(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6686977号
(P6686977)
(45)発行日　令和2年4月22日(2020.4.22)　(24)登録日　令和2年4月6日(2020.4.6)

(51)Int.Cl.　F I
G10L 21/0272 (2013.01)　G10L 21/0272 100Z
G10L 15/00 (2013.01)　G10L 15/00 200H
G10L 15/25 (2013.01)　G10L 15/25

請求項の数 12　(全 25 頁)

(21)出願番号　特願2017-123643(P2017-123643)
(22)出願日　平成29年6月23日(2017.6.23)
(65)公開番号　特開2019-8134(P2019-8134A)
(43)公開日　平成31年1月17日(2019.1.17)
審査請求日　平成30年6月1日(2018.6.1)

(73)特許権者　000001443
カシオ計算機株式会社
東京都渋谷区本町１丁目６番２号
(74)代理人　100074099
弁理士　大菅　義之
(74)代理人　100121083
弁理士　青木　宏義
(74)代理人　100138391
弁理士　天田　昌行
(72)発明者　山谷　崇史
東京都羽村市栄町３丁目２番１号　カシオ計算機株式会社　羽村技術センター内
(72)発明者　中込　浩一
東京都羽村市栄町３丁目２番１号　カシオ計算機株式会社　羽村技術センター内
最終頁に続く

(54)【発明の名称】音源分離情報検出装置、ロボット、音源分離情報検出方法及びプログラム

(57)【特許請求の範囲】
【請求項１】
音声を取得するために所定の指向性を有する音声取得手段と、
前記音声取得手段により取得された前記音声から、所定の対象の信号音声の到来方向である第１方向を検出する第１方向検出手段と、
前記音声取得手段により取得された前記音声から、雑音音声の到来方向である第２方向を検出する第２方向検出手段と、
前記第１方向と前記第２方向とに基づいて、音源分離方向又は音源分離位置を検出する検出手段と、
前記音声取得手段が前記音声を取得するタイミングで前記所定の対象の口唇画像を取得する画像取得手段と、
前記口唇画像に基づいて、前記所定の対象の口唇の開口又は前記口唇の閉口を判定する判定手段と、
を備え、
前記第１方向検出手段は、前記判定手段による前記口唇の開口の判定時に、前記音声取得手段により取得された前記音声を前記信号音声とし、
前記第２方向検出手段は、前記判定手段による前記口唇の閉口の判定時に、前記音声取得手段により取得された前記音声を前記雑音音声とする、
ことを特徴とする音源分離情報検出装置。
【請求項２】

前記検出手段は、前記信号音声と前記雑音音声とから算出された信号対雑音比が閾値以下である場合、前記第１方向と前記第２方向とに基づいて、前記信号／雑音比が前記閾値を超える前記音源分離方向又は前記音源分離位置を検出する、
ことを特徴とする請求項１に記載の音源分離情報検出装置。
【請求項３】
前記画像取得手段は、顔部画像を更に取得し、
前記口唇画像から前記所定の対象の前記口唇の移動量を取得する口唇移動量取得手段と、
前記顔部画像から前記所定の対象の顔部の回転量を取得する顔部回転量取得手段と、
を更に備え、
前記判定手段は、前記口唇の移動量と前記顔部の回転量とに基づいて、前記所定の対象の前記口唇の開口又は前記口唇の閉口を判定する、
ことを特徴とする請求項１又は２に記載の音源分離情報検出装置。
【請求項４】
前記判定手段は、前記口唇の移動量のうちの前記口唇の開閉方向の移動量が第１の閾値を超え、且つ、前記口唇の移動量のうちの前記口唇の延伸方向の移動量が第２の閾値未満であり、且つ、前記顔部の回転量が第３の閾値未満であるときに、前記口唇の開口又は前記口唇の閉口を判定する、
ことを特徴とする請求項３に記載の音源分離情報検出装置。
【請求項５】
前記第１方向検出手段は、前記判定手段による前記口唇の開口の判定時に、前記信号音声の信号音声パワーに基づいて、前記第１方向を検出し、
前記第２方向検出手段は、前記判定手段による前記口唇の閉口の判定時に、前記雑音音声の雑音音声パワーに基づいて、前記第２方向を検出する、
ことを特徴とする請求項１乃至４の何れか１項に記載の音源分離情報検出装置。
【請求項６】
前記所定の対象にメッセージを報知する報知手段を更に備え、
前記報知手段は、前記所定の対象に現在位置から前記音源分離位置まで移動させるために、前記音源分離位置までの移動方向及び移動距離を含む前記メッセージを報知する、
ことを特徴とする請求項１乃至５の何れか１項に記載の音源分離情報検出装置。
【請求項７】
請求項１乃至６の何れか１項に記載の音源分離情報検出装置と、
自装置を移動する移動手段と、
前記自装置を動作する動作手段と、
前記音源分離情報検出装置、前記移動手段及び前記動作手段を制御する制御手段と、
を備える、
ことを特徴とするロボット。
【請求項８】
前記制御手段は、前記移動手段を制御して、前記音源分離位置に前記自装置を移動させる、
ことを特徴とする請求項７に記載のロボット。
【請求項９】
前記制御手段は、前記自装置が前記所定の対象とアイコンタクトを取りながら、又は、前記自装置が前記所定の対象の方を向きながら、前記音源分離位置に移動するように、前記動作手段を制御する、
ことを特徴とする請求項８に記載のロボット。
【請求項１０】
前記制御手段は、前記自装置が前記音源分離位置に一気に移動するのではなく、少しだけ動いたり、回転のみをしたりして、前記音源分離位置に移動するように、前記移動手段及び前記動作手段を制御する、

ことを特徴とする請求項８又は９に記載のロボット。
【請求項１１】
音声を取得するために所定の指向性を有する音声取得手段により取得された前記音声から、所定の対象の信号音声の到来方向である第１方向を検出し、
前記音声取得手段により取得された前記音声から、雑音音声の到来方向である第２方向を検出し、
前記第１方向と前記第２方向とに基づいて、音源分離方向又は音源分離位置を検出し、
前記音声を取得するタイミングで前記所定の対象の口唇画像を取得し、
前記口唇画像に基づいて、前記所定の対象の口唇の開口又は前記口唇の閉口を判定し、
前記口唇の開口の判定時に、前記取得された前記音声を前記信号音声とし、
前記口唇の閉口の判定時に、前記取得された前記音声を前記雑音音声とする、
ことを含む、
ことを特徴とする音源分離情報検出方法。
【請求項１２】
音源分離情報検出装置のコンピュータを、
音声を取得するために所定の指向性を有する音声取得手段により取得された前記音声から、所定の対象の信号音声の到来方向である第１方向を検出し、
前記音声取得手段により取得された前記音声から、雑音音声の到来方向である第２方向を検出し、
前記第１方向と前記第２方向とに基づいて、音源分離方向又は音源分離位置を検出し、
前記音声を取得するタイミングで前記所定の対象の口唇画像を取得し、
前記口唇画像に基づいて、前記所定の対象の口唇の開口又は前記口唇の閉口を判定し、
前記口唇の開口の判定時に、前記取得された前記音声を前記信号音声とし、
前記口唇の閉口の判定時に、前記取得された前記音声を前記雑音音声とする、
ように機能させる、
ことを特徴とするプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、音源分離情報検出装置、ロボット、音源分離情報検出方法及びプログラムに関する。
【背景技術】
【０００２】
人間、動物等に模した形態を有し、人間と会話等のコミュニケーションをすることができるロボットが知られている。このようなロボットには、自装置に搭載されたマイクの出力に基づいてロボットの周囲に発生した音を検出し、その音が対象者の発声した音声であると判別すると、その対象者がいる方向にロボットの顔の向きあるいは体の向きを変え、その対象者に話しかけたり手を振ったりする等の動作をするものもある。
【０００３】
かかるロボットの動作を実現するためには、対象者が発声する音声である信号音声（信号源）の方向又は位置を検出するために、ロボットの周囲に発生した音から信号音声以外の音声である不要な雑音音声（雑音源）を取り除いて対象者が発声する信号音声のみを分離させる、音源分離技術が必要となる。
【０００４】
従来、信号音声対雑音音声比（Ｓ／Ｎ比）を高めるために音源分離技術の一種であるビームフォーミングをする技術（例えば特許文献１）が知られている。
【先行技術文献】
【特許文献】
【０００５】
【特許文献１】特開２００５－２５３０７１号公報

