(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第5572445号
(P5572445)
(45)発行日　平成26年8月13日(2014.8.13)　(24)登録日　平成26年7月4日(2014.7.4)

(51)Int.Cl.　F I
G10L 21/0208 (2013.01)　G10L 21/0208 100B
G10L 21/0272 (2013.01)　G10L 21/0272 100B

請求項の数 8　(全 22 頁)

(21)出願番号　特願2010-105369 (P2010-105369)
(22)出願日　平成22年4月30日 (2010.4.30)
(65)公開番号　特開2011-232691 (P2011-232691A)
(43)公開日　平成23年11月17日 (2011.11.17)
審査請求日　平成24年11月27日 (2012.11.27)

(73)特許権者　000005326
本田技研工業株式会社
東京都港区南青山二丁目１番１号
(74)代理人　100064908
弁理士　志賀　正武
(74)代理人　100108578
弁理士　高橋　詔男
(74)代理人　100146835
弁理士　佐伯　義文
(74)代理人　100094400
弁理士　鈴木　三義
(74)代理人　100107836
弁理士　西　和哉
(74)代理人　100108453
弁理士　村山　靖彦

最終頁に続く

(54)【発明の名称】残響抑圧装置、及び残響抑圧方法

(57)【特許請求の範囲】
【請求項１】
音声信号を取得する音声取得部と、
音声信号を出力する音声出力部と、
前記音声出力部から既知の音声信号を出力させたときに、前記音声取得部によって取得された残響音を含む前記既知の音声信号と残響音を含まない当該既知の音声信号とに基づいて、反響音をキャンセルする分離行列である反響音キャンセル分離行列を算出し、当該音声取得部によって取得された当該残響音を含む既知の音声信号と当該残響音を含まない既知の音声信号と算出した前記反響音キャンセル分離行列とに基づいてブラインド音源分離における分離行列であるブラインド分離行列とブラインド残響除去における除去行列であるブラインド残響除去行列とを算出する残響データ演算部と、
前記残響データ演算部によって算出された前記反響音キャンセル分離行列に基づき残響抑圧を行うフィルタのフィルタ長を推定するフィルタ長推定部と、
前記フィルタ長推定部によって推定されたフィルタ長と前記残響データ演算部によって算出された反響音キャンセル分離行列と前記ブラインド分離行列と前記ブラインド残響除去行列とに基づき残響抑圧を行う残響抑圧部と、
を備えることを特徴とする残響抑圧装置。
【請求項２】
前記音声取得部によって取得された残響音を含む音声信号を周波数領域の第１信号に変換し、前記残響音を含まない既知の音声信号を周波数領域の第２信号に変換する短時間フ

ーリエ解析部
を備え、
前記残響データ演算部は、
前記短時間フーリエ解析部から入力された前記第１信号を周波数毎に空間球面化を行って第３信号を算出し、前記短時間フーリエ解析部から入力された前記第２信号を周波数毎にスケールの正規化を行うことで第４信号を算出し、算出された前記第３信号と前記第４信号と前記反響音キャンセル分離行列とを用いて前記ブラインド分離行列と前記ブラインド残響除去行列とを算出し、
前記残響抑圧部は、
前記フィルタ長推定部によって推定された前記フィルタ長に基づき、前記第３信号と前記第４信号と前記ブラインド分離行列と前記ブラインド残響除去行列と前記反響音キャンセル分離行列とを用いて、繰り返し独立成分分析処理を行い、独立成分分析処理が行われた演算結果に対してスケーリングを行い、スケーリングを行った信号からパワーが最大のものを選択することで前記音声取得部が取得した前記残響音を含む音声信号から前記残響音を抑圧する残響抑圧を行う
ことを特徴とする請求項１に記載の残響抑圧装置。
【請求項３】
前記算出された前記反響音キャンセル分離行列、前記ブラインド分離行列、及び前記ブラインド残響除去行列に基づき残響時間を推定する残響特性推定部、を備え、
前記フィルタ長推定部は、
前記推定された残響時間に基づき前記フィルタ長を推定する
ことを特徴とする請求項１または請求項２に記載の残響抑圧装置。
【請求項４】
前記フィルタ長推定部は、
直接音と間接音との比率に基づき前記フィルタ長を推定する
ことを特徴とする請求項１または請求項２に記載の残響抑圧装置。
【請求項５】
当該残響抑圧装置の位置が変化したことを検出する検出部、
を更に備え、
残響データ演算部は、
前記残響抑圧装置の位置が変化したことを検出した場合に前記反響音キャンセル分離行列、前記ブラインド分離行列、及び前記ブラインド残響除去行列を演算する
ことを特徴とする請求項１から請求項４のいずれか１項に記載の残響抑圧装置。
【請求項６】
前記検出部は、
前記残響抑圧装置の位置が変化したことを検出した場合に、前記残響抑圧部が残響抑圧に用いるパラメータ、あるいは、前記フィルタ長推定部がフィルタ長推定に用いるパラメータの少なくとも一方のパラメータを前記残響抑圧装置の位置が変化したことに基づき切り替える
ことを特徴とする請求項５に記載の残響抑圧装置。
【請求項７】
テスト音声信号を出力する音声出力部、
を更に備え、
前記音声取得部は、前記出力されたテスト音声信号を取得し、残響データ演算部は、前記取得されたテスト音声信号から前記反響音キャンセル分離行列、前記ブラインド分離行列、及び前記ブラインド残響除去行列を演算する
ことを特徴とする請求項１から請求項６のいずれか１項に記載の残響抑圧装置。
【請求項８】
残響抑圧装置の残響抑圧方法において、
音声取得部が、音声信号を取得する音声取得工程と、

残響データ演算部が、テスト音声信号を出力する音声出力部から既知の音声信号を出力させたときに、前記音声取得工程によって取得された残響音を含む前記既知の音声信号と残響音を含まない当該既知の音声信号とに基づいて、反響音をキャンセルする分離行列である反響音キャンセル分離行列を算出し、当該音声取得部によって取得された当該残響音を含む既知の音声信号と当該残響音を含まない既知の音声信号と算出した前記反響音キャンセル分離行列とに基づいてブラインド音源分離における分離行列であるブラインド分離行列とブラインド残響除去における除去行列であるブラインド残響除去行列とを算出する残響データ演算工程と、

フィルタ長推定部が、前記残響データ演算工程によって算出された前記反響音キャンセル分離行列に基づき残響抑圧を行うフィルタのフィルタ長を推定するフィルタ長推定工程と、

残響抑圧部が、前記フィルタ長推定工程によって推定されたフィルタ長と前記残響データ演算工程によって算出された反響音キャンセル分離行列と前記ブラインド分離行列と前記ブラインド残響除去行列とに基づき残響抑圧を行う残響抑圧工程と、

を備えることを特徴とする残響抑圧方法。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、残響抑圧装置、及び残響抑圧方法に関する。

【背景技術】

【０００２】

残響抑圧処理は，遠隔会議通話または補聴器における明瞭度の向上およびロボットの音声認識（ロボット聴覚）に用いられる自動音声認識の認識率の向上を目的として、自動音声認識の前処理として利用されている重要な技術である。残響抑圧処理において、所定のフレーム毎に、取得した音声信号から残響成分を算出し、取得した音声信号から算出した残響成分を除去することで残響を抑圧していた（例えば、特許文献１参照）。

【先行技術文献】

【特許文献】

【０００３】

【特許文献１】特開平９－２６１１３３号公報

【発明の概要】

【発明が解決しようとする課題】

【０００４】

しかしながら、特許文献１の従来技術では、所定のフレーム長さにおいて残響抑圧を行っていたため、フレーム長が長い場合は処理に時間がかかりすぎるという問題点があり、フレーム長が短すぎると十分な残響抑圧の効果が得られにくいという問題点があった。

【０００５】

本発明は、上記の問題点に鑑みてなされたものであって、精度良く残響抑圧を行える残響抑圧装置及び残響抑圧方法を提供することを課題としている。

【課題を解決するための手段】

【０００６】

上記目的を達成するため、本発明に係る残響抑圧装置は、音声信号を取得する音声取得部と、音声信号を取得する音声取得部と、音声信号を出力する音声出力部と、前記音声出力部から既知の音声信号を出力させたときに、前記音声取得部によって取得された残響音を含む前記既知の音声信号と残響音を含まない当該既知の音声信号とに基づいて、反響音をキャンセルする分離行列である反響音キャンセル分離行列を算出し、当該音声取得部によって取得された当該残響音を含む既知の音声信号と当該残響音を含まない既知の音声信号と算出した前記反響音キャンセル分離行列とに基づいてブラインド音源分離における分離行列であるブラインド分離行列とブラインド残響除去における除去行列であるブラインド残響除去行列とを算出する残響データ演算部と、前記残響データ演算部によって算出さ

