, \phi | s)$ を算出する。音源の位置の推定方法は、任意である。例えば、音源の位置の推定は、CSP (Cross-Power Spectrum Phase Analysis) 法又は MUSIC (Multiple Signal Classification) 法を使用して行うことができる。以下、CSP法を使用して、音源の位置を推定する例について説明する。

【0045】

図9は、人の音声（音波）がマイクアレイ20のマイクロホン20_i及び20_jに到来する状態を模式的に示している。マイクロホン20_i及び20_j間の距離dに応じて、音波がマイクロホン20_i及び20_jに到来するときに時間差 τ が生じる。

【0046】

音源位置推定部32は、水平角 θ において、音源が存在する確率 $P(\theta | s)$ を、CSP係数（相関関数）を用いた下記式（3）により算出する。

【0047】

【数3】

$$P(\theta|s) = CSP(\tau) \quad \cdots(3)$$

【0048】

ここで、CSP係数は、下記式（4）によって求めることができる（電子情報通信学会論文誌 D-I I Vol. J83-D-II No. 8 pp. 1713-1721、「マイクロホンアレーを用いたCSP法に基づく複数音源位置推定」を参照）。式（4）において、nは時間、 $S_i(n)$ はマイクロホン20_iで受音した音響信号、 $S_j(n)$ はマイクロホン20_jで受音した音響信号を示している。式（4）において、DFTは、離散フーリエ変換を示す。また、*は共役複素数を示す。

【0049】

【数4】

$$CSP_{i,j}(\tau) = DFT^{-1} \left[\frac{DFT[S_i(n)] DFT[S_j(n)]^*}{|DFT[S_i(n)]| |DFT[S_j(n)]|} \right] \quad \cdots(4)$$

【0050】

時間差 τ は、音速c、マイクロホン20_i、20_j間の距離d、及びサンプリング周波数 F_s を用いて、下記式（5）によって表せる。

【0051】

【数5】

$$\tau = \frac{dF_s}{c} \cos(\theta) \quad \cdots(5)$$

【0052】

よって、下記式（6）に示すように、式（3）のCSP係数を式（5）によって時間軸から方向軸に変換することによって、水平角 θ において音源が存在する確率 $P(\theta | s)$ を算出できる。

【0053】

【数6】

10

20

30

40

50

$$P(\theta|s) = CSP \left(\frac{dF_s}{c} \cos(\theta) \right) \quad \dots(6)$$

【0054】

なお、上記で示した確率 $P(\theta|s)$ の算出方法では、2つのマイクロホン20i及び20jを使用した例を示しているが、2つ以上のマイクロホンを使用して、音源が存在する確率 $P(\theta|s)$ を算出してもよい。また、垂直角 ϕ における音源が存在する確率 $P(\phi|s)$ は、水平角 θ における確率 $P(\theta|s)$ と同様に、CSP係数と時間差 τ によって算出できる。また、確率 $P(\theta|s)$ 及び確率 $P(\phi|s)$ に基づいて、確率 $P(\theta, \phi|s)$ を算出できる。

10

【0055】

図10は、水平角 θ における音源が存在する確率 $P(\theta|s)$ を例示している。確率 $P(\theta|s)$ が高いほど、その水平角 θ に目的音の音源が存在する可能性が高いことを表している。

【0056】

図4に示すように、音源位置推定部32によって音源の位置が推定されると、音響確度判定部34は、音声の音響信号 s に含まれている確からしさを示す音響確度 CM_s を判定する(S4)。具体的には、音響確度判定部34は、マイクアレイ20から出力される音響信号 s に基づいて、音響確度 CM_s を設定する。なお、音響確度 CM_s の判定(S4)は、音源の位置を推定する(S3)よりも前に行ってもよい。

20

【0057】

図11は、音響確度 CM_s の判定方法(S4の詳細)の一例を示している。図11の例では、音響確度判定部34は、マイクアレイ20が受信した音の音声らしさを判定する。具体的には、音響確度判定部34は、人の音声GMM(Gaussian Mixture Model)と非音声GMMを用いて、音響確度 CM_s を算出する。音声GMMと非音声GMMは、予め学習して生成されたものである。音声GMMと非音声GMMを示す情報は記憶部40に格納されている。

【0058】

音響確度判定部34は、まず、音響信号 s における音声GMMによる尤度 L_v を算出する(S401)。次に、音響確度判定部34は、音響信号 s における非音声GMMによる尤度 L_n を算出する(S402)。それから、音響確度判定部34は、音響確度 CM_s を「 $CM_s = L_v / L_n$ 」に設定する(S403)。

30

【0059】

図4に示すように、重み設定部35は、画像確度 CM_v に基づいて顔の確率 $P(\theta, \phi|v)$ に重み W_v を設定し、且つ音響確度 CM_s に基づいて音源の確率 $P(\theta, \phi|s)$ に重み W_s を設定する(S5)。図12は、画像確度 CM_v 、音響確度 CM_s と、重み W_v 、 W_s との対応付けを示している。このように、重み W_v 、 W_s は、単調増加関数によって、確度 CM_v 、 CM_s に応じて決定される。

40

【0060】

図4に示すように、重み W_v 、 W_s が設定されると、收音方向決定部36は、收音方向を決定する(S6)。具体的には、收音方向決定部36は、顔の確率 $P(\theta, \phi|v)$ とその重み W_v 、及び音源の確率 $P(\theta, \phi|s)$ とその重み W_s を用いて、音源である人物がいる確率 $P(\theta, \phi)$ を下記式(7)によって算出する。それから、下記式(8)により、確率 $P(\theta, \phi)$ が最大となる水平角 θ 、垂直角 ϕ を收音方向として決定する。なお、確率 $P(\theta, \phi|s)$ の対数については下記式(9)によって表現できる。

【0061】

【数7】

50

$$P(\theta, \varphi) = P(\theta, \varphi|v)^{Wv} \times P(\theta, \varphi|s)^{Ws} \quad \cdots(7)$$

【 0 0 6 2 】

【数 8】

$$\hat{\theta}, \hat{\varphi} = \arg \max(P(\theta, \varphi)) \quad \cdots(8)$$

10

【 0 0 6 3 】

【数 9】

$$\log P(\theta, \varphi) = WvP(\theta, \varphi|v) + WsP(\theta, \varphi|s) \quad \cdots(9)$$