【発明の概要】
【発明が解決しようとする課題】
【０００６】
しかしながら、従来の音源分離技術では、信号音声と雑音音声とが同方向から到来する場合には音源分離が困難であるという課題があり、そのような場合に、信号音声と雑音音声とを同時に検出してしまい、対象者の信号音声（信号源）の方向又は位置の検出において誤検出が発生するという問題点があった。
【０００７】
本発明は、以上のような課題を解決するためのものであり、信号音声を雑音音声から分離できる音源分離情報を検出することが可能な音源分離情報検出装置、ロボット、音源分離情報検出方法及びプログラムを供給することを目的とする。
【課題を解決するための手段】
【０００８】
前記目的を達成するため、本発明に係る音源分離情報検出装置の一様態は、
音声を取得するために所定の指向性を有する音声取得手段と、
前記音声取得手段により取得された前記音声から、所定の対象の信号音声の到来方向である第１方向を検出する第１方向検出手段と、
前記音声取得手段により取得された前記音声から、雑音音声の到来方向である第２方向を検出する第２方向検出手段と、
前記第１方向と前記第２方向とに基づいて、音源分離方向又は音源分離位置を検出する検出手段と、
前記音声取得手段が前記音声を取得するタイミングで前記所定の対象の口唇画像を取得する画像取得手段と、
前記口唇画像に基づいて、前記所定の対象の口唇の開口又は前記口唇の閉口を判定する判定手段と、
を備え、
前記第１方向検出手段は、前記判定手段による前記口唇の開口の判定時に、前記音声取得手段により取得された前記音声を前記信号音声とし、
前記第２方向検出手段は、前記判定手段による前記口唇の閉口の判定時に、前記音声取得手段により取得された前記音声を前記雑音音声とする、
ことを特徴とする。
【０００９】
前記目的を達成するため、本発明に係るロボットの一様態は、
前記音源分離情報検出装置と、
自装置を移動する移動手段と、
前記自装置を動作する動作手段と、
前記音源分離情報検出装置、前記移動手段及び前記動作手段を制御する制御手段と、
を備える、
ことを特徴とする。
【００１０】
前記目的を達成するため、本発明に係る音源分離情報検出方法の一様態は、
音声を取得するために所定の指向性を有する音声取得手段により取得された前記音声から、所定の対象の信号音声の到来方向である第１方向を検出し、
前記音声取得手段により取得された前記音声から、雑音音声の到来方向である第２方向を検出し、
前記第１方向と前記第２方向とに基づいて、音源分離方向又は音源分離位置を検出し、
前記音声を取得するタイミングで前記所定の対象の口唇画像を取得し、
前記口唇画像に基づいて、前記所定の対象の口唇の開口又は前記口唇の閉口を判定し、
前記口唇の開口の判定時に、前記取得された前記音声を前記信号音声とし、
前記口唇の閉口の判定時に、前記取得された前記音声を前記雑音音声とする、

ことを含む、
ことを特徴とする。
【００１１】
前記目的を達成するため、本発明に係るプログラムの一様態は、
音源分離情報検出装置のコンピュータを、
音声を取得するために所定の指向性を有する音声取得手段により取得された前記音声から、所定の対象の信号音声の到来方向である第１方向を検出し、
前記音声取得手段により取得された前記音声から、雑音音声の到来方向である第２方向を検出し、
前記第１方向と前記第２方向とに基づいて、音源分離方向又は音源分離位置を検出し、
前記音声を取得するタイミングで前記所定の対象の口唇画像を取得し、
前記口唇画像に基づいて、前記所定の対象の口唇の開口又は前記口唇の閉口を判定し、
前記口唇の開口の判定時に、前記取得された前記音声を前記信号音声とし、
前記口唇の閉口の判定時に、前記取得された前記音声を前記雑音音声とする、
ように機能させる、
ことを特徴とする。
【発明の効果】
【００１２】
本発明によれば、信号音声を雑音音声から分離できる音源分離情報を検出することが可能な音源分離情報検出装置、ロボット、音源分離情報検出方法及びプログラムを供給することが可能になる。
【図面の簡単な説明】
【００１３】
【図１】本発明の実施の形態にかかるロボットの外観図である。
【図２】ロボットの構成を示すブロック図である。
【図３】ロボット制御機能の構成を示すブロック図である。
【図４】図３のブロック図の構成の処理例を示すフローチャートである。
【図５】ラベル付けされた顔パーツ検出結果のフォーマット例を示す図である。
【図６】頭部の回転の自由度を模式的に表した図である。
【図７】雑音音声の音源到来方向推定処理の例を示すフローチャートである。
【図８】仮の音源位置とマイクの配置との一例を示す図である。
【図９】マイクアレイの指向特性の例を示す図である。
【図１０】音源分離情報検出の説明図（その１）である。
【図１１】音源分離情報検出の説明図（その２）である。
【図１２】音源分離情報検出の説明図（その３）である。
【発明を実施するための形態】
【００１４】
以下、本発明を実施するための形態について図面を参照しながら詳細に説明する。図１は、実施の形態に係るロボット１００を正面から見た場合の外観を模式的に示した図である。ロボット１００は、頭部１０１と胴体１０２とを備えた人型のコミュニケーションロボットである。ロボット１００は、例えば住宅内に設置され、所定の対象である住人等（以下「対象者」と記載）に呼びかけられると、呼びかけた対象者と会話する。
【００１５】
図１に示すように、ロボット１００の頭部１０１には、カメラ１０４（画像取得手段）と、マイクアレイ１０３（音声取得手段）と、スピーカ１０５（報知手段）と、センサ群１０６と、首関節駆動部１０７と、足回り駆動部１０８と、が設けられている。
【００１６】
カメラ１０４は、頭部１０１の前面の下側、人の顔でいうところの鼻の位置に設けられている。カメラ１０４は、後述する制御部１２７の制御の下、撮像を行う。
【００１７】

マイクアレイ１０３は、例えば１３個のマイクからなる。１３個のマイクのうちの８個のマイクが、人の顔でいうところの額の高さの位置であって、頭部１０１の周周りに等間隔で配置されている。これら８個のマイクよりも上側に、４個のマイクが頭部１０１の周回りに等間隔で配置されている。更に、１個のマイクが頭部１０１の頭頂部に配置されている。マイクアレイ１０３はロボット１００の周囲で発生した音を検出する。

【００１８】

スピーカ１０５は、カメラ１０４よりも下側、人の顔でいうところの口の位置に設けられている。スピーカ１０５は、後述する制御部１２７の制御の下、各種の音声を出力する。

【００１９】

センサ群１０６は、人の顔でいうところの目の位置と耳の位置とに設けられている。センサ群１０６は、加速度センサ、障害物検知センサ等を含み、ロボット１００の姿勢制御や、安全性の確保のために使用される。

【００２０】

首関節駆動部１０７は、頭部１０１と胴体１０２とを連結する部材である。頭部１０１は、破線で示される首関節駆動部１０７によって、胴体１０２に連結されている。首関節駆動部１０７は、複数のモータを含む。後述する制御部１２７がこれら複数のモータを駆動すると、ロボット１００の頭部１０１が回転する。首関節駆動部１０７は、ロボット１００の頭部１０１を回転させると共にその回転量を取得する、顔部回転量取得手段としての役割を有する。

【００２１】

足回り駆動部１０８は、ロボット１００を移動させる移動手段としての役割を有する。特には図示しないが、足回り駆動部１０８は、胴体１０２の下側に設けられた４つの車輪（ホイール）を含む。４つの車輪のうちの、２つが胴体１０２の前側に、残り２つが後ろ側に配置されている。車輪として、例えば、オムニホイール、メカナムホイールが使用される。後述の制御部２０１は、足回り駆動部１０８の車輪を回転させることにより、ロボット１００を移動させる。

【００２２】

図２は、図１の外観を有するロボット１００の制御系であるロボット制御システム２００を示すブロック図である。図２において、図１と同じ参照番号を付した部分は図１と同じものである。図２において、胴体１０２内に設置される制御部２０１は、ＣＰＵ（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ：中央演算処理装置）、ＲＡＭ（Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ：ランダムアクセスメモリ）等を含む。制御部２０１は、頭部１０１内のマイクアレイ１０３、カメラ１０４、スピーカ１０５、センサ群１０６、胴体１０２内の首関節駆動部１０７及び足回り駆動部１０８と、それぞれ電気的に接続され、ＲＡＭを作業領域として、後述する記憶部２０２に記憶されている制御プログラム２０５を読み出して実行することにより、前記各部を制御する。

【００２３】

記憶部２０２は、ソリッドステートディスクドライブ、ハードディスクドライブ、フラッシュメモリ等を含み、胴体１０２の内部に設けられている。記憶部２０２は、制御部２０１によって実行される制御プログラム２０５、マイクアレイ１０３が集音した音声データ、カメラ１０４が撮像した画像データ等を含む各種データを記憶する。記憶部２０２が記憶する制御プログラム２０５には、後述する音源分離情報検出プログラム、移動プログラム、及び対話プログラム等が含まれる。

【００２４】

操作ボタン２０３は、胴体１０２の背中に設けられている（図１において不図示）。操作ボタン２０３は、ロボット１００を操作するための各種のボタンであり、電源ボタン、スピーカ１０５の音量調節ボタン等を含む。

【００２５】

電源部２０４は、胴体１０２に内蔵された充電池であり、ロボット制御システム２００

の各部に電力を供給する。

【００２６】

図３は、図２の制御部２０１が記憶部２０２内の制御プログラム２０５を実行する機能の一部として実現される対話機能の構成を示すブロック図である。なお、図３に示される各機能部は、制御部２０１内のＦＰＧＡ（Ｆｉｅｌｄ　Ｐｒｏｇｒａｍｍａｂｌｅ　Ａｒｒａｙ）等のハードウェアによって実現されてもよい。