れた前記反響音キャンセル分離行列に基づき残響抑圧を行うフィルタのフィルタ長を推定するフィルタ長推定部と、前記フィルタ長推定部によって推定されたフィルタ長と前記残響データ演算部によって算出された反響音キャンセル分離行列と前記ブラインド分離行列と前記ブラインド残響除去行列とに基づき残響抑圧を行う残響抑圧部と、を備えることを特徴としている。

また、本発明に係る残響抑圧装置は、前記音声取得部によって取得された残響音を含む音声信号を周波数領域の第１信号に変換し、前記残響音を含まない既知の音声信号を周波数領域の第２信号に変換する短時間フーリエ解析部を備え、前記残響データ演算部は、前記短時間フーリエ解析部から入力された前記第１信号を周波数毎に空間球面化を行って第３信号を算出し、前記短時間フーリエ解析部から入力された前記第２信号を周波数毎にスケールの正規化を行うことで第４信号を算出し、算出された前記第３信号と前記第４信号と前記反響音キャンセル分離行列とを用いて前記ブラインド分離行列と前記ブラインド残響除去行列とを算出し、前記残響抑圧部は、前記フィルタ長推定部によって推定された前記フィルタ長に基づき、前記第３信号と前記第４信号と前記ブラインド分離行列と前記ブラインド残響除去行列と前記反響音キャンセル分離行列とを用いて、繰り返し独立成分分析処理を行い、独立成分分析処理が行われた演算結果に対してスケーリングを行い、スケーリングを行った信号からパワーが最大のものを選択することで前記音声取得部が取得した前記残響音を含む音声信号から前記残響音を抑圧する残響抑圧を行うようにしてもよい。

【０００７】

また、本発明に係る残響抑圧装置において、前記算出された前記反響音キャンセル分離行列、前記ブラインド分離行列、及び前記ブラインド残響除去行列に基づき残響時間を推定する残響特性推定部、を備え、前記フィルタ長推定部は、前記推定された残響時間に基づき前記フィルタ長を推定するようにしてもよい。

【０００８】

また、本発明に係る残響抑圧装置において、前記フィルタ長推定部は、直接音と間接音との比率に基づき前記フィルタ長を推定するようにしてもよい。

【０００９】

また、本発明に係る残響抑圧装置において、当該残響抑圧装置の位置が変化したことを検出する検出部、を更に備え、残響データ演算部は、前記残響抑圧装置の位置が変化したことを検出した場合に前記反響音キャンセル分離行列、前記ブラインド分離行列、及び前記ブラインド残響除去行列を演算するようにしてもよい。

【００１０】

また、本発明に係る残響抑圧装置において、前記検出部は、前記残響抑圧装置の位置が変化したことを検出した場合に、前記残響抑圧部が残響抑圧に用いるパラメータ、あるいは、前記フィルタ長推定部がフィルタ長推定に用いるパラメータの少なくとも一方のパラメータを前記残響抑圧装置の位置が変化したことに基づき切り替えるようにしてもよい。

【００１１】

また、本発明に係る残響抑圧装置において、テスト音声信号を出力する音声出力部、を更に備え、前記音声取得部は、前記出力されたテスト音声信号を取得し、残響データ演算部は、前記取得されたテスト音声信号から前記反響音キャンセル分離行列、前記ブラインド分離行列、及び前記ブラインド残響除去行列を演算するようにしてもよい。

【００１２】

上記目的を達成するため、本発明に係る残響抑圧装置における残響抑圧方法は残響抑圧装置の残響抑圧方法において、音声取得部が、音声信号を取得する音声取得工程と、残響データ演算部が、テスト音声信号を出力する音声出力部から既知の音声信号を出力させたときに、前記音声取得工程によって取得された残響音を含む前記既知の音声信号と残響音を含まない当該既知の音声信号とに基づいて、反響音をキャンセルする分離行列である反響音キャンセル分離行列を算出し、当該音声取得部によって取得された当該残響音を含む既知の音声信号と当該残響音を含まない既知の音声信号と算出した前記反響音キャンセル

分離行列とに基づいてブラインド音源分離における分離行列であるブラインド分離行列とブラインド残響除去における除去行列であるブラインド残響除去行列とを算出する残響データ演算工程と、フィルタ長推定部が、前記残響データ演算工程によって算出された前記反響音キャンセル分離行列に基づき残響抑圧を行うフィルタのフィルタ長を推定するフィルタ長推定工程と、残響抑圧部が、前記フィルタ長推定工程によって推定されたフィルタ長と前記残響データ演算工程によって算出された反響音キャンセル分離行列と前記ブラインド分離行列と前記ブラインド残響除去行列とに基づき残響抑圧を行う残響抑圧工程と、を備えることを特徴としている。

【発明の効果】

【００１３】

本発明によれば、取得された音声信号から残響データを演算して、演算された残響データに基づいて残響特性を推定して、推定された残響特性に基づいて残響抑圧を行うフィルタのフィルタ長を推定するため、残響特性に応じた残響抑圧を精度良く効率的に行うことが可能になる。

【００１４】

本発明によれば、推定された残響特性の残響時間に基づいてフィルタ長を推定するようにしたので、さらに精度が良く効率の良い残響抑圧を行うことが可能になる。

【００１５】

本発明によれば、直接音と反射音との比率に基づいてフィルタ長を推定するようにしたので、さらに精度が良く効率の良い残響抑圧を行うことが可能になる。

【００１６】

本発明によれば、当該残響抑圧装置が設置されている位置が変化したか否かを検出し、設置位置が変化して設置されている環境が変化した場合、残響データの演算と残響特性の推定を行い、推定された残響特性に基づいて残響抑圧を行うフィルタのフィルタ長を推定するため、さらに精度が良く効率の良い残響抑圧を行うことが可能になる。

【００１７】

本発明によれば、残響抑圧部が残響抑圧に用いるパラメータ、あるいは、フィルタ長を推定するためのパラメータの少なくともどちらか一方のパラメータを予め設定されている位置に関する情報に基づいて切り替えるため、さらに精度が良く効率の良い残響抑圧を行うことが可能になる。

【００１８】

本発明によれば、音声出力部が残響データを演算するためのテスト音声信号を出力して、音声取得部が、出力されたテスト音声信号を取得して、取得された音声信号から残響データを演算して、演算された残響データに基づいて残響特性を推定して、推定された残響特性に基づいて残響抑圧を行うフィルタのフィルタ長を推定するため、さらに精度が良く効率の良い残響抑圧を行うことが可能になる。

