

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3780516号

(P3780516)

(45) 発行日 平成18年5月31日(2006.5.31)

(24) 登録日 平成18年3月17日(2006.3.17)

(51) Int. Cl.	F I	
G 1 0 L 15/28 (2006.01)	G 1 0 L 15/28	4 0 0
G 1 0 L 15/00 (2006.01)	G 1 0 L 15/00	2 0 0 H
G 1 0 L 15/20 (2006.01)	G 1 0 L 15/20	3 7 0 E
G 1 0 L 15/24 (2006.01)	G 1 0 L 15/24	Q

請求項の数 15 (全 24 頁)

(21) 出願番号	特願2002-502769 (P2002-502769)	(73) 特許権者	503360115
(86) (22) 出願日	平成13年6月8日(2001.6.8)		独立行政法人科学技術振興機構
(86) 国際出願番号	PCT/JP2001/004858		埼玉県川口市本町4丁目1番8号
(87) 国際公開番号	W02001/095314	(74) 代理人	100082876
(87) 国際公開日	平成13年12月13日(2001.12.13)		弁理士 平山 一幸
審査請求日	平成14年12月9日(2002.12.9)	(72) 発明者	中臺 一博
(31) 優先権主張番号	特願2000-173915 (P2000-173915)		千葉県佐倉市白井86
(32) 優先日	平成12年6月9日(2000.6.9)	(72) 発明者	奥乃 博
(33) 優先権主張国	日本国(JP)		東京都渋谷区西原2-10-9
		(72) 発明者	北野 宏明
			埼玉県川越市西小仙波町2-18-3
		審査官	荏原 雄一
		(56) 参考文献	特開平11-41577 (JP, A)
			最終頁に続く

(54) 【発明の名称】 ロボット聴覚装置及びロボット聴覚システム

## (57) 【特許請求の範囲】

## 【請求項1】

内部にノイズ発生源を備えたロボットにおいて、  
少なくともロボットの一部を覆う防音性の外装と、  
上記外装の外側に設けられ、主として外部の音を集音する少なくとも二つの外部マイクと、

上記外装の内側に設けられ、主として内部のノイズ発生源からのノイズを集音する少なくとも一つの内部マイクと、

上記外部マイク及び内部マイクからの信号に基づいて、それぞれ外部マイクからの音響信号から、内部のノイズ発生源からのノイズ信号をキャンセルするとともに、上記内部マイクからの信号からノイズ発生源によるバーストノイズを検出してこのバーストノイズを含む帯域の信号部分を除去する処理部と、

上記処理部からの左右の音響信号から、音の方向付けを行なう方向情報抽出部と、を含み、

上記処理部は、内部マイクからの音響信号のパワーが外部マイクからの音響信号のパワーより十分大きく、所定値以上のピークが所定値以上の帯域に亘って検出されたとき、バーストノイズとしてこの帯域の信号部分を除去することを特徴とするロボット聴覚装置。

## 【請求項2】

内部にノイズ発生源を備えたロボットにおいて、  
少なくともロボットの一部を覆う自己認識用の防音性の外装と、

10

20

上記外装の外側に設けられ、主として外部の音を集音する少なくとも二つの外部マイクと、

上記外装の内側に設けられ、主として内部のノイズ発生源からのノイズを集音する少なくとも一つの内部マイクと、

上記外部マイク及び内部マイクからの信号に基づいて、それぞれ外部マイクからの音響信号から、内部のノイズ発生源からのノイズ信号をキャンセルするとともに、上記内部マイクからの信号からノイズ発生源によるバーストノイズを検出してこのバーストノイズを含む帯域の信号部分を除去する処理部と、

上記処理部からの左右の音響信号から、音の方向付けを行なう方向情報抽出部と、を含み、

10

上記処理部は、内部マイクからの音響信号のパワーが外部マイクからの音響信号のパワーより十分大きく、所定値以上のピークが所定値以上の帯域に亘って検出されたとき、バーストノイズとしてこの帯域の信号部分を除去することを特徴とするロボット聴覚装置。

#### 【請求項 3】

前記方向情報抽出部が、聴覚用のエピポーラ幾何により音の方向情報を計算して、音の方向付けを行なうことを特徴とする、請求項 1 又は 2 に記載のロボット聴覚装置。

#### 【請求項 4】

前記方向情報抽出部が、聴覚用のエピポーラ幾何に基づいた方法により、音の方向情報を計算して音の方向付けを行い、

調波構造を有する音に関しては、調波構造を利用した音源分離と音の強度差情報とを利用して音の方向付けを行なうことを特徴とする、請求項 1 又は 2 に記載のロボット聴覚装置。

20

#### 【請求項 5】

内部にノイズ発生源を備えたロボットにおいて、

少なくともロボットの一部を覆う防音性の外装と、

上記外装の外側に設けられ、主として外部の音を集音する少なくとも二つの外部マイクと、

上記外装の内側に設けられ、主として内部のノイズ発生源からのノイズを集音する少なくとも一つの内部マイクと、

上記外部マイク及び内部マイクからの信号に基づいて、外部マイクからの音響信号から内部のノイズ発生源からのノイズ信号をキャンセルするとともに、上記内部マイクからの信号からノイズ発生源によるバーストノイズを検出してこのバーストノイズを含む帯域の信号部分を除去する処理部と、

30

上記処理部からの左右の音響信号から、それぞれ周波数解析を行なって、当該音響データが示すピッチを伴う調波構造から、時間、周波数及びパワーに関する音響データを取り出すピッチ抽出部と、

上記ピッチ抽出部で取り出された左右の音響データに基づいて、音の方向付けを行なう左右チャンネル対応部と、

音の調波構造又は上記左右チャンネル対応部で取り出された音の方向付け情報に基づいて、上記音響データから音源毎の音響データに分離する音源分離部と、を含み、

40

上記処理部は、内外マイクの強度差がテンプレートの駆動機構のノイズの強度差と近く、内外マイクの入力音のスペクトルの強度とパターンとがテンプレートの駆動機構のノイズ周波数応答に近く、さらに駆動機構が動作しているときに、ノイズをバーストノイズとしてこの帯域の信号部分を除去することを特徴とするロボット聴覚システム。

#### 【請求項 6】

内部にノイズ発生源を備えたロボットにおいて、

少なくともロボットの一部を覆う自己認識用の防音性の外装と、

上記外装の外側に設けられ、主として外部の音を集音する少なくとも二つの外部マイクと、

上記外装の内側に設けられ、主として内部のノイズ発生源からのノイズを集音する少な

50

くとも一つの内部マイクと、

上記外部マイク及び内部マイクからの信号に基づいて、外部マイクからの音響信号から内部のノイズ発生源からのノイズ信号をキャンセルするとともに、上記内部マイクからの信号からノイズ発生源によるバーストノイズを検出してこのバーストノイズを含む帯域の信号部分を除去する処理部と、

上記処理部からの左右の音響信号からそれぞれ周波数解析を行なって、当該音響データが示すピッチを伴う調波構造から時間、周波数及びパワーに関する音響データを取り出すピッチ抽出部と、

上記ピッチ抽出部で取り出された左右の音響データに基づいて、音の方向付けを行なう左右チャンネル対応部と、

音の調波構造又は上記左右チャンネル対応部で取り出された音の方向付け情報に基づいて、上記音響データから音源毎の音響データに分離する音源分離部と、を含み、

上記処理部は、内外マイクの強度差がテンプレートの駆動機構のノイズの強度差と近く、内外マイクの入力音のスペクトルの強度とパターンとがテンプレートの駆動機構のノイズ周波数応答に近く、さらに駆動機構が動作しているときに、ノイズをバーストノイズとしてこの帯域の信号部分を除去することを特徴とするロボット聴覚システム。

【請求項 7】

内部に駆動機構等のノイズ発生源を備えた人型または動物型のロボットにおいて、少なくともロボットの頭部を覆う防音性の外装と、

上記外装の外側にて両側の耳に対応する両耳位置に設けられ、主として外部の音を集音する少なくとも一対の外部マイクと、

上記外装の内側に設けられ、主として内部のノイズ発生源からのノイズを集音する少なくとも一つの内部マイクと、

上記外部マイク及び内部マイクからの信号に基づいて、外部マイクからの音響信号から内部のノイズ発生源からのノイズ信号をキャンセルするとともに、上記内部マイクからの信号からノイズ発生源によるバーストノイズを検出してこのバーストノイズを含む帯域の信号部分を除去する処理部と、

上記処理部からの左右の音響信号から、それぞれ周波数解析を行なって、当該音響データが示すピッチを伴う調波構造から、時間、周波数及びパワーに関する音響データを取り出すピッチ抽出部と、

上記ピッチ抽出部で取り出された左右の音響データに基づいて、音の方向付けを行なう左右チャンネル対応部と、

音の調波構造又は上記左右チャンネル対応部で取り出された音の方向付け情報に基づいて、上記音響データから音源毎の音響データに分離する音源分離部と、を含み、

上記処理部は、内外マイクの強度差がテンプレートの駆動機構のノイズの強度差と近く、内外マイクの入力音のスペクトルの強度とパターンとがテンプレートの駆動機構のノイズ周波数応答に近く、さらに駆動機構が動作しているときに、ノイズをバーストノイズとしてこの帯域の信号部分を除去することを特徴とするロボット聴覚システム。

【請求項 8】

内部に駆動機構等のノイズ発生源を備えた人型または動物型のロボットにおいて、少なくともロボットの頭部を覆う自己認識用の防音性の外装と、

上記外装の外側にて両側の耳に対応する両耳位置に設けられ、主として外部の音を集音する少なくとも一対の外部マイクと、

上記外装の内側に設けられ、主として内部のノイズ発生源からのノイズを集音する少なくとも一つの内部マイクと、

上記外部マイク及び内部マイクからの信号に基づいて、外部マイクからの音響信号から内部のノイズ発生源からのノイズ信号をキャンセルするとともに、上記内部マイクからの信号からノイズ発生源によるバーストノイズを検出してこのバーストノイズを含む帯域の信号部分を除去する処理部と、