【００２７】

図３において、音声取得手段として機能する音声入力部３０１は、図１のマイクアレイ１０３を構成する各マイクから、音声を入力する。

【００２８】

画像取得手段として機能する画像入力部３０４、顔検出部３０５、及び口パーツ検出部３０６は、音声入力部３０１が音声を取得するタイミングで、所定の対象である対象者の口唇画像を取得する。具体的には、画像入力部３０４が、図１のカメラ１０４から画像を入力する。次に、顔検出部３０５が、音声入力部３０１が例えば所定の閾値以上のパワーを有する音声を入力するタイミングで、入力された画像から顔領域を検出する。そして、口パーツ検出部３０６が、検出された顔領域から口パーツを検出し、口唇画像とする。

【００２９】

判定手段として機能する口開閉判定部３０７は、口パーツ検出部３０６が出力する口唇画像に基づいて、対象者の口唇の開口又は口唇の閉口を判定する。

【００３０】

音源到来方向推定部３０２は、口開閉判定部３０７が口唇の開口（口唇が開いている状態）を判定しているときには、第１方向検出手段として機能し、音声入力部３０１が入力する音声を信号音声として、口パーツ検出部３０６が出力する口唇画像及びその信号音声の信号音声パワーに基づいて、信号音声の到来方向である第１方向を推定する。

【００３１】

一方、音源到来方向推定部３０２は、口開閉判定部３０７が口唇の閉口（口唇が閉じている状態）を判定しているときには、第２方向検出手段として機能し、音声入力部３０１が入力する音声を雑音音声として、その雑音音声の雑音音声パワーに基づいて、雑音音声の到来方向である第２方向を推定する。

【００３２】

音源到来方向推定部３０２は、第２方向検出手段として機能するときの処理例として、音源定位手法の一手法であるＭＵＳＩＣ（ＭＵｌｔｉｐｌｅ　ＳＩｇｎａｌ　Ｃｌａｓｓｉｆｉｃａｔｉｏｎ）法に基づく処理を実行することにより、対象者以外の音源からの雑音音声の音源定位（雑音源の位置）を推定する。この処理の詳細については後述する。

【００３３】

音源分離部３０３は、例えば下記文献１で示されているビームフォーミング技術に基づく演算処理を実行することにより、音源到来方向推定部３０２により現在得られている信号音声の到来方向である第１方向又は雑音音声の到来方向である第２方向を入力として、対象者が発声する信号音声を強調し又は信号音声以外の雑音音声を抑圧する音源分離の処理を実行する。

【００３４】

＜文献１＞

浅野　太、“音源分離”、[online]、2011年11月受領、電子情報通信学会『知識の森』、[2017年6月15日検索]、インターネット

<URL:http://www.ieice-hbkb.org/files/02/02gun_06hen_02.pdf>

【００３５】

具体的には、音源分離部３０３は、口開閉判定部３０７が口唇の開口を判定しているときには、上記ビームフォーミングの演算処理により、信号音声を音源到来方向推定部３０２により現在得られている第１方向にビームステアリング（強調）するビームステアリング演算処理を実行することにより、強調された信号音声を得て、それを音量算出部３０８

に出力する。
【００３６】
一方、音源分離部３０３は、口開閉判定部３０７が口唇の閉口を判定しているときには、上記ビームフォーミングの演算処理により、雑音音声を音源到来方向推定部３０２により現在得られている第２方向にヌルステアリング（抑圧）するヌルステアリング演算処理を実行することにより、抑圧された雑音音声を得て、それを音量算出部３０８に出力する。
【００３７】
なお、音源分離部３０３が実行する前記処理は、マイクアレイ１０３として所定の指向性を有する物理的な指向性マイクを用いて実現されてもよい。
【００３８】
音量算出部３０８は、音源分離部３０３が出力するビームステアリング（強調）された信号音声又はヌルステアリング（抑圧）された雑音音声のそれぞれの音量を算出する。
【００３９】
Ｓ／Ｎ比算出部３０９は、音量算出部３０８が算出する信号音声の音量と雑音音声の音量とに基づいて、信号対雑音比（以下「Ｓ／Ｎ比」と記載）を算出し、そのＳ／Ｎ比が閾値よりも大きいか否かを判定する。音源分離部３０３、音量算出部３０８、及びＳ／Ｎ比算出部３０９は、第１方向と第２方向とに基づいて、音源分離方向又は音源分離位置を検出する検出手段として機能する。
【００４０】
Ｓ／Ｎ比算出部３０９での判定の結果、Ｓ／Ｎ比が閾値以下である場合には、図２の制御部２０１は、音声認識のための十分なＳ／Ｎ比が得られていないと判定する。この場合、制御部２０１は例えば、図１又は図２の足回り駆動部１０８を制御することにより、例えば対象者に対して一定の関係（例えば一定の距離又は一定の角度等）を維持しながら、ロボット１００を移動させる。
【００４１】
ロボット１００の移動の後、制御部２０１は再び、図３のロボット制御機能を動作させ、上述と同様のＳ／Ｎ比の判定動作を実行させる。この結果、Ｓ／Ｎ比算出部３０９が算出するＳ／Ｎ比が閾値よりも大きくなると、図２の制御部２０１は、音声認識のための十分なＳ／Ｎ比が得られ、対象者に対するロボット１００の位置関係が、信号音声を雑音音声から最も良く分離できる最適化された位置である音源分離位置になったと判定する（又は、対象者に対するロボット１００の方向関係が、信号音声を雑音音声から最も良く分離できる最適化された方向である音源分離方向になったと判定する）。この場合、制御部２０１は、図３の音声認識部３１０に、音源分離部３０３が出力するビームステアリング（強調）された信号音声に対する音声認識処理を実行させることにより、対象者の発話内容を理解する。更に、制御部２０１は、この音声認識結果に応じて、対話アルゴリズムに従って、図３の発声部３１１から図１又は図２のスピーカ１０５を介して、対象者に対して発声を行って対話をする。
【００４２】
図３において、音声認識部３１０は、既知の音声認識技術を使って音声認識処理を実行する。また、発声部３１１は、既知の音声合成技術を使って音声合成による発声処理を実行する。
【００４３】
図４は、図３のブロック図の構成の処理例を示すフローチャートである。このフローチャートの処理例は、図３のブロック図の構成を実現する制御部２０１のハードウェアが実行する処理として、又は図２の制御部２０１が実行する制御プログラム２０５の処理として実現される。
【００４４】
まず、図３の顔検出部３０５が、顔検出処理を実行する（ステップＳ４０１）。この顔検出処理では、音声入力部３０１が例えば所定の閾値以上のパワーを有する音声を入力す

るタイミングで、カメラ１０４から画像入力部３０４を介して入力された画像から、顔領域が検出される。顔検出処理としては、既知の顔検出技術を使用することができる。例えば、下記文献２に記載されている何れかの顔検出技術が適用されてよい。

【００４５】

＜文献２＞

堀田　一弘、“小特集　顔認識技術　1.顔認識の研究動向”、[online]、2012年3月28日公開、映像情報メディア学会誌、Vol.64,No.4(2010),p.459-462、[2017年6月15日検索]、インターネット

<URL: https://www.jstage.jst.go.jp/article/itej/64/4/64_4_455/_pdf>

【００４６】

次に、図３の口パーツ検出部３０６が、口パーツ検出処理を実行する（ステップＳ４０２）。口パーツ検出処理としては、既知の顔パーツ検出技術を使用することができる。例えば、下記文献３に記載されている何れかの顔パーツ検出技術が適用されてよい。

【００４７】

＜文献３＞

littlewing、“WEBカメラで利用できる顔認識技術まとめ-その2”、[online]、2015年4月7日公開、[2017年6月15日検索]、インターネット

<URL: http://littlewing.hatenablog.com/entry/2015/04/07/221856>

【００４８】

ステップＳ４０２の口パーツ検出処理により、まず例えばラベル付けされた座標値である顔パーツ検出結果が得られる。ラベル付けされた顔パーツ検出結果のフォーマット例としては、例えば図５に示されるように、下記文献４のFigure2として記載されている例を採用することができる。

【００４９】

＜文献４＞

C.sagonas,”Facial point annotations”、[online]、[2017年6月15日検索]、インターネット

<URL: https://ibug.doc.ic.ac.uk/resources/facial-point-annotations/>

【００５０】

ステップＳ４０２の口パーツ検出処理では、図５に例示される顔パーツ検出結果のうちの例えば、ラベル４９から６８までが口パーツとして検出され、またラベル２８から３６までが鼻パーツとして検出される。

【００５１】

次に、図３の口開閉判定部３０７は、ステップＳ４０２で算出された口パーツと鼻パーツとのラベル付けされた座標値（例えば図５のラベル４９～６８、ラベル２８～３６）を用いて、口唇の開口（口唇が開いているか）又は口唇の閉口（口唇が閉じているか）を検出する口開閉検出処理を実行する（ステップＳ４０３）。

【００５２】

ステップＳ４０３で口開閉判定部３０７はまず、口唇の縦座標（顔の上下方向）の変化Δｙを算出する。今、ある時刻のフレームＦ（ｔ）にて、下記（１）式の演算により、ｙ座標量差分総計ｙ（ｔ）が算出される。