【図面の簡単な説明】

【００１９】

【図１】本実施形態に係る残響抑圧装置を組み込んだロボットが取得する音声信号の一例を説明する図である。

【図２】同実施形態に係る残響抑圧装置１００のブロック図の一例を示す図である。

【図３】同実施形態に係るＳＴＦＴ処理を説明する図である。

【図４】同実施形態に係るＭＣＳＢ－ＩＣＡ部１１４の内部構成を説明する図である。

【図５】同実施形態に係る残響強度を検出する処理手順を説明する図である。

【図６】同実施形態に係るロボットのみが発話してマイクから音声信号を取得している状態を説明する図である。

【図７】同実施形態に係る残響強度の一例を示す図である。

【図８】同実施形態に係るＭＣＳＢ－ＩＣ処理の変化の一例を示す図である。

【図９】同実施形態に係る実験に用いたデータ及び残響抑圧装置の設定条件である。

【図１０】同実施形態に係る音声認識の設定を説明する図である。

【図１１】同実施形態に係る音声認識の設定を説明する図である。
【図１２】同実施形態に係る推定されたフィルタ長を用いた音声認識率の一例を示す図である。
【図１３】同実施形態に係るケースＢ（バージ・インの発生なし）且つ場所１の場合の音声認識率を示すグラフである。
【図１４】同実施形態に係るケースＢ（バージ・インの発生なし）且つ場所２の場合の音声認識率を示すグラフである。
【図１５】同実施形態に係るケースＣ（バージ・インの発生あり）且つ場所１の場合の音声認識率を示すグラフである。
【図１６】同実施形態に係るケースＣ（バージ・インの発生あり）且つ場所２の場合の音声認識率を示すグラフである。
【図１７】第２実施形態に係る残響抑圧装置１００ａのブロック図の一例を示す図である。
【発明を実施するための形態】
【００２０】
以下、図１～図１７を用いて本発明の実施形態について詳細に説明する。なお、本発明は斯かる実施形態に限定されず、その技術思想の範囲内で種々の変更が可能である。
【００２１】
［第１実施形態］
図１は、本実施形態における残響抑圧装置を組み込んだロボットが取得する音声信号の一例を説明する図である。ロボット１は、図１に示すように、基体部１１と、基体部１１にそれぞれ可動連結される頭部１２（可動部）と、脚部１３（可動部）と、腕部１４（可動部）とを備えている。また、ロボット１は、背負う格好で基体部１１に収納部１５を装着している。なお、基体部１１には、スピーカ２０（音声出力部１４０）が収納され、頭部１２にはマイク３０が収納されている。なお、図１は、ロボット１を側面から見た図であり、マイク３０およびスピーカ２０はそれぞれ複数収納されている。
【００２２】
まず、本実施形態の概略を説明する。
図１のように、ロボット１のスピーカ２０から出力される音声信号を、ロボット１の発話S_rとして説明する。
ロボット１が発話している時に、ヒト２が割り込んで発話することをバージ・イン（Ｂａｒｇｅ－ｉｎ）と呼ぶ。バージ・インが発生しているとき、ロボット１には、当該ロボット１の発話のために、割り込んできたヒト２の発話を聞き分けることが困難である。
そして、ヒト２およびロボット１が発話している場合、ロボット１のマイク３０には、ヒト２の発話S_uが空間を経由して伝達する残響音を含むヒト２の音声信号h_uと、ロボット１の発話S_rが空間を経由して伝達する残響音を含むロボット１の音声信号h_rとが入力される。
【００２３】
図１において、ロボット１のマイク３０が集音する音声信号をモデル化すると、$h_u+h_r=H_u\cdot S_u+H\cdot S_r$のように表せる。$H_u$とHは周波数領域の関数である。$H_u\cdot S_u+H\cdot S_r$において、$S_r$はロボット１の発話のため、当該ロボット１にとって既知である。マイク３０が集音した音声信号の中で$H_u\cdot S_u$には、ヒト２が発話してからロボット１に伝播する間に残響音（エコー）が付加されてしまっているため、$H_u\cdot S_u$を用いて音声認識するより、S_uを用いて音声認識を行えれば認識率が高いことが予測される。また、Hは、ロボット１が単独でスピーカ２０を介して発話し、発話した音声データを、マイク３０を介して取得し、当該ロボット１がいる環境の残響特性を解析することで算出する。さらに、本実施形態では、ＩＣＡ（ｉｎｄｅｐｅｎｄｅｎｔ　ｃｏｍｐｏｎｅｎｔ　ａｎａｌｙｓｉｓ；独立成分分析）をベースにしたＭＣＳＢ－ＩＣＡ（ｍｕｌｔｉ－ｃｈａｎｎｅｌ　ｓｅｍｉ－ｂｌｉｎｄ　ＩＣＡ）を用いて残響音をキャンセル、すなわち抑圧する。さらに、ＭＣＳＢ－ＩＣＡの分離フィルタのフレーム数を、算出した残

響特性に基づいて推定することで、ロボット１がいる環境に合わせたフレーム数を算出する。そして、最終的には、算出されたフレーム数を用いて残響成分を抑圧することでヒト２の発話の音声信号S_{H}を算出する。

【００２４】

図２は、本実施形態における残響抑圧装置１００のブロック図の一例を示す図である。図２のように、残響抑圧装置１００にはマイク３０、スピーカ２０が接続され、マイク３０は複数のマイク３１、３２・・・を備えている。また、残響抑圧装置１００は、制御部１０１と、音声生成部１０２と、音声出力部１０３と、音声取得部１１１と、残響データ算出部１１２と、ＳＴＦＴ部１１３と、ＭＣＳＢ－ＩＣＡ部１１４と、記憶部１１５と、フィルタ長推定部１１６と、分離データ出力部１１７とを備えている。

【００２５】

制御部１０１は、残響特性を測定するための音声を生成して出力する指示を音声生成部１０２に出力し、ロボット１が残響特性を測定するための発話中を示す信号を音声取得部１１１とＭＣＳＢ－ＩＣＡ部１１４に出力する。

【００２６】

音声生成部１０２は、制御部１０１からの指示に基づき、残響特性測定用の音声信号（テスト信号）を生成し、生成した音声信号を音声出力部１０３に出力する。

【００２７】

音声出力部１０３には、生成された音声信号が入力され、入力された音声信号を所定のレベルに増幅してスピーカ２０に出力する。

【００２８】

音声取得部１１１は、マイク３０が集音した音声信号を取得し、取得した音声信号をＳＴＦＴ部１１３に出力する。また、音声取得部１１１は、制御部１０１から残響特性を測定するための音声を生成して出力する指示が入力された時、残響特性を測定するための音声信号を取得し、取得した音声信号を残響データ算出部１１２に出力する。

【００２９】

残響データ算出部（残響データ演算部）１１２には、取得された音声信号と生成された音声信号が入力され、取得された音声信号と生成された音声信号、および記憶部１１５に記憶されている演算式を用いて反響音キャンセル分離行列W_rを算出する。また、残響データ算出部１１２には、算出した反響音キャンセル分離行列W_rを記憶部１１５に書き込んで記憶させる。

【００３０】

ＳＴＦＴ（ｓｈｏｒｔ－ｔｉｍｅ　Ｆｏｕｒｉｅｒ　ｔｒａｎｓｆｏｒｍａｔｉｏｎ；短時間フーリエ解析）部１１３には、取得された音声信号と生成された音声信号が入力され、入力された各音声信号にハニング等の窓関数を音声信号に乗じて有限期間内で、解析位置をシフトしながら解析を行う。そして、ＳＴＦＴ部１１３は、取得された音声信号を、フレームｔ毎にＳＴＦＴ処理して時間－周波数領域の信号ｘ（ω、ｔ）に変換し、また、生成された音声信号を、フレームｔ毎にＳＴＦＴ処理して時間－周波数領域の信号s_r（ω、ｔ）に変換し、変換した信号ｘ（ω、ｔ）と信号s_r（ω、ｔ）を周波数ωごとにＭＣＳＢ－ＩＣＡ部１１４に出力する。図３（ａ）と図３（ｂ）は、ＳＴＦＴ処理を説明する図である。図３（ａ）は、取得された音声信号の波形であり、図３は、この取得された音声信号に乗じられる窓関数である。図３（ｂ）において、記号Ｕはシフト長であり、記号Ｔは解析を行う期間（窓長）を示している。

【００３１】

ＭＣＳＢ－ＩＣＡ部（残響抑圧部）１１４には、ＳＴＦＴ部１１３から変換された信号ｘ（ω、ｔ）と信号s_r（ω、ｔ）が周波数ωごとに入力され、制御部１０１からロボット１が残響特性を測定するための発話中を示す信号が入力され、フィルタ長推定部１１６から推定されたフィルタ長データが入力される。また、ＭＣＳＢ－ＩＣＡ部１１３は、残響特性を測定するための発話中を示す信号が入力されていない場合、入力された信号ｘ（ω、ｔ）と信号s_r（ω、ｔ）と記憶部１１４に記憶されている反響音キャンセル分離行

列W_r、各モデル及び各係数を用いて、分離フィルタW_{1u}とW_{2u}を算出する。分離フィルタW_{1u}とW_{2u}算出後、マイク３０が取得した音声信号からヒト２の直接発話信号を分離し、分離した直接発話信号を分離データ出力部１１７に出力する。

【００３２】

図４は、ＭＣＳＢ－ＩＣＡ部１１４の内部構成を説明する図である。図４のように、ＳＴＦＴ部１１３から入力された信号ｘ（ω、ｔ）はバッファ２０１を介して強制空間球面化部２１１に入力され、ＳＴＦＴ部１１３から入力された信号s_r（ω、ｔ）はバッファ２０２を介して分散正規化部２１２に入力される。そして、ＩＣＡ部２２１には、強制空間球面化部２１１から空間球面化された信号と、分散正規化部２１２から正規化された信号とが入力され、入力された信号を用いて繰り返しＩＣＡ処理を行い、演算結果をスケーリング部２３１に出力し、スケーリングされた信号を直接発話分離部２４１に出力する。なお、スケーリング部２３１は、ｐｒｏｊｅｃｔｉｏｎ　Ｂａｃｋ処理を用いてスケーリングを行い、直接発話分離部２４１は、入力された信号からパワーが最大のものを選択して出力する。

【００３３】

記憶部１１５には、ロボット１がマイク３０を介して取得する音声信号のモデル、解析するための分離モデル、解析するために必要なパラメータ等が予め書き込まれて記憶され、さらに、算出された反響音キャンセル分離行列W_r、分離フィルタW_{1u}及び分離フィルタW_{2u}が書き込まれて記憶されている。

