(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2010-130144
(P2010-130144A)
(43)公開日　平成22年6月10日(2010.6.10)

(51)Int.Cl.		F I			テーマコード（参考）
H04R 1/40	(2006.01)	H04R	1/40	320A	5D015
G10L 15/00	(2006.01)	G10L	15/00	200H	5D018
G10L 15/28	(2006.01)	G10L	15/28	400	5D020
H04R 3/00	(2006.01)	H04R	3/00	320	

審査請求 未請求　請求項の数 12　ＯＬ　（全 18 頁）

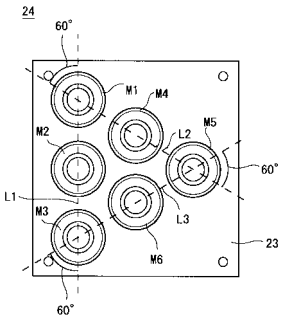
(21)出願番号　特願2008-300558 (P2008-300558)
(22)出願日　平成20年11月26日 (2008.11.26)

(71)出願人　000003207
トヨタ自動車株式会社
愛知県豊田市トヨタ町１番地
(74)代理人　100103894
弁理士　家入　健
(72)発明者　松井　禮朗
愛知県豊田市トヨタ町１番地　トヨタ自動車株式会社内

Ｆターム(参考)　5D015 DD02
5D018 BB22
5D020 BB04

(54)【発明の名称】ロボット、収音装置、及び音声処理方法

(57)【要約】
【課題】高いＳ／Ｎ比で収音することができるロボット、収音装置、及び音声処理方法を提供する。
【解決手段】本発明の一態様にかかるロボット１０は、複数のマイクが配列されたマイクユニット２４を有するロボット１０であって、胴体部１１と、胴体部１１に上に取り付けられた頭部１２と、上方を向いた複数のマイク素子Ｍ１～Ｍ６を有し、頭部１２に設けられたマイクユニット２４と、を備え、複数のマイクＭ１～Ｍ６のうち、最外周にあるマイクが、三角形を構成する三辺上に配列されているものである。
【選択図】図４



【特許請求の範囲】
【請求項１】
複数のマイクが配列されたマイクユニットを有するロボットであって、
胴体部と、
前記胴体部の上に取り付けられた頭部と、
上方を向いた複数のマイクを有し、前記頭部に設けられたマイクユニットと、を備え
前記複数のマイクのうち、最外周にあるマイクが、三角形を構成する三辺上に配列されているロボット。
【請求項２】
前記三角形が正三角形であり、各辺にマイクが配列されていることを特徴とする請求項１に記載のロボット。
【請求項３】
発話者が近接しているか否かを判定するために設けられたセンサをさらに備え、
前記センサからの出力に応じて、前記マイクからの入力のＯＮ／ＯＦＦを制御していることを特徴とする請求項１、又は２に記載のロボット。
【請求項４】
前記最外周にあるマイクが前記三辺に沿って等間隔で配置されていることを特徴とする請求項１乃至３のいずれか１項に記載のロボット。
【請求項５】
上方を向いて配列された複数のマイクを有するマイクユニットと、
前記マイクからの音声信号に対して処理を行う処理部と、を備え、
前記複数のマイクのうち、最外周にあるマイクが、三角形を構成する三辺上に配列されている収音装置。
【請求項６】
前記三角形が正三角形であり、各辺にマイクが配列されていることを特徴とする請求項５に記載の収音装置。
【請求項７】
発話者が近接しているか否かを判定するために設けられたセンサをさらに備え、
前記センサからの出力に応じて、前記マイクからの入力のＯＮ／ＯＦＦを制御していることを特徴とする請求項５又は６に記載の収音装置。
【請求項８】
前記最外周にあるマイクが前記三辺に沿って等間隔で配置されていることを特徴とする請求項５乃至７のいずれか１項に記載の収音装置。
【請求項９】
上方を向いた複数のマイクを備え、前記複数のマイクのうち、最外周にあるマイクが三角形を構成する三辺上に配列された収音装置で収音した音声に対する音声処理方法であって、
前記複数のマイクで音声を収音するステップと、
前記収音した音声に基づいて、前記三角形の三辺のうち、音源の方位に近い辺を推定するステップと、
前記音源の方位に近い辺上に配列されたマイクからの音声信号に対して、遅延処理を行った後、加算処理を行うステップと、を備える音声処理方法。
【請求項１０】
前記三角形が正三角形であり、各辺にマイクが配列されていることを特徴とする請求項９に記載の音声処理方法。
【請求項１１】
発話者が近接しているか否かを判定するために設けられたセンサからの出力に応じて、前記マイクからの入力のＯＮ／ＯＦＦを制御していることを特徴とする請求項９、又は１０に記載の音声処理方法。
【請求項１２】

前記最外周にあるマイクが前記三辺に沿って等間隔で配置されていることを特徴とする請求項９乃至１１のいずれか１項に記載の音声処理方法。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、ロボット、収音装置、及び音声処理方法に関し、特に詳しくは、複数のマイクを用いたロボット、収音装置、及び音声処理方法に関する。

【背景技術】

【０００２】

利用者が発した音声信号を検出するために複数のマイクロフォン（以下、マイク、又はマイク素子ともいう）を用いた技術が開示されている（特許文献１乃至８）。特許文献１、４、８では、複数のマイクによって、音源方向を特定している。特許文献２では、各空間軸（Ｘ，Ｙ，Ｚ）上に、少なくとも３つのマイクを配置している。また、特許文献３では、ロボットの頭部に指向性マイクを取り付け、胴体にアレイマイクを取り付けている。また、特許文献５のマイクロフォンシステムでは、話者の方向を判別するためのセンサを設けている。そして、複数のマイクのうち、音源に最も近いマイクのみをＯＮにして、その他のマイクをＯＦＦしている。特許文献６では、複数チャネルのマイクによって収音して、発話区間を判定している。

【０００３】

【特許文献１】特開２００６－１８１６５１号公報
【特許文献２】特開平１２－１３４６８８号公報
【特許文献３】特開２００７－２２１３００号公報
【特許文献４】特開２００２－１８６０８４号公報
【特許文献５】特開２００６－２４５７２５号公報
【特許文献６】国際公開２００４／０７１１３０号パンフレット
【特許文献７】特開２００４－２７４７６３号公報
【特許文献８】特開２００６－２４５７２５号公報

【発明の開示】

【発明が解決しようとする課題】

【０００４】

マイクロフォンアレイでは、複数のマイクによって収音した音を加算処理することができる。これにより、単一マイクと比較して、より鮮明な音の収録を可能としている。具体的には、マイクで収音した音声データに対して、ＤＳＰ遅延制御をした後、加算処理を行っている。これにより、Ｓ／Ｎ比を向上することができるため、より明瞭で聞きやすい収音を実現することができる。

【０００５】

マイクから水平方向の発話者に収音ビームを形成することで、収音能力を高めることができる。それと同時に、余分な周囲ノイズを排除して、エコーの原因ともなる回り込み音を取りにくい設計としている。

