

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5550456号  
(P5550456)

(45) 発行日 平成26年7月16日 (2014. 7. 16)

(24) 登録日 平成26年5月30日 (2014. 5. 30)

(51) Int. Cl.	F I
<b>G 1 O L 21/028 (2013.01)</b>	G 1 O L 21/028 C
<b>G 1 O L 21/0308 (2013.01)</b>	G 1 O L 21/0308 C
	G 1 O L 21/028 Z

請求項の数 8 (全 32 頁)

(21) 出願番号	特願2010-124873 (P2010-124873)	(73) 特許権者	000005326
(22) 出願日	平成22年5月31日 (2010. 5. 31)		本田技研工業株式会社
(65) 公開番号	特開2010-282193 (P2010-282193A)		東京都港区南青山二丁目1番1号
(43) 公開日	平成22年12月16日 (2010. 12. 16)	(74) 代理人	100064908
審査請求日	平成24年11月27日 (2012. 11. 27)		弁理士 志賀 正武
(31) 優先権主張番号	61/217, 760	(74) 代理人	100108578
(32) 優先日	平成21年6月4日 (2009. 6. 4)		弁理士 高橋 詔男
(33) 優先権主張国	米国 (US)	(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】

音声信号を生成し、生成した音声信号を出力する音声信号出力部と、  
前記音声信号出力部と音源とから発せられた音源信号を集音する複数のマイクロホンによって、前記生成された音声信号と前記生成された音声信号以外を含む入力信号を取得する音声取得部と、

前記生成された音声信号と前記入力信号と前記音源信号との相関関係を示す分離行列を算出するとともに、前記音源信号の分離度を前記入力信号に基づいて評価する第1評価関数を算出する第1評価関数算出部と、

前記第1評価関数に基づき音源信号の分離度が所定値より高くなる場合の前記分離行列である最適分離行列を算出し、前記算出した最適分離行列に基づいて、前記生成された音声信号以外の音声信号であって、残響成分が抑圧された当該音声信号を1人の人の音声信号として前記入力信号から分離することにより残響成分を抑圧する残響成分抑圧部と、

前記分離行列の更新量の大きさを定めるステップ・サイズ関数を、複数の区間に分け当該複数の区間毎に線形関数を用いて近似し、前記近似した線形関数に基づき前記ステップ・サイズを算出し、前記算出されたステップ・サイズを用いて、前記第1評価関数に基づき音源信号の前記分離度が前記所定値より高くなるように前記分離行列の更新を繰り返す分離行列更新部と、

を備えることを特徴とする残響抑圧装置。

【請求項 2】

前記分離行列更新部は、

前記線形関数を用いて近似した各区間の両端のステップ・サイズを用いて前記ステップ・サイズを前記複数の区間について区間毎に算出し、前記ステップ・サイズを評価するための第2評価関数を最小にするように前記ステップ・サイズを算出する

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】

前記分離行列更新部は、

前記線形関数を用いて近似した各区間の両端のステップ・サイズは、互いに隣り合う区間において連続するように前記複数の区間毎に、前記線形関数を用いて近似する

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。

10

【請求項4】

前記入力信号と前記生成された音声信号に対して球面化と分散正規化とを行うことで白色化する白色化部、

を更に備え、

前記第1評価関数算出部は、前記白色化された入力信号に基づき前記分離行列を算出する

ことを特徴とする請求項1から請求項3のいずれか1項に記載の残響抑圧装置。

【請求項5】

前記白色化部は、

前記白色化された信号に対して時間相関を強制的に除去し、当該時間相関を強制的に除去した信号から、前記白色化された信号に対して前記生成された音声信号と前記入力信号との相関を強制的に除去する

ことを特徴とする請求項4に記載の残響抑圧装置。

20

【請求項6】

前記残響成分抑圧部は、

前記複数の区間の数である区間数を算出するためのパラメータを球面化して算出し、算出したパラメータを用いて前記区間数を算出する

ことを特徴とする請求項1から請求項5のいずれか1項に記載の残響抑圧装置。

【請求項7】

前記残響成分抑圧部は、

独立成分分析法を用いて前記分離行列の更新と、前記生成された音声信号以外の音声信号であって、前記入力信号から残響成分が抑圧された当該音声信号を分離することにより残響成分を抑圧する

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。

30

【請求項8】

音声信号出力部が、音声信号を生成し、生成した音声信号を出力する音声信号出力工程と、

音声取得部が、前記音声信号出力部と音源とから発せられた音源信号を集音する複数のマイクロホンによって、前記生成された音声信号と前記生成された音声信号以外を含む入力信号を取得する音声取得工程と、

第1評価関数算出部が、前記生成された音声信号と前記入力信号と前記音源信号との相関関係を示す分離行列を算出するとともに、前記音源信号の分離度を前記入力信号に基づいて評価する第1評価関数を算出する第1評価関数算出工程と、

40

残響成分抑圧部が、前記第1評価関数算出工程により算出された前記第1評価関数に基づき音源信号の分離度が所定値より高くなる場合の前記分離行列である最適分離行列を算出し、前記算出した最適分離行列に基づいて、前記生成された音声信号以外の音声信号であって、残響成分が抑圧された当該音声信号を1人の人の音声信号として前記入力信号から分離することにより残響成分を抑圧する残響成分抑圧工程と、

分離行列更新部が、前記分離行列の更新量の大きさを定めるステップ・サイズの関数を、複数の区間に分け当該複数の区間毎に線形関数を用いて近似し、前記近似した線形関数

50

に基づき前記ステップ・サイズを算出し、前記算出されたステップ・サイズを用いて、前記第1評価関数に基づき音源信号の前記分離度が前記所定値より高くなるように前記分離行列の更新を繰り返す分離行列更新工程と、

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のため、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）。

【0003】

また、独立成分分析（ICA；Independent Component Analysis）を用いた残響抑圧手法として、フィルタの更新を、ステップ・サイズを固定して行う手法（例えば、非特許文献1参照）が提案されている。さらに、独立成分分析を用いた残響抑圧手法として、フィルタの更新において、ステップ・サイズを適応的に変更して行う手法（例えば、特許文献2参照）が提案されている。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開平9 261133号公報

【特許文献2】特開2008 306712号公報

【非特許文献】

【0005】

【非特許文献1】武田龍、中臺一博、駒谷和範、尾形哲也、奥乃博、“ロボット音声対話のためのMFTとICAによるバリエーション許容機能の評価”、情報処理学会、情報処理学会第70回全国大会、2008年、p2-135

【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、非特許文献1の従来技術では、分離フィルタの更新を、フレーム毎に固定されたステップ・サイズで行っていたため、ステップ・サイズを細かくした場合、残響音成分を抑圧する処理に時間がかかり、ステップ・サイズが適正で無い場合は残響抑制を精度良く行えないこともあるという問題点があった。また、特許文献2の従来技術では、多チャンネルのセミブラインド独立成分分析において、評価関数の非線形性が高いため、独立成分分析に用いる分離フィルタが高次になり、演算に時間がかかり、演算コストが高くなり実用的ではないという問題点があった。

【0007】

本発明は、上記の問題点に鑑みてなされたものであって、演算処理を低減しつつ、精度良く残響成分を抑圧す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】

【0008】

上記目的を達成するため、本発明に係る残響抑圧装置は、音声信号を生成し、生成した音声信号を出力する音声信号出力部と、前記音声信号出力部と音源とから発せられた音声信号を集音する複数のマイクロホンによって、前記生成された音声信号と前記生成された

10

20

30

40

50

音声信号以外を含む入力信号を取得する音声取得部と、前記生成された音声信号と前記入力信号と前記音源信号との相関関係を示す分離行列を算出するとともに、前記音源信号の分離度を前記入力信号に基づいて評価する第1評価関数を算出する第1評価関数算出部と、前記第1評価関数に基づき音源信号の分離度が所定値より高くなる場合の前記分離行列である最適分離行列を算出し、前記算出した最適分離行列に基づいて、前記生成された音声信号以外の音声信号であって、残響成分が抑圧された当該音声信号を1人の人の音声信号として前記入力信号から分離することにより残響成分を抑圧する残響成分抑圧部と、前記分離行列の更新量の大きさを定めるステップ・サイズの間接関数を、複数の区間に分け当該複数の区間毎に線形関数を用いて近似し、前記近似した線形関数に基づき前記ステップ・サイズを算出し、前記算出されたステップ・サイズを用いて、前記第1評価関数に基づき音源信号の前記分離度が前記所定値より高くなるように前記分離行列の更新を繰り返す分離行列更新部と、を備えることを特徴としている。

10

## 【0009】

また、本発明に係る残響抑圧装置において、前記分離行列更新部は、前記線形関数を用いて近似した各区間の両端のステップ・サイズを用いて前記ステップ・サイズを前記複数の区間について区間毎に算出し、前記ステップ・サイズを評価するための第2評価関数を最小にするように前記ステップ・サイズを算出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、前記分離行列更新部は、前記線形関数を用いて近似した各区間の両端のステップ・サイズは、互いに隣り合う区間において連続するように前記複数の区間毎に、前記線形関数を用いて近似するようにしてもよい。

20

## 【0011】

また、本発明に係る残響抑圧装置において、前記入力信号と前記生成された音声信号に対して球面化と分散正規化とを行うことで白色化する白色化部を更に備え、前記第1評価関数算出部は、前記白色化された入力信号に基づき前記分離行列を算出するようにしてもよい。