【 0 0 6 4 】

図 1 3 A ～ 図 1 3 C は、水平角 θ を決定するときの一例を示している。図 1 3 A は、顔の確率の一例を示している。図 1 3 B は、音源の確率の一例を示している。図 1 3 C は、図 1 3 A と図 1 3 B とにより算出される收音方向を示している。画像データ v の平均輝度 Y_{ave} が高い場合、画像確度 CM_v は高くなり、物体位置情報の重み W_v は大きくなる。また、音響信号 s において雑音が大きいと、音響確度 CM_s は小さくなり、音源位置情報の重み W_s は小さくなる。よって、平均輝度 Y_{ave} が高い場合及び雑音が大きき場合は、重みの大きい顔の確率 $P(\theta | v)$ が優先されて、水平角 θ が決定される。

20

【 0 0 6 5 】

図 1 4 A ～ 図 1 4 C は、水平角 θ を決定するときの他の例を示している。図 1 4 A は、顔の確率の他の例を示している。図 1 4 B は、音源の確率の他の例を示している。図 1 4 C は、図 1 4 A と図 1 4 B とにより算出される收音方向を示している。画像データ v の平均輝度 Y_{ave} が低い場合、画像確度 CM_v は低くなり、物体位置情報の重み W_v は小さくなる。また、音響信号 s において雑音が小さいと、音響確度 CM_s は大きくなり、音源位置情報の重み W_s は大きくなる。よって、平均輝度 Y_{ave} が低い場合及び雑音が小さい場合は、重みの大きい音源の確率 $P(\theta | s)$ が優先されて、水平角 θ が決定される。

30

【 0 0 6 6 】

図 4 に示すように、ビームフォーム部 3 7 は、決定された收音方向 θ, ϕ に基づいて、音響信号 s に含まれる目的音である音声を受音する (S 7)。

【 0 0 6 7 】

なお、図 4 のステップ S 1 ～ S 4 を実行する順序は、任意である。

【 0 0 6 8 】

3. 効果及び補足

40

本開示の收音装置 1 は、音源となる物体から出力される目的音を收音するための收音装置である。收音装置 1 は、カメラ 1 0 によって生成された画像データ v の入力を受ける第 1 の入力部 (制御部 3 0、または入出力インタフェース部 5 0) と、マイクアレイ 2 0 から出力される音響信号 s の入力を受ける第 2 の入力部 (制御部 3 0、または入出力インタフェース部 5 0) と、画像データ v 及び音響信号 s に基づいて收音する方向 θ, ϕ を決定する制御部 3 0 と、を有する。制御部 3 0 は、物体位置検出部 3 1 と、音源位置推定部 3 2 と、重み付け部 3 0 0 と、收音方向決定部 3 6 とを含む。物体位置検出部 3 1 は、画像データ v に基づいて、物体の水平方向又は垂直方向の少なくともいずれか一方における位置を検出し、検出した物体の位置を示す情報である物体位置情報 $P(\theta, \phi | v)$ を出力する。音源位置推定部 3 2 は、音響信号 s に基づいて、音源の水平方向又は垂直方向の少

50

なくともいずれか一方における位置を推定し、推定した音源の位置を示す情報である音源位置情報 $P(\theta, \phi | s)$ を出力する。重み付け部 300 は、物体位置情報に対してカメラ 10 の撮影状況に応じた重み W_v を設定し、且つ音源位置情報に対してマイクアレイ 20 の受音状況に応じた重み W_s を設定する。收音方向決定部 36 は、物体位置情報 $P(\theta, \phi | v)$ と物体位置情報の重み W_v と音源位置情報 $P(\theta, \phi | s)$ と音源位置情報の重み W_s とに基づいて、收音する方向 θ, ϕ を決定する。

【0069】

このように、カメラ 10 の撮影状況に応じた重み W_v とマイクアレイ 20 の受音状況に応じた重み W_s を使用して收音する方向 θ, ϕ を決定しているため、物体位置情報 $P(\theta, \phi | v)$ と音源位置情報 $P(\theta, \phi | s)$ の信頼性が高いほうの情報を優先させることができる。よって、物体位置情報 $P(\theta, \phi | v)$ 又は音源位置情報 $P(\theta, \phi | s)$ の精度がよくない場合であっても、精度良く目的音の收音方向を決定することができる。これにより、マイクアレイ 20 から出力された音響信号 s から、決定された收音方向の音を抽出することによって、目的音を精度良く收音することができる。

【0070】

具体的には、重み付け部 300 は、物体が画像データに含まれている確からしさを示す画像確度 CM_v に基づいて物体位置情報の重み W_v を設定し、且つ目的音が音響信号に含まれている確からしさを示す音響確度 CM_s に基づいて音源位置情報の重み W_s を設定する。画像確度 CM_v は、カメラ 10 の撮影状況に応じた値である。また、音響確度 CM_s は、マイクアレイ 20 の受音状況に応じた値である。よって、画像確度 CM_v 及び音響確度 CM_s に基づいて重み W_v, W_s を設定することによって、カメラ 10 の撮影状況及びマイクアレイ 20 の受音状況に応じた重み W_v, W_s を設定することができる。これにより、精度良く目的音の收音方向を決定することができる。

【0071】

(第 2 実施形態)

本実施形態の收音装置 1 は、物体位置情報の重み W_v をセンサの出力に基づいて設定する。

【0072】

図 15 は、第 2 実施形態の收音装置の構成を示すブロック図である。第 2 実施形態の收音装置 1 は、センサ部 70 を備える。センサ部 70 は、周囲の明るさを検知する輝度センサ、物体までの距離を検知する距離センサ、及び加速度センサのうちの少なくとも 1 つを含む。第 2 実施形態の收音装置 1 は、画像確度判定部 33 を備えない。重み設定部 35 は、センサ部 70 の出力に基づいて、物体位置情報の重み W_v を設定する。例えば、重み設定部 35 は、輝度センサが検知した輝度に基づいて、輝度が高いほど物体位置情報の重み W_v が大きくなるように、重み W_v を設定してもよい。重み設定部 35 は、距離センサが検知した物体までの距離に基づいて、距離に近いほど物体位置情報の重み W_v が大きくなるように、重み W_v を設定してもよい。加速度が大きい場合は画像がブレている可能性があるため、重み設定部 35 は、加速度センサが検知した加速度に基づいて、加速度が大きいほど物体位置情報の重み W_v が小さくなるように、重み W_v を設定してもよい。

【0073】

センサ部 70 の出力はカメラ 10 の撮影状況に応じた値であるため、本実施形態においても、精度良く目的音の收音方向を決定することができる。

【0074】

(第 3 実施形態)

本実施形態の收音装置 1 は、画像データ v から発話区間を検出して收音を制御する。

【0075】

図 16 は、第 3 実施形態における制御部 30 の機能を示している。第 3 実施形態の制御部 30 は、発話区間検出部 38 を備える。発話区間検出部 38 は、例えば、口唇が開いている状態の特徴量を示す情報を備え、その情報に基づいて画像データ v において口唇が開いているか否かを判定する。これにより、発話区間検出部 38 は、発話の開始時点及び終