【００５３】

$$y(t)=yy1+yy2 \quad \cdots (1)$$

【００５４】

（１）式において、ｙｙ１は、上口唇（下側）と下口唇（上側）とのｙ座標量差分総計であり、図５の関係より、下記（２）式から（７）式までの累算演算により算出される。これらの式において、演算「＋＝」は、左辺の値に右辺の値を累算する演算を示す。また、関数「ｆａｂｓ（）」は、括弧内の数値に対する絶対値を浮動小数で算出する関数である。また例えば、「ｄａｔａ．ｙ［６１］（ｔ）」は、時刻ｔのフレーム画像Ｆ（ｔ）内における図５のラベル６１番のｙ座標データ値を示す。他も同様である。

【００５５】

ｙｙ１＋＝ｆａｂｓ（ｄａｔａ．ｙ［６１］（ｔ）
－ｄａｔａ．ｙ［６７］（ｔ））・・・（２）
ｙｙ１＋＝ｆａｂｓ（ｄａｔａ．ｙ［６１］（ｔ）
－ｄａｔａ．ｙ［５８］（ｔ））・・・（３）
ｙｙ１＋＝ｆａｂｓ（ｄａｔａ．ｙ［６２］（ｔ）
－ｄａｔａ．ｙ［６６］（ｔ））・・・（４）
ｙｙ１＋＝ｆａｂｓ（ｄａｔａ．ｙ［６２］（ｔ）
－ｄａｔａ．ｙ［５７］（ｔ））・・・（５）
ｙｙ１＋＝ｆａｂｓ（ｄａｔａ．ｙ［６３］（ｔ）
－ｄａｔａ．ｙ［６５］（ｔ））・・・（６）
ｙｙ１＋＝ｆａｂｓ（ｄａｔａ．ｙ［６３］（ｔ）
－ｄａｔａ．ｙ［５６］（ｔ））・・・（７）

【００５６】

（１）式において、ｙｙ２は、鼻下と下口唇（上側）とのｙ座標量差分総計であり、図５の関係より、下記（８）式から（１２）式までの演算により算出される。

【００５７】

ｙｙ２＋＝ｆａｂｓ（ｄａｔａ．ｙ［３１］（ｔ）
－ｄａｔａ．ｙ［６０］（ｔ））・・・（８）
ｙｙ２＋＝ｆａｂｓ（ｄａｔａ．ｙ［３２］（ｔ）
－ｄａｔａ．ｙ［６１］（ｔ））・・・（９）
ｙｙ２＋＝ｆａｂｓ（ｄａｔａ．ｙ［３３］（ｔ）
－ｄａｔａ．ｙ［６２］（ｔ））・・・（１０）
ｙｙ２＋＝ｆａｂｓ（ｄａｔａ．ｙ［３４］（ｔ）
－ｄａｔａ．ｙ［６３］（ｔ））・・・（１１）
ｙｙ２＋＝ｆａｂｓ（ｄａｔａ．ｙ［３４］（ｔ）
－ｄａｔａ．ｙ［６４］（ｔ））・・・（１２）

【００５８】

図４のステップＳ４０３で口開閉判定部３０７は次に、下記（１３）式により、時刻ｔのフレーム画像Ｆ（ｔ）に対して（１）式の演算で算出したｙ座標量差分総計ｙ（ｔ）と、１フレーム時刻前の時刻（ｔ－１）のフレーム画像Ｆ（ｔ－１）に対して（１）式と同様の演算で算出したｙ座標量差分総計ｙ（ｔ－１）との差分絶対値Δｙを求める。ここで、関数「ａｂｓ（）」は、括弧内の数値に対する絶対値を整数で算出する関数である。

【００５９】

Δｙ＝ａｂｓ（ｙ（ｔ）－ｙ（ｔ－１））・・・（１３）

【００６０】

（１３）式で算出されるΔｙは、口唇の移動量を示しており、上口唇と下口唇とが離れる方向もしくは近づく方向に移動している時に大きくなる。即ち、口開閉判定部３０７は、口唇移動量取得手段として動作する。

【００６１】

図４のステップＳ４０３で口開閉判定部３０７は、口唇の横座標（顔の左右方向）の変化Δｘについても、前記Δｙの場合と同様の演算で算出する。

【００６２】

即ち今、ある時刻のフレームＦ（ｔ）にて、下記（１４）式の演算によって、ｘ座標量差分総計ｘ（ｔ）が算出される。（１４）式で例えば、「ｄａｔａ．ｘ［６１］（ｔ）」は、時刻ｔのフレーム画像Ｆ（ｔ）内における図５のラベル６１番のｘ座標データ値を示す。他も同様である。

【００６３】

ｘ（ｔ）＝ｄａｔａ．ｘ［６１］（ｔ）＋ｄａｔａ．ｘ［６２］（ｔ）
＋ｄａｔａ．ｘ［６３］（ｔ）＋ｄａｔａ．ｘ［６７］（ｔ）

＋ｄａｔａ．ｘ［６６］（ｔ）＋ｄａｔａ．ｘ［６５］（ｔ）
・・・（１４）

【００６４】
次に、下記（１５）式により、時刻ｔのフレーム画像Ｆ（ｔ）に対して（１４）式の演算で算出したｘ座標量差分総計ｘ（ｔ）と、１フレーム時刻前の時刻（ｔ－１）のフレーム画像Ｆ（ｔ－１）に対し（１４）式と同様の演算で算出したｘ座標量差分総計ｘ（ｔ－１）との差分絶対値Δｘが算出される。

【００６５】
Δｘ＝ａｂｓ（ｘ（ｔ）－ｘ（ｔ－１））・・・（１５）

【００６６】
（１５）式で算出されるΔｘの値は、Δｙの場合と同様に口唇の移動量を示しており、口唇が左右どちらかに移動している時に大きくなる。この場合も口開閉判定部３０７は、口唇移動量取得手段として動作する。

【００６７】
図４のステップＳ４０３で口開閉判定部３０７は続いて、図１の頭部１０１の回転判定を行う。口開閉判定部３０７は、図１又は図２の首関節駆動部１０７から制御部２０１に入力する信号に基づいて、フレーム時刻ｔのフレーム画像Ｆ（ｔ）と、その１時刻前のフレーム時刻（ｔ－１）のフレーム画像Ｆ（ｔ－１）における、頭部姿勢の差分Δｒｏｌｌ、Δｙａｗ及びΔｐｉｔｃｈを、下記（１６）式、（１７）式及び（１８）式により算出する。

【００６８】
Δｒｏｌｌ＝ａｂｓ（Ｆ（ｔ）ｒｏｌｌ－Ｆ（ｔ－１）ｒｏｌｌ）・・・（１６）
Δｙａｗ＝ａｂｓ（Ｆ（ｔ）ｙａｗ－Ｆ（ｔ－１）ｙａｗ）・・・（１７）
Δｐｉｔｃｈ＝ａｂｓ（Ｆ（ｔ）ｐｉｔｃｈ－Ｆ（ｔ－１）ｐｉｔｃｈ）
・・・（１８）

【００６９】
ここで例えば、Ｆ（ｔ）ｒｏｌｌは、時刻ｔのフレーム画像Ｆ（ｔ）に対応して図１又は図２の首関節駆動部１０７から制御部２０１に入力するロール角度値Ｆ（ｔ－１）ｒｏｌｌは、時刻（ｔ－１）のフレーム画像Ｆ（ｔ－１）に対応して図１又は図２の首関節駆動部１０７から制御部２０１に入力するロール角度値である。ヨー角度値Ｆ（ｔ）ｙａｗ及びＦ（ｔ－１）ｙａｗ、ピッチ角度値Ｆ（ｔ）ｐｉｔｃｈ及びＦ（ｔ－１）ｐｉｔｃｈについても、それぞれ同様である。図６は、図１のロボット１００の頭部１０１の回転の自由度を模式的に表した図である。図１又は図２の首関節駆動部１０７により、ロボット１００の頭部１０１は、胴体１０２に対して、ピッチ軸Ｘｍの軸回り、ロール軸Ｚｍの軸回り、ヨー軸Ｙｍの軸回りにそれぞれ回転可能である。首関節駆動部１０７は、ピッチ軸Ｘｍの軸回りのピッチ角度値、ロール軸Ｚｍの軸回りのロール角度値及びヨー軸Ｙｍの軸回りのヨー角度値をそれぞれ、上記のようにして制御部２０１に出力する。

【００７０】
図４のステップＳ４０３で口開閉判定部３０７は、上記（１６）式、（１７）式及び（１８）式の演算の結果、ロール角度差分値Δｒｏｌｌ、ヨー角度差分値Δｙａｗ及びピッチ角度差分値Δｐｉｔｃｈを、頭部１０１の回転角度として算出する。この場合、口開閉判定部３０７は、頭部１０１＝口唇画像の回転量を取得する口唇回転量取得手段として動作する。

【００７１】
なお、頭部１０１の回転角度の推定方式としては様々な手法が知られており、上記以外の技術が採用されてもよい。

【００７２】
図４のステップＳ４０３で口開閉判定部３０７は、以上のようにして、口唇の縦座標の変化Δｙと、横座標の変化Δｘと、ロボット１００の頭部１０１の回転角度としてロール角度差分値Δｒｏｌｌ、ヨー角度差分値Δｙａｗ及びピッチ角度差分値Δｐｉｔｃｈに基