## 【0012】

また、本発明に係る残響抑圧装置において、前記白色化部は、前記白色化された信号に対して時間相関を強制的に除去し、当該時間相関を強制的に除去した信号から、前記白色化された信号に対して前記生成された音声信号と前記入力信号との相関を強制的に除去するようにしてもよい。

30

## 【0013】

また、本発明に係る残響抑圧装置において、前記残響成分抑圧部は、前記複数の区間の数である区間数を算出するためのパラメータを球面化して算出し、

算出したパラメータを用いて前記区間数を算出するようにしてもよい。

## 【0014】

また、本発明に係る残響抑圧装置において、前記残響成分抑圧部は、独立成分分析法を用いて前記分離行列の更新と、前記生成された音声信号以外の音声信号であって、前記入力信号から残響成分が抑圧された当該音声信号を分離することにより残響成分を抑圧するようにしてもよい。

40

## 【0015】

上記目的を達成するため、本発明に係る残響抑圧装置の残響抑圧方法は、音声信号出力部が、音声信号を生成し、生成した音声信号を出力する音声信号出力工程と、音声取得部が、前記音声信号出力部と音源とから発せられた音源信号を集音する複数のマイクロホンによって、前記生成された音声信号と前記生成された音声信号以外を含む入力信号を取得する音声取得工程と、第1評価関数算出部が、前記生成された音声信号と前記入力信号と前記音源信号との相関関係を示す分離行列を算出するとともに、前記音源信号の分離度を前記入力信号に基づいて評価する第1評価関数を算出する第1評価関数算出工程と、残響成分抑圧部が、前記第1評価関数算出工程により算出された前記第1評価関数に基づき音源信号の分離度が所定値より高くなる場合の前記分離行列である最適分離行列を算出し、

50

前記算出した最適分離行列に基づいて、前記生成された音声信号以外の音声信号であって、残響成分が抑圧された当該音声信号を1人の人の音声信号として前記入力信号から分離することにより残響成分を抑圧する残響成分抑圧工程と、分離行列更新部が、前記分離行列の更新量の大きさを定めるステップ・サイズの間数を、複数の区間に分け当該複数の区間毎に線形関数を用いて近似し、前記近似した線形関数に基づき前記ステップ・サイズを算出し、前記算出されたステップ・サイズを用いて、前記第1評価関数に基づき音源信号の前記分離度が前記所定値より高くなるように前記分離行列の更新を繰り返す分離行列更新工程と、を備えることを特徴としている。

【発明の効果】

【0016】

10

本発明によれば、音声信号出力部が、音声信号を生成して出力し、音声取得部が、マイクロホンからの入力信号を取得し、第1評価関数算出部が、音声信号と入力信号と音源信号との相関関係を示す分離行列を算出し、且つ入力信号から音源信号の分離度を評価する第1評価関数を算出し、残響成分抑圧部が、最適分離行列を算出し最適分離行列に基づき入力信号から音声信号以外の音源信号を分離することで残響成分を抑圧し、分離行列更新部が、分離行列の更新量を定めるステップ・サイズの間数を複数の区間に分け、複数の区間毎に線形関数で近似しステップ・サイズを算出し第1評価関数値に基づき音源信号の分離度が所定値より高くなるようにステップ・サイズに基づき分離行列の更新を繰り返して残響成分を抑圧するようにしたので、演算処理を低減しつつ、精度良く残響成分の抑圧を行うことができる。

20

【0017】

本発明によれば、マイクロホンが取得した音声信号を白色化し、第1評価関数算出部が白色化された信号を用いて分離行列を算出するようにしたので、さらに演算処理を低減しつつ、精度良く残響成分の抑圧を行うことができる。

【0018】

本発明によれば、ステップ・サイズを算出するための区間数を算出するためのパラメータを球面化して算出し、算出されたパラメータを用いて区間数を算出するようにしたので、演算処理を低減しつつ、精度良く残響成分の抑圧を行うことができる。

【図面の簡単な説明】

【0019】

30

【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MCSE-ICA部113の内部構成を説明する図である。

【図5】同実施形態に係るステップ・サイズの更新を説明する図である。

【図6】同実施形態に係るステップ・サイズの区間線形モデリングを説明する図である。

【図7】同実施形態に係る処理手順のフローチャートである。

【図8】第2実施形態に係る処理手順のフローチャートである。

【図9】第1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。

40

【図10】第1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】第1実施形態に係る実験条件を説明する図である。

【図12】第1実施形態に係る実験A（バージ・インの発生なし）且つ環境1の場合の音声認識率を示すグラフである。

【図13】第1実施形態に係る実験A（バージ・インの発生なし）且つ環境2の場合の音声認識率を示すグラフである。

【図14】第1実施形態に係る実験B（バージ・インの発生あり）且つ環境1の場合の音声認識率を示すグラフである。

【図15】第1実施形態に係る実験B（バージ・インの発生あり）且つ環境2の場合の音声認識率を示すグラフである。

50

【図 16】第 1 実施形態に係る実験 A と実験 B の P T R の結果をまとめた図である。

【発明を実施するための形態】

【0020】

以下、図 1 ~ 図 16 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 にはマイクロホン 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、マイクロホン 30 およびスピーカ 20 はそれぞれ複数収納されている。

10

【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 (B a r g e - i n) 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。

20

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 のマイクロホン 30 には、ヒト 2 の発話  $S_u$  が空間を經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  $S_r$  が空間を經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】

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、すなわち抑圧して、ヒト 2 の発話  $S_u$  を分離する。この MCSB-ICA において、ロボット 1 がマイクロホン 30 により取得した音声信号と、ロボット 1 による既知の発話とを白色化し、白色化された値を用いて、ICA により第 1 評価関数が最小値になるように反復処理を行うことで分離フィルタを算出し、マイクロホン 30 が取得した音声信号から残響成分を抑圧することでヒト 2 の音声信号を分離し、ヒト 2 の発話の音声信号  $S_u$  を算出する。また、ICA において、更新量の大きさであるステップ・サイズをいくつかの区間に分けた区間線形モデルにより算出する。なお、音源信号の分離度が、所定の値より高くなるまで反復処理を行う処理は、使用する評価関数によっては、ICA により第 1 評価関数の値が最小値になるように反復処理を行い、または、最大値になるように反復処理を行うようにしてもよい。

30

【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 にはマイクロホン 30、スピーカ 20 が接続され、マイクロホン 30 は複数のマイクロホン 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、STFT 部 112 と、MCSB-ICA 部 113 と、記憶部 114 と、分離データ出力部 115 とを備えている。

40

【0025】

制御部 101 は、ロボット 1 の発話である音声信号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 が発話中であることを示す信号を STFT 部 112 と MCSB-ICA 部 113 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、ロボット 1 の発話である音声

50

信号を生成し、生成した音声信号を音声出力部 103 と STFT 部 112 とに出力する。

【0027】

音声出力部 103 には、生成された音声信号が入力される。音声出力部 103 は、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、マイクロホン 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 112 に出力する。

【0029】

STFT (Short-time Fourier Transformation; 短時間フーリエ解析: 白色化部) 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る。STFT 部 112 は、取得された音声信号と生成された音声信号とにハニング等の窓関数を音声信号に乘じ、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 112 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  $x(\cdot, t)$  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  $s_r(\cdot, t)$  に変換し、変換した信号  $x(\cdot, t)$  と信号  $s_r(\cdot, t)$  を周波数ごとに MCSB-ICA 部 113 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(b) は、この取得された音声信号に乘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  $T$  は解析を行う期間を示している。なお、このシフト長さは、所定の値を用いる。

10

20

【0030】

MCSB-ICA 部 (第 1 評価関数算出部、残響成分抑圧部、分離行列更新部) 113 には、STFT 部 112 から変換された信号  $x(\cdot, t)$  と信号  $s_r(\cdot, t)$  が周波数ごとに入力される。また、MCSB-ICA 部 113 は、入力された信号  $x(\cdot, t)$  と信号  $s_r(\cdot, t)$  と記憶部 114 に記憶されている各モデル及び各係数を用いて、分離フィルタ  $W_r$ 、分離フィルタ  $W_{1u}$  および  $W_{2u}$  を、取得された音声信号からヒト 2 の直接音声信号  $S_u$  の分離度を示す第 1 評価関数に基づき音源信号の分離度が所定値より高くなるときの各分離フィルタを選択するようにして、反復処理により算出する。分離フィルタ  $W_r$ 、 $W_{1u}$  および  $W_{2u}$  算出後、算出された分離フィルタ  $W_r$ 、 $W_{1u}$  および  $W_{2u}$  を用いて、マイクロホン 30 が取得した音声信号からヒト 2 の直接発話信号  $s(t)$  を推定して分離し、分離した直接発話信号を分離データ出力部 115 に出力する。

30

【0031】

図 4 は、MCSB-ICA 部 113 の内部構成を説明する図である。図 4 のように、MCSB-ICA 部 113 は、強制空間球面化部 210 と、ICA 部 221 と、スケーリング部 231 と、直接音選択部 241 とを備え、強制空間球面化部 210 は、空間球面化部 211 と分散正規化部 212 を備えている。