上記処理部からの左右の音響信号からそれぞれ周波数解析を行なって、当該音響データ

10

20

30

40

50

が示すピッチを伴う調波構造から時間、周波数及びパワーに関する音響データを取り出すピッチ抽出部と、

上記ピッチ抽出部で取り出された左右の音響データに基づいて、音の方向付けを行なう左右チャンネル対応部と、

音の調波構造又は上記左右チャンネル対応部で取り出された音の方向付け情報に基づいて、上記音響データから音源毎の音響データに分離する音源分離部と、を含み、

上記処理部は、内外マイクの強度差がテンプレートの駆動機構のノイズの強度差と近く、内外マイクの入力音のスペクトルの強度とパターンとがテンプレートの駆動機構のノイズ周波数応答に近く、さらに駆動機構が動作しているときに、ノイズをバーストノイズとしてこの帯域の信号部分を除去することを特徴とするロボット聴覚システム。

10

【請求項 9】

視覚、触覚等の知覚システムをさらに備えており、前記左右チャンネル対応部が、これらの知覚システムからの情報及び駆動機構の制御信号を参照して音の方向付け及び画像との対応付けを行なうことを特徴とする、請求項 5 ~ 8 のいずれかに記載のロボット聴覚システム。

【請求項 10】

視覚、触覚等の知覚システムをさらに備えており、前記左右チャンネル対応部が、これらの知覚システムからの情報及び駆動機構の制御信号を参照して音の方向付け及び画像との対応付けを行い、

前記左右チャンネル対応部が、前記知覚システムに対して音の方向付けに関する情報を出力することを特徴とする、請求項 5 ~ 8 のいずれかに記載のロボット聴覚システム。

20

【請求項 11】

視覚、触覚等の知覚システムをさらに備えており、前記左右チャンネル対応部が、これらの知覚システムからの情報及び駆動機構の制御信号を参照して音の方向付け及び画像との対応付けを行ない、

前記左右チャンネル対応部が、前記知覚システムに対して音の方向付けに関する情報を出力し、

前記左右チャンネル対応部が、聴覚用のエピソード幾何に基づいた方法により音の方向情報を計算して音の方向付けを行い、調波構造を有する音に関しては、調波構造を利用した音源分離と音の強度差情報とを利用して音の方向付けを行なうことを特徴とする、請求項 5 ~ 8 のいずれかに記載のロボット聴覚システム。

30

【請求項 12】

内部に駆動機構等のノイズ発生源を備えた人型または動物型のロボットにおいて、少なくともロボットの頭部を覆う防音性の外装と、

上記外装の外側にて両側の耳に対応する両耳位置に設けられ、主として外部の音を集音する少なくとも一対の外部マイクと、

上記外装の内側に設けられ、主として内部のノイズ発生源からのノイズを集音する少なくとも一つの内部マイクと、

上記外部マイク及び内部マイクからの信号に基づいて、外部マイクからの音響信号から内部のノイズ発生源からのノイズ信号をキャンセルするとともに、上記内部マイクからの信号からノイズ発生源によるバーストノイズを検出してこのバーストノイズを含む帯域の信号部分を除去する処理部と、

40

上記処理部からの左右の音響信号から、それぞれ周波数解析を行なって、当該音響データが示すピッチを伴う調波構造から、時間、周波数及びパワーに関する音響データを取り出すピッチ抽出部と、

上記ピッチ抽出部で取り出された左右の音響データに基づいて、音の方向付けを行なう左右チャンネル対応部と、

音の調波構造又は上記左右チャンネル対応部で取り出された音の方向付け情報に基づいて、上記音響データから音源毎の音響データに分離する音源分離部と、を含み、

上記処理部が、外部マイク及び内部マイクからの音響信号のスペクトルパワー差のパタ

50

ーンが前もって測定された駆動機構のノイズのスペクトルパワー差のパターンとほぼ同じであり、スペクトルの音圧及びパターンが前もって測定された駆動機構のノイズの周波数応答とほぼ同じであって、駆動機構の制御信号により駆動機構の連続動作を検出したとき、バーストノイズとしてこの帯域の信号部分を除去することを特徴とする、ロボット聴覚システム。

【請求項 1 3】

内部に駆動機構等のノイズ発生源を備えた人型または動物型のロボットにおいて、少なくともロボットの頭部を覆う自己認識用の防音性の外装と、上記外装の外側にて両側の耳に対応する両耳位置に設けられ、主として外部の音を集音する少なくとも一対の外部マイクと、

上記外装の内側に設けられ、主として内部のノイズ発生源からのノイズを集音する少なくとも一つの内部マイクと、

上記外部マイク及び内部マイクからの信号に基づいて、外部マイクからの音響信号から内部のノイズ発生源からのノイズ信号をキャンセルするとともに、上記内部マイクからの信号からノイズ発生源によるバーストノイズを検出してこのバーストノイズを含む帯域の信号部分を除去する処理部と、

上記処理部からの左右の音響信号からそれぞれ周波数解析を行なって、当該音響データが示すピッチを伴う調波構造から時間、周波数及びパワーに関する音響データを取り出すピッチ抽出部と、

上記ピッチ抽出部で取り出された左右の音響データに基づいて、音の方向付けを行なう左右チャンネル対応部と、

音の調波構造又は上記左右チャンネル対応部で取り出された音の方向付け情報に基づいて、上記音響データから音源毎の音響データに分離する音源分離部と、を含み、

上記処理部が、外部マイク及び内部マイクからの音響信号のスペクトルパワー差のパターンが前もって測定された駆動機構のノイズのスペクトルパワー差のパターンとほぼ同じであり、スペクトルの音圧及びパターンが前もって測定された駆動機構のノイズの周波数応答とほぼ同じであって、駆動機構の制御信号により駆動機構の連続動作を検出したとき、バーストノイズとしてこの帯域の信号部分を除去することを特徴とする、ロボット聴覚システム。

【請求項 1 4】

前記左右チャンネル対応部が、聴覚用のエピソード幾何により音の方向情報を計算して音の方向付けを行なうことを特徴とする、請求項 1 2 又は 1 3 に記載のロボット聴覚システム。

【請求項 1 5】

前記左右チャンネル対応部が、聴覚用のエピソード幾何に基づいた方法により音の方向情報を計算して音の方向付けを行い、調波構造を有する音に関しては、調波構造を利用した音源分離と音の強度差情報とを利用して音の方向付けを行なうことを特徴とする、請求項 1 2 又は 1 3 に記載のロボット聴覚システム。

【発明の詳細な説明】

技術分野

本発明はロボット、特に人型または動物型ロボットにおける聴覚装置に関するものである。

技術背景

近年、人型または動物型ロボットにおいては、視覚、聴覚の能動知覚が注目されてきている。能動知覚とは、ロボット視覚やロボット聴覚等の知覚を担当する知覚装置を、知覚すべき目標に追従するように、これらの知覚装置を支持する例えば頭部を駆動機構により姿勢制御するものである。

ここで、能動視覚に関しては、少なくとも知覚装置であるカメラが、駆動機構による姿勢制御によってその光軸方向が目標に向かって保持され、目標に対して自動的にフォーカシングやズームイン、ズームアウト等が行なわれることにより、目標がカメラによって撮像

10

20

30

40

50

されるようになっており、種々の研究が行なわれている。

これに対して、能動聴覚に関しては、少なくとも知覚装置であるマイクが、駆動機構による姿勢制御によってその指向性が目標に向くようにして保持され、目標からの音がマイクによって集音される。このとき、能動聴覚の不利な点として、駆動機構が作動している間は、マイクが駆動機構の作動音、特にバーストノイズを拾ってしまうため、目標からの音に大きなノイズが混入してしまい、目標からの音を正確に認識できなくなってしまうことがある。

しかしながら、駆動機構を停止させた状態での聴覚研究では、特に目標が移動しているような場合、目標の移動に追従しながらの所謂能動聴覚を行なうことはできない。

さらに、上述した駆動機構だけでなく、ロボット内部で発生する各種動作音及び定常的に発生するノイズも、聴覚装置としてのマイクが集音してしまうので、同様に完全な能動聴覚を得ることは困難であった。

ところで、ノイズキャンセルの方法としては、所謂アクティブノイズコントロール（ANC）という方法が知られている。

このANC法は、騒音源の近傍にマイクを設けて、騒音源からのノイズを集音し、騒音源からのノイズをキャンセルしたい場所におけるノイズを、IIR（無限インパルス応答）フィルタやFIR（有限インパルス応答）フィルタ等の適応フィルタによって予測して、ノイズをキャンセルしたい場所にて予想した騒音と逆位相の音をスピーカから出力してノイズを打ち消すことにより、ノイズを消滅させる方法である。

しかしながら、このようなANC法においては、ノイズを過去のデータに基づく予想によってキャンセルしているため、所謂バーストノイズに対応することは困難である。また、適応フィルタを利用してノイズをキャンセルしていることから、左右のチャンネル間の位相差情報が歪められ、あるいは消滅してしまうので、音の方向付けができなくなってしまう。

さらに、騒音源からのノイズを集音するマイクは、できるだけノイズのみを集音することが望ましいが、ロボット聴覚装置の場合にはノイズのみを集音することが困難である。

また、ノイズをキャンセルしたい場所におけるノイズを予想するための計算時間が必要であることから、騒音源とスピーカとの距離はある程度離れていることが前提条件となるが、ロボット聴覚装置の場合には、外部音を集音する外部マイクと内部の騒音を集音する内部マイクが比較的近くに設置されることから、計算時間が短くなるため、実質的にANC法を採用することが困難である。