づいて、以下のルールにより口唇の開閉判定を行う。即ち、口開閉判定部３０７は、下記（１９）式の論理式で示される条件が満たされたときに、口唇の開口（口唇が開いている状態）を判定し、その条件が満たされないときに、口唇の閉口（口唇が閉じている状態）を判定する。なお、（１９）式において、第１の閾値であるｙ_ｔｈ、第２の閾値であるｘ_ｔｈ、並びに、第３の閾値群であるｒｏｌｌ_ｔｈ、ｙａｗ_ｔｈ及びｐｉｔｃｈ_ｔｈはそれぞれ、Δｙ、Δｘ、Δｒｏｌｌ、Δｙａｗ及びΔｐｉｔｃｈの判定閾値である。

【００７３】

$$\Delta y > y\_th \ \&\& \ \Delta x < x\_th \ \&\& \ \Delta roll < roll\_th \ \&\& \ \Delta yaw < yaw\_th \ \&\& \ \Delta pitch < pitch\_th \quad \cdots (19)$$

【００７４】

即ち、口開閉判定部３０７は、上口唇と下口唇とが離れる方向もしくは近づく方向に移動しており、口唇の横方向移動量は少なく、かつロボット１００の頭部１０１があまり回転していない場合に、口唇の開口を判定する。Δｙだけでなく、Δｘ、Δｒｏｌｌ、Δｙａｗ、及びΔｐｉｔｃｈも口唇の開閉判定に用いることにより、イヤイヤ（左右に首を振る）、考えるために首を傾げるといった動作でも、誤判定を起こりにくくすることができる。

【００７５】

図４の説明に戻り、上記ステップＳ４０３での一連の処理により口開閉判定部３０７により口唇の開口が判定されると、以下のステップＳ４０４からステップＳ４０６までの一連の処理が実行される。

【００７６】

まず、図３の音源到来方向推定部３０２が、信号音声の到来方向の推定処理として、図３の顔検出部３０５により検出されている顔画像（＝口唇画像）の口唇方向に基づいて、ロボット１００（のカメラ１０４）に対する口唇方向角度Ｓ_ａｎｇを算出する処理を実行する（ステップＳ４０４）。

【００７７】

続いて、図３の音源分離部３０３が、例えば前述した文献１に記載されているビームフォーミングの演算処理により、ステップＳ４０４で算出された口唇方向角度Ｓ_ａｎｇの方向（第１方向）にビームステアリング（強調）するビームステアリング演算処理を実行することにより、強調された信号音声を得る（ステップＳ４０５）。

【００７８】

そして、図３の音量算出部３０８が、ステップＳ４０５で得られたビームステアリング（強調）された信号音声の音量Ｓpow を算出する（ステップＳ４０６）。

【００７９】

一方、ステップＳ４０３での一連の処理により口開閉判定部３０７により口唇の閉口が判定されると、以下のステップＳ４０７からステップＳ４０９までの一連の処理が実行される。

【００８０】

まず、図３の音源到来方向推定部３０２が、音源定位手法の一手法であるＭＵＳＩＣ法に基づく処理を実行することにより、対象者以外の音源からの雑音音声の音源定位（雑音源の位置）を推定してノイズ方向角度Ｎ_ａｎｇを決定する処理を実行する（ステップＳ４０７）。この処理の詳細については、後述する。

【００８１】

続いて、図３の音源分離部３０３が、例えば前述した文献１に記載されているビームフォーミングの演算処理により、ステップＳ４０７で算出されたノイズ方向角度Ｎ_ａｎｇの方向（第２方向）にヌルステアリング（抑圧）するヌルステアリング演算処理を実行す

ることにより、抑圧された雑音音声を得る（ステップＳ４０８）。

【００８２】

そして、図３の音量算出部３０８が、ステップＳ４０８で得られたヌルステアリング（抑圧）された雑音音声の音量Ｎpow を算出する（ステップＳ４０９）。

【００８３】

その後、図３のＳ／Ｎ比算出部３０９が、ステップＳ４０６で算出された信号音声の音量Ｓpow とステップＳ４０９で算出された雑音音声の音量Ｎpow とに基づいて、下記（２０）式の演算に基づいて、Ｓ／Ｎ比を算出する。

【００８４】

Ｓ／Ｎ比＝Ｓpow ／Ｎpow ・・・（２０）

【００８５】

更に、Ｓ／Ｎ比算出部３０９が、下記（２１）式の判定演算に基づいて、算出したＳ／Ｎ比が閾値ｓｎ_ｔｈよりも大きいか否かを判定する（ステップＳ４１０）。

【００８６】

Ｓ／Ｎ比＞ｓｎ_ｔｈ・・・（２１）

【００８７】

ステップＳ４１０の判定がＮＯの場合には、図２の制御部２０１は、音声認識のための十分なＳ／Ｎ比が得られていないと判定する。この場合、制御部２０１は例えば、図１又は図２の足回り駆動部１０８を制御することにより、例えば対象者に対して一定の関係（例えば一定の距離又は一定の角度等）を維持しながら、ロボット１００を移動させる（ステップＳ４１１）。移動処理の詳細については、後述する。

【００８８】

ロボット１００の移動の後再び、図４のステップＳ４０１からステップＳ４０９までの一連の制御処理が実行され、ステップＳ４１０のＳ／Ｎ比の判定が行われる。

【００８９】

やがて、ステップＳ４１０の判定がＹＥＳになると、図２の制御部２０１は、音声認識のための十分なＳ／Ｎ比が得られ、対象者に対するロボット１００の位置関係が、信号音声を雑音音声から最も良く分離できる最適化された位置である音源分離位置になったと判定する。この場合、制御部２０１は、図３の音声認識部３１０に、音源分離部３０３が出力するビームステアリング（強調）された信号音声に対する音声認識処理を実行させることにより、対象者の発話内容を理解する。更に、制御部２０１は、この音声認識結果に応じて、対話アルゴリズムに従って、図３の発声部３１１から図１又は図２のスピーカ１０５を介して、対象者に対して発声を行って対話をする（以上、ステップＳ４１２）。対話終了後、図２の制御部２０１は、図４のフローチャートで示される制御処理を終了する。

【００９０】

図７は、図４のステップＳ４０３での一連の処理により口開閉判定部３０７により口唇の閉口が判定された場合に、ステップＳ４０７で図３の音源到来方向推定部３０２によりＭＵＳＩＣ法に基づいて実行される、対象者以外の音源からの雑音音声の音源定位（雑音源の位置）を推定してノイズ方向角度Ｎ_ａｎｇを決定する処理の詳細例を示すフローチャートである。

【００９１】

まず、図１又は図２のマイクアレイ１０３に入力された音声が、時間周波数変換される（ステップＳ７０１）。ここでは例えば、時間周波数変換演算処理として、ＳＴＦＴ（Ｓｈｏｒｔ－Ｔｉｍｅ　Ｆｏｕｒｉｅｒ　Ｔｒａｎｓｆｏｒｍ：短時間フーリエ変換）が実行される。

【００９２】

音源数をＮとすると、第ｎ番目の音源の信号Ｓｎは、下記（２２）式で表せる。なお、ωは角周波数、ｆはフレーム番号である（以下の説明でも同様）。

【００９３】

Ｓｎ（ω，ｆ）（ｎ＝１，２，…，Ｎ）・・・（２２）

【００９４】
図１又は図２のマイクアレイ１０３の各マイクで観測される信号は、マイクアレイ１０３におけるマイクの数をＭとすると、下記（２３）式で表せる。
【００９５】
Ｘｍ（ω，ｆ）（ｍ＝１，２，・・・，Ｍ）・・・（２３）
【００９６】
音源から出た音は、空気を伝わってマイクアレイ１０３のマイクで観測されるが、そのときの伝達関数をＨｎｍ（ω）とすると、音源の信号を表す数式に、伝達関数を乗じることで、マイクアレイ１０３の各マイクで観測される信号を求めることができる。ｍ番目のマイクで観測される信号Ｘｍ（ω，ｆ）は下記（２４）式のように表される。
【００９７】
【数１】

$$X_m(\omega,f)=\sum_{n=1}^{N}S_n(\omega,f)H_{nm}(\omega) \quad \cdots(24)$$

【００９８】
ロボット１００は、マイクアレイ１０３としてマイクを複数有しているので、マイクアレイ１０３全体で観測される信号ｘ（ω，ｆ）は下記（２５）式で表すことができる。
【００９９】
【数２】

$$x(\omega,f)=\begin{bmatrix}X_1(\omega,f)\\X_2(\omega,f)\\\vdots\\X_M(\omega,f)\end{bmatrix} \quad \cdots(25)$$

【０１００】
同様に、全音源の信号ｓ（ω，ｆ）も下記（２６）式で表すことができる。
【０１０１】
【数３】

$$s(\omega,f)=\begin{bmatrix}S_1(\omega,f)\\S_2(\omega,f)\\\vdots\\S_N(\omega,f)\end{bmatrix} \quad \cdots(26)$$

【０１０２】
同様に、第ｎ番目の音源の伝達関数ｈｎ（ω）は下記（２７）式で表すことができる。
【０１０３】

【数４】

$$h_n(\omega)=\begin{bmatrix} H_{n1}(\omega) \\ H_{n2}(\omega) \\ \vdots \\ H_{nM}(\omega) \end{bmatrix} \quad \cdots(27)$$