了時点を検出する。発話の開始時点から発話の終了時点までが発話区間である。あるいは、音響信号 s を入力とし、音響特徴に基づいて発話区間の検出を実施するようにしても良い。または、画像データ v および音響信号 s の双方の特徴量に基づいて発話区間の検出を実施するようにしても良い。

【0076】

図16において、收音部80は、物体位置検出部31、音源位置推定部32、画像確度判定部33、音響確度判定部34、重み設定部35、收音方向決定部36、及びビームフォーム部37により構成される。收音部80は、発話区間検出部38によって検出された発話区間内において、目的音を收音する。例えば、物体位置検出部31は、発話区間に対応する画像データ v から顔の位置を検出してもよい。收音方向決定部36は、検出された発話区間においてのみ、收音方向を決定してもよい。ビームフォーム部37は、検出された発話区間に対応する音響信号 s のみから、目的音を收音してもよい。

【0077】

なお、口唇の開閉に基づく発話区間の検出は、物体位置検出部31が行ってもよい。例えば、物体位置検出部31は、顔を検出した領域 r (θ , ϕ) 内において、口唇の開閉を判定してもよい。

【0078】

(第4実施形態)

本実施形態の收音装置1は、画像データ v の複数の領域毎に、顔の確率 P に重み W_v を設定する。以下、図2および図17を参照して、本実施形態を説明する。

【0079】

物体位置検出部31は、画像データ v を複数の領域に分割し、各領域の画像が顔であるか否かを判定する。すなわち、物体位置検出部31は、各領域の画像が顔である確率 P を算出する。

【0080】

本実施形態では、図17に示すように、複数の領域のうち、2以上の領域で顔が検出されている。すなわち、領域 r_a には音声を発している顔が含まれ、領域 r_b には音声を発していない顔が含まれている。さらに、領域毎にカメラ10の撮影状況が異なっている。具体的には、複数の領域のうち、領域 r_a の画像においては、カメラ10の撮影状況に応じて、平均輝度が高すぎたり、低すぎたりしている。例えば、領域 r_a の周辺のみ光が強く当たったり、光がほとんど当たらなかったりする場合に、このようなことが起こる。一方、領域 r_b の画像においては、平均輝度は顔を検出するのに適切な値である。そのため、領域 r_a の画像に含まれる人物が音声を発しているにもかかわらず、領域 r_a の画像に含まれる顔の確率 P が、領域 r_b の画像に含まれる顔の確率 P よりも低くなってしまふ。これにより、音声を発していない領域 r_b の人物が音声を発していると誤検出されてしまうことがある。

【0081】

そこで、本実施形態の重み設定部35は、複数の領域のうち、検出した顔の位置に対応する領域 r_a の平均輝度に基づいて、領域 r_a の画像確度 CM_v を算出する。例えば、重み設定部35は、領域 r_a の平均輝度が高すぎたり、低すぎたりした場合には、画像確度 CM_v が低くなるように、領域 r_a の画像確度 CM_v を算出する。そして、重み設定部35は、領域 r_a の画像確度 CM_v が低い場合に、領域 r_a の物体位置情報の重み W_v が高くなるように、物体位置情報の重み W_v を設定する。つまり、重み設定部35は、領域 r_a の画像確度 CM_v に基づいて、領域 r_a の物体位置情報の重み W_v を設定する。

【0082】

以上のように、重み設定部35は、複数の領域のうち、人物の位置に対応する領域 r_a の平均輝度に基づいて画像確度 CM_v を算出する。そのため、例えば、領域 r_a の撮影状況に応じて、領域 r_a の平均輝度が高すぎたり、低すぎたりした場合でも、領域 r_a の画像に顔が含まれる確率が適切に算出される。

【0083】

なお、重み設定部 35 は、平均輝度に基づく以外に、領域 r_a の画像の輝度の分散に基づいて、画像確度 CM_v を算出してもよい。

【0084】

また、重み設定部 35 は、領域 r_a に顔の一部のみが含まれている場合に、画像確度 CM_v を低く設定してもよい。すなわち、重み設定部 35 は、領域 r_a に顔の一部のみが含まれている場合に、画像確度 CM_v に基づいて、領域 r_a の物体位置情報の重み W_v が高くなるように、物体位置情報の重み W_v を設定してもよい。これにより、画像 v 内で音声を発している人物にオクルージョンが発生しても、音声を発している人物の顔の確率が適切に算出される。

【0085】

(他の実施形態)

以上のように、本出願において開示する技術の例示として、第 1～第 4 実施形態を説明した。しかしながら、本開示における技術は、これに限定されず、適宜、変更、置き換え、付加、省略などを行った実施形態にも適用可能である。また、上記第 1～第 4 実施形態で説明した各構成要素を組み合わせ、新たな実施形態とすることも可能である。そこで、以下、他の実施形態を例示する。

【0086】

第 1 実施形態においては、重み設定部 35 は、物体位置情報の重み W_v および音源位置情報の重み W_s を設定している。しかし、本開示はこれに限定されない。重み設定部 35 は、物体位置情報の重み W_v または音源位置情報の重み W_s の一方のみを設定してもよい。すなわち、重み付け部 300 は、画像確度判定部 33 または音響確度判定部 34 の一方のみを含んでいてもよい。この例について、図 18 と図 19 を参照して説明する。

【0087】

図 18 に示すように、他の実施形態の重み付け部 300 は、音響確度判定部 34 を含んでいない。すなわち、重み設定部 35 は、物体位置情報の重み W_v のみを設定する。この場合、收音方向決定部 36 は、物体位置情報 $P(\theta, \phi | v)$ と、物体位置情報の重み W_v と、音源位置情報 $P(\theta, \phi | s)$ とに基づいて、目的音を收音する方向を決定する。具体的には、收音方向決定部 36 は、第 1 実施形態における式 (7) の重み W_s を 1 に設定し、式 (7) を用いることで、目的音を收音する方向を決定できる。このような構成であっても、第 1 実施形態の收音装置 1 と同様に、物体位置情報の重み W_v を用いることで、撮影状況に応じた誤検出を抑制することができる。

【0088】

また、図 19 に示すように、他の実施形態の別の重み付け部 300 は、画像確度判定部 33 を含んでいない。すなわち、重み設定部 35 は、音源位置情報の重み W_s のみを設定する。この場合、收音方向決定部 36 は、物体位置情報 $P(\theta, \phi | v)$ と、音源位置情報 $P(\theta, \phi | s)$ と、音源位置情報の重み W_s とに基づいて、目的音を收音する方向を決定する。具体的には、收音方向決定部 36 は、第 1 実施形態における式 (7) の重み W_v を 1 に設定し、式 (7) を用いることで、目的音を收音する方向を決定できる。このような構成であっても、第 1 実施形態の收音装置 1 と同様に、音源位置情報の重み W_s を用いることで、マイクアレイ 20 の受信状況に応じた誤検出を抑制することができる。