STFT 部 112 から入力された信号  $x(\cdot, t)$  は、バッファ 201 を介して強制空間球面化部 210 の空間球面化部 211 に入力され、STFT 部 112 から入力された信号  $s_r(\cdot, t)$  は、バッファ 202 を介して強制空間球面化部 210 の分散正規化部 212 に入力される。そして、強制空間球面化部 210 は、入力された信号  $x(\cdot, t)$  と信号  $s_r(\cdot, t)$  を、球面化と分散正規化により白色化し、白色化した信号を ICA 部 221 に出力する。そして、ICA 部 221 には、強制空間球面化部 210 から白色化された信号が入力される。ICA 部 221 は、入力された信号を用いて、第 1 評価関数を最小値にするように繰り返し ICA 処理を行って分離フィルタを算出し、算出した分離フィルタを用いて分離する音声信号を算出する。そして、ICA 部 221 は、算出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音選択部 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、直接音選択部 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、選択した信号を分離データ出力部 115 に出力する。

40

【0032】

50

記憶部 114 には、ロボット 1 がマイクロホン 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、MCSB-ICA 部 113 の更新ルール等が予め書き込まれて記憶され、さらに、算出された分離フィルタ  $W_r$ 、分離フィルタ  $W_{1u}$  及び分離フィルタ  $W_{2u}$  が、更新中に書き込まれて記憶される。

【0033】

分離データ出力部 115 には、MCSB-ICA 部 113 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。

【0034】

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 114 には、ロボット 1 がマイクロホン 30 を介して取得する音声信号は、式 (1) の FIR (Finite Impulse Response; 有限インパルス応答) のモデルのように定義できる。

10

【0035】

【数 1】

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

【0036】

20

式 (1) において、記号  $x_1(t) \dots x_L(t)$  は、複数のマイクロホン 30 の各スペクトル ( $L$  はマイクロホン番号)、 $x(t)$  はベクトルであり  $[x_1(t), x_2(t), \dots, x_L(t)]^T$ 、 $s_u(t)$  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 がマイクロホン 30 を介して取得する時刻  $t$  におけるモデル化である。

【0037】

また、記憶部 114 には、ロボット 1 のマイクロホン 30 が集音した音声信号が、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 114 には、ロボット 1 の発話の音声信号が、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。

30

【0038】

【数 2】

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

【0039】

【数 3】

40

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

【0040】

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。

50



【 0 0 4 1 】

次に、独立成分分析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、MCSB-ICAの分離モデルを次式(4)のように定義し、記憶部114に記憶されている。

【 0 0 4 2 】

【数4】

$$\begin{pmatrix} \hat{s}(t) \\ X(t-d) \\ S_r(t) \end{pmatrix} = \begin{pmatrix} W_{1u} & W_{2u} & W_r \\ 0 & I_2 & 0 \\ 0 & 0 & I_r \end{pmatrix} \begin{pmatrix} x(t) \\ X(t-d) \\ S_r(t) \end{pmatrix} \quad \dots (4) \quad 10$$

【 0 0 4 3 】

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。また、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が含まれている。

【 0 0 4 4 】

【数5】

$$\hat{s}(t) \quad \dots (5) \quad 20$$

【 0 0 4 5 】

また、式(4)において、 $W_{1u}$ は、 $L \times L$ のブラインド分離行列(分離フィルタ)であり、 $W_{2u}$ は、 $L \times L(N+1)$ の分離行列(分離フィルタ)であり、 $W_r$ は、 $L \times (M+1)$ の残響音キャンセルの分離行列である。また、 $I_2$ と $I_r$ は、それぞれに対応する大きさの単位行列である。

【 0 0 4 6 】

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  $S_r(t)$  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。明示的な第1評価関数は、次式(6)のように表される。

【 0 0 4 7 】

【数6】

$$J(W) = -\sum_{i=1}^L E[\log p_i(s_i(t))] - \log |\det W_{1u}| + H \quad \dots (6) \quad 40$$

【 0 0 4 8 】

式(6)において、 $p_i(s_i)$ は確率変数  $s_i$  の結合確率密度関数であり、 $E[\cdot]$ は時間平均化演算子であり、 $H$ は  $\{x(t), X(t-d), S_r(t)\}$  の結合エントロピーである。

【 0 0 4 9 】

MCSB-ICA部113は、分離パラメータのセット  $W$  を、(6)式の第1評価関数が最小の値するように、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

を次式(7)~式(10)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(7)~式(10)は、記憶部114に予め書き込まれて記憶されている。

【0050】

【数7】

$$D = \Lambda - E[\phi(\hat{s}(t))\hat{s}^H(t)] \quad \dots(7)$$

【0051】

【数8】

10

$$W_{1u}^{[j+1]} = W_{1u}^{[j]} + \mu DW_{1u}^{[j]} \quad \dots(8)$$

【0052】

【数9】

$$W_{2u}^{[j+1]} = W_{2u}^{[j]} + \mu(DW_{2u}^{[j]} - E[\phi(\hat{s}(t))X^H(t-d)]) \quad \dots(9)$$

20

【0053】

【数10】

$$W_r^{[j+1]} = W_r^{[j]} + \mu(DW_r^{[j]} - E[\phi(\hat{s}(t))S_r^H(t)]) \quad \dots(10)$$

【0054】

なお、式(7)、式(9)~式(10)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(7)において、 $\Lambda$ は非ホロノミック拘束行列、すなわち、次式(11)の対角行列である。

30

【0055】

【数11】

$$E[\phi(\hat{s}(t))\hat{s}^H(t)] \quad \dots(11)$$

【0056】

また、式(8)~式(10)において、 $u$ は、更新量の大きさであるステップ・サイズのパラメータであり、 $\phi(x)$ は、非線形関数ベクトル $[\phi(x_1), \dots, \phi(x_L)]^H$ であり、次式(12)のように表される。

40

【0057】

【数12】

$$\phi(x) = -\frac{d}{dx} \log p(x) \quad \dots(12)$$

50

## 【0058】

さらに、音源の結合確率密度関数は、分散量  $\sigma^2$  であるとした場合、雑音に強い結合確率密度関数である  $p(x) = \exp(-|x|/\sigma^2) / (2\sigma^2)$  であり、 $\phi(x) = x^* / (2\sigma^2|x|)$  である。なお、 $x^*$  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  $|x| > 0$  において定義される。

## 【0059】

ICAの収束性の向上のため、強制空間球面化部210は、前処理として球面化の近似である強制空間球面化と分散正規化を行うことで白色化を行う。また、 $p(x) = \exp(-|x|/\sigma^2) / (2\sigma^2)$  において、球面化により入力信号を球面化(無相関化)し、分散を正規化するため、分散  $\sigma^2$  はほとんど1であると仮定する。

10

## 【0060】

次に、強制空間球面化について説明する。ユーザ発話に関する  $L(N+1) \times (K+J+1)$  の伝達特性行列  $H$  は、次式(13)のように表される。なお、 $L$  はロボット1のマイクロホンの本数である。 $N$  は伝達特性行列  $H$  の列数、 $K$  は伝達特性行列  $H$  の行、 $J$  は定数である。

## 【0061】

## 【数13】

$$H = \begin{pmatrix} h(0) & \cdots & \cdots & h(K) & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ 0 & \cdots & h(0) & \cdots & \cdots & h(K) \end{pmatrix} \quad \cdots(13)$$

20

## 【0062】

式(13)において、 $h(i) = [h_1(i), h_2(i), \dots, h_L(i)]^T$  である。そして、 $L(N+1) = K+J+1$  を満たすとき、伝達特性行列  $H$  は、 $L(N+1) \times L(N+1)$  の正方行列であるので、瞬時混合系で記述することができる。そして、既知のロボット発話を含む全体の過程は、次式(14)のように表すことができる。

## 【0063】

## 【数14】

$$\begin{pmatrix} X(t) \\ S_r(t) \end{pmatrix} = \begin{pmatrix} H & H_r \\ 0 & I \end{pmatrix} \begin{pmatrix} S_u(t) \\ S_r(t) \end{pmatrix} \quad \cdots(14)$$

30

## 【0064】

式(14)において、 $I$  は  $(M+1) \times (M+1)$  の単位行列であり、 $H_r$  は既知の行列  $L(N+1) \times (M+1)$  で式(15)のような伝達特性行列である。

## 【0065】

## 【数15】

$$H_r = \begin{pmatrix} h_r(0) & \cdots & h_r(M) \\ \vdots & \ddots & \vdots \\ 0 & \cdots & h_r(0) \end{pmatrix} \quad \cdots(15)$$

40

## 【0066】

式(15)において、 $h_r(i)$  は式(16)である。

## 【0067】

【数 1 6】

$$h_r(i) = \begin{pmatrix} h_{r_1}(i) \\ \vdots \\ h_{r_{L_r}}(i) \end{pmatrix} \quad \dots(16)$$

【0 0 6 8】

次に、強制空間球面化部 2 1 0 は、時空間相関行列 R の固有値、時空間相関行列 R の固有ベクトル E を用いて、次式 (17) ~ 式 (18) により球面化を行う。

10

【0 0 6 9】

【数 1 7】

$$R = \begin{pmatrix} E[X(t)X^H(t)] & E[X(t)S_r^H(t)] \\ E[S_r(t)X^H(t)] & E[S_r(t)\hat{S}_r^H(t)] \end{pmatrix} \quad \dots(17)$$

【0 0 7 0】

20

【数 1 8】

$$Z(t) = E\Lambda^{-\frac{1}{2}}E^H \begin{pmatrix} X(t) \\ S_r(t) \end{pmatrix} \quad \dots(18)$$

【0 0 7 1】

この固有値分解の計算量をさらに抑えるために、強制空間球面化部 2 1 0 は、強制的に時間相関の除去と、強制的に既知信号と観測信号との相関の除去を、次式 (19) ~ 式 (22) を用いて行う。