【０１０４】

全ての伝達関数を下記（２８）式のように表記する。

【０１０５】

ｈ（ω）＝［ｈ１（ω），ｈ２（ω），…ｈＮ（ω）］・・・（２８）

【０１０６】

（２８）式で表される伝達関数を、前述した（２４）式に適用すると、下記（２９）式のように表される。

【０１０７】

ｘ（ω，ｆ）＝ｈ（ω）ｓ（ω，ｆ）・・・（２９）

【０１０８】

ｈｎ（ω）は音源位置毎に独立であり、ある程度のフレーム数（例えば、フレーム数をＬとする））で見ればＳｎ（ω，ｆ）は無相関とみなせるので、ｘ（ω，ｆ）は音源数ＮをＲＡＮＫとする超平面を構成する。このとき、距離で正規化した音量が大きな音源の伝達関数方向に分布が広がりやすい。そこで、部分空間とゼロ空間とに分解することを考える。

【０１０９】

再び図７を参照する。次の（３０）式に示されるように、相関行列が計算される（ステップＳ７０２）。ここで、「＊」は複素共役転置を表す。

【０１１０】

【数５】

$$R(\omega,f)=\sum_{l=0}^{L-1} x(\omega,f+l)x^{*}(\omega,f+l) \quad \cdots(30)$$

【０１１１】

続いて、固有値分解が実行される（ステップＳ７０３）。ここで、固有値λｍ（ω，ｆ）と固有ベクトルｅｍ（ω，ｆ）とは固有値が降順になるように並べ替えられているものとする

。

【０１１２】

原理的には、ｈｎ（ω）は部分空間の固有ベクトルｅｍ（ω，ｆ）（ｍ＝１～Ｎ）の重み付け加算から復元できるが、実際には復元が困難であるためゼロ空間を構成する固有ベクトルｅｍ（ω，ｆ）（ｍ＝Ｎ＋１～Ｍ）がｈｎ（ω）と直交することを使って音源定位を実現する。

【０１１３】

しかし、雑音音声の音源は例えば建物室内を移動する可能性があるため、音源位置を予め知ることはできず、音源位置の伝達関数を予め取得しておくことは難しい。このため、仮の音源位置が決められ、仮の音源位置の伝達関数が予め用意されて、音源定位が行われ

る。

【０１１４】

図８は、仮の音源位置とマイクの配置との一例を示す図である。図８では、太線の円がロボット１００の頭１１０を表し、太線上の黒丸がマイクアレイ１０３のマイクを表す。なお、ここでは、便宜上図１のマイクアレイ１０３の１３個のマイクの全てを表示していない。ロボット１００の回りには４個の仮の音源位置があるものとする。

【０１１５】

マイクアレイ１０３の複数のマイクは、ロボット１００の頭１１０に配置されていることから、円周に沿って配置されているとみなすことができる。Ｘ軸の正の向きと、各マイクが成す円の中心（ロボット１００の頭１１０の中心位置に相当）と仮の音源１～４とをそれぞれ結んだ線と、がなす角度をθ１、θ２、θ３、θ４として、それぞれの伝達関数ｈθ（ω）を予め計算しておく。

【０１１６】

図８では、音源が４個の例を示したが、音源数がＮ個の場合、θ１、θ２、・・・、θＮのそれぞれの伝達関数ｈθ（ω）を予め計算しておけばよい。或いは、仮の音源位置の伝達関数を用意するのではなく、幾何的な情報をもとに予め伝達関数を計算しておいてもよい。

【０１１７】

再び図７を参照する。下記（３１）式を使用して、周波数帯毎のＭＵＳＩＣスペクトルが計算される（ステップＳ７０４）。

【０１１８】

【数６】

$$M_{\theta}(\omega,f)=\frac{h^{*}_{\theta}(\omega)h_{\theta}(\omega)}{\sum_{m=N+1}^{M}\left|h^{*}_{\theta}(\omega)e_{m}(\omega.f)\right|^{2}} \quad \cdots(31)$$

【０１１９】

ここで、（３１）式の分母は、ノイズや誤差、ＳＴＦＴの周波数帯間の信号漏洩の影響等からゼロにはならない。また、音源の方向と予め決めた角度θ（θ１、θ２、・・・、θＮ）の何れかが近い場合、つまりｈｎ（ω）とｈθ（ω）とが近い場合、（３１）式の値は極端に大きなものになる。図８に示す例では、雑音音声の音源と仮の音源の位置とが近いため、θ２の伝達関数を使用した場合、（３１）式の値が極端に大きくなることが想定される。

【０１２０】

次に、統合したＭＵＳＩＣのパワーを求めるため、下記（３２）式の演算により、周波数帯毎のＭＵＳＩＣスペクトルが重み付け加算される（ステップＳ７０５）。

【０１２１】

【数７】

$$M(f)=\sum_{\omega}w(\omega)M\ (\omega,f) \quad \cdots(32)$$

【０１２２】

重み付け係数は、固有値λｍ（ω，ｆ）が大きいほど大きくすれば、Ｓｎ（ω，ｆ）に含まれるパワーに応じた計算をすることもできる。この場合はＳｎ（ω，ｆ）に殆どパワーがない場合の悪影響を軽減できる。

【０１２３】

最後に、パワースペクトルから適切なピーク（極大値）が選択される（ステップＳ７０

６）。具体的には、まず、複数のピークが算出され、その中から適切なピークが選択されて、選択されたピークにおけるθが図４のステップＳ４０７で説明した雑音音声の音源方向のノイズ方向角度Ｎ_ａｎｇとされる。ここで、ピークを求めるのは以下のような理由による。本来の音源方向のθのパワーが必ずしも一番大きいとは限らず、本来の音源方向に近いθのパワーは総じて大きくなるので、音源方向は複数のピークの何れかに正解があるからである。その後、図７のフローチャートの処理が終了して、図４のステップＳ４０７の雑音音声の音源到来方向推定処理が終了する。

【０１２４】

以上の説明では、雑音音声の音源到来方向として平面を仮定して説明したが、３次元空間を仮定しても上記説明は成り立つ。

【０１２５】

図９は、図１又は図２のマイクアレイ１０３の指向特性の例を示す図、図１０から図１２は、音源分離方向検出の説明図である。図９において、マイクアレイ１０３は、１２０度付近で、各周波数においてまんべんなくマイナスゲインが得られている。従って、下記（３３）式のように、図４のステップＳ４０４で算出される口唇方向角度である対象者の信号音声方向Ｓ_ａｎｇと、ステップＳ４０７で算出されるノイズ方向角度Ｎ_ａｎｇとの差分の絶対値が１２０度付近になる音源分離方向が、最も良い音源分離が期待できる方向となる。

【０１２６】

ａｂｓ（Ｓ_ａｎｇ－Ｎ_ａｎｇ）・・・（３３）

【０１２７】

図４のステップＳ４１０の判定がＮＯ→Ｓ４１１→Ｓ４０１として実行される処理により実現されるアルゴリズムとしては、ロボット１００の位置毎に、前述した（２０）式により算出される信号音声対雑音音声のＳ／Ｎ比が前述した（２１）式の判定演算により閾値ｓｎ_ｔｈを超えたか否かが判定されながら、前述した図４のステップＳ４１１でのロボット１００の移動処理が繰り返され、Ｓ／Ｎ比が閾値ｓｎ_ｔｈを超えたと判定された地点が、信号音声と雑音音声との最適な分離位置、即ち音源分離位置とされる。

【０１２８】

なお、Ｓ／Ｎ比が閾値ｓｎ_ｔｈを超えた時点ではなく、閾値を超えた後にＳ／Ｎ比が最高となる地点が音源分離位置とされてもよい。

【０１２９】

例えば、図１０は、Ｓ_ａｎｇ＝０度、Ｎ_ａｎｇ＝１０度と算出された状態の例を示している。ロボット１００のカメラ１０４（図１）から見ると、対象者の右１０度方向にノイズの音源が存在していることになる。この状態から、図２の制御部２０１は、図１又は図２の足回り駆動部１０８を制御することにより、ロボット１００を、対象者を中心にして例えば右方向（図１０の方向Ａ）に移動させる。左方向への移動が行われてももちろん良い。ただし、図１０の例の場合は、右移動の方が音源分離位置に最短距離で近づくことができる。図１１は、上記移動後のロボット１００と対象者とノイズ音源の位置関係を示す図である。このような移動が繰り返されることにより、ロボット１００は最終的に、図１２に示される音源分離位置まで移動をして、移動を完了する。この音源分離位置は、マイクアレイ１０３が図９に示される指向特性を有する場合に、（３３）式で算出される信号音声方向Ｓ_ａｎｇとノイズ方向角度Ｎ_ａｎｇとの差分の絶対値が１２０度付近になる位置である。

【０１３０】

上述の動作において、図２の制御部２０１は、移動開始時に「聞き取りやすい位置に移動するね」などの音声を、図３の発声部３１１から発声させることにより、移動中は対話を中止してもらえるような文言を喋らせることが望ましい。また、移動中にも対話できるようにしてもよい。