従って、上述したANC法をロボット内部で発生するノイズをキャンセルするために採用することは不適である。

この発明は、以上の点にかんがみて、駆動機構等のロボット内部で発生するノイズに影響されることなく、外部の目標からの音を集音して能動知覚を行なうことができるようにした、ロボット聴覚装置及びロボット聴覚システムを提供することを目的としている。

#### 発明の開示

上記目的は、本発明の第一の構成によれば、内部にノイズ発生源を備えたロボットにおいて、少なくともロボットの一部を覆う防音性の外装と、上記外装の外側に設けられ主として外部の音を集音する少なくとも二つの外部マイクと、上記外装の内側に設けられ主として内部のノイズ発生源からのノイズを集音する少なくとも一つの内部マイクと、上記外部マイク及び内部マイクからの信号に基づいてそれぞれ外部マイクからの音響信号から内部のノイズ発生源からのノイズ信号をキャンセルする処理部と、上記処理部からの左右の音響信号から音の方向付けを行なう方向情報抽出部と、を含んでおり、上記処理部が、内部マイクからの信号からノイズ発生源によるバーストノイズを検出して、このバーストノイズを含む帯域の信号部分を除去することを特徴とするロボット聴覚装置により、達成される。

本発明によるロボット聴覚装置は、防音性の外装が自己認識用に構成されるのが望ましい。

本発明によるロボット聴覚装置は、内外マイクの強度差がテンプレートの駆動機構のノイ

10

20

30

40

50

ズの強度差と近く、内外マイクの入力音のスペクトルの強度とパターンとがテンプレートの駆動機構のノイズ周波数応答に近く、さらに駆動機構が動作しているときにノイズをバーストノイズとして、上記処理部が、好ましくはこの帯域の信号部分を除去する。

本発明によるロボット聴覚装置は、好ましくは、上記方向情報抽出部が、聴覚用のエピソード幾何に基づいた方法により、音の方向情報を計算して音の方向付け（音源定位）を行い、調波構造を有する音に関しては、調波構造を利用した音源分離と、音の強度差情報とを利用してロバスト（頑健）な音源定位を行う。

また、上記目的は、本発明の第二の構成によれば、内部にノイズ発生源を備えたロボットにおいて、少なくともロボットの一部を覆う、好ましくはロボットの自己認識用の防音性の外装と、上記外装の外側に設けられ主として外部の音を集音する少なくとも二つの外部マイクと、上記外装の内側に設けられ主として内部のノイズ発生源からのノイズを集音する少なくとも一つの内部マイクと、上記外部マイク及び内部マイクからの信号に基づいて、それぞれ外部マイクからの音響信号から内部のノイズ発生源からのノイズ信号をキャンセルする処理部と、上記処理部からの左右の音響信号からそれぞれ周波数解析を行なって当該音響データが示すピッチを伴う調波構造から時間、周波数及びパワーに関する音響データを取り出すピッチ抽出部と、上記ピッチ抽出部で取り出された左右の音響データに基づいて音の方向付けを行なう左右チャンネル対応部と、音の調波構造又は上記左右チャンネル対応部で取り出された音の方向付け情報に基づいて上記音響データから音源毎の音響データに分離する音源分離部と、を含んでおり、上記処理部が、内部マイクからの信号からノイズ発生源によるバーストノイズを検出して、このバーストノイズを含む帯域の信号部分を除去することを特徴とするロボット聴覚システムにより達成される。

さらに、上記目的は、本発明の第三の構成によれば、内部に駆動機構等のノイズ発生源を備えた人型または動物型のロボットにおいて、少なくともロボットの頭部を覆う、好ましくはロボットの自己認識用の防音性の外装と、上記外装の外側にて両側の耳に対応する両耳位置に設けられ主として外部の音を集音する少なくとも一対の外部マイクと、上記外装の内側に設けられ主として内部のノイズ発生源からのノイズを集音する少なくとも一つの内部マイクと、上記外部マイク及び内部マイクからの信号に基づいて外部マイクからの音響信号から内部のノイズ発生源からのノイズ信号をキャンセルする処理部と、上記処理部からの左右の音響信号からそれぞれ周波数解析を行なって当該音響データが示すピッチを伴う調波構造から時間、周波数及びパワーに関する音響データを取り出すピッチ抽出部と、上記ピッチ抽出部で取り出された左右の音響データに基づいて音の方向付けを行なう左右チャンネル対応部と、音の調波構造又は上記左右チャンネル対応部で取り出された音の方向付け情報に基づいて上記音響データから音源毎の音響データに分離する音源分離部と、を含んでおり、上記処理部が、内部マイクからの信号からノイズ発生源によるバーストノイズを検出して、このバーストノイズを含む帯域の信号部分を除去することを特徴とするロボット聴覚システムにより達成される。

この発明によるロボット聴覚システムは、好ましくは、ロボットが視覚、触覚等の知覚システムをさらに備えており、上記左右チャンネル対応部が、これらの知覚システムからの情報及び駆動機構の制御信号を参照して音の方向付け及び画像との対応付けを行なう。

この発明によるロボット聴覚システムは、好ましくは、上記左右チャンネル対応部が、上記知覚システムに対して音の方向付けに関する情報を出力する。

本発明によるロボット聴覚システムは、内外マイクの強度差がテンプレートの駆動機構のノイズの強度差と近く、内外マイクの入力音のスペクトルの強度とパターンとがテンプレートの駆動機構のノイズ周波数応答に近く、さらに駆動機構が動作しているときに、ノイズをバーストノイズとして、好ましくは上記処理部がこの帯域の信号部分を除去する。

本発明によるロボット聴覚システムは、内部マイクからの音響信号のパワーが外部マイクからの音響信号のパワーより十分大きく、所定周波数幅の複数個のサブバンドにて所定値以上のパワーを有しており、駆動機構の制御信号により駆動機構の連続動作を検出したとき、好ましくは、上記処理部がバーストノイズとしてこの帯域の信号部分を除去する。

本発明によるロボット聴覚システムは、外部マイク及び内部マイクからの音響信号のスペ

10

20

30

40

50

クトルパワー差のパターンが前もって測定された駆動機構のノイズのスペクトルパワー差のパターンとほぼ同じであり、スペクトルの音圧及びパターンが前もって測定された駆動機構のノイズの周波数応答とほぼ同じであって、駆動機構の制御信号により駆動機構の連続動作を検出したとき、好ましくは、上記処理部がバーストノイズとしてこの帯域の信号部分を除去する。

本発明によるロボット聴覚システムは、好ましくは、上記左右チャンネル対応部が、聴覚用のエピソード幾何に基づいた方法により音の方向情報を計算して音の方向付け（音源定位）を行い、調波構造を有する音に関しては、調波構造を利用した音源分離と音の強度差情報とを利用してロボスト（頑健）な音源定位を行う。

上記構成によれば、外部マイクが主として外部の目標からの音を集音し、内部マイクが主としてロボット内部の駆動機構等のノイズ発生源からのノイズを集音する。その際、外部マイクが集音した音響信号には、ロボット内部のノイズ発生源からのノイズ信号が混入しているが、この混入したノイズ信号は、処理部における演算処理によって、内部マイクが集音したノイズ信号によりキャンセルされて著しく低減される。その際、処理部が、内部マイクからの信号からノイズ発生源によるバーストノイズを検出して、外部マイクからの信号からこのバーストノイズを含む帯域の信号部分を除去する。すなわち、処理部はバーストノイズを含む外部マイクからの信号を全て取り除く。よって、方向情報抽出部または左右チャンネル対応部における音の方向付けを、バーストノイズに影響されることなくより正確に行なうことができる。

そして、ノイズがキャンセルされた音響信号から、ピッチ抽出部により周波数解析によって音響信号を取り出して、この音響信号から左右チャンネル対応部により音の方向付けを行なった後、この音の方向付けに基づいて音源分離部によって各音源毎の音響データに分離する。

したがって、外部マイクからの音響信号は、処理部での演算処理によってロボット内部の駆動機構等のノイズ発生源からのノイズが容易に著しく低減されると共に、特にバーストノイズを含む帯域の信号部分が除去されることによりそのS/N比が大幅に改善されるので、音源毎の音響データの分離をより一層良好に行なうことができる。

さらに、ロボットが視覚、触覚等の知覚システムを備えており、上記左右チャンネル対応部が、これらの知覚システムからの情報を参照して音の方向付けを行なう場合には、例えば視覚装置からの目標に関する視覚情報に基づいて、左右チャンネル対応部がより明確な方向付けを行なうことができる。

上記左右チャンネル対応部が、上記知覚システムに対して、音の方向付けに関する情報を出力する場合には、例えば視覚装置に対して目標に関する聴覚による方向付けの情報を出力するので、視覚装置がより一層正確な方向付けを行なうことができる。

処理部が、内外マイクの強度差がテンプレートの駆動機構のノイズの強度差と近く、内外マイクの入力音のスペクトルの強度とパターンとがテンプレートの駆動機構のノイズ周波数応答に近く、さらに駆動機構が動作しているときに、ノイズをバーストノイズとして、この帯域の信号部分を除去する場合、または内部マイクからの音響信号のパワーが外部マイクからの音響信号のパワーより十分大きく、所定周波数幅の複数個のサブバンドにて所定値以上のパワーを有しており、駆動機構の制御信号により駆動機構の連続動作を検出したとき、上記処理部がバーストノイズとしてこの帯域の信号部分を除去する場合には、バーストノイズの除去を容易に行なうことができる。