【００３４】

フィルタ長推定部（残響特性推定部、フィルタ長推定部）１１６は、記憶部１１５に記憶されている反響音キャンセル分離行列W_rを読み出し、読み出した反響音キャンセル分離行列W_rから後述する方法でフィルタ長を推定し、推定したフィルタ長データをＭＣＳＢ－ＩＣＡ部１１４に出力する。なお、フィルタ長とは、フレーム（窓）をサンプリングする数に関する値であり、フィルタ長が大きくなると時間方向に長い期間、サンプリングを行うことになる。

【００３５】

分離データ出力部１１７には、ＭＣＳＢ－ＩＣＡ部１１４から分離された直接発話信号が入力され、入力された直接発話信号を、例えば非図示の音声認識部に出力する。

【００３６】

次に、ロボット１が取得した音声から必要な音声信号を分離するための分離モデルについて説明する。記憶部１１５には、ロボット１がマイク３０を介して取得する音声信号は、式（１）ののＦＩＲ（ｆｉｎｉｔｅ　ｉｍｐｕｌｓｅ　ｒｅｓｐｏｎｓｅ；有限インパルス応答）のモデルのように定義する。

【００３７】

【数１】

$$x(t)=\sum_{n=0}^{N}h_u(n)s_u(t-n)+\sum_{m=0}^{M}h_r(m)s_r(t-n) \qquad \cdots(1)$$

【００３８】

式（１）において、記号x_1（ｔ）・・・x_L（ｔ）は、各マイク３１～３２の各スペクル（Ｌはマイク番号）、ｘ（ｔ）はベクトルであり［x_1（ｔ），x_2（ｔ），・・・，x_L（ｔ）］Ｔ、s_u（ｔ）はヒト２の発話、s_r（ｔ）は既知のロボット１のスペクトル、h_u（ｎ）はヒト２の音声スペクトルのＮ次元のＦＩＲ係数ベクトル、h_r（ｍ）は既知のロボット１のＭ次元のＦＩＲ係数ベクトルである。式（１）は、ロボット１がマイク３０を介して取得する時刻ｔにおけるモデル化である。

【００３９】

また、記憶部１１５には、ロボット１のマイク３０が集音した音声信号について、式（

２）のように残響成分を含んだベクトルＸ（ｔ）としてモデル化され予め記憶されている。さらに、記憶部１１５には、ロボット１の発話の音声信号について、式（３）のように残響成分を含んだベクトルＳ$_r$（ｔ）としてモデル化されて予め記憶されている。

【００４０】

【数２】

$$X(t)=[x(t),x(t-1),\cdots,x(t-N)]^T \quad \cdots(2)$$

【００４１】

【数３】

$$S_r(t)=[s_r(t),s_r(t-1),\cdots,s_r(t-M)]^T \quad \cdots(3)$$

【００４２】

式（３）において、ｓ$_r$（ｔ）はロボット１が発話した音声信号であり、ｓ$_r$（ｔ－１）は空間を伝達されて「１」遅延して音声信号が届くことを表し、ｓ$_r$（ｔ－Ｍ）は「Ｍ」遅延して届く音声信号が届くことを表している。すなわち、ロボット１から離れている距離が大きく、遅延量が大きいほど残響成分が大きくなることを表している。

【００４３】

次に、ＩＣＡを用いて既知の直接音Ｓ$_r$（ｔ）とＸ（ｔ－ｄ）と、ヒト２の直接発話信号ｓ$_u$とを独立となるように分離するため、記憶部１１５には、ＭＣＳＢ－ＩＣＡの分離モデルが次式（４）のように定義し、記憶部１１５に記憶されている。

【００４４】

【数４】

$$\begin{pmatrix}\hat{s}(t)\\ X(t-d)\\ S_r(t)\end{pmatrix}=\begin{pmatrix}W_{1u} & W_{2u} & W_r\\ 0 & I_2 & 0\\ 0 & 0 & I_r\end{pmatrix}\begin{pmatrix}x(t)\\ X(t-d)\\ S_r(t)\end{pmatrix} \quad \cdots(4)$$

【００４５】

式（４）において、ｄ（０より大きい）は、初期反射間隔であり、Ｘ（ｔ－ｄ）は、Ｘ（ｔ）をｄ遅延させたベクトルであり、式（５）は、Ｌ次元の推定された信号ベクトルである。

【００４６】

【数５】

$$\hat{s}(t) \quad \cdots(5)$$

【００４７】

また、Ｗ$_{1u}$は、Ｌ×Ｌのブラインド分離行列（分離フィルタ）、Ｗ$_{2u}$は、Ｌ×Ｌ（Ｎ＋１）のブラインド残響除去行列（分離フィルタ）、Ｗ$_r$は、Ｌ×（Ｍ＋１）の残響音キャンセルの分離行列（取得した残響特性に基づく残響要素）である。

また、Ｉ$_2$とＩ$_r$は、それぞれに対応した大きさの単位行列である。そして、式（５）

には、ヒト２の発話の直接発話信号といくつかの反射音信号とを含んでいる。

【００４８】

次に、式（４）を解くためのパラメータについて説明する。 式（４）において、分離パラメータのセットW＝｛W_{1u}、W_{2u}、W_r｝を、結合確率密度関数（ｐｒｏｂａｂｉｌｉｔｙ ｄｅｎｓｉｔｙ ｆｕｎｃｔｉｏｎ）とｓ（ｔ）、Ｘ（ｔ－ｄ）およびＳｒ（ｔ）の周辺確率密度関数（個々のパラメータの独立な確率分布を表わす周辺確率密度関数）の積との間の差の尺度としてＫＬ（ｋｕｌｌｂａｃｋ－Ｌｅｉｂｌｅｒ；カルバック・ライブラー）情報量を最小化するように推定する。また、周波数ωにおける分離行列の初期値W_{1u}（ω）は、周波数ω＋１において推定行列W_{1u}（ω＋１）にセットされている。

【００４９】

ＭＣＳＢ－ＩＣＡ部１１４は、分離パラメータのセットWを、ＫＬ情報量を自然勾配法により最小にするように各分離フィルタ次式（６）～式（９）のルールに従い繰り返し更新することで推定を行う。また、式（６）～式（９）は、記憶部１１５に予め書き込まれて記憶されている。

【００５０】

【数６】

$$D = \Lambda - E\left[\phi\left(\hat{s}(t)\right)\hat{s}^{H}(t)\right] \quad \cdots(6)$$

【００５１】

【数７】

$$W_{1u}^{[j+1]} = W_{1u}^{[j]} + \mu D W_{1u}^{[j]} \quad \cdots(7)$$

【００５２】

【数８】

$$W_{2u}^{[j+1]} = W_{2u}^{[j]} + \mu\left(DW_{2u}^{[j]} - E\left[\phi\left(\hat{s}(t)\right)X^{H}(t-d)\right]\right) \quad \cdots(8)$$

【００５３】

【数９】

$$W_{r}^{[j+1]} = W_{r}^{[j]} + \mu\left(DW_{r}^{[j]} - E\left[\phi\left(\hat{s}(t)\right)S_{r}^{H}(t)\right]\right) \quad \cdots(9)$$

【００５４】

なお、式（６）、式（８）～式（９）において、上付きＨは共役転置演算（エルミート転置）を表す。また、式（５）において、Λは非ホロノミック拘束行列、すなわち、式（１０）の対角行列である。