【０００６】

しかしながら、音源の方向が定まっていない場合、上記のようなマイクロフォンアレイでは、マイクの数を多数設ける必要がある。マイクロフォンアレイが大型化してしまうという問題点がある。特に、人型ロボットの場合、マイクロフォンアレイの搭載位置が制限されてしまう。例えば、マイクロフォンアレイをロボットの前側に搭載した場合、図１５に示すように、マイク指向領域がロボットの前方に制限されてしまう。また、マイク位置が発話者５２の口唇部の高さと水平な位置にならない。このため、収音ビームを強くしたい方向のノイズを強くしてしまい、収音したい音声が相対的に弱くなってしまう。すなわち、発話者の口唇高さが水平とならない事象については、性能が低下してしまう。このように、音源の方位とマイクロフォンアレイの位置の関係によって、Ｓ／Ｎ比（シグナルノイズ比）が低下してしまう。

【０００７】
本発明は、上記のような問題点に鑑みてなされたものであって、高いＳ／Ｎ比で収音することができるロボット、収音装置、及び音声処理方法を提供することを目的とする。
【課題を解決するための手段】
【０００８】
本発明の第１の態様に係るロボットは、複数のマイクが配列されたマイクユニットを有するロボットであって、胴体部と、前記胴体部の上に取り付けられた頭部と、上方を向いた複数のマイクを有し、前記頭部に設けられたマイクユニットと、を備え、前記複数のマイクのうち、最外周にあるマイクが、三角形を構成する三辺上に配列されているものである。
【０００９】
本発明の第２の態様に係るロボットは、上記のロボットであって、前記三角形が正三角形であり、各辺にマイクが配列されていることを特徴とするものである。
【００１０】
本発明の第３の態様に係るロボットは、上記のロボットであって、発話者が近接しているか否かを判定するために設けられたセンサをさらに備え、前記センサからの出力に応じて、前記マイクからの入力のＯＮ／ＯＦＦを制御していることを特徴とするものである。
【００１１】
本発明の第４の態様に係るロボットは、上記のロボットであって、前記最外周にあるマイクが前記三辺に沿って等間隔で配置されていることを特徴とするものである。
【００１２】
本発明の第５の態様に係る収音装置は、上方を向いて配列された複数のマイクを有するマイクユニットと、前記マイクからの音声信号に対して処理を行う処理部と、を備え、前記複数のマイクのうち、最外周にあるマイクが、三角形を構成する三辺上に配列されているものである。
【００１３】
本発明の第６の態様に係る収音装置は、上記の収音装置であって、前記三角形が正三角形であり、各辺にマイクが配列されていることを特徴とするものである。
【００１４】
本発明の第７の態様に係る収音装置は、上記の収音装置であって、発話者が近接しているか否かを判定するために設けられたセンサをさらに備え、前記センサからの出力に応じて、前記マイクからの入力のＯＮ／ＯＦＦを制御していることを特徴とするものである。
【００１５】
本発明の第８の態様に係る収音装置は、上記の収音装置であって、前記最外周にあるマイクが前記三辺に沿って等間隔で配置されていることを特徴とするものである。
【００１６】
本発明の第９の態様に係る音声処理方法は、上方を向いた複数のマイクを備え、前記複数のマイクのうち、最外周にあるマイクが三角形を構成する三辺上に配列された収音装置で収音した音声に対する音声処理方法であって、前記複数のマイクで音声を収音するステップと、前記収音した音声に基づいて、前記三角形の三辺のうち、音源の方位に近い辺を推定するステップと、前記音源の方位に近い辺上に配列されたマイクからの音声信号に対して、遅延処理を行った後、加算処理を行うステップと、を備えるものである。
【００１７】
本発明の第１０の態様に係る音声処理方法は、上記の音声処理方法であって、前記三角形が正三角形であり、各辺にマイクが配列されていることを特徴とするものである。
【００１８】
本発明の第１１の態様に係る音声処理方法は、上記の音声処理方法であって、発話者が近接しているか否かを判定するために設けられたセンサからの出力に応じて、前記マイクからの入力のＯＮ／ＯＦＦを制御していることを特徴とするものである。
【００１９】

本発明の第１２の態様に係る音声処理方法は、上記の音声処理方法であって、前記最外周にあるマイクが前記三辺に沿って等間隔で配置されていることを特徴とするものである。

【発明の効果】

【００２０】

本発明によれば、高いＳ／Ｎ比で収音することができるロボット、収音装置、及び音声処理方法を提供することができる。

【発明を実施するための最良の形態】

【００２１】

以下、本発明に係るロボットの実施形態を、図面に基づいて詳細に説明する。但し、本発明が以下の実施形態に限定される訳ではない。また、説明を明確にするため、以下の記載及び図面は、適宜、簡略化されている。

【００２２】

本発明の実施形態に係るロボット１０の全体構成について、図１を用いて説明する。図１は、ロボットの全体構成を模式的に示す正面図であり、発話者が発話した音声をロボット１０が収音する様子を示している。図１に示すロボット１０は、人型ロボット（ヒューマノイドロボット）であり、胴体部１１と胴体部１１の上に設けられた頭部１２とを備えている。なお、頭部１２が胴体部１１に対して回転するようにしてもよい。胴体部１１には、腕や脚の関節を駆動するためのモータやバッテリなどが内蔵されている。なお、ロボット１は、完全な人型ロボットに限られるものではない。例えば、脚部の代わりに車輪で移動するロボットであってもよい。このように一部のみが人型のロボットであってもよい。また、胴体部１１には、近接センサ４１が設けられている。近接センサ４１は、発話者５２が近接したか否かを判定する。すなわち、近接センサ４１は、発話者５２がロボット１０から一定の距離以内に接近したことを知らせるセンサ信号を出力する。

【００２３】

頭部１２には、収音装置１３が内蔵されている。収音装置１３は、図２に示すように、頭部１２の上方に設けられている。すなわち、頭頂部に収音装置１３が取り付けられている。収音装置１３には、後述するように複数のマイク素子が設置されている。それぞれのマイク素子が上方を向いている。これにより、図１に示すような、マイク指向領域５１を得ることができる。マイク指向領域５１は、頭部１２の上に配置される。これにより、発話者５２の口唇高さをマイク指向領域５１に合わせることができ、発話者５２が発生した音声を効率よく収音することができる。

【００２４】

次に、収音装置１３の構成に付いて図３を用いて説明する。図３は、収音装置１３の全体構成を模式的に示す図である。図３に示すように、収音装置１３は、マイク素子２０が複数設けられている。そして、複数のマイク素子２０がマイク基板２３の固定されている。マイク基板２３は、水平に配置されている。マイク基板２３と複数のマイク素子２０がマイクユニット２４を構成する。

【００２５】

そして、複数のマイク素子２０を覆うように、カバー２２が設けられている。カバー２２は、マイクユニット２４の上方に設けられる。カバー２２は、ロボット１０の頭頂部に取り付けられる。したがって、カバー２２が頭頂部分を構成することにある。マイク素子２０及びマイク基板２３にカバー２２を被せることで、マイク素子２０及びマイク基板２３が露出しなくなる。これにより、マイクユニット２４が頭部１２に収納される。また、カバー２２とマイク素子２０との間には、ＥＭＣシールド材２１が設けられている。ＥＭＣシールド材２１は、マイク素子２０やマイク基板２３などに生じるＥＭＣノイズを低減するために電気的に接地されている。よって、Ｓ／Ｎ比を向上することができる。なお、マイク素子２０は上方を向いているため、カバー２２の方向に向いていることになる。