外部マイク及び内部マイクからの音響信号のスペクトルパワー差のパターンが前もって測定された駆動機構のノイズのスペクトルパワー差のパターンとほぼ同じであり、スペクトルの音圧及びパターンが前もって測定された駆動機構のノイズの周波数応答とほぼ同じであって、駆動機構の制御信号により駆動機構の連続動作を検出したとき、上記処理部がバーストノイズとしてこの帯域の信号部分を除去する場合には、バーストノイズの除去をより正確に行なうことができる。

上記方向情報抽出部が、エピソード幾何に基づいた方法により、音の方向情報を計算して音の方向付け（音源定位）を行い、調波構造を有する音に関しては、調波構造を利用した

10

20

30

40

50



音源分離と、音の強度差情報とを利用してロバスト（頑健）な音源定位を行う場合には、従来の視覚システムで行なわれているエピポーラ幾何による計算方法を聴覚システムに応用することによって、ロボットの外装や音響環境による影響を受けずに、音の方向付けを一層正確に行なうことができる。

ここで、従来の両聴システムにおいて一般的である頭部伝達関数（HRTF）を利用することは、本発明では必要としない。頭部伝達関数は音環境の変化に弱いことが知られているが、本発明では、音環境が変化したとしても頭部伝達関数の再計算や調整を行なう必要がないので、汎用性の高い聴覚システムを構成することができる。

発明を実施するための最良の形態

以下、本発明のロボット聴覚装置及びロボット聴覚システムにおける実施の形態を図面を用いて詳細に説明する。 10

図1乃至図2はこの発明によるロボット聴覚システムの一実施形態を備えた実験用の人型ロボットの全体構成を示している。

図1において、人型ロボット10は、4DOF（自由度）のロボットとして構成されており、ベース11と、ベース11上にて一軸（垂直軸）周りに回動可能に支持された胴体部12と、胴体部12上にて三軸方向（垂直軸，左右方向の水平軸及び前後方向の水平軸）の周りに揺動可能に支持された頭部13と、を含んでいる。

上記ベース11は固定配置されていてもよく、脚部として動作可能としてもよく、或いは、移動可能な台車等の上に載置されていてもよい。

上記胴体部12は、ベース11に対して垂直軸の周りに、図1にて矢印Aで示すように回動可能に支持されており、図示しない駆動手段によって回転駆動されると共に、図示の場合、防音性の外装によって覆われている。 20

上記頭部13は胴体部12に対して連結部材13aを介して支持されており、この連結部材13aに対して前後方向の水平軸の周りに、図1にて矢印Bで示すように揺動可能に、また左右方向の水平軸の周りに、図2にて矢印Cで示すように揺動可能に支持されていると共に、上記連結部材13aが、胴体部12に対してさらに前後方向の水平軸の周りに、図1にて矢印Dで示すように揺動可能に支持されており、それぞれ図示しない駆動手段によって各矢印A，B，C，D方向に回転駆動される。

ここで、上記頭部13は、図3に示すように全体が防音性の外装14により覆われていると共に、前側にロボット視覚を担当する視覚装置としてのカメラ15を、また両側にロボット聴覚を担当する聴覚装置としての一对の外部マイク16（16a，16b）を備えている。 30

さらに、上記頭部13は、図3に示すように、外装14の内側にて左右に離隔して配設された一对の内部マイク17（17a，17b）を備えている。

上記外装14は、例えばウレタン樹脂等の吸音性の合成樹脂から構成されており、頭部13の内部をほぼ完全に密閉することにより、頭部13の内部の遮音を行なうように構成されている。なお、胴体部12の外装も、同様にして吸音性の合成樹脂から構成されている。なお、この外装14は、ロボットが自己を認識するためのものでもあり、自己認識用にロボット内外から発せられる音を仕切る役割をするようになっている。ここで、自己認識とは、ロボットが、自分の内部の駆動手段から発したノイズやロボットの口（図示省略）から発せられた声などの音声とロボット外部からの音とを区別することを意味する。本発明では上記外装14は、ロボットの内部と外部の音が区別できる程度にロボット内部が密閉されるように構成されている。 40

上記カメラ15は公知の構成であって、例えば所謂パン，チルト，ズームの3DOF（自由度）を有する市販のカメラが適用され得る。

上記外部マイク16は、それぞれ頭部13の側面において、前方に向かって指向性を有するように取り付けられている。

ここで、外部マイク16の左右の各外部マイク16a，16bは、それぞれ図1及び図2に示すように、外装14の両側にて前方に向いた段部14a，14bにて内側に取り付けられ、段部14a，14bに設けられた貫通穴を通して前方の音を集音すると共に、外装 50

14の内部の音をできるだけ拾わないように、音反射材や吸音材などを配した適宜の手段により遮音されている。これにより、外部マイク16a, 16bは所謂バイノーラルマイクとして構成されている。なお、外部マイク16a, 16bの取付位置の近傍において、段部14a, 14bは人間の外耳形状或いは碗状に形成されていてもよい。

上記内部マイク17は、外装14の内側において、上述した各外部マイク16a, 16bの近傍に、図示の場合には、前記カメラ15の両端付近の上方にそれぞれ左右一対で配設されている。すなわち、一方の外部マイク16aの近傍には一方の内部マイク17aが、また、他方の外部マイク16bの近傍には他方の内部マイク17bが配設されている。尤も、上記内部マイク17は、上記した位置に限定されることなく外装14の内側で任意の位置に備えられていてよい。

図4は、上記外部マイク16及び内部マイク17を含む音響処理のための聴覚システムの電氣的構成を示している。図4において、聴覚システム20は、各外部マイク16a, 16b及び内部マイク17a, 17bからの音響信号をそれぞれ増幅するアンプ21a, 21b, 21c, 21dと、これらの増幅信号をAD変換によりデジタル音響信号SOL, SOR, SIL, SIRに変換するADコンバータ22a, 22b, 22c, 22dと、これらのデジタル音響信号が入力される処理部としての左右のノイズ除去回路23, 24と、ノイズ除去回路23, 24からのデジタル音響信号SR, SLが入力されるピッチ抽出部25, 26と、ピッチ抽出部25, 26からの音響データが入力される左右チャンネル対応部27と、左右チャンネル対応部27からのデータが入力される音源分離部28と、から構成されている。

上記ADコンバータ22a乃至22dは、例えば量子化ビット数16または24で、48kHzサンプリングした信号を取り出すように構成されている。

そして、左側の外部マイク16aからのデジタル音響信号SOL及び左側の内部マイク17aからのデジタル音響信号SILはノイズ除去回路23に入力され、また右側の外部マイク16bからのデジタル音響信号SOR及び左側の内部マイク17bからのデジタル音響信号SIRはノイズ除去回路24に入力される。これらのノイズ除去回路23, 24は同一の構成であり、それぞれ外部マイク16からの音響信号から内部マイク17のノイズ信号によりノイズキャンセルするように構成されている。即ち、ノイズ除去回路23は、外部マイク16aからのデジタル音響信号SOLを、内部マイク17aが集音したロボット内部のノイズ発生源からのノイズ信号SILに基づいて、例えば最も簡便には、外部マイク16aからのデジタル音響信号SOLから内部マイク17aからの音響信号SILを減算する等の適宜の演算処理によるノイズキャンセル処理によって、外部マイク16aからの音響信号SOLに混入したロボット内部の各駆動機構（駆動手段）等のノイズ発生源からのノイズを除去して左の音響信号SLを生成する。また、ノイズ除去回路24は、外部マイク16bからのデジタル音響信号SORを、内部マイク17bが集音したロボット内部のノイズ発生源からのノイズ信号SIRに基づいて、例えば最も簡便には、外部マイク16bからのデジタル音響信号SORから内部マイク17bからの音響信号SIRを減算する等の適宜の演算処理によるノイズキャンセル処理によって、外部マイク16bからの音響信号SIRに混入したロボット内部の各駆動機構等のノイズ発生源からのノイズを除去して右の音響信号SRを生成する。

ここで、上記ノイズ除去回路23, 24は、内部マイク17a, 17bからの音響信号SIL, SIRから所謂バーストノイズを検出して、外部マイク16a, 16bからの音響信号SOL, SORからこのバーストノイズの帯域に対応する信号部分を除去することにより、バーストノイズの混入による音の方向付けの精度を高めるようになっている。このバーストノイズの除去は、ノイズ除去回路23, 24内にて、以下のようにして行なわれる。

まず、ノイズ除去回路23, 24は、第一のバーストノイズ除去方法として、内部マイク17a, 17bからの音響信号SIL, SIRと外部マイク16a, 16bからの音響信号SOL, SORとを比較して、音響信号SIL, SIRのパワーが音響信号SOL, SORのパワーより十分大きく、また音響信号SIL, SIRのパワーが所定値（例えば3

10

20

30

40

50

0 dB)以上のピークが所定周波数幅、例えば47Hzのサブバンドに関して、一定数(例えば20)以上のサブバンドに亘って連続して、さらに駆動手段が作動を続けている場合に、バーストノイズがあると判断して、音響信号SOL, SORの当該サブバンドに対応する信号部分を除去する。このために、ノイズ除去回路23, 24には駆動機構の制御信号が入力されている。

なお、バーストノイズの除去やその検出の判断方法としては、好ましくは後述する第二のバーストノイズ除去方法を行うのが望ましい。

このようなバーストノイズの除去は、例えばアダプティブフィルタによって行なわれる。このアダプティブフィルタは、リニアフェーズフィルタであって、例えば100のオーダーのFIRフィルタにより構成される。各FIRフィルタのパラメータは、適応アルゴリズムとしての最小二乗法により計算される。