30

【0 0 7 2】

【数 1 9】

$$E[X(t)X^H(t)] = \begin{pmatrix} R(0) & \dots & 0 \\ \dots & \ddots & \vdots \\ 0 & \dots & R(0) \end{pmatrix} \quad \dots(19)$$

【0 0 7 3】

40

【数 2 0】

$$E[S_r(t)S_r^H(t)] = \begin{pmatrix} \lambda_r & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_r \end{pmatrix} \quad \dots(20)$$

【0 0 7 4】

【数 2 1】

$$E[X(t)S_r^H(t)] = 0 \quad \dots(21)$$

【0075】

【数 2 2】

$$E[S_r(t)X^H(t)] = 0 \quad \dots(22)$$

10

【0076】

式(19)～式(22)において、空間相関行列  $R(0)$  は、 $E[x(t)x^H(t)]$  であり、分散  $\sigma_r$  は式(23)である。

【0077】

【数 2 3】

$$E[s_r(t)s_r^H(t)] \quad \dots(23)$$

20

【0078】

式(19)～式(22)を用いて、式(17)と式(18)の演算を行うことは、観測信号(ロボット1のマイクロホン30が取得したロボット1の発話とヒト2の発話)に空間球面化を行うことと、既知信号(ロボット1の発話)にスケールの正規化を行うことを意味している。そして、強制的な時間相関の除去と、強制的な既知信号と観測信号との相関の除去とにより、観測信号  $X(t)$  は式(24)により変換され、既知信号  $S_r(t)$  は式(26)により変換され、強制空間球面化が行われる。

すなわち、MCSB-ICA部113の空間球面化部211には、変換された信号  $x(\omega, t)$  が周波数  $\omega$  ごとにされ、周波数  $\omega$  をインデックスとして順次、次式(24)を用いて空間球面化を行い、 $z(t)$  を算出する。

30

【0079】

【数 2 4】

$$z(t) = V_u x(t) \quad \dots(24)$$

【0080】

ただし、 $V_u$  は式(25)である。

40

【0081】

【数 2 5】

$$V_u = E_u \Lambda^{-\frac{1}{2}} E_u^H \quad \dots(25)$$

【0082】

さらに、式(25)において、 $E_u$  と  $\Lambda$  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。

50

さらに、MCSB-ICA部113の分散正規化部212には、変換された信号 $s_r(\tau, t)$ が周波数ごとにされ、周波数をインデックスとして順次、次式(26)を用いてスケールの正規化を行う。

【0083】

【数26】

$$\tilde{s}_r(t) = \lambda_r^{-\frac{1}{2}} s_r(t) \quad \dots(26)$$

10

【0084】

強制空間球面化後、式(4)、式(7)~式(10)の $x$ と $s_r$ が、 $z$ と式(5)に代入される。そして、空間球面化部211は、このように演算された $z(\tau, t)$ をICA部221に出力し、分散正規化部212は、スケール正規化された式(26)の値をICA部221に出力する。

【0085】

次に、分離プロセスの再起表現について説明する。まず、処理を行う上で、式(4)を次式(27)のように書き換える。

【0086】

【数27】

$$\hat{s}_r(t) = \sum_{n=0}^N W_u(n)x(t-n) + \sum_{m=0}^M W_r(m)s_r(t-m) \quad \dots(27)$$

20

【0087】

式(27)において、 $W_u(n)$ は $L \times L$ の分離行列であり、 $W_r(m)$ は $L \times 1$ の分離行列である。演算を簡単にするため、 $W_u(n)$ の範囲( $n$ は、1以上かつ $d$ 未満)を省略しているため、 $W_u(n)$ はこの範囲内でゼロに等しいと仮定される。また、 $W_{1u}$ は $W_u(0)$ に対応し、 $W_{2u}$ は $[W_u(d), \dots, W_u(N)]$ に対応し、および  
 $W_r$ は $[W_r(0), \dots, W_r(M)]$ に対応している。さらに、式(8)~式(10)は、増分記号 $W_x$ を用いることで、次式(28)のように書き換えることができる。なお、 $W_x$ とは、 $W_r$ と $W_{1u}$ および $W_{2u}$ を略して表している。

30

【0088】

【数28】

$$W_x^{[j+1]} = W_x^{[j]} + \mu \Delta W_x^{[j]} \quad \dots(28)$$

40

【0089】

式(27)、 $W_x$ により、推定された式(5)の再帰的表現は、次式(29)のように表される。

【0090】

【数 2 9】

$$\hat{s}^{[j+1]}(t) = \sum_{n=0}^N (W_u^{[j]}(n) + \mu \Delta W_u^{[j]}(n)) x(t-n) \\ + \sum_{m=0}^M (W_r^{[j]}(m) + \mu \Delta W_r^{[j]}(m)) s_r(t-m) \quad \dots(29)$$

【0091】

次式(30)～式(31)のように置くと、式(29)は式(32)のように再帰表現として表すことができる。 10

【0092】

【数 3 0】

$$y_u^{[j]}(t-n) = \Delta W_u^{[j]}(n) x(t-n) \quad \dots(30)$$

【0093】

【数 3 1】

20

$$y_r^{[j]}(t-m) = \Delta W_r^{[j]}(m) s_r(t-m) \quad \dots(31)$$

【0094】

【数 3 2】

$$\hat{s}^{[j+1]}(t) = \hat{s}^{[j]}(t) + \sum_{n=0}^N \mu y_u^{[j]}(t-n) + \sum_{m=0}^M \mu y_r^{[j]}(t-m) \quad \dots(32)$$

30

【0095】

次に、更新量の大きさであるステップ・サイズ  $u$  の推定を一般化するために、ステップ・サイズ  $\mu$  を、式(33)で表されるそれぞれの音源、遅延フレーム  $m$  と  $n$ 、それぞれの繰り返し係数  $j$  とが、各々異なるように係数を修正する。

【0096】

【数 3 3】

$$\hat{s}_i^{[j+1]} \quad \dots(33)$$

40

【0097】

次に、フレームが異なる  $L \times L$  の式(34)～式(35)の対角ステップ・サイズ行列を用いて、式(33)は式(36)のように書き換えることができる。

【0098】

【数 3 4】

$$\mu_r^{[j]}(n) = \text{diag}(\mu_{1,u}^{[j]}(n), \dots, \mu_{L,u}^{[j]}(n)) \quad \dots(34)$$

【0 0 9 9】

【数 3 5】

$$\mu_r^{[j]}(m) = \text{diag}(\mu_{1,r}^{[j]}(m), \dots, \mu_{L,r}^{[j]}(m)) \quad \dots(35) \quad 10$$

【0 1 0 0】

【数 3 6】

$$\hat{s}^{[j+1]}(t) = \hat{s}^{[j]}(t) + \sum_{n=0}^N \mu_u^{[j]}(n) y_u(t-n) + \sum_{m=0}^M \mu_r^{[j]}(m) y_r(t-m) \quad \dots(36)$$

20

【0 1 0 1】

なお、式(34)～式(35)において、記号  $\text{diag}$  は、対角成分を表している。式(35)の更新にともない更新規則である式(28)は、次式(37)のように書き換わる。

【0 1 0 2】

【数 3 7】

$$W_x^{[j+1]}(n) = W_x^{[j]}(n) + \mu_x^{[j]}(n) \Delta W_x^{[j]}(n) \quad \dots(37)$$

30

【0 1 0 3】

また、最適なステップ・サイズのパラメータである式(38)は、次式(39)のステップ・サイズのための第2評価関数を最小化することにより推定する。

【0 1 0 4】

【数 3 8】

$$\mu^{[j]} = \{\mu_u^{[j]}, \mu_r^{[j]}\} \quad \dots(38)$$

40

【0 1 0 5】

【数 3 9】

$$J(\mu^{[j]}) = J(W^{[j+1]}) - J(W^{[j]}) \propto J(W^{[j+1]}) \quad \dots(39)$$

【0 1 0 6】

式(39)は、式(6)を用いて、次式(40)のように置き換えられる。

50



【 0 1 0 7 】

【 数 4 0 】

$$J(\mu^{[j]}) = -\sum_{i=1}^L E[\log p_i(s_i^{[j+1]}(t))] - \log |\det W_u^{[j+1]}(0)| \quad \dots(40)$$

【 0 1 0 8 】

式(40)において、第2項の式(41)は、第1項の式(42)に比べて値が非常に小さいと仮定する。この仮定により、式(40)は、式(43)のように近似される。

10

【 0 1 0 9 】

【 数 4 1 】

$$-\log |\det W_u^{[j+1]}(0)| \quad \dots(41)$$

【 0 1 1 0 】

【 数 4 2 】

$$-\sum_{i=1}^L E[\log p_i(s_i^{[j+1]}(t))] \quad \dots(42)$$

20

【 0 1 1 1 】

【 数 4 3 】

$$J(\mu^{[j]}) \approx -\sum_{i=1}^L E[\log p_i(s_i^{[j+1]}(t))] \quad \dots(43)$$