【００５５】

【数１０】

$$E\left[\phi\left(\hat{s}(t)\right)\hat{s}^{H}(t)\right] \quad \cdots(10)$$

【００５６】
また、式（７）～式（９）において、ｕは、ステップ・サイズのパラメータであり、φ（ｘ）は、非線形関数ベクトル［φ（ｘ$_{1}$）、・・・、φ（ｘ$_{L}$）］^{H}であり、次式（１１）のように表され、記憶部１１５に書き込まれて記憶されている。
【００５７】
【数１１】

$$\phi(x) = -\frac{d}{dx}\log p(x) \qquad \cdots(11)$$

【００５８】
さらに、音源のＰＤＦは、分散量σ^{2}であるとした場合、雑音に強いＰＤＦであるｐ（ｘ）＝ｅｘｐ（－｜ｘ｜／σ^{2}）／（２σ^{2}）であり、φ（ｘ）＝ｘ*／（２σ^{2}｜ｘ｜）であり、ｘ*はｘの共役であると仮定する。この２つの関数は、連続領域である｜ｘ｜＞εにおいて定義される。
【００５９】
次に、音声を分離する処理手順を、図５～図８を用いて説明する。図５は、本実施形態における残響強度を検出する処理手順を説明する図である。なお、残響強度の検出は、ロボット１がいる環境が変わった場合、例えば、別の部屋に移動した後、室外に出た後毎に行う。また、ロボット１は、例えば、当該ロボット１に組み込まれている非図示のカメラで撮像された画像データを用いて、環境が変化したか否かを判定する。あるいは、ロボット１が水平方向または垂直方向に移動し、当該ロボット１がいた位置が変化した場合にも残響強度を検出する処理を行うようにしてもよい。
【００６０】
［ステップＳ１；Ｅｍｉｓｓｉｏｎ　ｏｆ　ｓｅｌｆ　ｓｐｅｃｈ］
まず、図６のように、ロボット１は、当該ロボット１が現在いる環境で、制御部１０１は、残響強度を測定するための所定の音声信号を生成する指示を音声生成部１０２に出力する。音声生成部１０２には、所定の音声信号を生成する指示が入力され、入力された生成指示に基づき所定の音声信号を生成し、生成した所定の音声信号を音声出力部１０３に出力する。音声出力部１０３には、生成された所定の音声信号が入力され、入力された所定の音声信号を所定のレベルに増幅してスピーカ２０に出力する。なお、残響強度を測定するための所定の音声信号は、例えば、１つの母音または１つの子音であってもよい。図６は、ロボットのみが発話してマイクから音声信号を取得している状態を説明する図である。
【００６１】
次に、音声取得部１１１には、マイク３０が集音した音声信号が入力され、入力された音声信号を残響データ算出部１１２に出力する。マイク３０が集音する音声信号は、音声生成部１０２が生成した音声信号Ｓ$_{r}$に、スピーカ２０から発せられた音声が壁、天井、床などで反響した残響成分を含む音声信号ｈ$_{r}$である。
【００６２】
次に、残響データ算出部１１２には、取得された音声信号が入力され、入力された音声信号を記憶部１１５に記憶されている式（９）を用いて反響音キャンセル分離行列Ｗ$_{r}$を算出する。また、残響データ算出部１１２は、演算した残響特性データを記憶部１１５に書き込んで記憶させる。なお、式（９）を演算するとき、入力値はＷ$_{r}$のみなのでフィルタ長を１に設定する。
【００６３】
［ステップＳ２；Ｃａｌｃｕｌａｔｉｏｎ　ｏｆ　ｅｃｈｏ　ｉｎｔｅｎｓｉｔｉｅｓ］
ステップＳ２では、ステップＳ１で算出されたＷｒを使って、フィルタ長を推定するための残響強度のグラフを生成する。
まず、フィルタ長推定部１１６は、記憶部１１５に記憶されている反響音キャンセル分

離行列W_rを読み出す。フィルタ長推定部１１６は、読み出した反響音キャンセル分離行列W_rを、パラメータW_rを式（１２）のような行列に置き直す。

【００６４】

$W_r = [w_r(0)\ w_r(1) \cdots w_r(M)]$　・・・（１２）

【００６５】

なお、式（１２）のW_rにおいて、$w_r(m)$は、Ｌ×１ベクトルであり式（１３）のように表される。

【００６６】

【数１２】

$$W_r(m) = \left[w_r^1(m) w_r^2(m) \cdots w_r^L(M)\right]^T \quad \cdots(13)$$

【００６７】

そして、周波数ωにおけるこのフィルタの正規化されたパワー関数は、次式（１４）のように定義する。

【００６８】

【数１３】

$$p_r^i(\omega, m) = \frac{\left|\omega_r^i(\omega, m)\right|^2}{\max_m \left|\omega_r^i(\omega, m)\right|^2} \quad \cdots(14)$$

【００６９】

式（１４）において、ｉはマイク３０の番号（マイク３１、３２、・・・）であり、ｍはフィルタのインデックスである。式（１４）のパワー関数は、残響強度を反映し、また、環境の残響時間に関係しているので、このパワー関数に基づいて残響時間を推定する。

次に、平均化された周波数のパワー関数と平均化されたマイクのパワー関数Ｐと、関数Ｐの対数値Ｌは、次式（１５）と式（１６）のように残響時間のための基準として定義する。

【００７０】

【数１４】

$$p(m) = \frac{\sum_i \sum_{\omega \in \Omega} p_r^i(\omega, m)}{\max_m \sum_i \sum_{\omega \in \Omega} p_r^i(\omega, m)} \quad \cdots(15)$$

【００７１】

【数１５】

$$L(m) = 20 \log_{10} P(m) \quad \cdots(16)$$

【００７２】

式（１５）において、Ωは周波数バンド・セットに基づく値である。フィルタ長推定部１１６は、この式（１５）と式（１６）を用いて、図７のように残響強度を仮想的にプロットする。図７において、縦軸は音声レベルであり、横軸は時間軸を表している。図７のように、生成された音声信号をスピーカ３０から発した時（時刻０）の音声レベルが一番高くロボット１がいる環境の残響特性に応じて、音声レベルは下がっていく。

【００７３】

［ステップＳ３；Ｅｓｔｉｍａｔｉｏｎ ｏｆ ｄｅｒｅｖｅｒｂｅｒａｔｉｏｎ ｆｉｌｔｅｒ ｌｅｎｇｔｈ］

ステップＳ３では、図７のプロットされた残響強度のグラフを用いて、フィルタ長Ｍを検定する。

まず、図７のように、フィルタ長推定部１１６は、フィルタ長の推定のため式（１７）を用いて線形回帰解析を行う。

【００７４】

ｙ＝ａ×ｍ＋ｂ　・・・（１７）

【００７５】

式（１７）において、ａとｂは係数であり、ｍはフィルタ長のインデックス、そしてｙはＬ（ｍ）と等価である。次に、図７のように、フィルタ長推定部１１６は、Ｐ（ｍ）のピーク値からいくつかのサンプルを抽出し、最小二乗平均（ＬＭＳ；ｌｅａｓｔ ｍｅａｎ ｓｑｕａｒｅ）法を用いてａとｂを推定する。

次に、フィルタ長推定部１１６は、残響除去のフィルタ長を、次式（１８）において、ｍがＬ（ｍ）＝L_dの値を満足するように算出し、算出した残響除去のフィルタ長をＩＣＡ部２２１に出力する。

【００７６】

【数１６】

$$\hat{N}=\frac{L_d-b}{a} \qquad \cdots(18)$$

【００７７】

一例として、図７において、RT_{20}＝２４０ｍｓｅｃ（RT_{20}は残響時間）、そして線形回帰線２５１を式（１７）により推定する。そして、推定されたフィルタ長は、式（１８）においてL_d＝－６０（ライン２５２）との交点２５３の値、Ｍ＝約１３である。

【００７８】

［ステップＳ４；Ｉｎｃｒｅｍｅｎｔａｌ ｓｅｐａｒａｔｉｏｎ ｐｏｌｉｎｇ ｎｏｔｉｆｉｃａｔｉｏｎ］

ヒト２の発話が発声した場合、このステップＳ４を行い、式（４）を用いて式（５）を求めることで、マイク３０から取得された音声信号からヒト２の残響成分除去した音声信号を算出する。

【００７９】

音声取得部１１１には、マイク３０が集音した音声信号が入力され、入力された音声信号をＳＴＦＴ部１１３に出力する。また、音声生成部１０２は、音声を生成している場合、生成した音声信号をＳＴＦＴ部１１３に出力する。

【００８０】

次に、ＳＴＦＴ部１１３には、マイク３０が取得した音声信号と、音声生成部１０２が生成した音声信号とが入力され、取得された音声信号をフレームｔ毎にＳＴＦＴ処理して時間－周波数領域の信号ｘ（ω、ｔ）に変換し、変換した信号ｘ（ω、ｔ）を周波数ωごとにＭＣＳＢ－ＩＣＡ部１１４に出力する。また、ＳＴＦＴ部１１３は、生成された音声

信号を、フレームｔ毎にＳＴＦＴ処理して時間－周波数領域の信号s_r（ω、ｔ）に変換し、変換した信号s_r（ω、ｔ）を周波数ωごとにＭＣＳＢ－ＩＣＡ部１１４に出力する。

【００８１】

ＭＣＳＢ－ＩＣＡ部１１４の強制空間球面化部２１１には、変換された信号ｘ（ω、ｔ）が周波数ωごとに入力され、周波数ωをインデックスとして順次、次式（１９）を用いて空間球面化を行い、ｚ（ｔ）を算出する。また、式（１９）と式（２１）は、式（５）を解く上で高速化を行うために用いている。