かくして、ノイズ除去回路23, 24は、図6に示すように、それぞれバーストノイズ除去部として機能してバーストノイズの検出及び除去を行なう。

上記ピッチ抽出部25, 26は同一の構成であり、ノイズ除去回路23, 24からの左右の音響信号SL, SRをそれぞれ周波数解析して、時間, 周波数及びパワーの三軸から成る音響データを取り出すように構成されている。即ち、ピッチ抽出部25は、ノイズ除去回路23からの左の音響信号SLを周波数解析することにより、時間とパワーで成る二軸の音響信号SLから時間, 周波数及びパワーの三軸から成るスペクトログラムと呼ばれる左の音響データDLを取り出す。また、ピッチ抽出部26も同様にして、ノイズ除去回路24からの右の音響信号SRを周波数解析することにより、時間とパワーから成る二軸の音響信号SRから時間, 周波数及びパワーの三軸で成る右の音響データDRを取り出す。ここで、上記周波数解析は、例えば20m秒の窓長で、7.5m秒づつずらしながら、FFT(高速フーリエ変換)を行なうことにより実行される。なお、このような周波数解析は、FFTに限らず一般的な種々の方法で行なわれるようにしてもよい。

このようにして得られた音響データDLにおいては、音声や音楽における各音はスペクトログラム上にてピークの連なりとして示されると共に、一般的に調波構造を有しており、整数倍の周波数値に規則的にピークを有するようになっている。

ピークの抽出は次のように行われる。スペクトルは、例えば48KHzのサンプリング速度で1024のサブバンドについてフーリエ変換することにより計算される。次に、スペクトルの内、閾値以上のパワーを持ったローカルピークを抽出する。閾値は周波数ごとに異なり、部屋の暗騒音を一定時間計測することによって自動的に求められる。この際、バンドパスフィルタを用いて、ノイズの大きい90Hz以下の周波数域とパワーの小さい3KHz以上の高周波数域とは計算量低減のためにカットする。これにより十分高速なピーク抽出が実現できる。

上記左右チャンネル対応部27は、左右のピッチ抽出部25, 26からの音響データDL, DRにてそれぞれピークにより示される調波構造のピッチに関して、その位相差, 時間差等に基づいて同じ音から派生するピッチを左右チャンネルで対応付けることにより音の方向付けを行なう。この音の方向付けは、例えばエピポラ幾何に基づいた方法により、音の方向情報を計算して音の方向付け(音源定位)を行い、調波構造を有する音に関しては、調波構造を利用した音源分離と音の強度差情報とを利用してロバスト(頑健)な音源定位を行う。

ここで、視覚によるエピポラ幾何の場合には、図6(A)に示すように、互いに光軸が平行であって画面が同一面に位置し、同じ焦点距離を有する二つのカメラから成る簡単なステレオカメラにおいて、点P(X, Y, Z)が各カメラの画面にて点P1l(x1, y1)及び点P2(xr, yr)に投影されるとき、以下の関係式

$$X = \frac{b(x_1 + x_r)}{2d}, \quad Y = \frac{b(y_1 + y_r)}{2d}, \quad Z = \frac{bf}{d}$$

が成立する。ここで、fは各カメラの焦点距離, bはベースライン, dは(x1 - xr)で定義される。

10

20

30

40

50

このエピソード幾何の概念を聴覚に導入することにより、図6(B)に示すように、外部マイク16a, 16bの中心から音源Pへの角度  $\theta$  について、以下の方程式

$$\cos \theta = \frac{v}{2\pi f b} \Delta \phi$$

が成立する。ここで、 $v$ は音速、 $f$ は音の周波数である。

そして、左右の外部マイク16a, 16bから音源までの距離差  $l$  に基づいて、外部マイクからの左右の音響信号  $SOL$ ,  $SOR$  の間に位相差  $IPD = \frac{2\pi f l}{v} \sin \theta$  が発生する。

音の方向付けは、例えば各サブバンドの帯域幅が  $47\text{Hz}$  になるようなFFT(高速フーリエ変換)によりピークを抽出し、位相差  $IPD$  を計算することにより行なわれる。なお、ピークの抽出は、例えば  $48\text{kHz}$  のサンプリング速度で  $1024$  のサブバンドについてフーリエ変換を計算することにより、HRTFを利用した場合に比較して十分速く、正確に計算される。

これにより、頭部伝達関数(HRTF)によらない音の方向付け(音源定位)が実現できる。ピークの抽出においては、例えば  $48\text{kHz}$  のサンプリング速度で  $1024$  ポイントのFFTを使用して、スペクトルサブトラクションによる方法を用いる。これにより、実時間の処理が精度よく行える。なお、このスペクトルサブトラクションは、FFTの窓関数の性質をも考慮に入れたスペクトル補間を伴っている。

かくして、左右チャンネル対応部27は、図5に示すように、方向情報抽出部として機能して、方向情報を抽出する。図示の場合、左右チャンネル対応部27は、聴覚システム20以外にロボット10に備えられた図示しない他の知覚システム30から目標に関する情報、具体的には例えば視覚システムによる目標の位置、方向、形状、移動の有無等に関する情報や、触覚システムによる目標の柔軟度、振動の有無、触感等の情報が入力されることにより、上述した目標からの音の方向付けが正確に行なわれる。例えば、左右チャンネル対応部27は、上記(聴覚による)方向情報と、カメラ15からの視覚情報に基づく(視覚による)方向情報と比較して、これらの整合性をチェックし関連付ける。

さらに、左右チャンネル対応部27は、人型ロボット10の各駆動手段の制御信号に基づいて、頭部13の方向情報(ロボット座標)を取得することにより、目標との相対位置を計算することができる。これにより、人型ロボット10が移動している場合であっても、目標からの音の方向付けが一層正確に行なわれる。

上記音源分離部28は、公知の構成により、左右チャンネル対応部27からの方向付けの情報及び音響データ  $DL$ ,  $DR$  が入力されることにより、ディクシオンパスフィルタを使用することによって、方向付けの情報に基づいて音響データ  $DL$ ,  $DR$  から音源を同定すると共に、各音源毎に音響データを分離するように構成されている。

このディクシオンパスフィルタは、例えば以下のようにしてサブバンドを収集する。即ち、特定の方向  $\theta$  を、各サブバンド( $47\text{Hz}$ )に関して  $\theta$  に変換した後、ピークを抽出して、位相差( $IPD$ )及び  $\Delta \phi$  を計算する。そして、位相差が  $\Delta \phi = \frac{2\pi f l}{v} \sin \theta$  であるとき、当該サブバンドを収集する。このようにして全てのサブバンドについて上記処理をして、収集されたサブバンドから成る波形を構成する。

ここで、同時のFFTにより得られる左右チャンネルのスペクトルを  $S_p^{(l)}$  及び  $S_p^{(r)}$  とし、 $S_p(f_p)$  のピーク周波数を  $f_p$  としたとき、左右のチャンネルのスペクトル  $S_p^{(l)}(f_p)$  及び  $S_p^{(r)}(f_p)$  は、それぞれ実数部  $R[S_p^{(r)}(f_p)]$ ,  $R[S_p^{(l)}(f_p)]$  及び虚数部  $I[S_p^{(r)}(f_p)]$ ,  $I[S_p^{(l)}(f_p)]$  で表わされる。

従って、上記  $\Delta \phi$  は、

$$\Delta \phi = \tan^{-1} \left( \frac{I[S_p^{(r)}(f_p)]}{R[S_p^{(r)}(f_p)]} \right) - \tan^{-1} \left( \frac{I[S_p^{(l)}(f_p)]}{R[S_p^{(l)}(f_p)]} \right)$$

により求められる。

かくして、視覚(カメラ15)によるエピソード平面から聴覚(外部マイク16)によるエピソード平面への変換は図6に示すように容易に行なわれるので、目標の方向( $\theta$ )は

10

20

30

40

50

、聴覚によるエピソード幾何に基づいて、上記の方程式 [ 数 2 ] により、 $f = f_p$ として容易に求められる。

このように、左右チャンネル対応部 27 で音源定位が行われ、その後音源分離部 28 で音源が分離される。なお、図 7 はそれらの処理の概念図である。

また、音の方向付けと音源定位とについては、調波構造抽出によって音源分離を実現する方法により、調波構造を有する音に対してロバストな音源定位は行える。すなわち、これは、図 4 に示すモジュールの内、左右チャンネル対応部 27 と音源分離部 28 とを入れ替えて、音源分離部 28 からのデータが左右チャンネル対応部 27 へ入力されることにより実現できる。

ここで、調波構造を有する音に対する音源分離と音源定位について説明する。図 8 に示すように、まず音源分離においては、ピーク抽出で抽出されたピークは、周波数が低いものから順番に取りだされる。その周波数  $F_0$  と一定以内の誤差 (例えば心理学的実験によって得られた 6%) で整数倍とみなせる周波数  $F_n$  とを持つローカルピークが、周波数  $F_0$  の倍音としてクラスタリングを行われる。このクラスタリングによって集められた最終的なピークの集合が一つの音と見なされる。これによって、音源分離が行われる。

次に、音源定位について説明する。一般に両耳聴における音源定位には、頭部伝達関数 (HRTF) から求められる両耳間位相差 (IPD) と両耳間強度差 (IID) とが使用される。しかし、HRTF は頭部の形状や環境に大きく依存し、環境が変わる都度、計測が必要であるため実環境アプリケーションには不向きである。

そこで、本発明は、HRTF に依らない IPD を利用した音源定位法として、ステレオ視におけるエピソード幾何の概念を聴覚に拡張した聴覚エピソード幾何に基づく方法を適用している。

この際、(1) 音の倍音構造の利用、(2) IPD を用いた聴覚エピソード幾何による定位結果と IID を用いた定位結果とのデンプスター - シェイファー理論を用いた統合、(3) モータ動作中でも正確な音源定位を可能とするアクティブオーディションの導入、によって音源定位のロバスト性を向上させている。