【0089】

第 1 実施形態においては、人の顔を検出する例について説明したが、人の音声を收音する場合、検出対象の物体は、人の顔に限らず、人として認識できる部分であればよい。例えば、検出される物体は、人の身体又は唇であってもよい。

【0090】

第 1 実施形態においては、画像データ v の平均輝度 Y_{ave} に基づいて、画像確度 CM_v を判定したが、画像確度 CM_v の判定を別の方法で行ってもよい。例えば、画像データ v の輝度の分散に基づいて、画像確度 CM_v を判定してもよい。具体的には、例えば、画像データ v の輝度の分散 Y_{std} が推奨分散 Y_{std_base} よりも小さければ、画像確度判定部 33 は、画像確度 CM_v を「 Y_{std} / Y_{std_base} 」に設定する。分散 Y

10

20

30

40

50

s_{std} が推奨分散 Y_{std_base} 以上であれば、画像確度判定部33は、画像確度 CM_v を「1」に設定する。

【0091】

画像確度 CM_v の判定のさらに別の方法として、画像データ v の輝度ヒストグラムを用いても良い。例えば、輝度ヒストグラムが低輝度又は高輝度に集中しているときは画像確度 CM_v を低く設定し、輝度ヒストグラムが低輝度から高輝度まで幅広く分散しているときは画像確度 CM_v を高く設定する。

【0092】

また、画像データ v のフォーカスが合っているかどうかに基づいて、画像確度 CM_v を判定してもよい。

【0093】

第1実施形態においては、音響確度 CM_s を、音声GMMによる尤度 L_v 及び非音声GMMによる尤度 L_n に基づいて判定したが、音響確度 CM_s の判定を別の方法で行ってもよい。例えば、音響確度判定部34は、音響信号 s を音源分離技術によって音声と非音声に分離し、音声と非音声のパワー比に基づいて、音響確度 CM_s を算出してもよい。音源分離技術として、例えば、ブラインド音源分離 (Blind Source Separation) を用いてもよい。

【0094】

第1実施形態では、確度 CM_v 、 CM_s に基づいて重み W_v 、 W_s を設定したが、重み W_v 、 W_s は確度 CM_v 、 CM_s と同一の値であってもよい。

【0095】

上記実施形態において、画像確度 CM_v と音響確度 CM_s の両方が、例えば、所定値以下であれば、前回設定した收音方向 θ 、 ϕ を今回の收音方向として決定してもよい。

【0096】

上記実施形態では、收音方向として、水平角 θ 及び垂直角 ϕ を決定する場合について説明したが、水平角 θ 及び垂直角 ϕ の少なくともいずれか一方のみで、目的音を発する物体の位置を特定できる場合は、水平角 θ 及び垂直角 ϕ の少なくともいずれか一方を決定するだけであってもよい。

【0097】

上記実施形態においては、人の音声を目的音として收音したが、目的音は人の音声に限らない。例えば、目的音は、車の音又は動物の鳴き声であってもよい。目的音が車の音である場合は、車の特徴を示す情報に基づいて、領域 $r(\theta, \phi)$ 内の画像が車である確率 $P(\theta, \phi | v)$ を算出してもよい。また、車の音のGMMに基づいて、音響確度 CM_s を設定してもよい。

【0098】

(実施形態の概要)

(1) 本開示の收音装置は、音源となる物体から出力される目的音を收音するための收音装置であって、カメラによって生成された画像データ(v)の入力を受ける第1の入力部と、マイクアレイから出力される音響信号(s)の入力を受ける第2の入力部と、画像データ(v)及び音響信号(s)に基づいて、目的音を收音する方向(θ, ϕ)を決定する制御部と、を有する。制御部は、検出部と、推定部と、重み付け部と、決定部とを含む。検出部は、画像データ(v)に基づいて、物体の水平方向又は垂直方向の少なくともいずれか一方における位置を検出し、検出した物体の位置を示す情報である物体位置情報($P(\theta, \phi | v)$)を出力する。推定部は、音響信号(s)に基づいて、音源の水平方向又は垂直方向の少なくともいずれか一方における位置を推定し、推定した音源の位置を示す情報である音源位置情報($P(\theta, \phi | s)$)を出力する。重み付け部は、物体位置情報に対してカメラの撮影状況に応じた重み(W_v)を設定する。決定部は、物体位置情報($P(\theta, \phi | v)$)と物体位置情報の重み(W_v)と音源位置情報($P(\theta, \phi | s)$)とに基づいて、目的音を收音する方向(θ, ϕ)を決定する。

【0099】

10

20

30

40

50

これにより、物体位置情報 $P(\theta, \phi | v)$ の信頼性を考慮することによって、精度良く目的音の收音方向を決定することができる。よって、マイクアレイ 20 から出力された音響信号 s から、決定された收音方向の音を抽出することによって、目的音を精度良く收音することができる。

【0100】

(2) (1) の收音装置において、重み付け部は、物体が画像データに含まれている確からしさを示す画像確度 (CM_v) に基づいて物体位置情報の重みを設定してもよい。

【0101】

これにより、物体位置情報に対してカメラの撮影状況に応じた重み (W_v) を精度良く設定することができる。

【0102】

(3) (2) の收音装置において、重み付け部は、画像データの平均輝度又は輝度の分散に基づいて、画像確度 (CM_v) を算出してもよい。

【0103】

これにより、撮影時の照明環境に応じた重み (W_v) を精度良く設定することができる。

【0104】

(4) (2) の收音装置において、検出部は、画像データを複数の領域に分割し、重み付け部は、複数の領域のうち、検出した物体の位置に対応する一の領域の平均輝度又は輝度の分散に基づいて、画像確度を算出してもよい。

【0105】

これにより、領域の撮影状況に応じて、領域の平均輝度が高すぎたり、低すぎたりした場合でも、領域の画像に物体が含まれる確率を適切に算出することができる。

【0106】

(5) (1) ~ (4) の收音装置において、重み付け部は、マイクアレイの受信状況に応じて音源位置情報の重みを設定し、決定部は、物体位置情報と物体位置情報の重みと音源位置情報と音源位置情報の重みに基づいて、目的音を收音する方向を決定するとしてもよい。