【００８２】

【数１７】

$$z(t) = V_u x(t) \qquad \cdots(19)$$

【００８３】

ただし、V_uは式（２０）である。

【００８４】

【数１８】

$$V_u = E_u \Lambda^{-\frac{1}{2}} E_u^H \qquad \cdots(20)$$

【００８５】

さらに、式（２０）において、E_uとΛ_uは、固有ベクトル行列であり、固有対角行列$R_u = E|x(t)x^H(t)|$である。

さらに、ＭＣＳＢ－ＩＣＡ部１１４の分散正規化部２１２には、変換された信号s_r（ω，ｔ）が周波数ωごとに入力され、周波数ωをインデックスとして順次、次式（２１）を用いてスケールの正規化を行う。

【００８６】

【数１９】

$$\tilde{s}_r(t) = \lambda_r^{-\frac{1}{2}} s_r(t) \qquad \cdots(21)$$

【００８７】

なお、スケーリングの正規化において、逆変換法（ｐｒｏｊｅｃｔｉｏｎ　ｂａｃｋ　ｍｅｔｈｏｄ）を用い、逆分離行列の要素は、分離信号に従って乗算される。そして、式（２２）のｉ番目の列、ｊ番目の行の要素c_jは、式（５）のｊ番目の要素のスケーリングは、式（２３）～式（２４）の式の関係に従って行う。

【００８８】

【数２０】

$$\hat{H}_u = \left(W_{1u} V_0\right)^{-1} \qquad \cdots(22)$$

【００８９】

【数２１】

$$l_j = \arg\max_l \left| \hat{H}_u(l, j) \right| \quad \cdots(23)$$

【００９０】

【数２２】

$$c_j = \hat{H}_u(l_j, j) \quad \cdots(24)$$

【００９１】

強制空間球面化部２１１は、このように演算されたｚ（ω，ｔ）をＩＣＡ部２２１に出力し、分散正規化部２１２は、このように演算された式（２１）の値をＩＣＡ部２２１に出力する。

【００９２】

次に、ＩＣＡ部２２１には、演算されたｚ（ω，ｔ）と式（２１）の値とが入力され、さらに、記憶部１１５に記憶されている分離モデル（分離フィルター）を読み出す。次に、ＩＣＡ部２２１は、式（４）、式（６）～式（９）のｘに式（１９）を代入し、s_rに式（２１）を代入して、W_{1u}とW_{2u}を算出し、すでにステップＳ１で算出されたW_rを用いて、ＭＣＳＢ－ＩＣＡ部１１４が式（５）のデータを算出する。

【００９３】

図８は、ＭＣＳＢ－ＩＣＡ処理の変化の一例を示す図である。通常の分離モードにおいて、ＭＣＳＢ－ＩＣＡのブロック幅増加分離を行う。ＩＣＡは、分離行列を安定して推測するために、所定の持続時間、データをバッファする。このようにバッファを使用するため、時間ｔの分離を行うため先行するブロックサイズI_bを利用する。図８においては、シフト量I_sが増加する場合、遅れ時間も増加する。また、シフト量I_sが減少する場合、算出処理が増加する。このように、本実施形態では、オーバーラップ・パラメータ係数I_sを使用する。

【００９４】

次に、本実施形態の残響抑圧装置を備えるロボット１で行った実験方法と実験結果の例を説明する。図９～図１２は、実験条件である。図９は、実験に用いたデータ及び残響抑圧装置の設定条件である。図９のように、インパルス応答は１６ＫＨｚサンプル、残響時間は２４０ｍｓと６７０ｍｓ、ロボット１とヒト２との距離は１．５ｍ、ロボット１とヒト２の角度は０度、４５度、９０度、－４５度、－９０度、使用したマイク３０の本数は２本（ロボット１の頭部に設置）、ＳＴＦＴ分析はハニング窓のサイズ３２ｍｓ（５１２ポイント）かつシフト量１２ｍｓ（１９２ポイント）、入力信号データは［－１．０　１．０］に正規化されたものである。

【００９５】

図１０は、音声認識の設定を説明する図である。図１０のように、テスト・セットは２００の文章（日本語）、訓練セットは２００人（それぞれ１５０の文章））、音響モデルはＰＴＭ－ｔｒｉｐｈｏｎｅ、３値のＨＭＭ（隠れマルコフモデル）、言語モデルは語彙サイズ２０ｋ、発話解析はハニング窓のサイズ３２ｍｓ（５１２ポイント）、シフト量１０ｍｓ、特徴量はＭＦＣＣ（Ｍｅｌ－Ｆｒｅｑｕｅｎｃｙ　Ｃｅｐｓｔｒｍ　Ｃｏｅｆｆｉｃｉｅｎｔ；スペクトル包絡）は２５次（１２次＋Δ１２次＋Δパワー）である。また、他のＳＴＦＴ設定条件は、フレーム間隔係数ｄ＝２、反響キャンセルのフィルタ長Ｎと通常の分離モードの残響除去のフィルタ長Ｍは同じ値、適応ステップ・サイズのための係数は予め設定され、推定されたフィルタ係数は、Ω＝｛５，６，・・・、２００｝かつＬ

$_{d}$＝－６０、直線回帰のためのサンプル数は6に設定してある。また、音声認識エンジンは、公知のＪｕｌｉｕｓ（http://julius.sourceforge.jp/）を使用している。

【００９６】

次に、実験結果を図11～図１６に示す　図１１は、推定されたフィルタ長の設定を示した図である。図１１のように、ノイズあり且つ残響時間が２４０ｍｓの場合、ノイズあり且つ残響時間６７０ｍｓの場合、ノイズなし且つ残響時間が２４０ｍｓの場合、ノイズなし且つ残響時間６７０ｍｓの場合、各々についてＭ$_{ｍａｘ}$が２０，３０，５０についての推定されたフィルタ長の平均値と偏差を示している。場所１（Ｅｎｖ．Ｉ）は、通常の部屋（残響時間ＲＴ$_{２０}$＝２４０ｍｓ）、場所２（Ｅｎｖ．ＩＩ）は、ホールのような部屋（残響時間ＲＴ$_{２０}$＝６７０ｍｓ）である。

【００９７】

図１２は、推定されたフィルタ長を用いた音声認識率の一例を示す図である。図１２のように、ケースＢは、バージ・インが発生していない場合、ケースＣは、バージ・インが発生している場合、各々について音声分離無しでの認識率（ｎｏ　ｐｒｏｃ）、ブロックサイズＩ$_{ｂ}$が１６６（２秒）、２０８（２．５秒）、２５５（３秒）、残響時間２４０ｍｓと６７０ｍｓの各音声認識率を示している。また、シフト量Ｉ$_{ｓ}$は、ブロックサイズＩ$_{ｂ}$の半分に設定されている。一例として、残響音がないクリーンな音声信号による認識率は、実験に用いた残響抑圧装置では約９３％である。

【００９８】

図１２の結果をグラフにしたのが図１３～図１６である。図１３は、ケースＢ（バージ・インの発生なし）且つ場所１の場合の音声認識率を示すグラフであり、図１４は、ケースＢ（バージ・インの発生なし）且つ場所２の場合の音声認識率を示すグラフである。図１５は、ケースＣ（バージ・インの発生あり）且つ場所１の場合の音声認識率を示すグラフであり、図１６は、ケースＣ（バージ・インの発生あり）且つ場所２の場合の音声認識率を示すグラフである。各グラフの横軸はフィルタ長（Ｎ）であり、縦軸は音声認識率（％）である。

図１３のように、残響時間が短い部屋（場所１）且つバージ・インが発生していない場合、推定されたフィルタ長（Ｎ＝１４）３０１より不適切なフィルタ長（Ｎ＝３５）の方が認識率（正答率）は低くかつブロックサイズＩ$_{ｂ}$を変えると認識率の差が大きくなる。ルタ長（Ｎ＝３５）３０２の場合はブロックサイズＩ$_{ｂ}$により認識率に差が生じている。一方、残響時間が長い部屋（場所２）且つバージ・インが発生していない場合、推定されたフィルタ長（Ｎ＝３５）で認識率は６０％以上である。そして、図１３と図１４のように、残響時間が短い場合のフィルタ長はＮ＝１４で短く、残響時間が長い場合のフィルタ長はＮ＝３６で長い。このように、ロボット１が取得した環境の残響特性に基づき、適切なフィルタ長（フレーム長）を推定することで、音声認識率を改善できる。

図１５のように、残響時間が短い部屋（場所１）且つバージ・インが発生している場合、推定されたフィルタ長（Ｎ＝１４）より不適切なフィルタ長（Ｎ＝３５）の方が認識率（正答率）は低くかつブロックサイズＩ$_{ｂ}$を変えると認識率の差が大きくなる。一方、残響時間が長い部屋（場所２）且つバージ・インが発生している場合、推定されたフィルタ長（Ｎ＝３５）で認識率は４０％以上である。