図 8 に示すように、この音源定位は、音源分離によって分離された調波構造を有した各音に対して行う。ロボットでは、左右のマイクのベースラインから 1.5 KHz 以下の周波数域に対しては IPD、それ以上の周波数域では IID、による音源定位が有効である。このため、入力音のうち 1.5 KHz 以上の倍音成分と 1.5 KHz 以下の倍音成分との 2 つに分けて処理を行う。まず、入力音のうち 1.5 KHz 以下の周波数  $f_k$  を持った各倍音成分に対して、聴覚エピソード幾何を使用して、ロボット正面に対して  $\pm 90^\circ$  の範囲で  $5^\circ$  おきに IPD 仮説 ( $P_h(\theta, f_k)$ ) を生成する。

次に、次式に示す距離関数により、入力の各倍音における IPD ( $P_s(f_k)$ ) と各仮説間の距離 ( $d(\theta)$ ) とを計算する。ここで、 $n_{f < 1.5 \text{ KHz}}$  は周波数が 1.5 KHz 以下である倍音数である。

$$d(\theta) = \frac{1}{n_{f < 1.5 \text{ KHz}}} \sum_{k=0}^{n_{f < 1.5 \text{ KHz}} - 1} \frac{(P_h(\theta, f_k) - P_s(f_k))^2}{f_k}$$

次に、得られた距離に対し、次式によって定義される確率密度関数を適用し、距離を IPD を用いた場合の音源方向を支持する確信度  $B_{F_{IPD}}$  に変換する。ここで、 $m$  と  $s$  とは、それぞれ  $d(\theta)$  の平均と分散とであり、 $n$  は  $d$  の個数である。

$$B_{F_{IPD}}(\theta) = \frac{\frac{s}{\sqrt{n}}}{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx}$$

入力音のうち 1.5 KHz 以上の周波数を持った倍音に関しては、IID の総和の正負に応じて下記表 1 に示す値を IID を用いた場合の音源方向を支持する確信度  $B_{F_{IID}}$  として与える。

10

20

30

40

表1. IID確信度 ( $BF_{IID}(\theta)$ ) を示す表

$\theta$		$90^\circ \sim 35^\circ$	$30^\circ \sim -30^\circ$	$-35^\circ \sim 90^\circ$
IIDの総和	+	0.35	0.5	0.65
	-	0.65	0.5	0.35

IPD, IIDそれぞれの処理によって得られた音源方向を支持する値から、これらを次式で示されるデンプスター-シェイファー理論によって統合し、IPDとIIDとの両方から音源方向を支持する新しい確信度を生成する。

$$BF_{IPD+IID}(\theta) = BF_{IPD}(\theta) BF_{IID}(\theta) + (1 - BF_{IPD}(\theta)) BF_{IID}(\theta) + BF_{IPD}(\theta)(1 - BF_{IID}(\theta))$$

10

このような確信度  $BF_{IPD+IID}$  を各角度ごとに生成し、それらのうち最大の確信度値を持った方向が音源の方向と成る。

本発明の実施形態による人型ロボット10は以上のように構成されており、外部マイク16a, 16bが集音しようとする目標からの音は、以下のようにして集音され、ノイズキャンセルされることにより音源として知覚される。

まず、外部マイク16a, 16bが、主として目標からの外部の音を集音してそれぞれアナログ音響信号を出力する。ここで、外部マイク16a, 16bはロボット内部からのノイズも集音するが、外装14自体が頭部13の内部を密閉しており、また外部マイク16a, 16bが頭部13の内部に対して遮音されていることにより、混入するノイズが比較的低いレベルに抑えられている。

20

これに対して、内部マイク17a, 17bは、主としてロボット内部からのノイズ、例えば前述した各駆動機構の作動音や冷却ファンの作動音等のノイズ発生源からのノイズを集音する。ここで、内部マイク17a, 17bは外部からの音も集音するが、外装14が内部を密閉していることからそのレベルは比較的低く抑えられている。

このようにして集音された外部マイク16a, 16bからのアナログ音響信号及び内部マイク17a, 17bからのアナログ音響信号は、それぞれアンプ21a乃至21dにより増幅された後、ADコンバータ22a乃至22dによりデジタル音響信号SOL, SOR, SIL, SIRにAD変換されて、ノイズ除去回路23及び24に入力される。

30

ノイズ除去回路23, 24は、それぞれ外部マイク16a, 16bからの音響信号SOL, SORから内部マイク17a, 17bからの音響信号SIL, SIRを減算する等の演算処理を行なうことにより、外部マイク16a, 16bからの音響信号SOL, SORからロボット内部のノイズ発生源からのノイズ信号を除去すると共に、バーストノイズを検出して、外部マイク16a, 16bからの音響信号SOL, SORからバーストノイズを含むサブバンドの信号部分を除去することにより、それぞれノイズ特にバーストノイズを除去した真の音響信号SL, SRを外部へ出力する。

そして、ピッチ抽出部25, 26が、それぞれ音響信号SL, SRに基づいて周波数解析により、音響信号SL, SRに含まれる全ての音に関して当該音に関するピッチを抽出し、このピッチに対応する当該音の調波構造、開始時、終了時と共に、音響データDL, DRとして左右チャンネル対応部27に出力する。

40

続いて、左右チャンネル対応部27が、これらの音響データDL, DRに基づいて音源毎の音の方向付けを行なう。

この場合、左右チャンネル対応部27は、例えば各ピッチ抽出部25, 26で抽出された音響データDL, DRに基づいて左右のチャンネルの調波構造を比較して、最も近いピッチを対応付ける。その際、左右のチャンネルのピッチを一对一で比較するだけでなく、一方のチャンネルの複数のピッチを他方のチャンネルの一つのピッチと比較する等により、より正確な対応付けを行なうことが好ましい。

そして、左右チャンネル対応部27は、対応付けられたピッチの位相を比較すると共に、前述したエピソード幾何に基づいた方法により、音の方向情報を計算することにより、音

50

の方向付けを行なう。

これにより、音源分離部28が、左右チャンネル対応部27からの音の方向付け情報に基づいて、音響データDL, DRから、音源毎の音に関する音響データを抽出して、音源毎の音に分離する。かくして、聴覚システム20は音源毎の音の分離によって音響認識を行ない、能動聴覚を行なうことができる。

このようにして、本発明実施形態による人型ロボット10によれば、ノイズ除去回路23, 24により、外部マイク16a, 16bからの音響信号SOL, SORから、内部マイク17a, 17bからの音響信号SIL, SIRに基づいてノイズキャンセルを行なうと共に、外部マイク16a, 16bからの音響信号SOL, SORから、バーストノイズを含むサブバンドの信号成分を除去することにより、各駆動機構によって外部マイク16a, 16bの指向性の方向を目標に向けながら、バーストノイズの影響を受けることなく、音の方向付けを行なうことができると共に、方向付けが、従来のようにHRTFを使用することなく、エピソード幾何に基づいた方法により計算することによって行なわれ得るので、音環境の変化によりHRTFの調整または再測定を行なう必要がなく、計算時間が短縮され得ると共に、未知の音環境においても、各音源からの音を分離して、より正確な音声認識を行なうことができる。

従って、例えば目標が移動している場合であっても、各駆動機構によって、外部マイク16a, 16bの指向性の方向を常に目標に対して追従させながら、目標の音響認識を行なうことができる。このとき、左右チャンネル対応部27が、例えば他の知覚システム30として、視覚システムからの視覚による目標の方向付け情報を参照して音の方向付けを行なうことにより、より正確な音の方向付けを行なうことができる。

また、他の知覚システム30として、視覚システムが使用される場合、左右チャンネル対応部27が視覚システムに対して音の方向付け情報を出力するようにしてもよい。この場合、視覚システムが画像認識により目標の方向付けを行なう際に、聴覚システム20の音に関する方向付け情報を参照することによって、例えば目標が移動して障害物の裏に隠れてしまった場合でも、目標からの音を参照して、目標の方向付けをより一層正確に行なうことができる。

以下に、具体的な実験例を示す。

図9に示すように、上記人型ロボット10が、10平方mの居住用の部屋40内にて、二つの音源としてのスピーカ41, 42に対向している。ここで、人型ロボット10は、その頭部13を53度(右を0度とし、左回りの角度)の方向(回転前の方向)に向けている。

一方のスピーカ41は、500Hzの単調音を再生し、人型ロボット10の前方5度左(58度)の位置に在る。これに対して、他方のスピーカ42は600Hzの単調音を再生し、人型ロボット10から見てスピーカ41の69度左(127度)の位置に在る。人型ロボット10から各スピーカ41, 42までの距離は、約210cmである。

ここで、人型ロボット10のカメラ15の視野は、水平方向に関して約45度であるから、人型ロボット10は、スピーカ42をカメラ15により見ることができない。

この状態で、スピーカ41が音を再生し、スピーカ42が約3秒遅れて音を再生する場合について、人型ロボット10が聴覚によりスピーカ42の音の方向付けを行なって、スピーカ42の方向に向けて頭部13を回転し、音源としてのスピーカ42と視覚対象物としてのスピーカ42とを関連付ける、という実験を行なった。なお、回転後の頭部13の方向は131度の方向になる。

実験は、人型ロボット10の頭部13の回転速度について、速い動作(68.8度/秒)及び遅い動作(14.9度/秒)と、S/N比0dBの弱いノイズ(内部の待機状態の音と同様のパワー)とS/N比約50dBの強いノイズ(バーストノイズ)の条件で行なわれ、以下のような結果が得られた。

図10は、人型ロボット10の内部で発生する内部音(ノイズ)のスペクトログラムで、(A)は速い動作の場合、(B)は遅い動作の場合を示している。これらのスペクトログラムによれば、駆動モータによるバーストノイズが明白に示されている。