【0107】

これにより、物体位置情報 $P(\theta, \phi | v)$ 及び音源位置情報 $P(\theta, \phi | s)$ の信頼性が高いほうの情報を優先させることによって、精度良く目的音の收音方向を決定することができる。よって、マイクアレイ 20 から出力された音響信号 s から、決定された收音方向の音を抽出することによって、目的音を精度良く收音することができる。

【0108】

(6) (5) の收音装置において、重み付け部は、目的音が音響信号に含まれている確からしさを示す音響確度 (CM_s) に基づいて音源位置情報の重みを設定するとしてもよい。

【0109】

これにより、また、音源位置情報に対してマイクアレイの受信状況に応じた重み (W_s) を精度良く設定することができる。

【0110】

(7) (6) の收音装置において、重み付け部は、音響信号に含まれる目的音と非目的音の尤度比又はパワー比に基づいて、音響確度 (CM_s) を算出してもよい。

【0111】

これにより、受信時の周囲の状況に応じた重み (W_s) を精度良く設定することができる。

【0112】

(8) (1) の收音装置は、周囲の明るさを検知する輝度センサ、物体までの距離を検知する距離センサ、又は加速度センサのうちの少なくとも1つを含むセンサ部をさらに有してもよい。重み付け部は、センサ部の出力に基づいて、物体位置情報の重み (W_s) を設定してもよい。

10

20

30

40

50

【0113】

これにより、撮影時の状況に応じた重み (W_v) を精度良く設定することができる。

【0114】

(9) (1) の收音装置は、カメラ及びマイクアレイのうちの少なくとも一方を備えてもよい。

【0115】

これにより、例えば、カメラを備えた收音装置にマイクアレイのみを外付けすることによって、目的音を精度良く收音することができる。

【0116】

(10) (1) の收音装置において、検出部は、物体として、人の身体、顔、又は唇を検出してもよい。

【0117】

これにより、人の音声を精度良く收音することができる。

【0118】

(11) また、本開示の別の收音装置は、音源となる物体から出力される目的音を收音するための收音装置であって、カメラによって生成された画像データ (v) の入力を受ける第1の入力部と、マイクアレイから出力される音響信号 (s) の入力を受ける第2の入力部と、画像データ (v) 及び音響信号 (s) に基づいて、目的音を收音する方向 (θ , ϕ) を決定する制御部と、を有する。制御部は、検出部と、推定部と、重み付け部と、決定部とを含む。検出部は、画像データ (v) に基づいて、物体の水平方向又は垂直方向の少なくともいずれか一方における位置を検出し、検出した物体の位置を示す情報である物体位置情報 ($P(\theta, \phi | v)$) を出力する。推定部は、音響信号 (s) に基づいて、音源の水平方向又は垂直方向の少なくともいずれか一方における位置を推定し、推定した音源の位置を示す情報である音源位置情報 ($P(\theta, \phi | s)$) を出力する。重み付け部は、マイクアレイの受音状況に応じて音源位置情報の重みを設定する。決定部は、物体位置情報 ($P(\theta, \phi | v)$) と音源位置情報 ($P(\theta, \phi | s)$) と音源位置情報の重み (W_s) とに基づいて、目的音を收音する方向 (θ , ϕ) を決定する。

【0119】

これにより、音源位置情報 ($P(\theta, \phi | s)$) の信頼性を考慮することによって、精度良く目的音の收音方向を決定することができる。よって、マイクアレイ20から出力された音響信号 s から、決定された收音方向の音を抽出することによって、目的音を精度良く收音することができる。

【0120】

(12) 本開示の收音方法は、音源となる物体から出力される目的音を收音するための收音方法であって、カメラによって生成された画像データ (v) の入力を受けるステップと、マイクアレイから出力される音響信号 (s) の入力を受けるステップと、画像データ (v) に基づいて、物体の水平方向又は垂直方向の少なくともいずれか一方における位置を検出し、検出した物体の位置を示す情報である物体位置情報 ($P(\theta, \phi | v)$) を出力するステップと、音響信号 (s) に基づいて、音源の水平方向又は垂直方向の少なくともいずれか一方における位置を推定し、推定した音源の位置を示す情報である音源位置情報 ($P(\theta, \phi | s)$) を出力するステップと、カメラの撮影状況に応じて物体位置情報の重み (W_v) を設定するステップと、物体位置情報 ($P(\theta, \phi | v)$) と物体位置情報の重み (W_v) と音源位置情報 ($P(\theta, \phi | s)$) とに基づいて、目的音を收音する方向 (θ , ϕ) を決定するステップと、を含む。

【0121】

これにより、物体位置情報 $P(\theta, \phi | v)$ の信頼性を考慮することによって、精度良く目的音の收音方向を決定することができる。よって、マイクアレイ20から出力された音響信号 s から、決定された收音方向の音を抽出することによって、目的音を精度良く收音することができる。

【0122】

(13) また、本開示の別の收音方法は、音源となる物体から出力される目的音を收音するための收音方法であって、カメラによって生成された画像データ (v) の入力を受けるステップと、マイクアレイから出力される音響信号 (s) の入力を受けるステップと、画像データ (v) に基づいて、物体の水平方向又は垂直方向の少なくともいずれか一方における位置を検出し、検出した物体の位置を示す情報である物体位置情報 ($P(\theta, \phi | v)$) を出力するステップと、音響信号 (s) に基づいて、音源の水平方向又は垂直方向の少なくともいずれか一方における位置を推定し、推定した音源の位置を示す情報である音源位置情報 ($P(\theta, \phi | s)$) を出力するステップと、マイクアレイの受信状況に応じて音源位置情報の重み (W_s) を設定するステップと、物体位置情報 ($P(\theta, \phi | v)$) と音源位置情報 ($P(\theta, \phi | s)$) と音源位置情報の重み (W_s) とに基づいて、目的音を收音する方向 (θ, ϕ) を決定するステップと、を含む。

10

【0123】

これにより、音源位置情報 $P(\theta, \phi | s)$ の信頼性を考慮することによって、精度良く目的音の收音方向を決定することができる。よって、マイクアレイ 20 から出力された音響信号 s から、決定された收音方向の音を抽出することによって、目的音を精度良く收音することができる。

【0124】

(14) 本開示のプログラムは、コンピュータに上記收音方法を実行させる。

【0125】

本開示の全請求項に記載の收音装置及び收音方法は、ハードウェア資源、例えば、プロセッサ、メモリ、及びプログラムとの協働などによって、実現される。

20

【産業上の利用可能性】

【0126】

本開示の收音装置は、例えば、会話中の人の音声を收音する装置として、有用である。

【符号の説明】

【0127】

- 1 收音装置
- 10 カメラ
- 20 マイクアレイ
- 30 制御部
- 31 物体位置検出部
- 32 音源位置推定部
- 33 画像確度判定部
- 34 音響確度判定部
- 35 重み設定部
- 36 收音方向決定部
- 37 ビームフォーム部
- 38 発話区間検出部
- 40 記憶部
- 50 入出力インタフェース部
- 60 バス
- 70 センサ部
- 300 重み付け部