【００２６】

各マイク素子２０は、マイク基板２３を介して、アンプ基板２５に接続されている。す

なわち、マイク素子２０で収音した音声信号は、アンプ基板２５に設けられたアンプによって、増幅される。ここで、アンプ基板２５には、マイク素子２０と同数のアンプが設けられている。そして、アンプで増幅された音声信号は、Ａ／Ｄ変換基板２８によって、アナログ信号からデジタル信号に変換される。そして、デジタル信号に変換された音声信号は、ＰＣ（Ｐｅｒｓｏｎａｌ　Ｃｏｍｐｕｔｅｒ）４０に入力される。ＰＣ４０は、入力された音声信号に対して音声認識処理を行う。なお、これらの処理については、後述する。例えば、アンプ基板２５やマイクユニット２４等は、頭部１２に収納されている。ＰＣ４０は、胴体部１１に収納されていてもよい。アンプ基板２５、Ａ／Ｄ変換基板２８、ＰＣ４０等が、それぞれ、マイクユニットで収音された音声信号に対してよりを行う処理部となる。なお、図３では、後述するビーム幅演算基板などが省略されている。

【００２７】

次に、マイクユニット２４におけるマイク素子２０の配置について、図４を用いて説明する。図４は、マイクユニット２４の構成を示す上面図である。図４では、６個のマイク素子２０をマイク素子Ｍ１～Ｍ６と識別している。すなわち、マイク素子Ｍ１、マイク素子Ｍ２、マイク素子Ｍ３、マイク素子Ｍ４、マイク素子Ｍ５、及びマイク素子Ｍ６がマイク基板２３の上に設置されている。マイク素子Ｍ１～Ｍ６は、上を向いてマイク基板２３に固定されている。すなわち、マイク素子Ｍ１～Ｍ６は、それぞれ頭部の頭頂に向けられている。６つのマイク素子Ｍ１～Ｍ６が同一平面上に配置されている。すなわち、６つのマイク素子Ｍ１～Ｍ６は、水平に配列されている。

【００２８】

ここでは、６個のマイク素子Ｍ１～Ｍ６は、その中心が正三角形の三辺上に配置されるように配列されている。図４では、正三角形の各辺を仮想直線Ｌ１、仮想直線Ｌ２、及び仮想直線Ｌ３と示している。すなわち、仮想直線Ｌ１～Ｌ３が三角形を構成する。仮想直線Ｌ１、仮想直線Ｌ２、及び仮想直線Ｌ３のそれぞれは、水平面内にある。マイク素子Ｍ１～Ｍ３の中心点は、仮想直線Ｌ１上に配置されている。また、マイク素子Ｍ１～Ｍ３は、等間隔に配列されている。そして、マイク素子Ｍ１とマイク素子Ｍ３の間に、マイク素子Ｍ２が配置されている。したがって、マイク素子Ｍ１の中心点とマイク素子Ｍ３の中心点は、それぞれ正三角形の頂点に配置される。このように、マイク素子Ｍ１～Ｍ３は仮想直線Ｌ１上に一列に配列されている。仮想直線Ｌ１上のマイク素子Ｍ１～Ｍ３がＣＨ（チャネル）１構成する。なお、仮想直線Ｌ１の方向が、ロボット１０の前後方向になる。

【００２９】

仮想直線Ｌ２は、仮想直線Ｌ１から６０°傾いている。そして、仮想直線Ｌ１と仮想直線Ｌ２とは、マイク素子Ｍ１の中心点で交差する。仮想直線Ｌ２上には、マイク素子Ｍ１、Ｍ４、Ｍ５が配列されている。すなわち、マイク素子Ｍ１、Ｍ４、Ｍ５の中心点は、仮想直線Ｌ２上に配置される。マイク素子Ｍ１、Ｍ４、Ｍ５は、等間隔で配置されている。そして、マイク素子Ｍ１とマイク素子Ｍ５の間に、マイク素子Ｍ４が配置されている。したがって、マイク素子Ｍ５の中心点は、正三角形の頂点に配置されている。このように、マイク素子Ｍ１、Ｍ４、Ｍ５は仮想直線Ｌ２上に一列に配列されている。仮想直線Ｌ２上のマイク素子Ｍ１、Ｍ４、Ｍ５がＣＨ（チャネル）２を構成する。

【００３０】

仮想直線Ｌ３は、仮想直線Ｌ１、及び仮想直線Ｌ２から６０°傾いている。仮想直線Ｌ１と仮想直線Ｌ３とは、マイク素子Ｍ３の中心点で交差する。仮想直線Ｌ２と仮想直線Ｌ３とは、マイク素子Ｍ５の中心点で交差する。仮想直線Ｌ３上には、マイク素子Ｍ３、Ｍ６、Ｍ５が配列されている。すなわち、マイク素子Ｍ３、Ｍ６、Ｍ５の中心点は、仮想直線Ｌ３上に配置される。マイク素子Ｍ３、Ｍ６、Ｍ５は、等間隔で配置されている。そして、マイク素子Ｍ３とマイク素子Ｍ５の間に、マイク素子Ｍ６が配置されている。したがって、マイク素子Ｍ１、Ｍ５の中心点は、正三角形の頂点に配置されている。このように、マイク素子Ｍ３、Ｍ６、Ｍ５は仮想直線Ｌ３上に一列に配列されている。仮想直線Ｌ３上のマイク素子Ｍ３、Ｍ６、Ｍ５がＣＨ（チャネル）３を構成する。

【００３１】

各チャネルとも、３つのマイク素子から構成されている。仮想直線Ｌ１、仮想直線Ｌ２、及び仮想直線Ｌ３は、複数のマイク素子Ｍ１～Ｍ６の最外周上にある。仮想直線Ｌ１、仮想直線Ｌ２、及び仮想直線Ｌ３は、マイクユニットの外形となる正三角形を構成する。従って、最外周にあるマイク素子Ｍ１～Ｍ６を結ぶと正三角形になる。このように、１列目（仮想直線Ｌ１に平行でＭ５の中心を通る列）に１個、２列目（仮想直線Ｌ１に平行でＭ４、Ｍ６の中心を通る列）に２個、３列目（仮想直線Ｌ１）に３個のマイク素子が配置されている。

【００３２】

仮想直線Ｌ１は、ロボット１０の前後方向に延びている。ここでは、マイク素子Ｍ１がロボットの前側に配置され、マイク素子Ｍ３がロボットの後ろ側に配置されている。そして、音声の到来方向に応じて、図５乃至図７に示すように、使用するチャネルを選択する。すなわち、音源の方位に応じて、最適なチャネルが使用される。なお、図５乃至７は、マイクユニット２４の配置を示す上面図であり、図中の上側がロボット１０の前側、下側が後ろ側になっている。

【００３３】

例えば、前方、又は後方からの音声は、チャネルＣＨ１で検知される。図５の矢印に示すように、後方から音声が到達した場合、ＣＨ１のマイク素子Ｍ１～Ｍ３が使用される。すなわち、マイク素子Ｍ１～Ｍ３を用いて収音した音声データに対して、ＤＳＰ遅延制御を行った後、加算処理を行う。音源の方位が仮想直線Ｌ１と平行な場合、ＣＨ１が使用される。