【０１３１】

上述した図４のフローチャートで例示される制御処理において、Ｓ／Ｎ比算出部３０９

でのステップＳ４１０の判定の結果、Ｓ／Ｎ比が閾値ｓｎ_ｔｈ以下である場合に、制御部２０１は例えば、図３の発声部３１１を介して図１又は図２のスピーカ１０５から、対象者に対して、「僕を中心にして・・・度ほど回転するように移動してください。」というような意味の発声を行って、対象者に移動を促すような制御が行われてもよい。

【０１３２】

また、上述のような発声を行いながら、継続的に取得したノイズ方向角度Ｎ_ａｎｇが都合の良い角度になるまで、「もう少し」や「ストップ」などの発声を行って対象者に指示をするような制御が行われてもよい。

【０１３３】

例えば建物室内のマップ情報を利用できる場合には、対象者やノイズの２次元又は３次元の音源位置をマップ上で推定し、その推定結果に基づいて音源分離位置に移動するような制御が実施されてもよい。音源位置のマップは、ノイズ音源になるべく近づいてその位置を特定して登録するようにしてもよい。

【０１３４】

一方、音源位置のマップが無い場合には、ロボット１００の移動中に獲得したノイズ方向とその時の位置とロボット１００本体の向きとから、ノイズの音源位置を推定するようにしてもよい。この場合、観測点が２点以上あれば音源位置が決められる。推定方向にある程度の誤差を持たせて、より多くの観測点から推定が行われるようにしてもよい。

【０１３５】

更に、上記のようなマップ情報を使ったノイズの音源位置の推定結果に基づいて、「あと・・・度回転して」というような発声を行って対象者に指示をするような制御が行われてもよい。

【０１３６】

上述した実施形態において、ロボット１００が移動するときに、ロボット１００がそっぽを向きながら移動したり、ロボット１００が勝手に動いていると、対象者が違和感を感じてしまうため、対象者が違和感を感じないように、移動することが望ましい。例えば、対象者とアイコンタクトを取ったり、対象者の方を向きながら移動することが望ましい。また、音源分離位置まで一気に移動するのではなく、少しだけ動いたり、回転のみをしてもよい。

【０１３７】

以上の実施形態によれば、信号音声が雑音音声から最も良く分離した状態で音源分離が行える最適化された音源分離情報（音源分離方向又は音源分離位置）を検出することが可能となる。これにより、対象者の音声以外の他の音声を排除して、音声認識の誤認識を減らすことが可能となる。

【０１３８】

以上説明した実施形態において、図２の制御部２０１が記憶部２０２に記憶され図４や図７のフローチャートの処理例で示される制御プログラム２０５を実行することにより図３で示される機能を実現する場合、制御プログラム２０５は、例えば外部記憶装置や可搬記録媒体に記録して配布してもよく、あるいは特には図示しない無線や有線の通信インタフェースを介してネットワークから取得できるようにしてもよい。

【０１３９】

以上の実施形態に関して、更に以下の付記を開示する。

（付記１）

音声を取得するために所定の指向性を有する音声取得手段と、

前記音声取得手段により取得された所定の対象の信号音声から、前記信号音声の到来方向である第１方向を検出する第１方向検出手段と、

前記音声取得手段により取得された雑音音声から、前記雑音音声の到来方向である第２方向を検出する第２方向検出手段と、

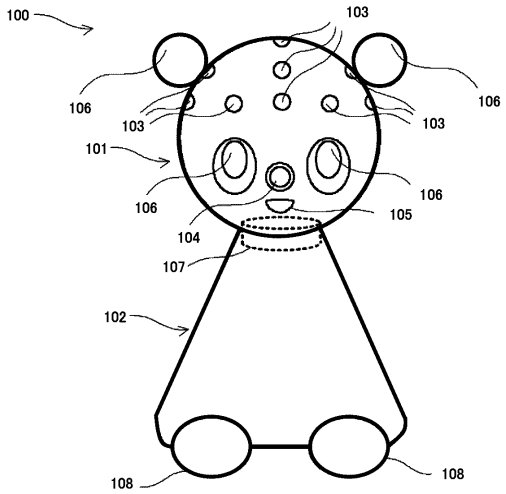
前記第１方向と前記第２方向とに基づいて、音源分離方向又は音源分離位置を検出する検出手段と、

を備える、
ことを特徴とする音源分離情報検出装置。
（付記２）
前記検出手段は、前記信号音声と前記雑音音声とから算出された信号対雑音比が閾値以下である場合、前記第１方向と前記第２方向とに基づいて、前記信号対雑音比が前記閾値を超える前記音源分離方向又は前記音源分離位置を検出する、
ことを特徴とする付記１に記載の音源分離情報検出装置。
（付記３）
前記音声取得手段が前記音声を取得するタイミングで前記所定の対象の口唇画像を取得する画像取得手段と、
前記口唇画像に基づいて、前記所定の対象の口唇の開口又は前記口唇の閉口を判定する判定手段と、
を更に備え、
前記第１方向検出手段は、前記判定手段による前記口唇の開口の判定時に、前記音声取得手段により取得された前記音声を前記信号音声とし、
前記第２方向検出手段は、前記判定手段による前記口唇の閉口の判定時に、前記音声取得手段により取得された前記音声を前記雑音音声とする、
ことを特徴とする付記１又は２に記載の音源分離情報検出装置。
（付記４）
前記画像取得手段は、顔部画像を更に取得し、
前記口唇画像から前記所定の対象の前記口唇の移動量を取得する口唇移動量取得手段と、
前記顔部画像から前記所定の対象の顔部の回転量を取得する顔部回転量取得手段と、
を更に備え、
前記判定手段は、前記口唇の移動量と前記顔部の回転量とに基づいて、前記所定の対象の前記口唇の開口又は前記口唇の閉口を判定する、
ことを特徴とする付記３に記載の音源分離情報検出装置。
（付記５）
前記判定手段は、前記口唇の移動量のうちの前記口唇の開閉方向の移動量が第１の閾値を超え、且つ、前記口唇の移動量のうちの前記口唇の延伸方向の移動量が第２の閾値未満であり、且つ、前記顔部の回転量が第３の閾値未満であるときに、前記口唇の開口又は前記口唇の閉口を判定する、
ことを特徴とする付記４に記載の音源分離情報検出装置。
（付記６）
前記第１方向検出手段は、前記判定手段による前記口唇の開口の判定時に、前記信号音声の信号音声パワーに基づいて、前記第１方向を検出し、
前記第２方向検出手段は、前記判定手段による前記口唇の閉口の判定時に、前記雑音音声の雑音音声パワーに基づいて、前記第２方向を検出する、
ことを特徴とする付記３乃至５の何れか１つに記載の音源分離情報検出装置。
（付記７）
前記検出手段は、前記信号対雑音比が前記閾値を超えて最大となる方向を前記音源分離方向とする、又は、前記信号対雑音比が前記閾値を超えて最大となる位置を前記音源分離位置とする、
ことを特徴とする付記２乃至６の何れか１つに記載の音源分離情報検出装置。
（付記８）
前記検出手段は、前記信号対雑音比が前記閾値を超える場合、現在方向を前記音源分離方向とする、又は、現在位置を前記音源分離位置とする、
ことを特徴とする付記２乃至６の何れか１つに記載の音源分離情報検出装置。
（付記９）
前記所定の対象にメッセージを報知する報知手段を更に備え、

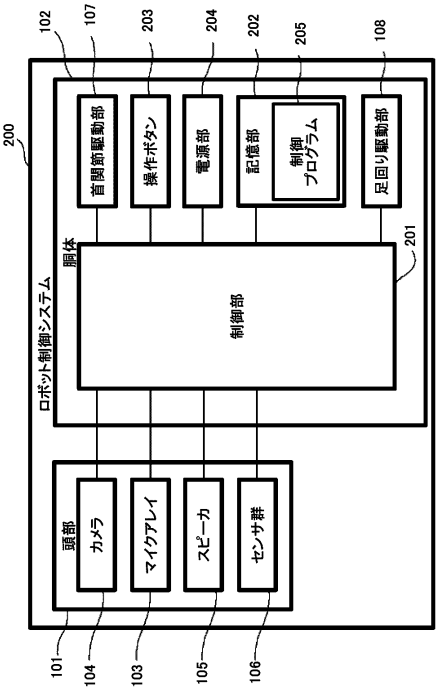
前記報知手段は、前記所定の対象に現在位置から前記音源分離位置まで移動させるために、前記音源分離位置までの移動方向及び移動距離を含む前記メッセージを報知する、
ことを特徴とする付記１乃至８の何れか１つに記載の音源分離情報検出装置。
（付記１０）
前記所定の対象は人又は動物である、
ことを特徴とする付記１乃至９の何れか１つに記載の音源分離情報検出装置。
（付記１１）
付記１乃至１０の何れかに記載の音源分離情報検出装置と、
自装置を移動させる移動手段と、
前記音源分離情報検出装置及び前記移動手段を制御する制御手段と、
を備える、
ことを特徴とするロボット。
（付記１２）
前記制御手段は、前記移動手段を制御して、前記音源分離位置に前記自装置を移動させる、
ことを特徴とする付記１１に記載のロボット。
（付記１３）
音声を取得するために所定の指向性を有する音声取得手段により取得された所定の対象の信号音声から、前記信号音声の到来方向である第１方向を検出し、
前記音声取得手段により取得された雑音音声から、前記雑音音声の到来方向である第２方向を検出し、
前記第１方向と前記第２方向とに基づいて、音源分離方向又は音源分離位置を検出する、
ことを含む、
ことを特徴とする音源分離情報検出方法。
（付記１４）
音源分離情報検出装置のコンピュータを、
音声を取得するために所定の指向性を有する音声取得手段により取得された所定の対象の信号音声から、前記信号音声の到来方向である第１方向を検出し、
前記音声取得手段により取得された雑音音声から、前記雑音音声の到来方向である第２方向を検出し、
前記第１方向と前記第２方向とに基づいて、音源分離方向又は音源分離位置を検出する、
ように機能させる、
ことを特徴とするプログラム。
【符号の説明】
【０１４０】
１００　ロボット
１０１　頭部
１０２　胴体
１０３　マイクアレイ
１０４　カメラ
１０５　スピーカ
１０６　センサ群
１０７　首関節駆動部
１０８　足回り駆動部
２００　ロボット制御システム
２０１　制御部
２０２　記憶部
２０３　操作ボタン