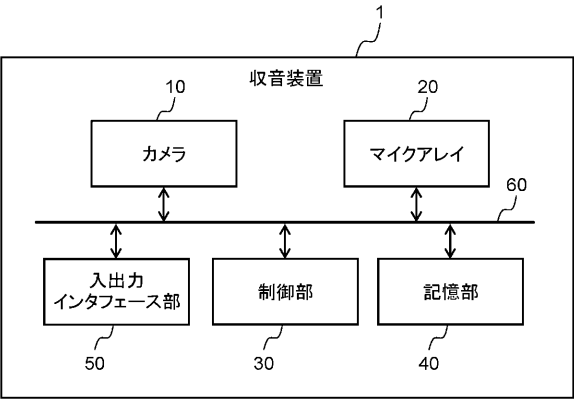
30

40

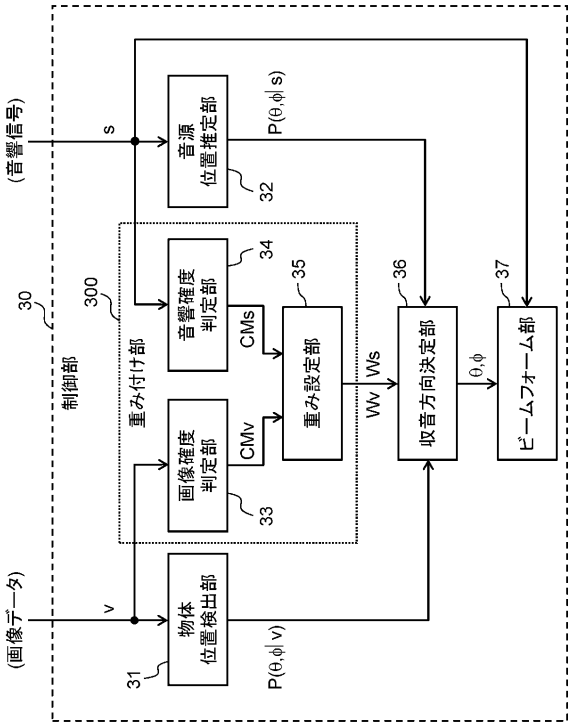
50

【図面】

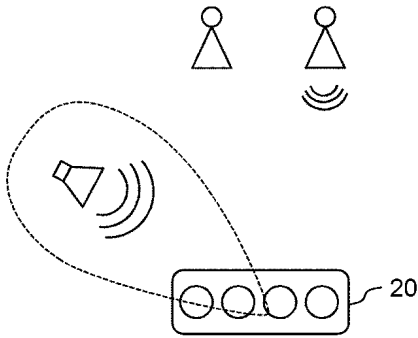
【図 1】



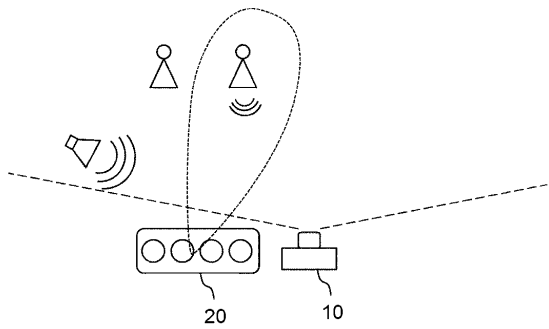
【図 2】



【図 3 A】



【図 3 B】



10

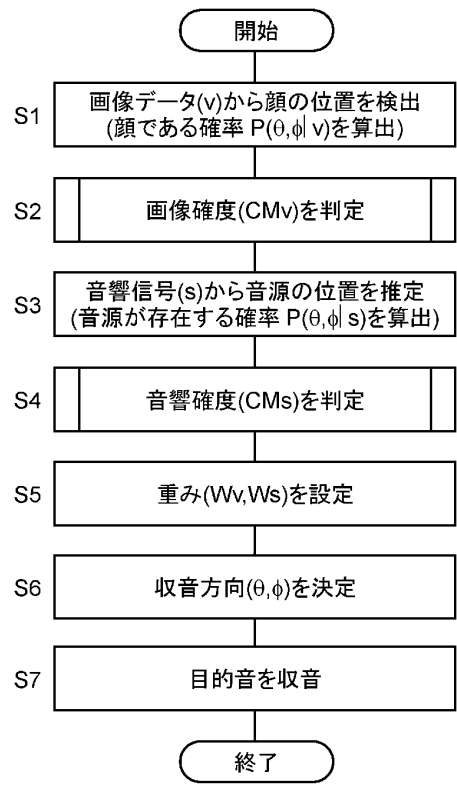
20

30

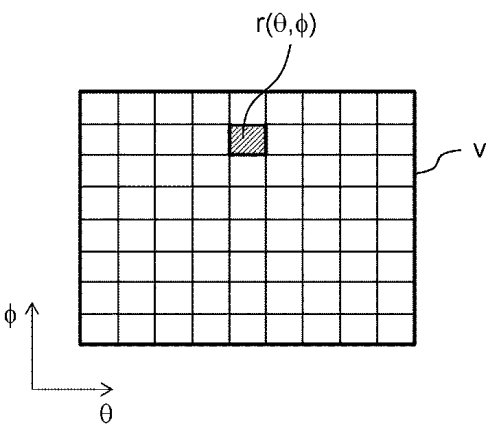
40

50

【図 4】



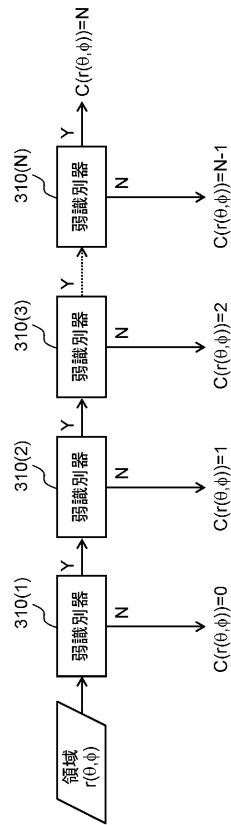
【図 5】



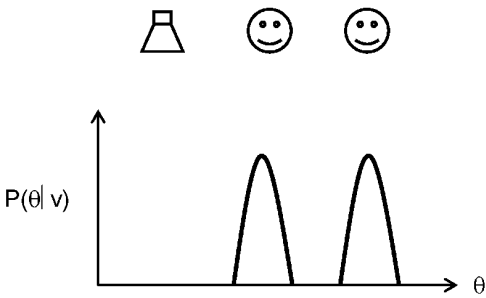
10

20

【図 6】



【図 7】

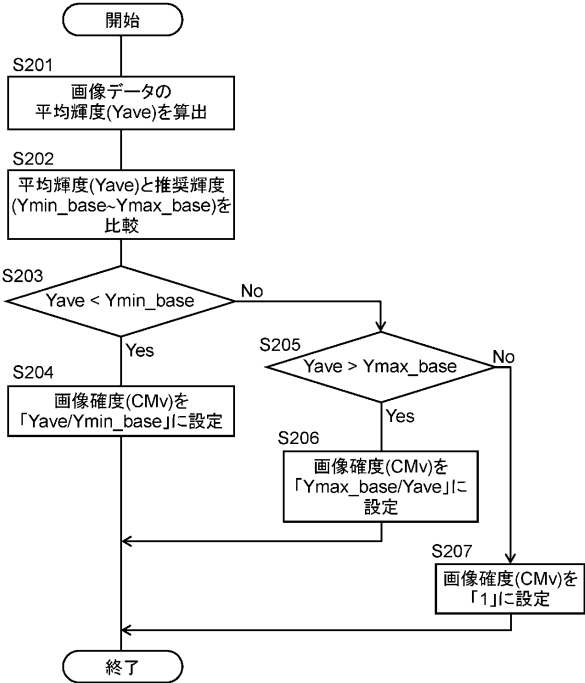


30

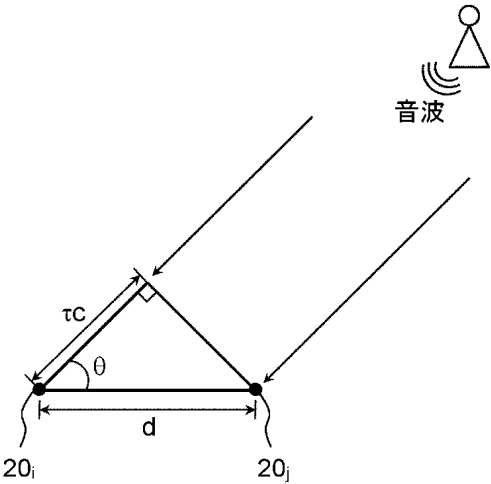
40

50

【図 8】