【００３４】

一方、斜め方向からの音声は、ＣＨ２又はＣＨ３を用いて検知される。例えば、図６の矢印に示すように左後方からの音声は、ＣＨ２で検知される。この場合、ＣＨ２のマイク素子Ｍ１、Ｍ４、Ｍ５が使用される。すなわち、マイク素子Ｍ１、Ｍ４、Ｍ５を用いて収音した音声データに対して、ＤＳＰ遅延制御を行った後、加算処理を行う。音源の方位が仮想直線Ｌ２と平行な場合、ＣＨ２が使用される。また、図７の矢印に示すように右後方からの音声は、ＣＨ３で検知される。この場合、ＣＨ３のマイク素子Ｍ３、Ｍ６、Ｍ５が使用される。マイク素子Ｍ３、Ｍ６、Ｍ５を用いて収音した音声データに対して、ＤＳＰ遅延制御を行った後、加算処理を行う。音源の方位が仮想直線Ｌ３と平行な場合、ＣＨ３が使用される。

【００３５】

各チャネルのポーラパターンを図８乃至１０に示す。図８はＣＨ１のポーラパターン、図９はＣＨ２のポーラパターン、図１０はＣＨ３のポーラパターンである。このように、各チャネルは、異なるポーラパターンを有している。すなわち、チャネル毎に指向領域が異なっている。いずれかのチャンネルを用いて音声を収音する。もちろん、音源の方位がチャネル方向と平行でない場合は、最も近い方位のチャネルを使用する。こうすることで、任意の方向からの音声に対しても、高いＳ／Ｎ比で収音することができる。

【００３６】

図１１は、各チャネルと音源の方位の関係を示すグラフである。音源方位が０°の場合（ＣＨ１）が左側に示され、音源方位が左６０°の場合（ＣＨ２）が中央に示され、音源方位が右６０°の場合（ＣＨ３）が右側に示されている。それぞれの場合において、３つの測定結果が示されている。すなわち、それぞれの場合に対して、マイク指向方向を変えたときの測定結果が示されている。図１１では、縦軸がＳ／Ｎ比を示している。図１１に示すように、音源方位が０°の場合、マイク指向方向を正面とすると、Ｓ／Ｎ比が高くなる。音源方位が左６０°の場合、マイク指向方向を左６０°とすると、Ｓ／Ｎ比が高くなる。音源方位が右６０°の場合、マイク指向方向を右６０°とすると、Ｓ／Ｎ比が高くなる。したがって、音源の方位に応じて、チャネルを使い分ければ、いずれの方位においても、高いＳ／Ｎ比での収音が可能になる。すなわち、音源の方位に最も近い方向の仮想直線を推定する。そして、音源の方位に近い仮想直線上のマイク素子２０を使用する。ここでは、選択された仮想直線上のマイク素子２０からの音声信号に対して音声処理が行われ

る。すなわち、音源の方位に基づいて、使用するチャネルを推定している。

【００３７】

次に、本実施の形態にかかる音声処理方法について、図１２、及び図１３を用いて説明する。図１２は、音声データに対してアナログ処理を行うための構成を示す回路ブロック図である。図１３は、音声認識処理を行うためのＰＣ４０の構成を示す図である。

【００３８】

図１２に示すようにアンプ基板２５には、アンプ３１と、バッファ３２が設けられている。アンプ３１は、各マイク素子２０からの音声信号を増幅する。バッファ３２は、アンプ３１で増幅された音声信号をバッファリングする。

【００３９】

バッファリングされた音声信号は、ビーム幅演算基板２６に入力される。ビーム幅演算基板２６は、各チャネルのビーム幅を演算して、使用するチャネルを選択する。そして、そのチャンネルに含まれる３つのマイク素子２０で検出された音声信号に対して、遅延処理、及び加算処理を行う。そのため、ビーム幅演算基板２６には、指向性演算部３４、及びフィルタ３５が設けられている。なお、ここでは指向性演算部３４及びフィルタ３５が一つの回路として示されている。

【００４０】

フィルタ３５は、電源ノイズ対策として、３００Ｈｚ以下の入力をカットする。フィルタ３５としては、ローパスフィルタを用いることができる。もちろん、ローパスフィルタ以外のフィルタをフィルタ３５として用いてもよい。例えば、ハイパスフィルタやバンドパスフィルタをフィルタ３５として用いて、所定の周波数帯の入力をカットしてもよい。このフィルタ３５は、加算処理前の音声信号に対してフィルタリング処理を行ってもよく、加算処理後の音声信号に対してフィルタリング処理を行ってもよい。

【００４１】

指向性演算部３４は、バッファ３２に蓄積された音声信号を比較して、音源の方位を推定する。そして、音源の方位に近いチャネルを使用チャネルとして選択する。具体的には、指向性演算部３４は、ＣＨ１～ＣＨ３の中で位相遅れが大きいチャネルを抽出する。すなわち、各チャネルに含まれている３つのマイク素子２０で検出された音声信号の時間遅れが大きくなっている方向を推定する。例えば、バッファ３２に記憶されている音声信号に対して、その遅延時間を算出する。ここで遅延時間は、３つのチャネル毎に算出される。遅延時間が大きくなるチャネルに設けられているマイク素子２０の配列方向を音源方向として推定する。換言すると、仮想直線Ｌ１～Ｌ３のうち、音源の方位に最も近い仮想直線を選択する。このように、音源の方位が正面（ＣＨ１）、左６０°（ＣＨ２）、右６０°（ＣＨ３）のいずれの方位に近いかを推定する。最も方位が近いチャネルを選択する。

【００４２】

そして、指向性演算部３４は、選択されたチャネルの音声信号に対してＤＳＰ遅延制御を行う。これにより、使用チャネルに含まれる３つのマイク素子２０からの音声信号が時間的に一致して重なる。遅延時間は、隣接するマイク素子２０の間隔に応じた値となっている。さらに、指向性演算部３４は、ＤＳＰ遅延制御によって遅延された音声信号に対して加算処理を行う。これにより、選択されたチャネルにおいて、人の音声を強調することができる。各マイク素子２０の音声信号では、ノイズがランダムに発生していると考えられるため、時間的にずれている３つの音声信号に対して、遅延した後、加算してとしても、ノイズが強調されない。よって、３つの音声信号を加算することにより、発話に基づく音声信号が強調される。すなわち、発話者５２の発話部分が重なるように、音声信号を遅延した後、加算する。ノイズ源で発生するノイズに対して、発話者の音声による音声信号を強調することができる。発話者の方向推定と超指向性マイクを組み合わせたことにより、目的音声と周辺雑音の切り分け性能を向上することができる。

【００４３】

ビーム幅演算基板２６には、ゲイン切替部３７、及び切替ＳＷ３８が設けられている。ここで、ゲイン切替部３７、及び切替ＳＷ３８は一つの回路として構成されている。ゲイ

ン切替部３７は、加算された音声信号のレベルに応じてゲインを切替える。すなわち、適当なゲインを選択して、加算後の音声信号を所定のレベルまで引き上げる。切替ＳＷ３８は、指向性演算部３４で選択されたチャネル信号が出力されるように、出力信号を切替える。これにより、選択されたチャネルの音声信号がビーム幅演算基板２６から出力される。すなわち、切替ＳＷ３８によって、ビーム幅演算基板２６は１つのチャネルにおける加算された音声信号を出力する。

【００４４】

図１３に示すように、ビーム幅演算基板２６からの音声信号は、Ａ／Ｄ変換基板２８に入力される。Ａ／Ｄ変換基板２８は、アナログの音声信号をデジタル信号に変換して、ＰＣ４０に出力する。これにより、加算処理後の音声信号がデジタルデータとなる。ＰＣ４０は、データ処理装置であり、デジタルの音声信号に対して音声認識処理を行う。