【００９９】

以上のように、残響特性に応じて分離フィルタ長であるフレーム長を設定するようにしたので、音声認識率が向上し、さらに音声認識にかかる演算量も適切にすることが可能になる。

【０１００】

また、本実施形態では、残響特性として残響時間を用いた例を説明したが、Ｄ値（音声の明瞭さを表す値であり、直接音が到達してから０～５０ｍｓｅｃまでのパワーと、０～音声が減衰するまでのパワーの比）を用いても良い。

【０１０１】

また、本実施形態では、残響特性の測定を制御部１０１から残響特性を測定するための

音声を生成して出力する指示が入力された時、残響特性を測定するための音声信号を取得して残響特性を測定する例を説明したが、音声取得部１１１は、音声生成部１０２が出力する生成された音声信号と比較しながら取得し、取得中にバージ・インが発生しているか否かを判別して、バージ・インが発生していないときに残響特性の測定用の音声信号を取得するようにしてもよい。

【０１０２】

［第２実施形態］

次に、第２実施形態について、図１７を用いて説明する。図１７は、本実施形態における残響抑圧装置１００ａのブロック図の一例を示す図である。第１実施形態では、ロボット１は、環境が変った場合に、発話を行い、当該ロボット１がいる環境の残響特性を測定する例を説明した。残響特性の測定は、例えば、ロボット１が移動する部屋毎に例えばマークが設置され、設置されているマークをロボット１のカメラ４０が撮像して公知の画像認識の手法を用いて、マークを検出して環境、例えば部屋を移動したことを検出した場合に行う。あるいは、ロボット１の記憶部１１４に予めマップを書き込んで記憶させておき、マップに基づき環境変化を検出した場合に行う。

【０１０３】

図１７のように、本実施形態における残響抑圧装置１００ａは、画像取得部３０１と、環境変化検出部３０２とをさらに備えている。また、残響抑圧装置１００ａには、カメラ４０が接続され、画像取得部３０１には、カメラにより撮像された画像信号が入力され、入力された画像信号を環境変化検出部３０２に出力する。環境変化検出部３０２は、入力された画像信号に基づき、残響抑圧装置１００ａが組み込まれているロボット１ａがいる位置が変化したか否かを判定し、位置が変化したことを検出した場合、位置が変化したことを示す信号を制御部１０１ａに出力する。制御部１０１ａは、位置が変化したことを示す信号が入力された場合、音声生成部１０２に残響特性測定用の音声信号（テスト信号）を生成する指示を出力する。以下、第１実施形態と同様の処理を行う。

【０１０４】

また、各パラメータを環境毎に予め記憶部１１５ａに書き込んで記憶させておき、マップ、マークとおのおの関連づけて記憶部１１５ａに記憶させておく。

そして、ロボット１ａが、環境が変ったことを検出した場合、制御部１０１ａは、残響特性を測定するとともに、各パラメータのセットを記憶部１１４ａから読み出して切り替えるようにしても良い。

【０１０５】

また、記憶部１１５ａに残響データが記憶されていない環境で、残響測定を行い、測定された残響特性と関連付けて、その環境に基づくパラメータを算出して、算出したパラメータを関連づけて新たに記憶部１１５ａに記憶させるようにしてもよい。

【０１０６】

また、例えば、各部屋にロボット１ａへ位置に関する情報を送信する非図示の位置情報送信装置を設置し、ロボット１ａはこの位置情報を受信した場合に環境が変化したと検出して、残響特性を測定するようにしてもよい。

【０１０７】

なお、第１、第２実施形態では、残響抑圧装置１００及び残響抑圧装置１００ａをロボット１（１ａ）に組み込んだ例を説明したが、残響抑圧装置１００及び残響抑圧装置１００ａは、例えば音声認識装置、音声認識装置を有する装置などに組み込んで用いることも可能である。

【０１０８】

なお、実施形態の図２及び図１７の各部の機能を実現するためのプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することにより各部の処理を行ってもよい。なお、ここでいう「コンピュータシステム」とは、ＯＳや周辺機器等のハードウェアを含むものとする。

また、「コンピュータシステム」は、WWWシステムを利用している場合であれば、ホームページ提供環境（あるいは表示環境）も含むものとする。

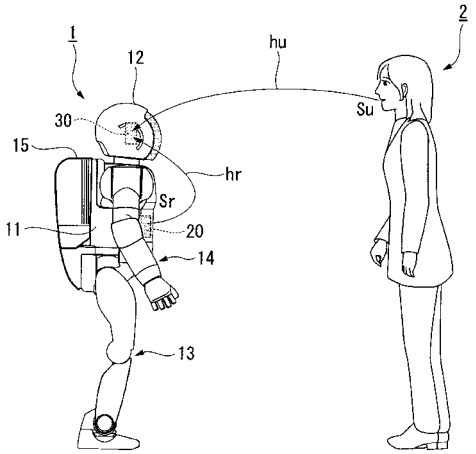
また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ＲＯＭ（Ｒｅａｄ　Ｏｎｌｙ　Ｍｅｍｏｒｙ）、ＣＤ－ＲＯＭ等の可搬媒体、ＵＳＢ（Ｕｎｉｖｅｒｓａｌ　Ｓｅｒｉａｌ　Ｂｕｓ）　Ｉ／Ｆ（インタフェース）を介して接続されるＵＳＢメモリー、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムを送信する場合の通信線のように、短時間の間、動的にプログラムを保持するもの、その場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリーのように、一定時間プログラムを保持しているものも含むものとする。また上記プログラムは、前述した機能の一部を実現するためのものであっても良く、さらに前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるものであっても良い。

【符号の説明】

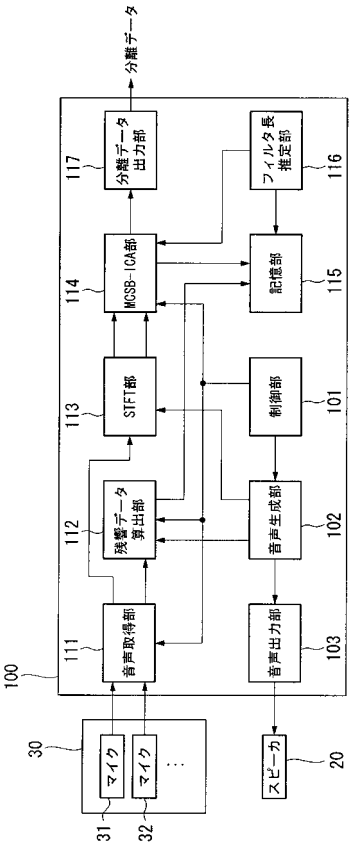
【０１０９】

１・・・ロボット
２０・・・スピーカ
３０、３１、３２・・・マイク
１００・・・残響抑圧装置
１０１・・・制御部
１０２・・・音声生成部
１１１・・・音声取得部
１１２・・・残響データ演算部
１１３・・・ＳＴＦＴ部
１１４・・・ＭＣＳＢ－ＩＣＡ部
１１５・・・記憶部
１１６・・・フィルタ長推定部
１１７・・・分離データ出力部
３０２・・・環境変化検出部

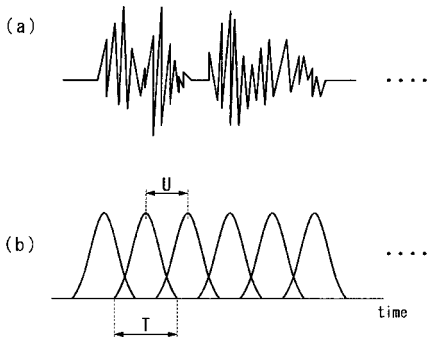
【図１】


1
12
hu
2
Su
30
15
hr
Sr
11
20
14
13

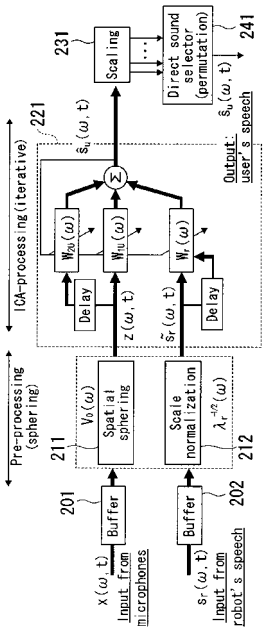
【図２】


分離データ
117
分離データ出力部
フィルタ長推定部
116
114
MCSB-ICA部
記憶部
115
113
STFT部
制御部
101
112
残響データ算出部
音声生成部
102
111
音声取得部
音声出力部
103
100
30
マイク
マイク
スピーカ
20
31
32