２０４　電源部
２０５　制御プログラム
３０１　音声入力部
３０２　音源到来方向推定部
３０３　音源分離部
３０４　画像入力部
３０５　顔検出部
３０６　口パーツ検出部
３０７　口開閉判定部
３０８　音量算出部
３０９　Ｓ／Ｎ比算出部
３１０　音声認識部
３１１　発声部

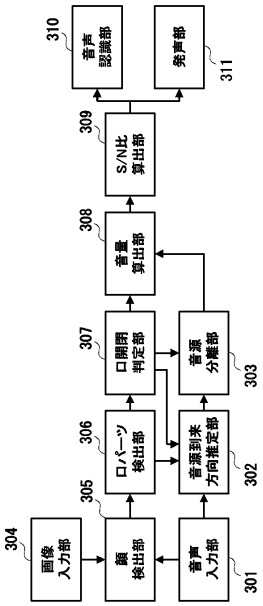
【図１】



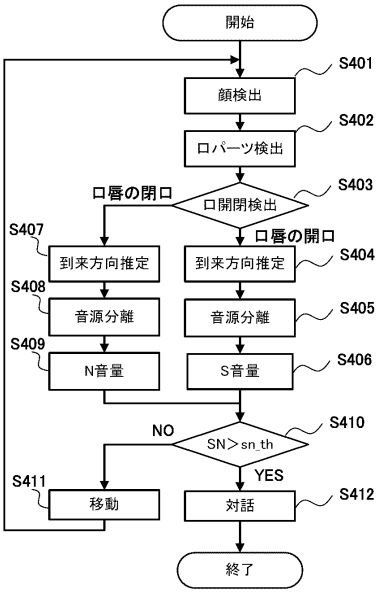
【図２】



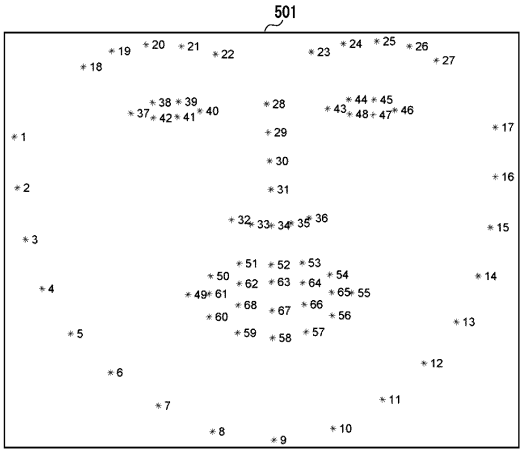
【図３】


310
音声認識部
発声部
311
309
S/N比算出部
308
音量算出部
307
口開閉判定部
音源分離部
303
306
口パーツ検出部
音源到来方向推定部
302
305
顔検出部
音声入力部
301
304
画像入力部

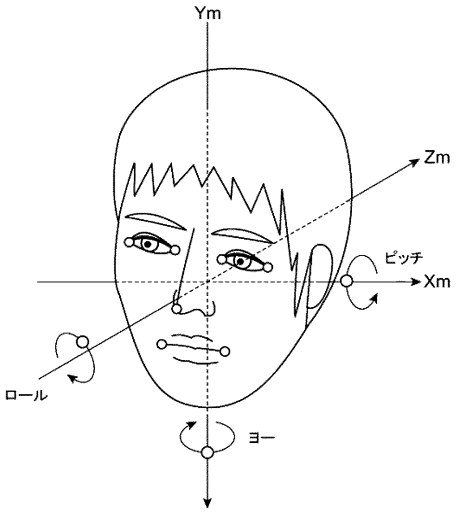
【図４】


開始
S401
顔検出
S402
口パーツ検出
S403
口開閉検出
口唇の閉口
口唇の開口
S407
到来方向推定
到来方向推定
S404
S408
音源分離
音源分離
S405
S409
N音量
S音量
S406
NO
S410
SN＞sn_th
YES
S411
移動
対話
S412
終了

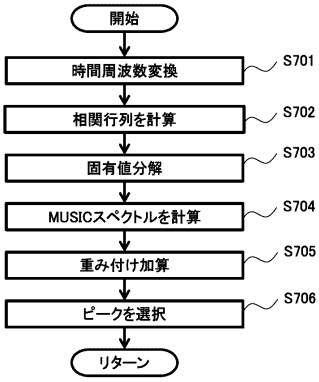
【図５】


501

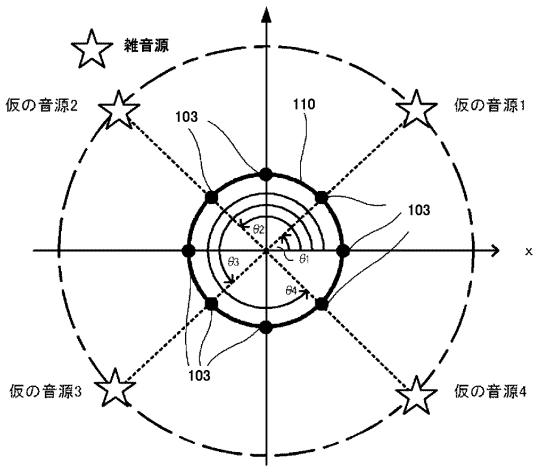
【図６】


Ym
Zm
ピッチ
Xm
ロール
ヨー

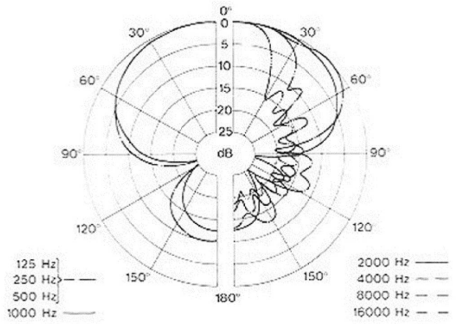
【図７】



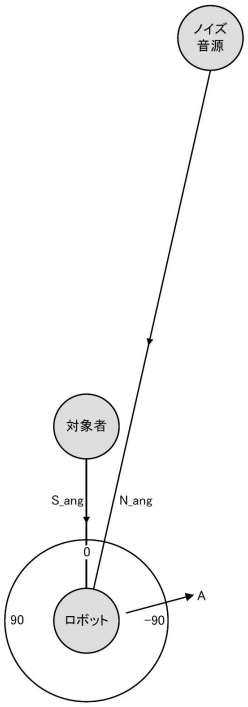
【図８】



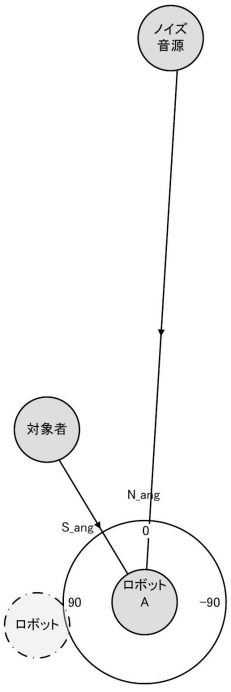
【図９】



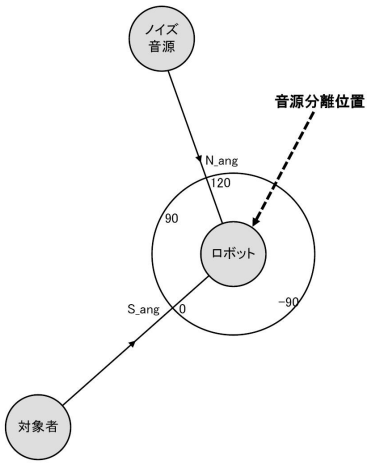
【図１０】



【図１１】



【図１２】



フロントページの続き

審査官　山下　剛史

(56)参考文献　特表２００５－５２９４２１（ＪＰ，Ａ）
特開２００６－１８１６５１（ＪＰ，Ａ）
特開２０１４－２０７５８９（ＪＰ，Ａ）
特開２００５－２５３０７１（ＪＰ，Ａ）
藤田善弘，パーソナルロボットＲ１００，日本ロボット学会誌，２０００年　3月，Vol.18，No.2，p.198-199

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　１３／００－９９／００