30

【 0 1 1 2 】

この近似の結果、 $J(\mu)$ を最小化する代わりに、式(44)をそれぞれ独立して最小化することができる。

【 0 1 1 3 】

【 数 4 4 】

$$-E[\log p_i(s_i^{[j+1]}(t))] \quad \dots(44)$$

40

【 0 1 1 4 】

式(44)は、まだ推定すべき多数のパラメータを有しているため、さらにパラメータ数を削減し演算コストを削減する。

次に、さらに演算コストを削減する方法について、図5を用いて説明する。図5は、ステップ・サイズの区間線形モデリングを説明する図である。図5(a)は、区間線形モデリングを行わない場合のステップ・サイズを示す図であり、図5(b)は、区間モデリングを行う場合のステップ・サイズを示す図である。すなわち、図5(b)のように、ステップ・サイズの関数を全て演算した場合、演算量が多い。このため、いくつかの区間に分け、分けた区間内で線形近似することで演算量を大幅に削減する。

50

観測された項である式(45)と既知の音源の項である式(46)に対する第*i*番目の要素のステップ・サイズ・パラメータは、遅延フレームの個数の項とほとんど同じであり、すなわち式(47)である。

【0115】  
【数45】

$$\mu_{i,u}^{[j]}(n) \quad \dots(45)$$

10

【0116】  
【数46】

$$\mu_{i,u}^{[j]}(m) \quad \dots(46)$$

【0117】  
【数47】

$$\mu_i^{[j]}(n) = \mu_{i,u}^{[j]}(n) = \mu_{i,r}^{[j]}(n) \quad \dots(47)$$

20

【0118】

ここで、式(48)の第*i*番目の要素を式(49)のように置き換え、式(50)の第*i*番目の要素を式(51)のように置き換えると、式(36)を次式(52)のように簡素化できる。

【0119】  
【数48】

$$y_u^{[j]} \quad \dots(48)$$

30

【0120】  
【数49】

$$y_{i,u}^{[j]} \quad \dots(49)$$

40

【0121】  
【数50】

$$y_r^{[j]} \quad \dots(50)$$

【0122】

50

【数 5 1】

$$y_{i,r}^{[j]} \quad \dots(51)$$

【0 1 2 3】

【数 5 2】

$$\hat{s}^{[j+1]}(t) = \hat{s}_i^{[j]} + \sum_{n=0}^N \mu_i^{[j]}(n) (y_{i,u}^{[j]}(t-n) + y_{i,r}^{[j]}(t-n)) \quad \dots(52) \quad 10$$

【0 1 2 4】

式(52)において、式(53)のように置くと、次式(54)のように表せる。

【0 1 2 5】

【数 5 3】

$$y_i^{[j]}(t-n) = y_{i,u}^{[j]}(t-n) + y_{i,r}^{[j]}(t-n) \quad \dots(53) \quad 20$$

【0 1 2 6】

【数 5 4】

$$\hat{s}^{[j+1]}(t) = \hat{s}_i^{[j]} + \sum_{n=0}^N \mu_i^{[j]}(n) y_{i,u}^{[j]}(t-n) \quad \dots(54)$$

【0 1 2 7】

次に、図5(b)のように、 $\mu_i^{[j]}(n)$ をP個に分割された区間線形関数で近似し、次式(55)のように定義する。

【0 1 2 8】

【数 5 5】

$$\mu_i^{[j]}(n) = \begin{cases} \mu_{i,p}^{[j]} + \frac{n - B_p}{B_{p+1} - B_p} (\mu_{i,p+1}^{[j]} - \mu_{i,p}^{[j]}) & (B_p \leq n \leq B_{p+1}) \\ 0 & \text{otherwise} \end{cases} \quad \dots(55)$$

30

40

【0 1 2 9】

式(55)において、式(56)と式(57)は、 $n$ の第 $p$ 番目の範囲 $[B_p, B_{p+1}]$ 内の端点301と302である。図5(b)において、最終区間Pの端点における値はゼロ、すなわち式(58)がゼロであると仮定する。また、図5(b)のように、線分を311~313に区切った場合、線分311と線分312の接続点である端点301は同じ値のため、線分311と線分312との接続点を連続するように近似している。

【0 1 3 0】

【数56】

$$\mu_{i,p}^{[j]} \quad \dots(56)$$

【0131】

【数57】

$$\mu_{i,p+1}^{[j]} \quad \dots(57)$$

10

【0132】

【数58】

$$\mu_{i,P}^{[j]} \quad \dots(58)$$

20

【0133】

次に、この線形区間モデルと式(55)を用いて、式(54)の第2項を次式(59)のように置き直す。

【0134】

【数59】

$$\sum_{n=0}^N \mu_i^{[j]}(n) y_i^{[j]}(t-n) = \sum_{p=0}^{P-1} \sum_{n=B_p}^{B_{p+1}-1} \left( \mu_{i,p}^{[j]} + \frac{n-B_p}{B_{p+1}-B_p} (\mu_{i,p+1}^{[j]} - \mu_{i,p}^{[j]}) \right) y_i^{[j]}(t-n)$$

$$\dots(59)$$

30

【0135】

さらに、式(58)の項でまとめて(summarizing)、そのパラメータを  $Y_i(t, p)$  に置き換えると、P個の項の和で表せるので、式(59)、すなわち式(54)は次式(60)のように表すことができる。

【0136】

【数60】

$$\hat{s}^{[j+1]}(t) = \hat{s}_i^{[j]} + \sum_{p=0}^{P-1} \mu_{i,p}^{[j]}(n) Y_i(t, p) \quad \dots(60)$$

40

【0137】

あるいは、行列表現を用いて、次式(61)のように表せる。

【0138】

【数 6 1】

$$\begin{pmatrix} \hat{s}_i^{[j+1]}(t) \\ Y_i(t) \end{pmatrix} = \begin{pmatrix} 1 & \mu_i^{[j]T} \\ 0 & I \end{pmatrix} \begin{pmatrix} \hat{s}_i^{[j]}(t) \\ Y_i(t) \end{pmatrix} \quad \dots(61)$$

【0 1 3 9】

式(61)において、式(62)は式(63)であり、 $Y_i(t)$ は式(64)である

10

【0 1 4 0】

【数 6 2】

$$\mu_i^{[j+1]} \quad \dots(62)$$

【0 1 4 1】

【数 6 3】

$$[\mu_{i,0}^{[j]}, \dots, \mu_{i,P-1}^{[j]}]^T \quad \dots(63)$$

20

【0 1 4 2】

【数 6 4】

$$[Y_i(t,0), \dots, Y_i(t,P-1)]^T \quad \dots(64)$$

【0 1 4 3】

このように、区間線形モデルにより、ステップ・サイズ推定する場合において、区間数  $P$  の数は、式(4)と比較して大幅に減少し、演算コストを下げるができる。

30

そして、1回目の繰り返しから(1+1)回目の繰り返しまでのステップ・サイズである式(65)の更新規則は、式(66)のように表される。

【0 1 4 4】

【数 6 5】

$$\mu_i^{[j]} \quad \dots(65)$$

【0 1 4 5】

【数 6 6】

$$\mu_i^{j,[l+1]} = \mu_i^{j,[l]} - \gamma E[\text{Re}[\phi(\hat{s}_i^{[j+1]})Y_i(t)]] \quad \dots(66)$$

40

【0 1 4 6】

式(66)において、 $\gamma$ はステップ・サイズのパラメータであり、 $\text{Re}[x]$ は $x$ の実数を表す。そして、この式(66)の適応において、パラメータ $\gamma$ と区間数 $P$ を設定する必要がある。

50

次に、区間数  $P$  を算出するために、次式 (67) ~ 式 (68) の規則を用いて  $Y_i(t)$  の球面化により収束速度を早める。

【0147】

【数67】

$$P_i(t) = V_i Y_i(t) \quad \dots(67)$$

【0148】

【数68】

$$V_i = E \Lambda_i^{-\frac{1}{2}} E_i^T \quad \dots(68)$$

【0149】

式 (68) において、 $E_i$  は、式 (69) の固有ベクトル行列であり、 $\Lambda_i$  は、式 (69) の固有値対向行列である。

【0150】

【数69】

$$R_i = E \left[ \text{Re} \left[ Y_i(t) Y_i^H(t) \right] \right] \quad \dots(69)$$

【0151】

式 (61) ~ 式 (64)、式 (66) の  $Y_i(t)$  は、全て  $P_i(t)$  に代入される。この結果、式 (37) におけるステップ・サイズ  $u$  は、式 (70) に置き換わる。

【0152】

【数70】

$$\mu_i^{[j]} V_i \quad \dots(70)$$

【0153】

また、この処理は、 $W$  の  $j$  回目の反復で式 (65) が収束するのを待たずに、式 (65) の反復を  $q$  回停止することができる。これは、この適応が  $W$  の推定の一部であるためであり、式 (65) を  $(j+1)$  回目の反復における式 (62) の初期値として再利用することができるためである。そして、 $j$  回目の反復でステップ・サイズのパラメータ  $\gamma$  をスケジューリングするために次式 (71) のアニーリング法を用いる。

【0154】

【数71】

$$\gamma^{[j]} = \frac{\alpha_\gamma}{j} + \beta_\gamma \quad \dots(71)$$

【0155】

10

20

30

40

50

なお、式(71)において、およびは定数パラメータである。

【0156】

次に、残響抑圧の処理手順を、図1、図5～図7を用いて説明する。図6は、ステップ・サイズの更新を説明する図である。図7は、本実施形態における処理手順のフローチャートである。まず、MCSB-ICA部113は、繰り返し処理で用いる係数*i*と係数*j*および1をそれぞれ初期化(=1)にする(ステップS1)。

【0157】

次に、音声取得部111には、マイクロホン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112に出力する(ステップS2)。マイクロホン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 $h_r$ 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。