10

20

30

40

50

従来のノイズ除去による方向情報は、図 1 1 ( A ) または ( B ) に示すように、頭部 1 3 が回転している間 ( 5 ~ 6 秒の間 )、ノイズにより大きな影響を受けており、人型ロボット 1 0 が音源を追跡するために頭部 1 3 を回転駆動している間は、聴覚がほぼ無効となるようなノイズを発生していることが分かる。

これに対して、本発明によるバーストノイズ除去による方向情報は、図 1 2 に示す弱いノイズ及び図 1 3 に示す強いノイズの場合に、それぞれ頭部 1 3 の回転駆動の間でも、バーストノイズの影響を受けずに、方向情報が正確に得られることが分かる。なお、図 1 4 ( A ) は図 1 3 ( A ) に対応するスペクトログラムを表し、また、図 1 4 ( B ) は図 1 3 ( B ) に対応するスペクトログラムを表すもので、いずれもノイズに比べて信号が強い場合を示している。

上記ノイズ除去回路 2 3 , 2 4 は、上述したように、各サブバンドについて、音響信号 S I L , S I R のパワーに基づいて、バーストノイズの有無の判定を行なってバーストノイズを除去しているが、バーストノイズの除去は、外装 1 4 の音響特性に基づいて、以下のようにして行なうようにしてもよい。

第二のバーストノイズ除去方法においては、マイクに入力されるノイズは、ある時刻に、下記の 3 要件 ( ( 1 ) ~ ( 3 ) ) を具備するとき、バーストノイズとして取り扱われる。

( 1 ) 内外のマイク 1 6 a , 1 6 b , 1 7 a , 1 7 b の強度差がテンプレートのモータなどの駆動手段のノイズの強度差と近い。

( 2 ) 内外マイクの入力音のスペクトル強度とパターンがテンプレートのモータノイズ周波数応答に近い。

( 3 ) モータなどの駆動手段が動作している。

即ち、第二のバーストノイズ除去方法においては、先ず、ノイズ除去回路 2 3 , 2 4 は、前以てロボット 1 0 の各駆動手段の動作時における音響測定データ ( 後述の図 1 5 ( A ) 及び ( B ) や図 1 6 ( A ) 及び ( B ) )、即ち外部マイク 1 6 及び内部マイク 1 7 からの音響信号データを測定し、テンプレートとして記憶しておく。

次に、ノイズ除去回路 2 3 , 2 4 は各サブバンド毎に、内部マイク 1 7 a , 1 7 b からの音響信号 S I L , S I R と外部マイク 1 6 a , 1 6 b からの音響信号 S O L , S O R に関して、上記記憶した音響測定データをテンプレートとしてバーストノイズの判定を行なう。即ち、ノイズ除去回路 2 3 , 2 4 は、外部マイクと内部マイクのスペクトルパワー差 ( 又は音圧差 ) のパターンが、測定された音響測定データにおける駆動手段のノイズのスペクトルパワー差のパターンとほぼ同じであり、スペクトルの音圧及びパターンが測定された駆動手段のノイズの周波数応答とほぼ同じであって、さらに駆動手段が作動を続けている場合に、バーストノイズがあると判断して当該サブバンドに対応する信号部分を除去する。

このようなバーストノイズの判定は、以下の理由に基づいている。外装 1 4 の音響特性は無響室において測定される。その際、測定される音響特性の項目は次の通りである。カバードロボット 1 0 の各駆動手段、即ち頭部 1 3 を前後方向に揺動させる第一のモータ ( モータ 1 ) , 頭部 1 3 を左右方向に揺動させる第二のモータ ( モータ 2 ) , 頭部 1 3 を垂直軸の周りに回動させる第三のモータ ( モータ 3 ) 及び胴体部 1 2 を垂直軸の周りに回動させる第四のモータ ( モータ 4 ) のノイズに関する内部マイク 1 7 及び外部マイク 1 6 による周波数対応は、図 1 5 ( A ) , ( B ) に示すようになっている。また、内部マイク 1 7 及び外部マイク 1 6 のスペクトルパワー差のパターンは、図 1 6 ( A ) に示すようになり、外部マイクの周波数応答から内部マイクの周波数応答を減算することにより得られる。同様に、外部音のスペクトルパワー差のパターンは、図 1 6 ( B ) に示すようになる。これはインパルス応答により得られる。インパルス応答は、水平方向及び垂直方向のマトリックス要素、即ち水平方向がロボット中心から 0 度 ,  $\pm 45$  度 ,  $\pm 90$  度及び  $\pm 180$  度 , そして垂直方向が 0 度及び 30 度である 12 点で測定される。

これらの図から、以下のことが観察される。即ち、

1 . 駆動手段 ( モータ ) のノイズは、広帯域であり、図 1 5 ( A ) , ( B ) に示すように、外部マイクによる信号よりも内部マイクによる信号が 10 dB 程度大きい。

10

20

30

40

50



2. 駆動手段(モータ)のノイズは、図16(A)に示すように、2.5kHz以上の周波数に関して、内部マイクより外部マイクの方がやや大きいかほぼ同等である。これは、外装14で外部音が遮断されることにより、内部マイクがより駆動手段のノイズを拾いやすいことを示している。

3. 駆動手段(モータ)のノイズは、2kHz以下の周波数では、外部マイクより内部マイクの方がやや大きく、この傾向は、図16(B)に示すように、特に700Hz以下の周波数で顕著である。これは、外装14内での共鳴を示しており、外装14の直径が約18cmであることから、周波数500Hzにおける $\lambda/4$ に対応している。また、図16(A)においても、同様の共鳴が発生している。

4. 内部音は、図15(A)、(B)を比較して、平均して外部音より約10dB大きい。従って、外装14の内部音及び外部音の分離効率は約10dBである。

このようにして、前以て外部マイクと内部マイクのスペクトルパワー差のパターン、そして共鳴によるピークを含むスペクトルの音圧及びパターンを記憶させておくことにより、駆動手段(モータ)のノイズに関する測定データを保持しておき、ノイズ除去回路23, 24が、各サブバンド毎に上述したバーストノイズの判定を行なって、バーストノイズがあると判定されたサブバンドに対応する信号部分を除去することにより、バーストノイズの影響を排除することができる。

前述したと同様の実験例を示す。

この場合、前述の実験例と同じ条件で、遅い動作(14.9度/秒)でのみ実験が行なわれ、以下のような結果が得られた。

図17は、人型ロボット10の内部で発生する内部音(ノイズ)のスペクトログラムを示している。このスペクトログラムによれば、駆動モータによるバーストノイズが明白に示されている。

ノイズ除去のない場合による方向情報は、図18に示すように、頭部13が回転している間(5~6秒の間)、ノイズにより影響を受けており、人型ロボット10が音源を追跡するために頭部13を回転駆動している間は、聴覚がほぼ無効となるようなノイズを発生していることが分かる。

また、前述した第一のバーストノイズ除去方法による方向情報は、図19に示すように、頭部13の回転駆動の間でもバーストノイズの影響による揺れがやや少なく、方向情報が比較的正確に得られる。

これに対して、上述した第二のバーストノイズ除去方法による方向情報は、図20に示すように、頭部13の回転駆動の間でもバーストノイズの影響による揺れが極めて少なく、方向情報がより一層正確に得られることが分かる。

なお、上記の実験と並行して、前述したANC法(適応フィルタとしてFIRフィルタを使用)によるノイズキャンセルも実験してみたが、バーストノイズを有効にキャンセルすることはできなかった。

上述した実施形態において、人型ロボット10は、4DOF(自由度)を有するように構成されているが、これに限らず、任意の動作を行なうように構成されたロボットに本発明によるロボット聴覚システムを組み込むことも可能である。

また、上述した実施形態においては、本発明によるロボット聴覚システムを人型ロボット10に組み込んだ場合について説明したが、これに限らず、犬型等の各種動物型ロボットや、その他の形式のロボットに組み込むことも可能であることは明らかである。

さらに、上述した実施形態においては、内部マイク17は一对のマイク17a, 17bから構成されているが、一つ以上のマイクから構成されていてもよい。また、外部マイク16は一对のマイク16a, 16bから構成されているが、二対以上のマイクから構成されていてもよい。

従来技術のANCは、位相に影響を与えるフィルタリング処理を行うことで位相のずれが生じるため、音源定位を正確に行う場合には適していなかった。これに対して、本願発明によれば、位相情報に影響を与えるフィルタリングを行わないため、即ちノイズが混入している部分のデータを使わないようにしていることで、位相のずれは生じず、それゆえ、

10

20

30

40

50

音源定位に適している。

産業上の利用可能性

以上のように、本発明によれば、駆動機構等のロボット内部で発生するノイズに影響されることなく、外部の目標からの音を集音して能動知覚を行なうことができるようにした、極めて優れたロボット聴覚装置及びロボット聴覚システムを提供できる。

【図面の簡単な説明】

本発明は、以下の詳細な説明及び本発明の実施例を示す添付図面に基づいて、より良く理解されるものとなろう。なお、添付図面に示す種々の実施例は本発明を特定又は限定することを意図するものではなく、単に本発明の説明及び理解を容易とするためだけのものである。

10

図中、

図1はこの発明によるロボット聴覚装置の第一の実施形態を組み込んだ人型ロボットの外観を示す正面図である。

図2は図1の人型ロボットの側面図である。

図3は図1の人型ロボットにおける頭部の構成を示す概略拡大図である。

図4は図1の人型ロボットにおけるロボット聴覚システムの電氣的構成を示すブロック図である。

図5は図4のロボット聴覚システムにおける要部を示すブロック図である。

図6(A)は視覚における、また、図6(B)は聴覚におけるエピポーラ幾何による方向付けを示す概略図である。

20

図7及び図8は、それぞれ音源定位と音源分離の処理を示す概念図である。

図9は図4のロボット聴覚システムの実験例を示す概略図である。

図10は図9の実験におけるロボット頭部の(A)速い動作及び(B)遅い動作における入力信号のスペクトログラムである。

図11(A)は図9の実験におけるバーストノイズを除去しない場合の速い動作における方向情報を示すグラフであり、図11(B)は遅い動作における方向情報を示すグラフである。