【００４５】

また、ＰＣ４０には、近接センサ４１からのセンサ信号が入力されている。近接センサ４１は、上記の通り、発話者５２が接近したか否かを判定している。すなわち、近接センサ４１は、ロボット１０から一定の範囲内に発話者５２が入っている場合に、センサ信号をＰＣ４０に出力する。

【００４６】

ＰＣ４０は、発話者５２が近接していることを示すセンサ信号が入力された場合のみ、音声認識処理を行う。すなわち、発話者５２がロボット１０から一定の距離以上離れている場合は、音声認識処理を行わない。また、発話者５２がロボット１０から一定の距離以上離れている場合は、マイクユニット２４、アンプ基板２５、ビーム幅演算基板２６、Ａ／Ｄ変換基板２８などをＯＦＦしてもよい。このように、近接センサ４１からの信号に応じて、マイクユニット２４などのＯＮ／ＯＦＦを制御する。これにより、必要時のみ音声に対する処理が行われるようになる。なお、近接センサ４１の代わりに、距離センサやカメラなどを用いて、接近しているか否かを判定してもよい。このように、発話者５２が遠くにいる場合は、マイク素子２０等をＯＦＦするように、収音装置１３をＯＮ／ＯＦＦ制御する。近接センサ４１からの出力に応じて、マイク素子２０からの入力のＯＮ／ＯＦＦを制御する。これにより、必要時のみ、音声に対する各処理が行われるようになる。

【００４７】

そして、ＰＣ４０は、Ａ／Ｄ制御ドライバを用いてＡ／Ｄ変換基板２８を制御する。これにより、選択されたチャネルからの音声信号が入力される。そして、雑音抑圧モジュールによって、雑音（ノイズ）の除去が行われる。ここまでの制御は、ディバイス制御によって行われる。すなわち、ハードウェア的な制御によって処理を行う。

【００４８】

雑音抑制された音声信号に対して、認識エンジンが音声認識処理を行う。認識エンジンはソフトウェア的な制御によって各処理を行う。セグメンタは、例えば、連続音声の音声信号に対してフレーム化処理を行う。また、セグメンタがノイズ処理後の発話区間前後の不要データを棄却している。これにより、発話者５２が発話した区間（フレーム）のみが抽出される。すなわち、音声信号の入力レベルが低い箇所では、音声認識処理が行われないようになる。よって、誤認識（誤挿入）の低減を図ることができる。

【００４９】

そして、音響モデルを用いて、音声信号のパターンがどの単語に対応しているかを判定する。これにより、連続音声に対する単語列が得られる。例えば、特徴ベクトルなどを用いて、音声信号に対応するパターンの単語を判定する。そして、ノイズモデルを用いてノイズを除去する。この後、発話辞書を参照して、意味を成す文章になっているかを判定する。例えば、言語モデルを用いて、文法的に正しいか否かを判定する。文法的に正しい文章になっている場合、この認識結果がロボット制御アプリに出力される。これにより、最も文法的に正しい文章が出力されることになる。なお、上記の音声認識処理以外の処理によって、音声を認識してもよい。すなわち、音声認識処理は特に限定されるものではなく、公知の音声認識処理方法を用いることができる。

【００５０】
このように、複数のマイク素子２０が上方に向けて配列されている。そして、配列された複数のマイク素子２０のうち、最外周にあるマイク素子２０が三角形の三辺に沿って配列されている。音源の方位に近い１辺上に配列されているマイク素子を用いて、音声を認識する。すなわち、１辺上のマイク素子２０で検出された音声信号に対して、遅延処理、及び加算処理を行う。こうすることで、Ｓ／Ｎ比を向上することができる。これにより、音声認識率を向上することができる。また、上面視において、マイク素子２０を三角形状に配置することで、縦横に配列する場合と比べて、マイク素子２０の数を少なくすることができる。さらに、音源の方位がいずれの方向であったとしても、少ない数のマイク素子でＳ／Ｎ比を向上することができる。

【００５１】
ここで、マイクユニット２４のタイプと取り付け位置によるＳ／Ｎ比に違いを図１４に示す。マイクＢが図４に示した三角形配置のマイクユニットである。マイクＡが従来配置のマイクユニットである。図１４に示すように、三角形配置のマイクユニット２４を頭部１２に取り付けることによって、Ｓ／Ｎ比を高くすることができる。また、マイクＡよりもマイクＢの方が、Ｓ／Ｎ比が高くなっている。マイクユニット２４を胸部（胴体部）よりも頭部に設けたほうが、Ｓ／Ｎ比が高くなっている。

【００５２】
このように、三角形の各辺を１つのチャネルとして、収音ビームを形成している。さらに、上方を向いたマイク素子２０が水平方向に沿って配列されている。したがって、いずれの方位からの音声であっても高いＳ／Ｎ比で収音することができる。また、マイク素子２０の高さを発話者の口唇部と同じ高さ、またはそれより低くすることで、マイク指向領域５１で音声を収音することができるようになる。これにより、確実に収音することができる。よって、音声を明瞭に収音することができる。さらに、最外周のマイク素子２０を三角形の３辺上に配列しているため、マイク素子２０の数を少なくすることができる。すなわち、マイク素子２０の数を少なくした場合でも高いＳ／Ｎ比を実現することができる。

【００５３】
例えば、胴体部１１には、関節を駆動するためのモータやＰＣ４０の冷却ファン等が収納されている。これらの各機器が動作すると、ノイズが発生する。すなわち、機械的に動作する機器がノイズ源となってしまう。これらのノイズ源で発生するノイズの入力レベルが目的とする音声の入力信号レベルよりも高い場合、ノイズに音声信号が埋もれてしまう。そこで、本実施の形態では、ノイズ源が存在しない、あるいはその数が少ない頭部１２に収音装置１３を収納している。頭部１２は胴体部１１に比べて、ノイズ源となる搭載機器が少ない。このため、ノイズ源からマイクユニット２４を遠ざけることができる。これにより、マイク素子２０へのノイズ混入を低減することができる。Ｓ／Ｎ比を高くすることができる。よって、明瞭に音声を収音することができ、音声認識率の低下を低減することができる。さらに、頭部１２の頭頂部にマイクユニット２４を設置することで、スペースを効率よく利用することができる。

【００５４】
なお、本実施の形態では、マイクユニット２４に配列するマイク素子２０を６個としたが、マイク素子２０の数はこれ以外の数であってもよい。たとえば、マイク素子２０の数を１０個にすると、１辺上に４つのマイク素子２０とすることができる。たとえば、１列目に１個、２列目に２個、３列目に３個、４列目に４個のマイク素子２０を配置する。この場合、図４に示した構成に対して、仮想直線Ｌ１の隣に、仮想直線Ｌ１と平行な方向に４つのマイク素子２０を近接配置する。すなわち、図４の構成の左隣に４つのマイク素子２０を追加した構成となる。また。この配置では、３列目の２個目のマイク素子（現在のマイク素子Ｍ２）が最外周のマイク素子とはならなくなるため、残りの９個のマイク素子２０が、三角形の三辺に沿って配列される。マイク素子１０の数を多くすることで、加算される音声信号の数が多くなるため、よりＳ／Ｎ比を高くすることができる。アレイ状に

配列する構成と比べて、少ない数のマイク素子２０で、Ｓ／Ｎ比を高くすることができる。よって、簡便な構成でＳ／Ｎ比を高くすることができる。また、三辺上のマイク素子２０を等間隔に配置することで、よりＳ／Ｎ比を向上することができる。