10

【0158】

次に、STFT部112には、マイクロホン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 $x(\cdot, t)$ に変換し、変換した信号 $x(\cdot, t)$ を周波数ごとにMCSB-ICA部113に出力する。また、STFT部112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 $s_r(\cdot, t)$ に変換し、変換した信号 $s_r(\cdot, t)$ を周波数ごとにMCSB-ICA部113に出力する。

【0159】

20

MCSB-ICA部113の強制空間球面化部210には、バッファ201とバッファ202を介して、周波数ごとに変換された信号 $x(\cdot, t)$ (マイクロホン入力に基づく信号)と変換された信号 $s_r(\cdot, t)$ (ロボット発話に基づく信号)とが入力され、入力された信号 $x(\cdot, t)$ と信号 $s_r(\cdot, t)$ に対して、強制空間球面化を行う(ステップS3)。

強制空間球面化は、MCSB-ICA部113の空間球面化部211と分散正規化部212により以下のように行われる。空間球面化部211には、変換された信号 $x(\cdot, t)$ が周波数ごとに入力され、周波数をインデックスとして順次、式(24)を用いて空間球面化を行い、 $z(t)$ に変換する。分散正規化部212には、変換された信号 $s_r(\cdot, t)$ が周波数ごとに入力され、周波数をインデックスとして順次、式(26)を用いてスケールの正規化を行う。

30

空間球面化部211は、このように演算された $z(\cdot, t)$ をICA部221に出力し、分散正規化部212は、このように正規化された式(26)の値をICA部221に出力する。

【0160】

以下、*i*番目の分離音 $s_i(t)$ について、ステップS4～ステップS12を*L*回(*L*はマイクロホン30の個数)繰り返すことで更新し、さらに、ステップS4～ステップS14を*L*回繰り返すことで残響成分を抑圧したヒト2の発話信号を抽出する。

まず、ICA部221は、*i*=1番目の分離音 $s^{[j]}(t)$ の予測値(式(5))を、式(36)を用いて算出する(ステップS4)。

40

【0161】

次に、ICA部221は、*j*番目の分離フィルタ $W_{1u}$ 、 $W_{2u}$ 及び $W_r$ の各勾配 $W^{[j]}$ を、算出された分離音 $s_r(t)$ の予測値と式(7)～式(10)を用いて算出する(ステップS5、図6の勾配 $W^{[j]}$ 算出工程401)。

【0162】

次に、ICA部221は、*l*番目のステップ・サイズ*u*演算用のパラメータ $Y_i(t)$ を、時間減衰を考慮して式(67)と式(68)の規則に従って球面化する(ステップS6)。

【0163】

次に、ICA部221は、ステップS6で算出された*l*番目の $Y_i(t)$ を用いて、式

50

(67)により区間線形モデルの区間数 $P$ を算出する。

次に、ICA部221は、ステップ・サイズのパラメータを、式(71)により算出する。

次に、ICA部221は、算出した区間数 $P$ とステップ・サイズのパラメータを用いて、式(66)により式(62)のステップ・サイズを算出して更新する(ステップS7、図6のステップ・サイズ $\mu$ 算出工程403)。

なお、ステップ・サイズの推定のための反復回数は、 $q$ 回( $q$ は1以上、 $L$ 未満の自然数)で停止するようにしても良い。

【0164】

次に、ICA部221は、 $l$ が所定の値の $L$ 未満か否かを判定し(ステップS8)、 $l$ が $L$ 未満の場合(ステップS8; Yes)、 $l$ に1を加算し(ステップS9)、ステップS6に戻り、 $l$ が $L$ 以上の場合(ステップS8; No)、ステップS10に進む。

10

【0165】

$l$ が $L$ 以上の場合(ステップS8; No)、ICA部221は、分離フィルタ $W_{1u}$ 、 $W_{2u}$ 及び $W_r$ の次( $j+1$ )の各分離フィルタ $W^{[j+1]}$ を、式(37)を用いて更新する(ステップS10、図6の勾配 $W^{[j]}$ 算出工程401、ステップ・サイズ $\mu$ 算出工程403、乗算工程404、加算工程405)。なお、ICA部221は、式(6)の $J(W)$ が小さくなるように、各分離フィルタを推定して更新していく。

【0166】

次に、ICA部221は、 $j$ が所定の値の $L$ 未満か否かを判定し(ステップS11)、 $j$ が $L$ 未満の場合(ステップS11; Yes)、 $j$ に1を加算し(ステップ12)、ステップS4に戻り、 $j$ が $L$ 以上の場合(ステップS11; No)、ステップS13に進む。

20

【0167】

$j$ が $L$ 以上の場合(ステップS11; No)、ICA部221は、 $i$ が所定の値の $L$ 未満か否かを判定し(ステップS13)、 $i$ が $L$ 未満の場合(ステップS13; Yes)、 $i$ に1を加算し(ステップS14)、ステップS6に戻り、 $i$ が $L$ 以上の場合(ステップS13; No)、ステップS15に進む。

$i$ が $L$ 以上の場合(ステップS13; No)、ICA部221は、算出された各分離音 $s(t)$ の推定値結果をスケール部231に出力し、スケールされた信号を直接音選択部241に出力する。なお、スケール部231は、projection Back処理等を用いてスケールを行い、直接音選択部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分離データ出力部115に出力する(ステップS15)。

30

すなわち、直接音選択部241は、複数のマイク30が取得した音声信号から、ヒト2の発話信号 $h_u$ から残響成分を抑圧して、ヒト2の直接発話である音声信号 $S_u$ を選択している。

【0168】

以上のように、マイクロホン30が取得した音声信号を白色化し、白色化された信号を区間線形モデルにより算出した区間数に基づきICA処理して分離フィルタ( $W_r$ 、 $W_{1u}$ 、 $W_{2u}$ )と分離音 $s(t)$ の推定値とを、区間線形モデルを用いて区間数 $P$ に分けて算出したステップ・サイズを用いて更新するようにしたので、残響抑圧における演算効率を上げることができ、少ない演算量で精度の良い残響抑圧を行うことができる。また、ステップ・サイズ算出時に、パラメータ $Y_i(t)$ を球面化して算出するようにしたので、より少ない演算量で残響抑圧を行うことができる。

40

【0169】

[第2実施形態]

第1実施形態との違いは、ステップS6aの処理である。第1実施形態では、ステップ・サイズ $u$ 演算用のパラメータ $Y_i(t)$ を球面化して算出したが、第2実施形態では、時間減衰を考慮せず、すなわち、球面化を行わずにステップ・サイズ $u$ のパラメータ $Y_i(t)$ を算出する(ステップS6a)。なお、時間減衰を考慮しないとは、 $Y_i(t)$ の要素が1の場合に相当する。すなわち、パラメータ $Y_i(t)$ の要素数に応じて、要素数

50



が1の場合は、第2実施形態のように球面化を行わずにパラメータ $Y_i(t)$ を算出してステップ・サイズ $u$ を算出し、要素数が1以上の場合は、第1実施形態のように球面化を行ってパラメータ $Y_i(t)$ を算出してステップ・サイズ $u$ を算出する。

【0170】

以上のように、パラメータ $Y_i(t)$ の要素数に応じて、要素数が1の場合は、球面化を行わずにパラメータ $Y_i(t)$ を算出してステップ・サイズ $u$ を算出するようにしたので、第1実施形態と同様に、残響抑圧における演算効率を上げることができ、少ない演算量で精度の良い残響抑圧を行うことができる。

【0171】

[実験結果例]

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一例を説明する。図9～図12は、実験条件である。音声データに対するこのインパルス応答は、通常の部屋(環境I;残響時間 $RT_{20} = 240\text{ msec}$ ,  $4.2\text{ m} \times 7.0\text{ m}$ )と、ホールに似た部屋(環境II;残響時間 $RT_{20} = 670\text{ msec}$ ,  $7.55\text{ m} \times 9.55\text{ m}$ )の2つの部屋において、 $16\text{ KHz}$ で記録された。

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 $16\text{ KHz}$ サンプル、残響時間は $240\text{ ms}$ と $670\text{ ms}$ 、ロボット1とヒト2との距離は $1.5\text{ m}$ 、ロボット1とヒト2の角度は $0$ 度、 $45$ 度、 $90$ 度、 $-45$ 度、 $-90$ 度、使用したマイクロホン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 $64\text{ ms}$ ( $512$ ポイント)かつシフト量 $24\text{ ms}$ ( $160$ ポイント)、入力信号データ( $16$ ビット、PCM(Pulse Code Modulation))は $[-1.0\ 1.0]$ に正規化されたものである。また、ロボット1の頭部に取り付けられたマイクロホン30の本数は2本である。

【0172】

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 $32\text{ ms}$ かつシフト量 $10\text{ ms}$ 、特徴量はMFCC(Mel-Frequency Cepstrum Coefficient;スペクトル包絡)は25次(12次MFCCと12次MFCCと1次元パワー)であり、次いで、ケプストラム平均正規化を用いた。なお、パワーは、例えば、フレーム中の信号のパワーを計算し、計算に用いたフレームと前後2フレームずつ計5フレームを使って線形回帰を行って算出する。

また、ユーザ(ヒト2)の音声およびロボットの音声について200個の日本語の文章を使用し、これらの文章を対応する記録されたインパルス応答に畳み込んだ。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。さらに、200人の男性と女性の話者によって発せられた明瞭な音声の150個の文章を使ってtriphoneベースの音響モデル(3状態および4混合)を学習した(閉じた語)。統計的言語モデルは、2万語からなり、これらの語は新聞から抽出した。