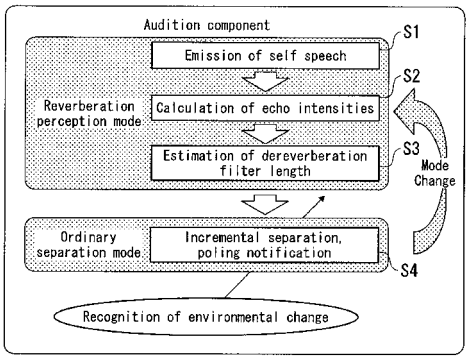
【図３】


(a)
(b)
U
T
time

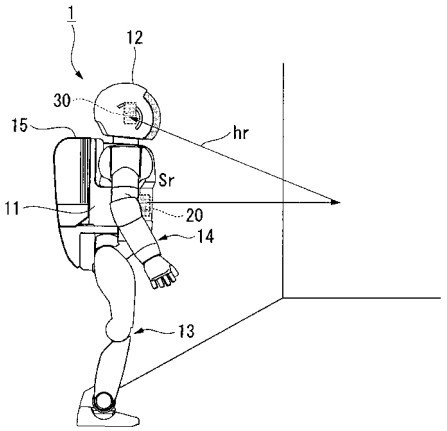
【図４】


231
Scaling
Direct sound selector (permutation)
241
Output: user's speech
221
ICA-processing(iterative)
Delay
Delay
Pre-processing (sphering)
211
Spatial sphering
Scale normalization
212
201
Buffer
Buffer
202
Input from microphones
Input from robot's speech

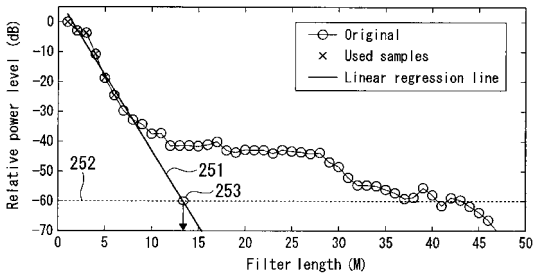
【図５】



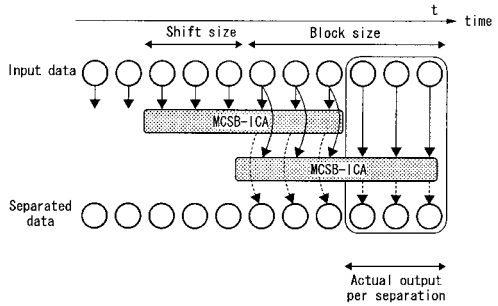
【図６】



【図７】



【図８】



【図９】

Impulse response	16-kHz sampling
Reverberation time (RT_{20})	240 and 670 ms
Distance and direction	1.5m and 0°, 45°, 90°, -45°, -90°
Number of microphones	Two (embedded in Robot's head)
STFT analysis	Manning: 32ms and shift: 12ms
Input wave data	[-1.0 1.0]normalized

【図１０】

Test set	200 sentences
Training set	200 persons (150 sentences each)
Acoustic model	PTM-triphone: 3-state, HMM
Language model	Statistical, vocabulary size 20k
Speech analysis	Manning: 32ms and shift: 10ms
Features	MFCC 25 dim. (12+Δ12+ΔPow)

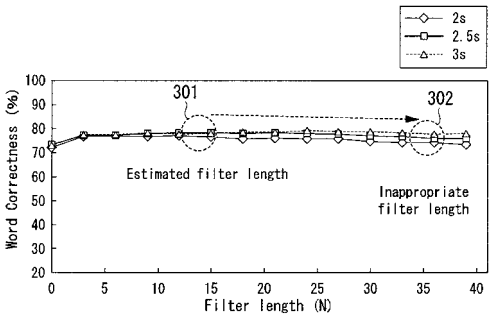
【図１１】

		Env. I (RT_{20} 240 ms)			Env. II (RT20 670ms)		
	Mmax	20	30	50	30	40	50
w/o noise	Mean	14.0	13.7	13.2	35.0	35.3	35.4
	Std.	0.43	0.46	0.53	1.22	1.24	1.28
with noise	Mean	14.2	14.0	13.6	36.1	36.3	36.2
	Std.	1.25	1.17	1.05	2.38	2.41	2.30

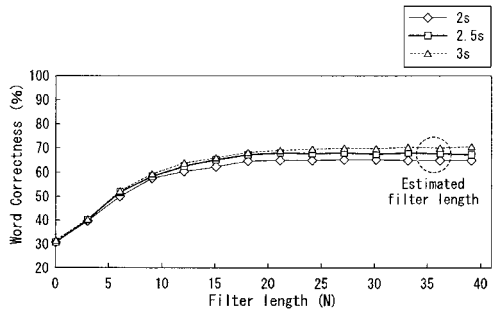
【図１２】

	Exp. B (non-barge-in)				Exp. C (barge-in)			
	no proc.	2s	2.5s	3s	no proc.	2s	2.5s	3s
Env.I (RT_{20} 240ms)	74.3	76.9	78.5	78.2	28.2	67.8	70.2	71.7
Env.II (RT_{20} 670ms)	26.1	63.9	66.8	69.2	11.0	37.1	41.2	43.3

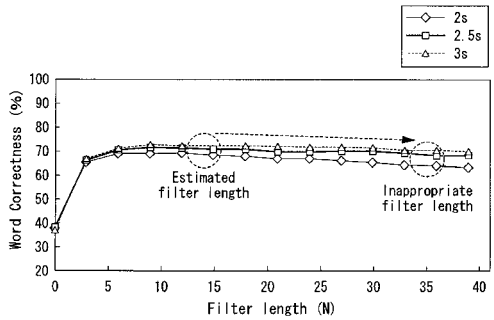
【図１３】



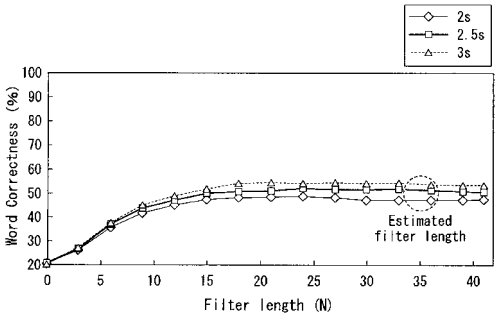
【図１４】



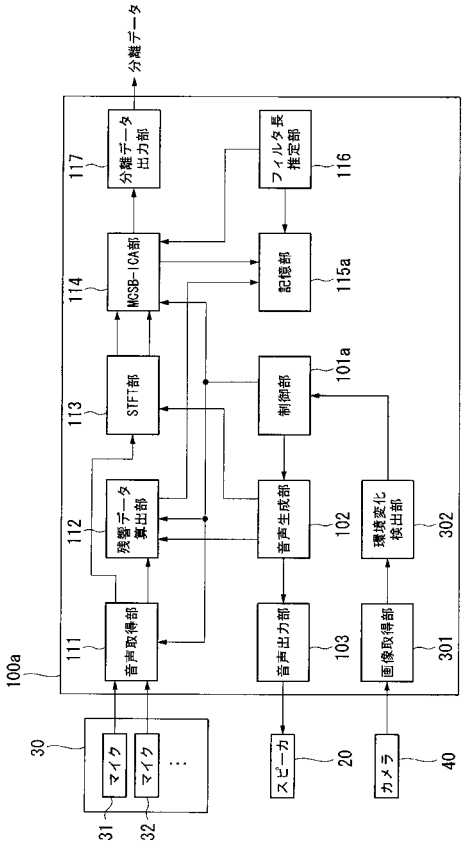
【図１５】



【図１６】



【図１７】



フロントページの続き

(72)発明者　中臺　一博
埼玉県和光市本町８－１　株式会社ホンダ・リサーチ・インスティチュート・ジャパン内
(72)発明者　武田　龍
埼玉県和光市本町８－１　株式会社ホンダ・リサーチ・インスティチュート・ジャパン内
(72)発明者　奥乃　博
埼玉県和光市本町８－１　株式会社ホンダ・リサーチ・インスティチュート・ジャパン内

審査官　山下　剛史

(56)参考文献　特開昭６４－２９０９４（ＪＰ，Ａ）
特開２００２－２３７７７０（ＪＰ，Ａ）
特開平１０－５６４０６（ＪＰ，Ａ）
特開昭６４－２９０９３（ＪＰ，Ａ）
特開２００９－２７６３６５（ＪＰ，Ａ）
特開２００９－１５９２７４（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ１０Ｌ　２１／０２－２１／０３０８
Ｈ０４Ｒ　３／００－　３／０２