【００５５】

上記の説明では、最外周のマイク素子２０を配置する三角形を、正三角形としたが、これ以外の三角形でもよい。たとえば、二等辺三角形や直角三角形としてもよい。もちろん、これ以外の三角形でもよい。また、三角形の頂点の角度は、鋭角とすることが好ましい。これにより、２辺の成す角度が小さくなるのを防ぐことができるため、音源がいずれの方位にあった場合でも、高いＳ／Ｎ比で音声信号を収音することができる。

【図面の簡単な説明】

【００５６】

【図１】本発明の実施形態にかかるロボットを模式的に示す正面図である。

【図２】本発明の実施形態にかかるロボットの頭部を示す図である。

【図３】ロボットに設けられた収音装置の構成を示す図である。

【図４】収音装置に設けられたマイクユニットの構成を示す上面図である。

【図５】収音装置に設けられたマイクユニットにおいて、チャンネル１の構成を示す図である。

【図６】収音装置に設けられたマイクユニットにおいて、チャンネル２の構成を示す図である。

【図７】収音装置に設けられたマイクユニットにおいて、チャンネル３の構成を示す図である。

【図８】収音装置に設けられたマイクユニットにおいて、チャンネル１のポーラパターンを示す図である。

【図９】収音装置に設けられたマイクユニットにおいて、チャンネル２のポーラパターンを示す図である。

【図１０】収音装置に設けられたマイクユニットにおいて、チャンネル３のポーラパターンを示す図である。

【図１１】音源方向別のマイク指向特性を示すグラフである。

【図１２】音声信号に対してアナログ処理を行う構成を示す回路ブロック図である。

【図１３】音声信号に対して音声認識処理を行う構成を示すブロック図である。

【図１４】マイク種別及び搭載位置による性能の比較結果を示すグラフである。

【図１５】ロボットに搭載されたマイクロフォンアレイのマイク指向領域を示す図である。

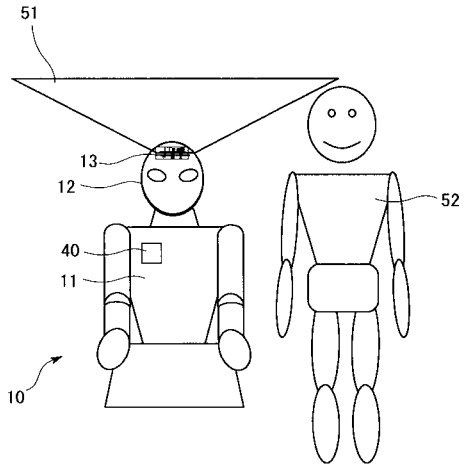
【符号の説明】

【００５７】

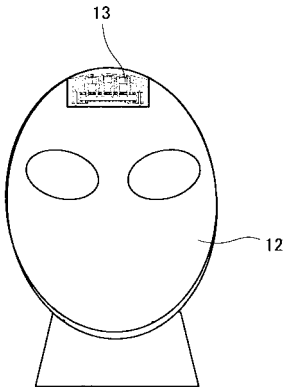
１０　ロボット
１１　胴体部
１２　頭部
１３　収音装置
１５　マイクユニット
１６　処理部
２１　シールド材
２２　カバー
２３　マイク基板
２４　マイクユニット
２５　アンプ基板
２６　ビーム幅演算基板
２８　Ａ／Ｄ変換基板
３１　アンプ
３２　バッファ

３４　指向性演算部
３５　フィルタ
３７　ゲイン切替部
３８　切替ＳＷ
４０　ＰＣ
４１　近接センサ
５１　マイク指向領域
５２　発話者
Ｍ１　マイク素子
Ｍ２　マイク素子
Ｍ３　マイク素子
Ｍ４　マイク素子
Ｍ５　マイク素子
Ｍ６　マイク素子