【0173】

実験は、バージ・インが発生していない場合(実験A)、バージ・インが発生している場合(実験B)の2種類である。バージ・インが発生していない場合は、ロボット1が取得する音声信号は、ユーザの音声信号(含む残響音)のみであり、バージ・インが発生している場合は、ユーザの音声信号とロボット1の音声信号とを含んでいる。また、全てのデータは、行列 $W_{1u}$ 、 $W_{2u}$ および $W_r$ を推定するために使用された(バッチ処理)。また、各実験では、PTR(Processing Time Ratio;処理時間比)も評価した。

【0174】

さらに、各実験において、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 $N$ と通常分離モードの残響除去のフィルタ長 $M$ は同じ値とし、環

10

20

30

40

50

境 I の  $N (= M)$  を 9 に設定し、環境 I I の  $N (= M)$  を 23 に設定した。  
 さらに、各実験において図 11 の 6 つの条件で実験を行った。図 11 は、実験条件を説明する図である。図 11 のように、実験条件 1 は、本実施形態の方法を用いた方法で、区間数  $P = 1$ 、パラメータは  $\mu = 0.2$  および  $\sigma = 5.0 \times 10^{-3}$  であり、実験条件 2 は、本実施形態の方法を用いた方法で、区間数  $P = 2$ 、パラメータは  $\mu = 0.2$  および  $\sigma = 5.0 \times 10^{-3}$  であり、実験条件 3 は、本実施形態の方法を用いた方法で、区間数  $P = 0$ 、パラメータは  $\mu = 0.2$  および  $\sigma = 5.0 \times 10^{-3}$  である。なお、区間数  $P = 2$  は  $(B_0, B_1, B_2) = (0, 4, N)$ 、区間数  $P = 1$  は  $(B_0, B_1) = (0, N)$  と設定した。ここで、区間数  $P = 0$  は、すべての  $\mu(n)$  について同じ適応ステップ・サイズを使用することを意味している。さらに、サブ ICA フィルタの最大反復回数

10

は、3 に設定した。  
 実験条件 4 は、ステップ・サイズをアニーリング・ベースで算出した値を用いる場合 ( $\mu = 0.9$ ) であり、実験条件 5 は、ステップ・サイズ  $\mu$  を 0.05 に固定した場合であり、実験条件 6 は、ステップ・サイズ  $\mu$  を 0.1 に固定した場合である。なお、アニーリング・ベースにおいてステップ・サイズの算出は、パラメータ  $\sigma = 0.6$  および  $\sigma = 5.0 \times 10^{-3}$  である。

【0175】

つぎに、ICA 処理による演算コストの比較である PTR について説明する。まず、使用する適応法は、それぞれの分離フィルタ  $W$  のそれぞれの反復において、 $(LP^2)$  ( $L$  はマイクロホンの本数、 $P$  はステップ・サイズ算出の区間数、 $\sigma$  は漸近記法におけるオーダーである。) を必要とする。そして、 $W$  を推定するために  $p_1$  回の反復を行い、ステップ・サイズ  $u_i$  を推定するために  $q_1$  回の反復を行う場合、総演算コストは  $p_1(L^2(N+M) + q_1(LP^2))$  である。一方、 $p_2$  回の反復を使用する標準的な ICA 処理との処理時間比 (PTR) は、おおよそ次式 (72) のように表すことができる。

20

【0176】

【数72】

$$PTR \propto \frac{p_1(L^2(N+M) + q_1(LP^2))}{p_2(L^2(N+M))} = \left(1 + \frac{q_1P^2}{L(N+M)}\right) \frac{p_1}{p_2}$$

30

...(72)

【0177】

式 (72) のように、多数のマイクロホン 30 を使用し、大きな  $N$  および  $M$  ( $N$  と  $M$  は分離行列の行数と列数) を用いた場合に、一般的な手法に比べ効率的であることを意味している。

【0178】

次に、実験結果を図 12 ~ 図 16 を用いて説明する。図 12 は、バージ・インが発生していない場合の環境 1 における音声認識率の一例を示す図であり、図 13 は、バージ・インが発生していない場合の環境 2 における音声認識率の一例を示す図である。また、図 14 は、バージ・インが発生している場合の環境 1 における音声認識率の一例を示す図であり、図 15 は、バージ・インが発生している場合の環境 2 における音声認識率の一例を示す図である。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。各グラフの横軸は反復回数であり、縦軸は音声認識率 (%) である。

40

【0179】

図 12 ~ 図 15 のように、ステップ・サイズを固定した場合、収束が速いのは、実験条件 6 の大きなステップ・サイズ ( $\mu = 0.1$ ) の方が実験条件 5 の小さなステップ・サイズ ( $\mu = 0.05$ ) より速いが、収束後の音声認識率は、実験条件 6 の大きなステップ・サイ

50

ズ ( $\mu = 0.1$ )の方が実験条件5の小さなステップ・サイズ ( $\mu = 0.05$ )より悪い。すなわち、ステップ・サイズを固定した手法では、分離性能 (音声認識率) と収束速度とを両立できないことが分かる。一方、実験条件4のステップ・サイズをアーニング・ベースで算出した場合、ステップ・サイズを固定した実験条件5と6より、全ての実験で分離性能と収束速度が勝っていた。

次に、本実施形態の方法において、実験条件1の区間数  $P = 2$  および実験条件2の区間数  $P = 1$  は、ほとんどすべての状況において、他のすべての実験条件4~6に勝っていた。なお、区間数  $P = 0$ の結果は、1つの適応ステップ・サイズでの性能を意味し、すなわち、区間線型モデルを適用していない場合の実験結果を意味している。この点において、区間数  $P = 2$ の実験条件1または  $P = 1$ の実験条件2の方が、区間数  $P = 0$ の実験条件3より収束速度が速く、例えば10程度の少ない反復回数でも音声認識率が高い (分離性能が良い)。

#### 【0180】

図12~図15において、点線501~504は、ステップ・サイズをアーニングにより算出し反復回数30回の場合の音声認識率である。また、各図のクロスポイント512~514は、実験条件1で実験を行ったときステップ・サイズをアーニングにより算出し反復回数30回の場合の音声認識率と同じ音声認識率になる反復回数である。すなわち、本実施形態の方法によれば、従来の手法を用いた他の実験条件に比べて、反復回数が半分または2/3程度に改善されている。

#### 【0181】

図16は、実験Aと実験BのPTRの結果をまとめた図である。図16のように、残響音が長い環境2においてバージ・インが発生している場合でも、PTRは1.4であり、例えば、従来手法では  $PTR = 2.0$ であったため、大幅に演算効率が改善されていることがわかる。一例として、残響音が長い環境2においてバージ・インが発生している場合、反復回数が16回では、本実施形態のリアルタイム係数 (= 処理時間 / データ持続時間) が1.0未満であった。

#### 【0182】

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。

#### 【0183】

なお、実施形態の図2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】

10

20

30

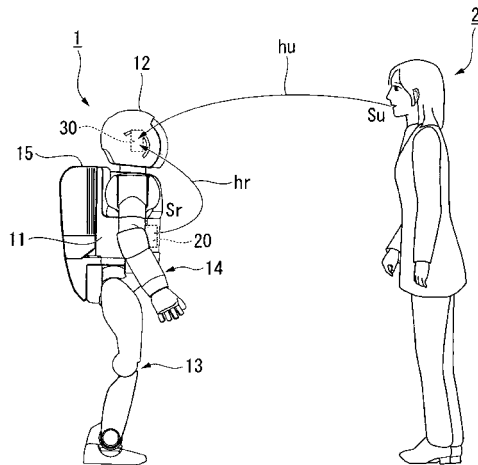
40

50

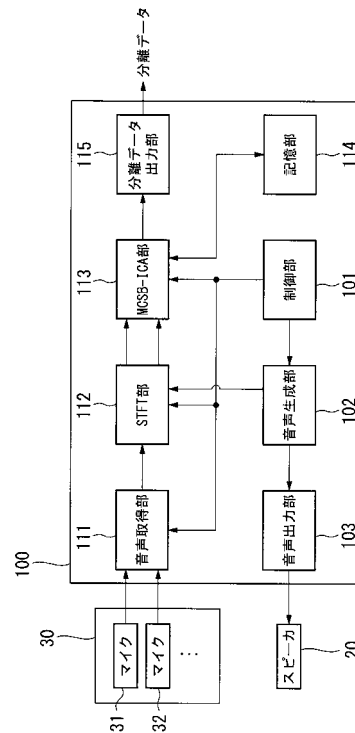
【 0 1 8 4 】

- 1・・・ロボット
- 20・・・スピーカ
- 30、31、32・・・マイクロホン
- 100・・・残響抑圧装置
- 101・・・制御部
- 102・・・音声生成部
- 111・・・音声取得部
- 112・・・STFT部(白色化部)
- 113・・・MCSB-ICA部(第1評価関数算出部、残響成分抑圧部、分離行列更新部)
- 114・・・記憶部
- 115・・・分離データ出力部
- 210・・・強制空間球面化部
- 211・・・空間球面化部
- 212・・・分散正規化部
- 221・・・ICA部
- 231・・・スケール部
- 241・・・直接音選択部