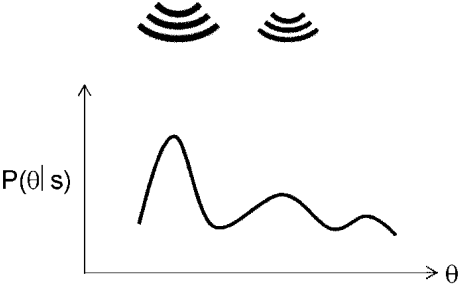
【図 9】



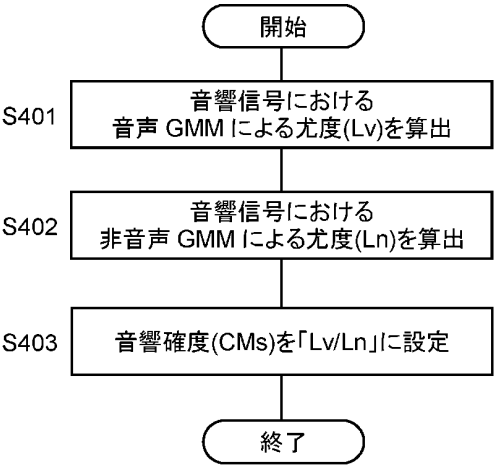
10

20

【図 10】



【図 11】

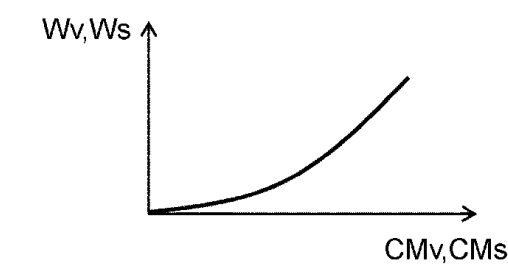


30

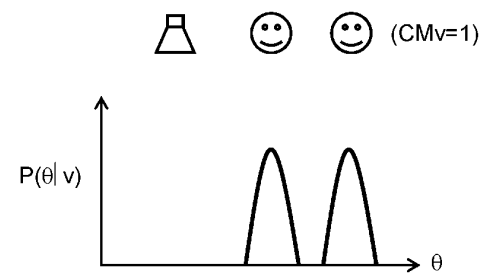
40

50

【図 1 2】

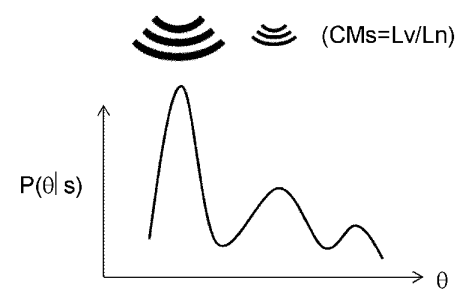


【図 1 3 A】



10

【図 1 3 B】

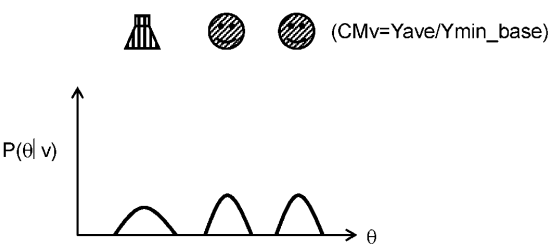


【図 1 3 C】

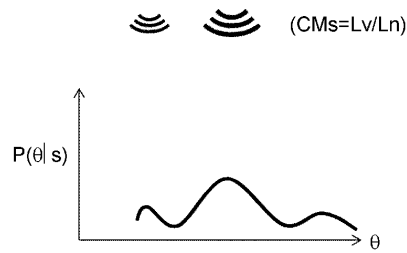


20

【図 1 4 A】



【図 1 4 B】

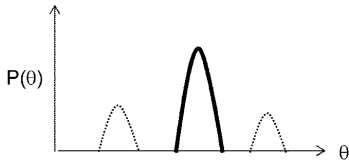


30

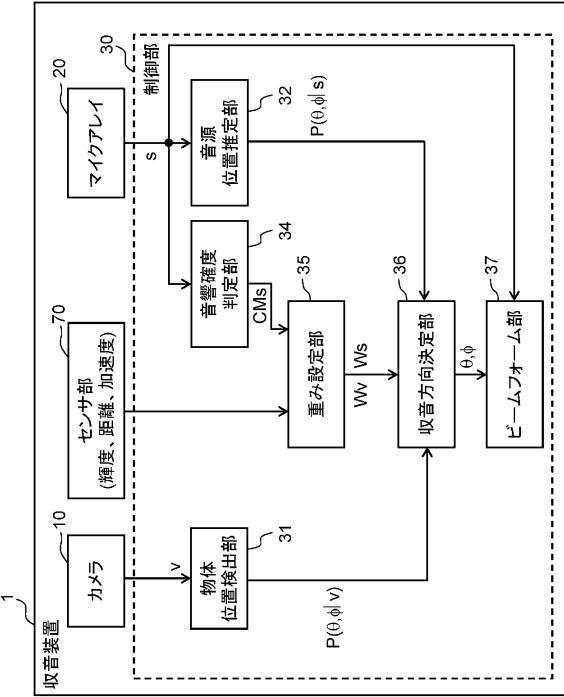
40

50

【図 1 4 C】



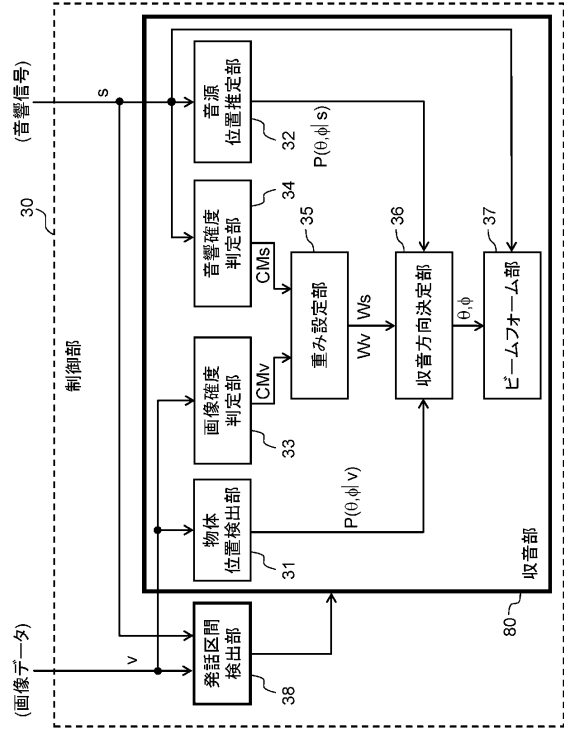
【図 1 5】