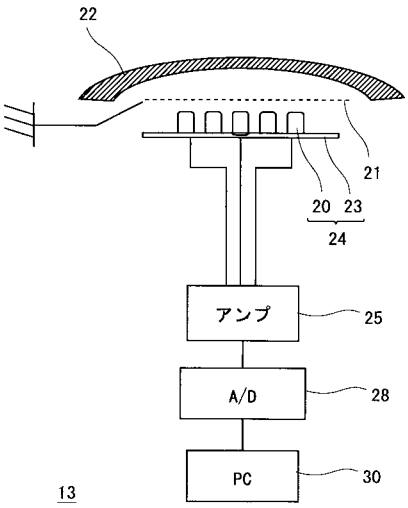
【図１】



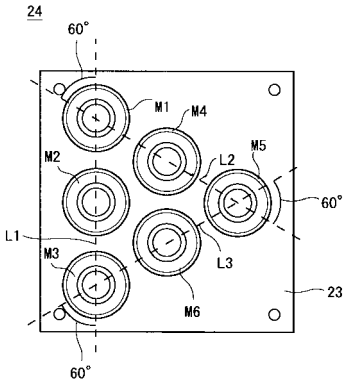
【図２】



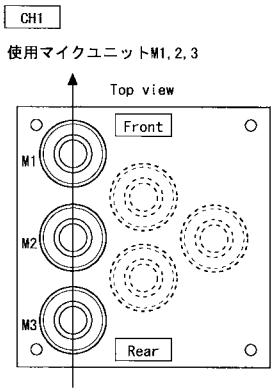
【図３】


22
21
20 23
24
アンプ
25
A/D
28
PC
30
13

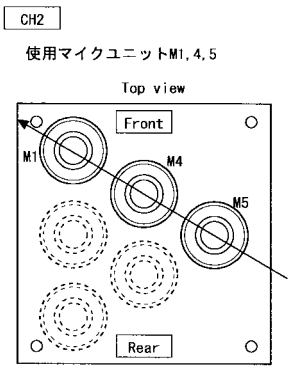
【図４】


24
60°
M1
M4
M2
L2
M5
60°
L1
M3
L3
23
M6
60°

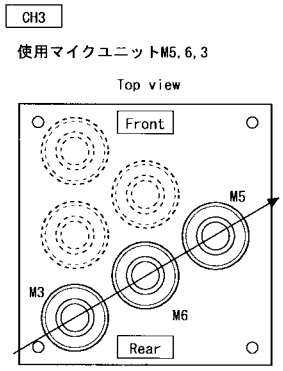
【図５】


CH1
使用マイクユニットM1,2,3
Top view
Front
M1
M2
M3
Rear

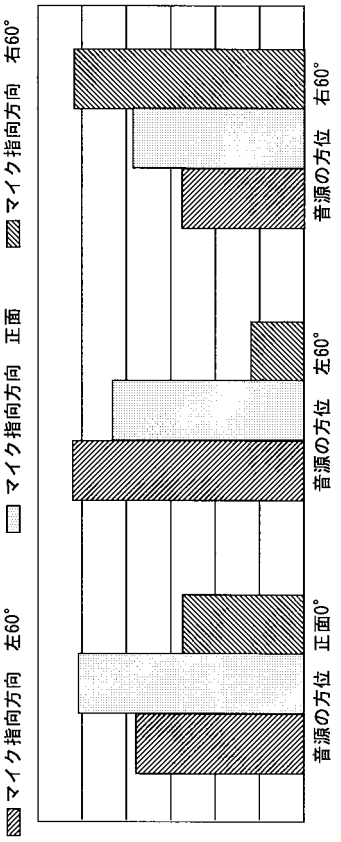
【図６】


CH2
使用マイクユニットM1,4,5
Top view
Front
M1
M4
M5
Rear

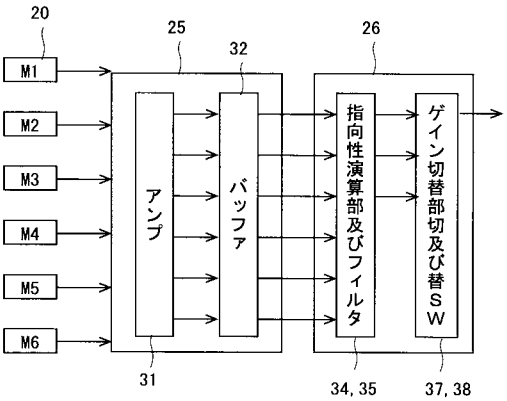
【図７】



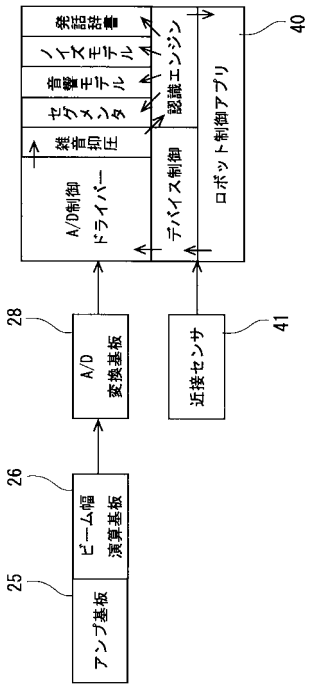
【図１１】