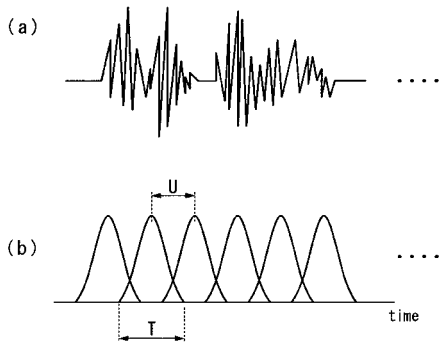
【 図 1 】



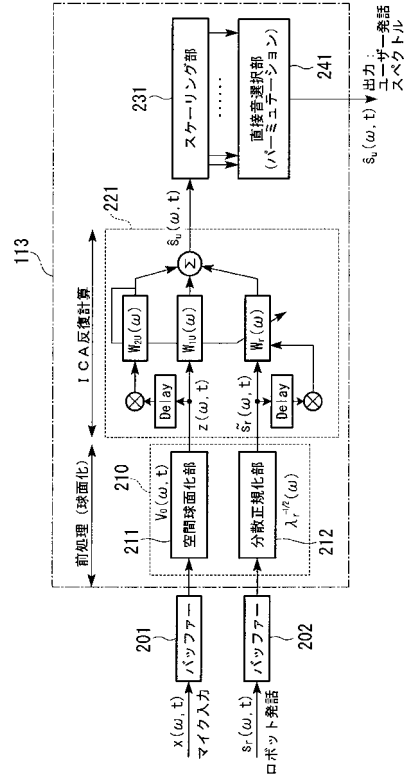
【 図 2 】



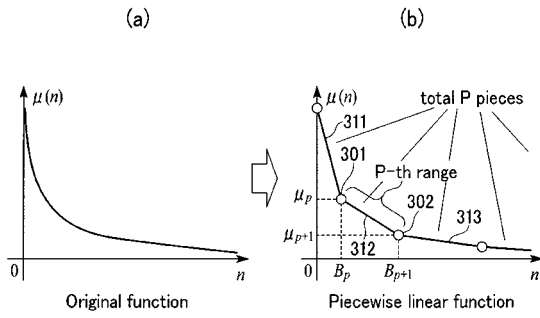
【図3】



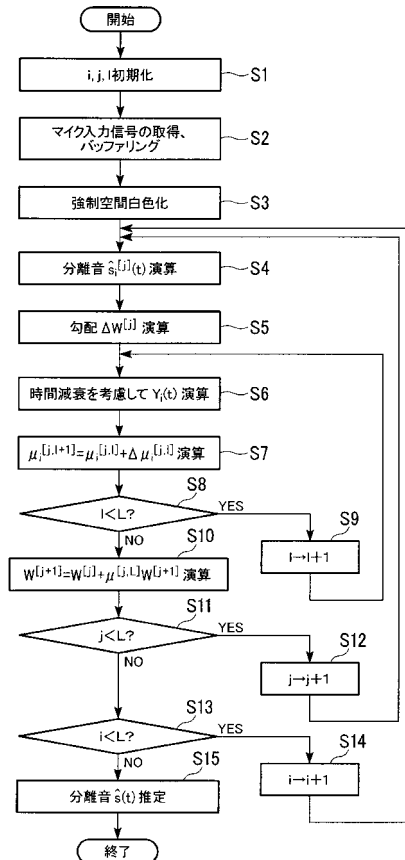
【図4】



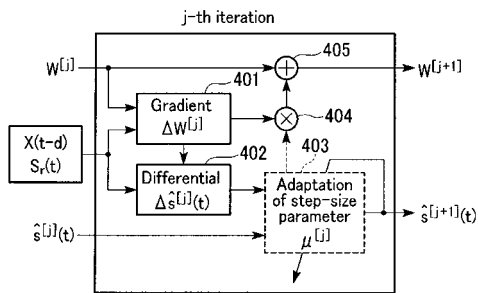
【図5】



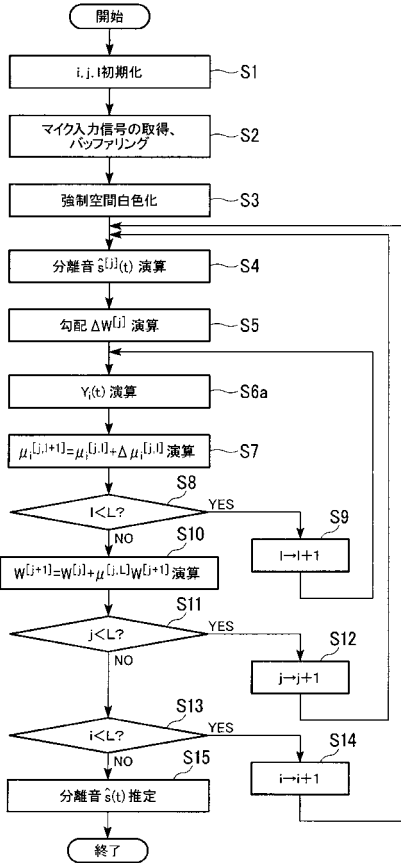
【図7】



【図6】



【図 8】



【図 9】

Impulse response	16-kHz sampling
Reverberation time (RT <sub>20</sub> )	240 and 670 ms
Distance and direction	1.5m and 0°, 45°, 90°, -45°, -90°
Number of microphones	Two (embedded in Robot's head)
SIFT analysis	Hanning: 64ms and shift: 24ms
Input wave data	[-1.0 1.0]normalized

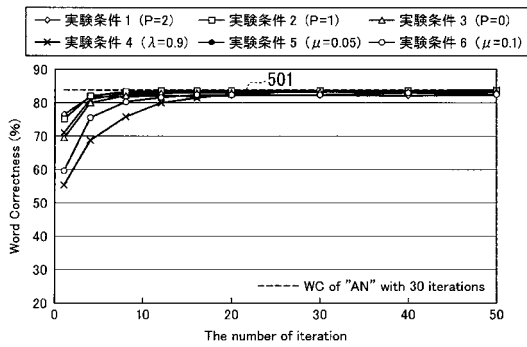
【図 10】

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	Hanning: 32ms and shift: 10ms
Features	MFCC 25 dim. (12+Δ12+ΔPow)

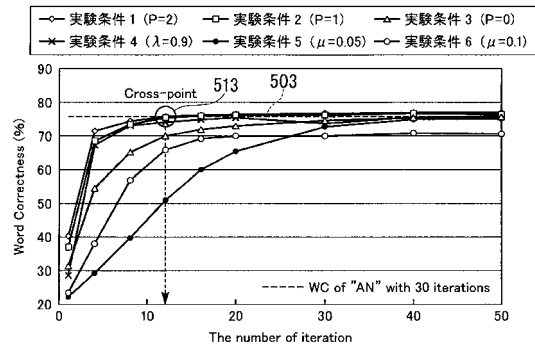
【図 11】

実験条件	説明
1	本実施形態の方法を用いた方法、区間数P=1
2	本実施形態の方法を用いた方法、区間数P=2
3	本実施形態の方法を用いた方法、区間数P=0
4	アニーリングベースで算出したステップ・サイズ、λ=0.9
5	ステップ・サイズμ=0.05に固定
6	ステップ・サイズμ=0.1に固定

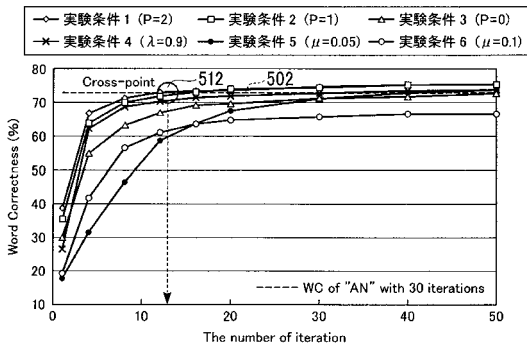
【図 12】



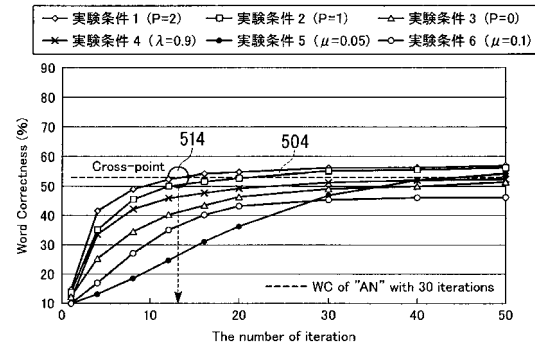
【図 14】



【図 13】



【図 15】



【 16 】

PROCESSING TIME RATIO (PTR)

Configuration	PTR	inverse PTR
Exp.A,N=9	1.983	0.504
Exp.A,N=23	1.564	0.609
Exp.B,N=9	1.727	0.579
Exp.B,N=23	1.406	0.711

---

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・インスティテュート・ジャパン内
- (72)発明者 中島 弘史  
埼玉県和光市本町 8 - 1 株式会社ホンダ・リサーチ・インスティテュート・ジャパン内
- (72)発明者 奥乃 博  
京都府京都市左京区吉田本町 国立大学法人京都大学 大学院情報学研究科内
- (72)発明者 武田 龍  
京都府京都市左京区吉田本町 国立大学法人京都大学 大学院情報学研究科内

審査官 五貫 昭一

- (56)参考文献 特開 2006 - 3617 (JP, A)  
特開 2006 - 330687 (JP, A)  
特開 2007 - 193035 (JP, A)  
特開 2008 - 306712 (JP, A)  
特開 2009 - 42754 (JP, A)  
特開 2010 - 193451 (JP, A)

- (58)調査した分野(Int.Cl., DB名)  
G10L 21/028  
G10L 21/0308