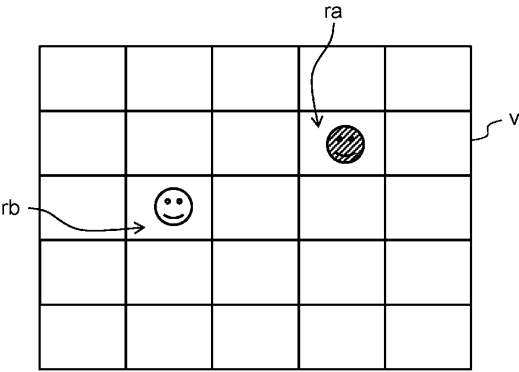
10

20

【図 1 6】



【図 1 7】

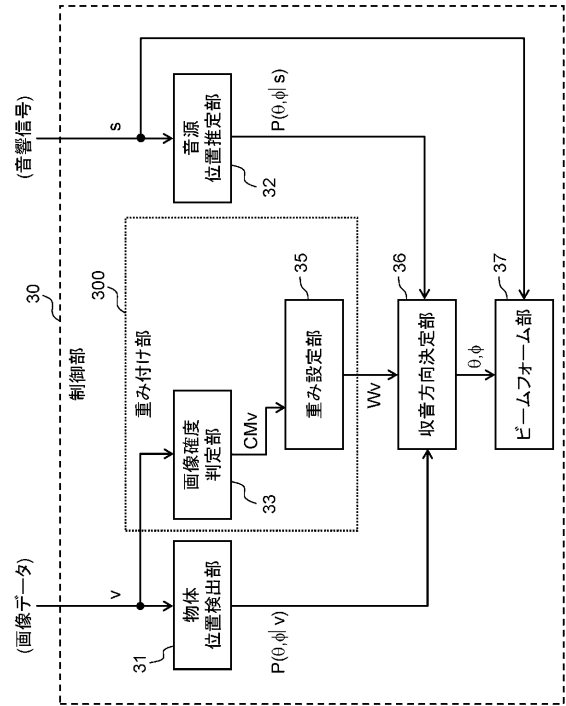


30

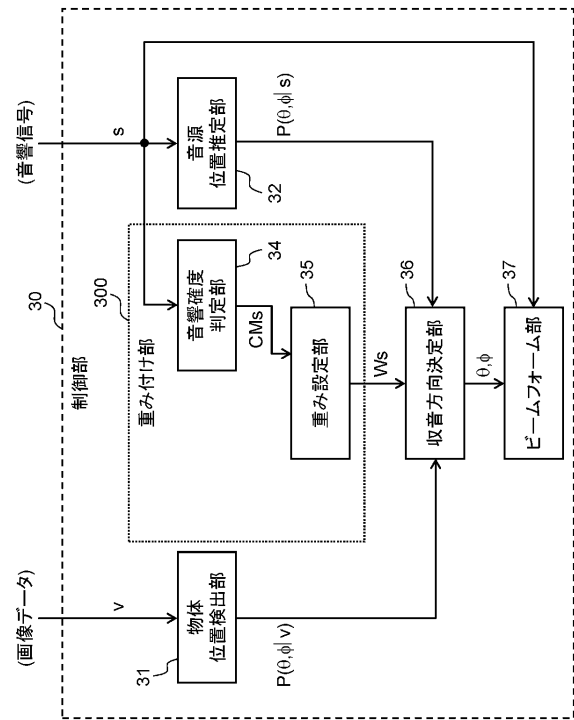
40

50

【図 18】



【図 19】



10

20

30

40

50

フロントページの続き

大阪府門真市大字門真1006番地 パナソニック株式会社内

審査官 渡邊 正宏

(56)参考文献 特表2012-512413 (JP, A)

特表2016-521894 (JP, A)

特開2012-042454 (JP, A)

(58)調査した分野 (Int.Cl., DB名)

G10K 11/00-13/00

G10L 13/00-13/10

G10L 19/00-19/26

G10L 21/00-21/18

G10L 25/00-25/93

G10L 99/00

H04N 7/10

H04N 7/14- 7/173

H04N 7/20- 7/56

H04N 19/89

H04N 21/438

H04N 19/895

H04N 21/00-21/858

H04R 3/00- 3/14