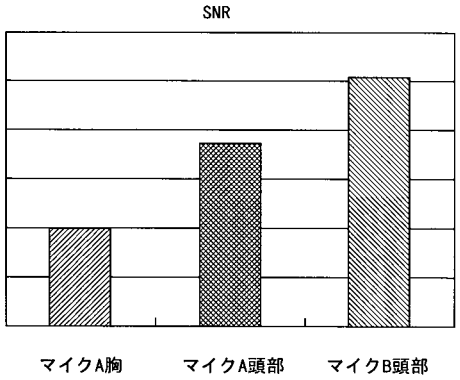
【図１２】



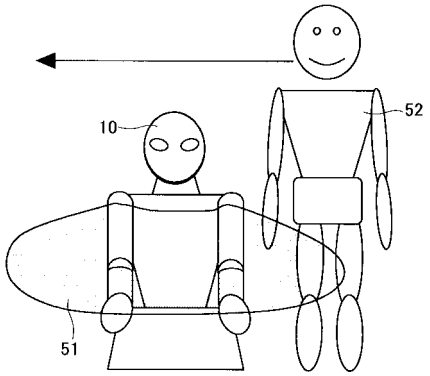
【図１３】



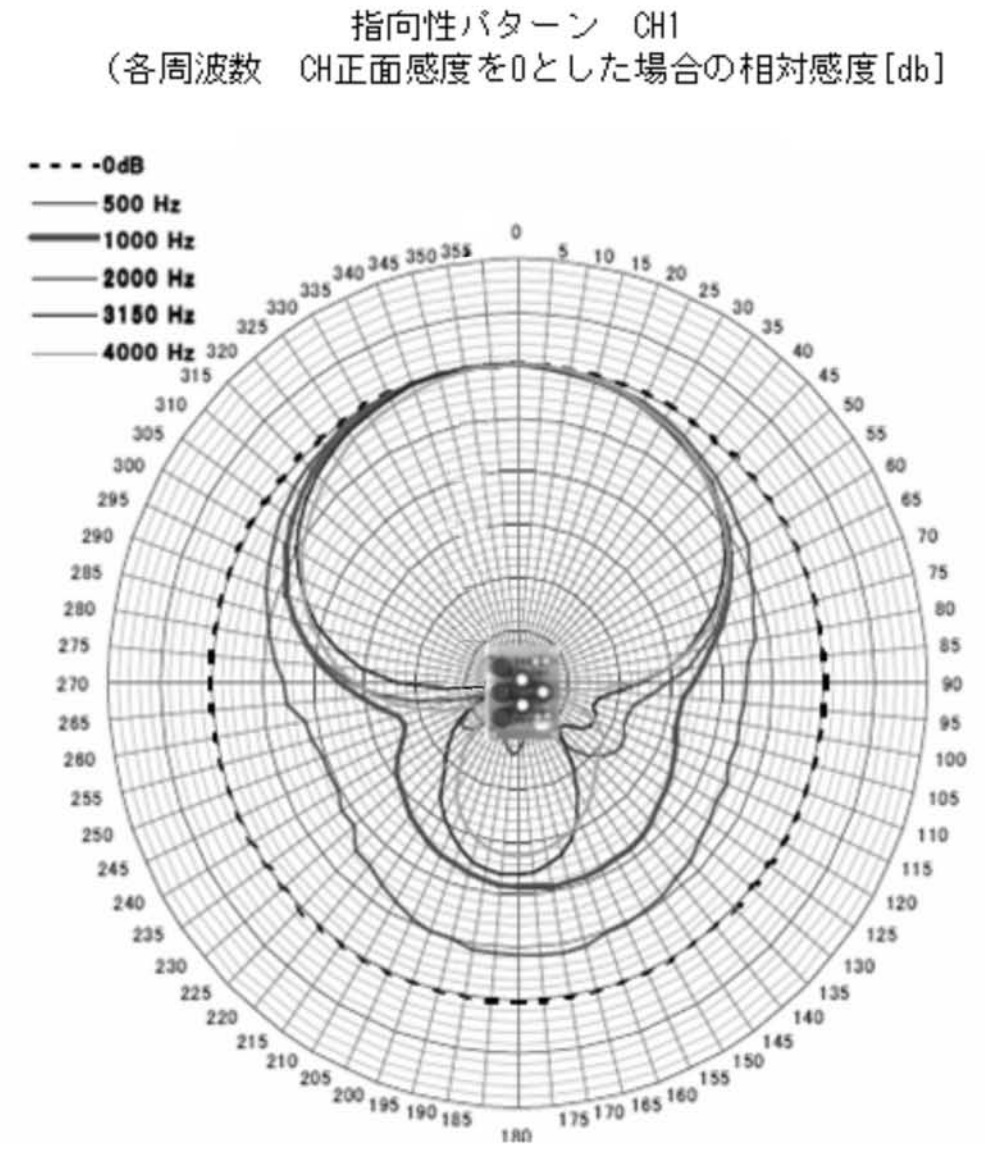
【図１４】



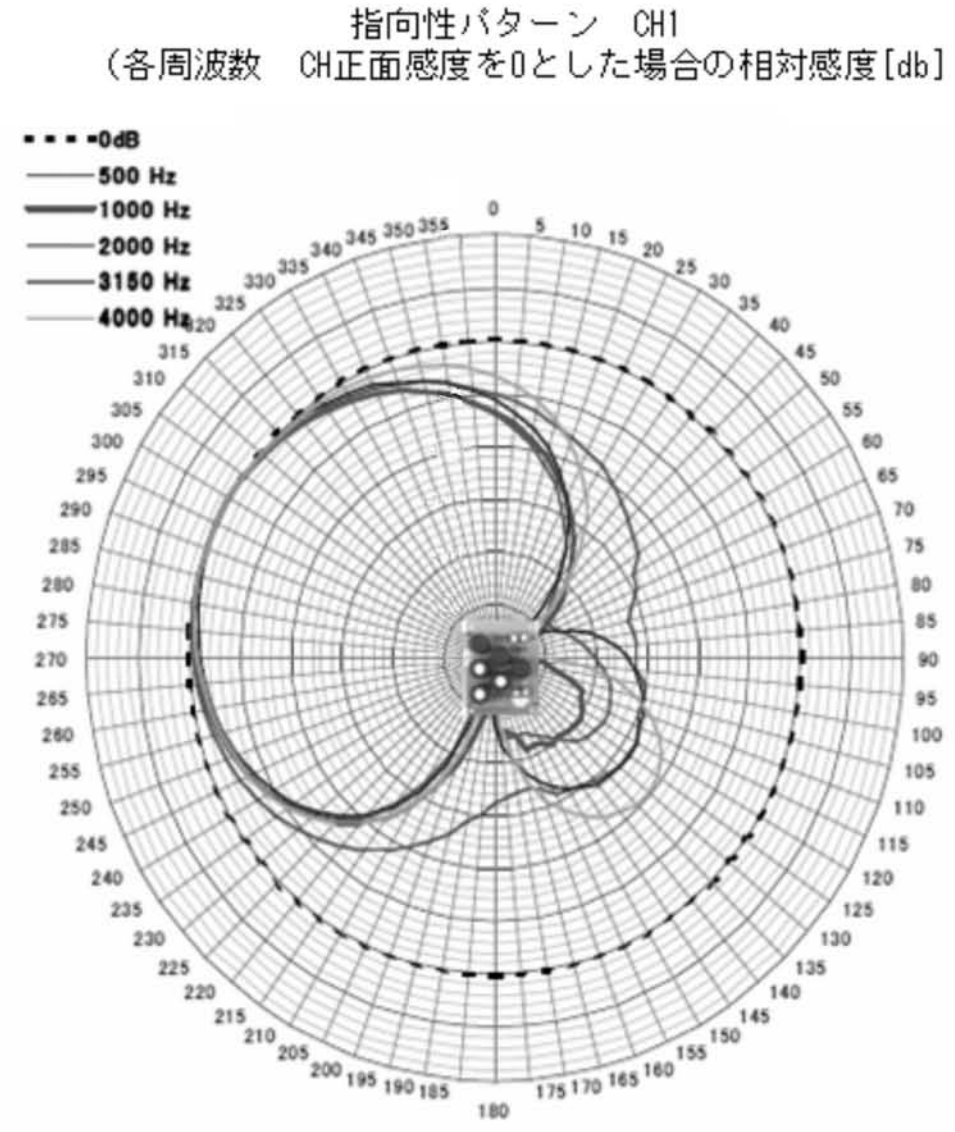
【図１５】



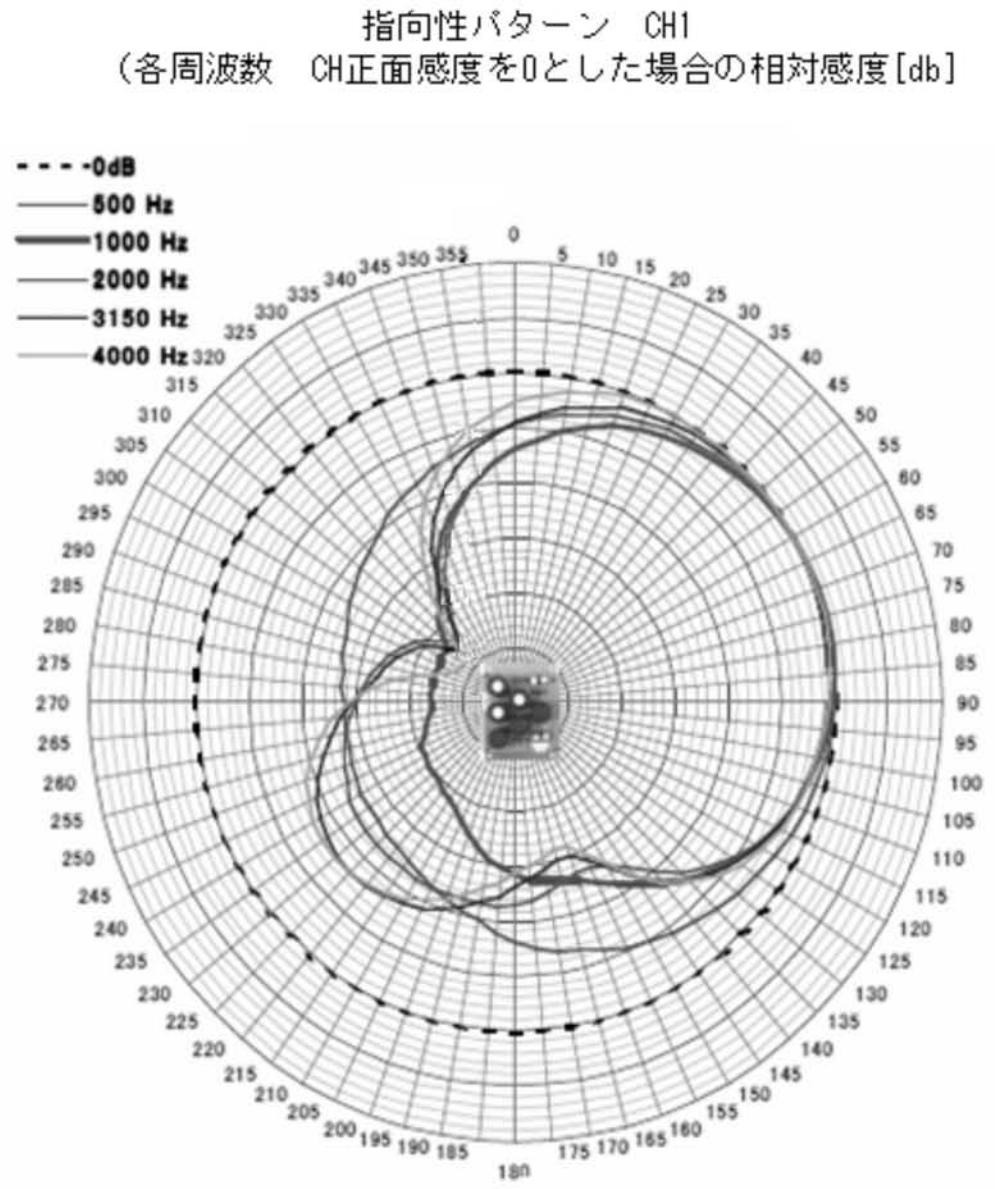
【図８】


指向性パターン　CH1
（各周波数　CH正面感度を0とした場合の相対感度[db]
-0dB
500 Hz
1000 Hz
2000 Hz
3150 Hz
4000 Hz
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355

【図９】


指向性パターン　CH1
（各周波数　CH正面感度を0とした場合の相対感度[db]
0dB
500 Hz
1000 Hz
2000 Hz
3150 Hz
4000 Hz

【図１０】


指向性パターン　CH1
（各周波数　CH正面感度を0とした場合の相対感度[db]
-0dB
500 Hz
1000 Hz
2000 Hz
3150 Hz
4000 Hz