図12(A)は図9の実験における弱いバーストノイズを除去した場合の速い動作における方向情報を示すグラフであり、図12(B)は遅い動作における方向情報を示すグラフである。

30

図13(A)は図9の実験における強いバーストノイズを除去した場合の速い動作における方向情報を示すグラフであり、図13(B)は遅い動作における方向情報を示すグラフである。

図14(A)は図13(A)に対応するスペクトログラムを、図14(B)は図13(B)に対応するスペクトログラムで、いずれもノイズに比べて信号が強い場合を示している。

図15(A)は駆動手段のノイズの内部マイクによる周波数応答を示すグラフであり、図15(B)は外部マイクによる周波数応答を示すグラフである。

図16(A)は図15の周波数応答における駆動手段のノイズのグラフ、図16(B)は外部音のスペクトルパワー差のパターンを示すグラフである。

40

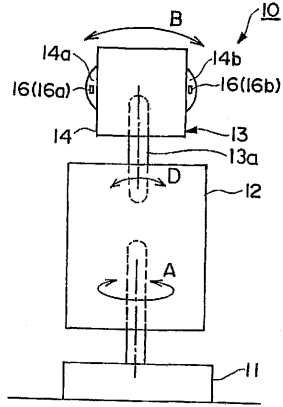
図17は、ロボット頭部の遅い動作における入力信号のスペクトログラムである。

図18はバーストノイズを除去しない場合の方向情報を示すグラフである。

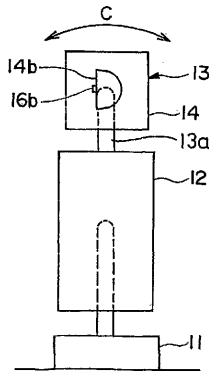
図19は図9の実験と同様の第一のバーストノイズ除去方法による方向情報を示すグラフである。

図20は第二のバーストノイズ除去方法による方向情報を示すグラフである。

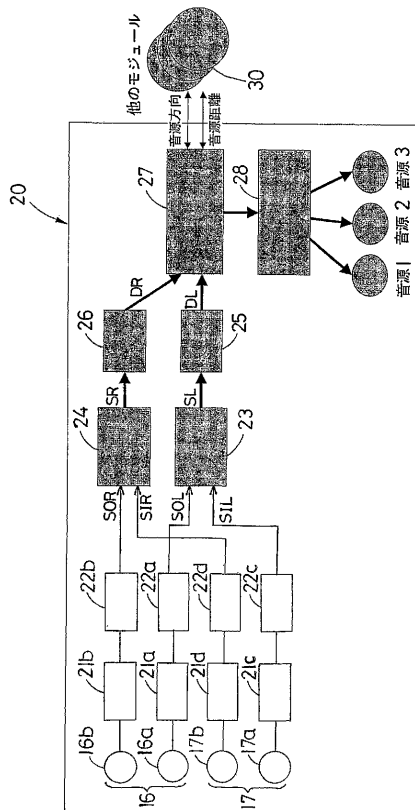
【図1】 第1図



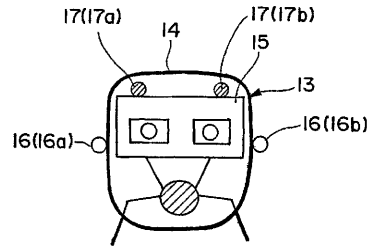
【図2】 第2図



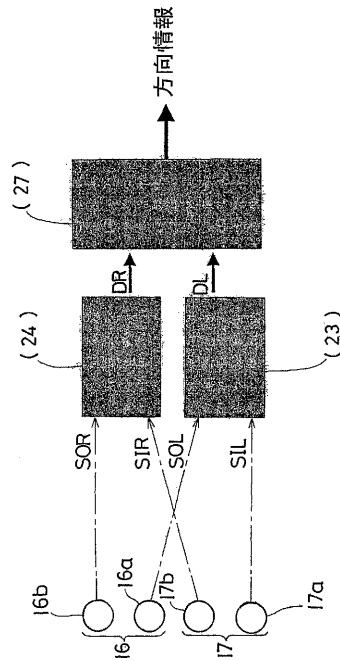
【図4】 第4図



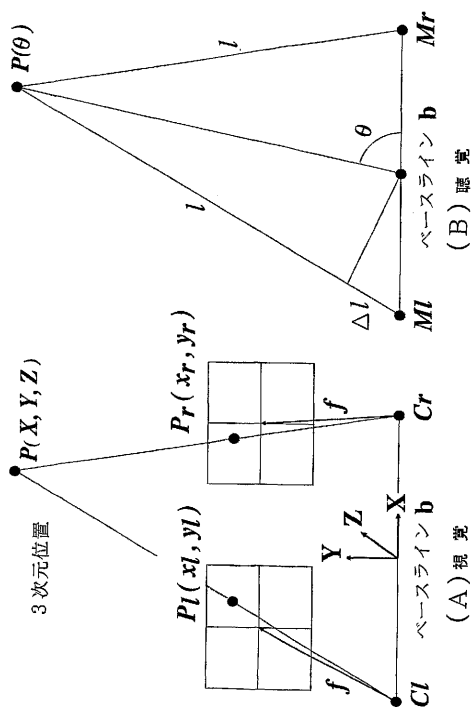
【図3】 第3図



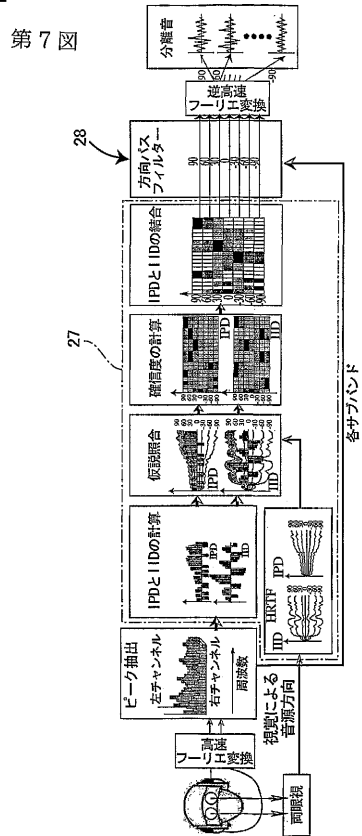
【図5】 第5図



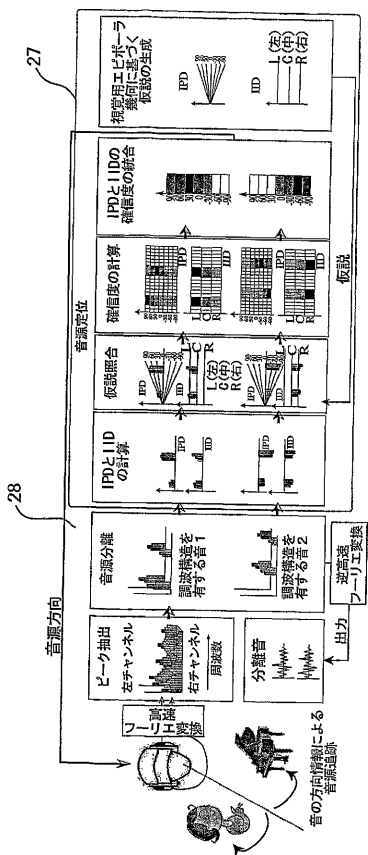
【図6】  
第6図



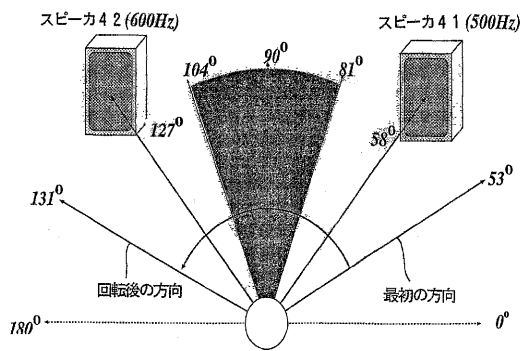
【図7】  
第7図



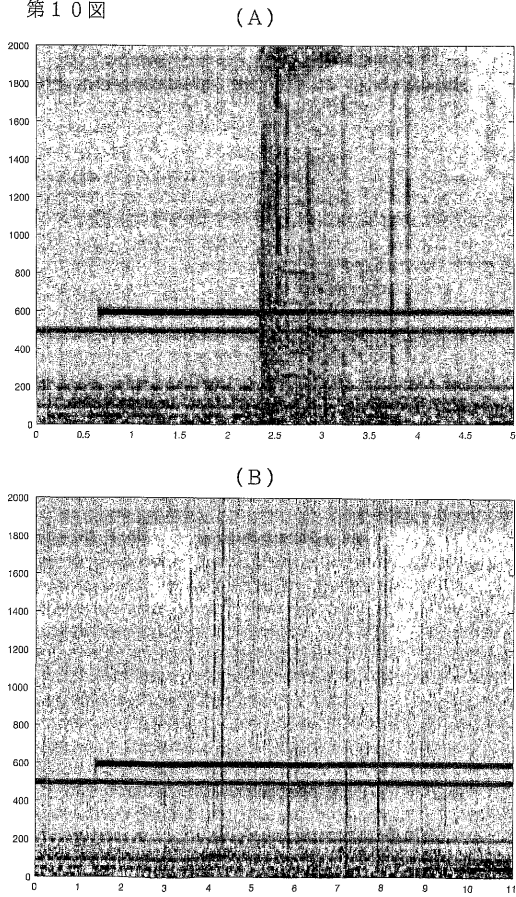
【図8】  
第8図



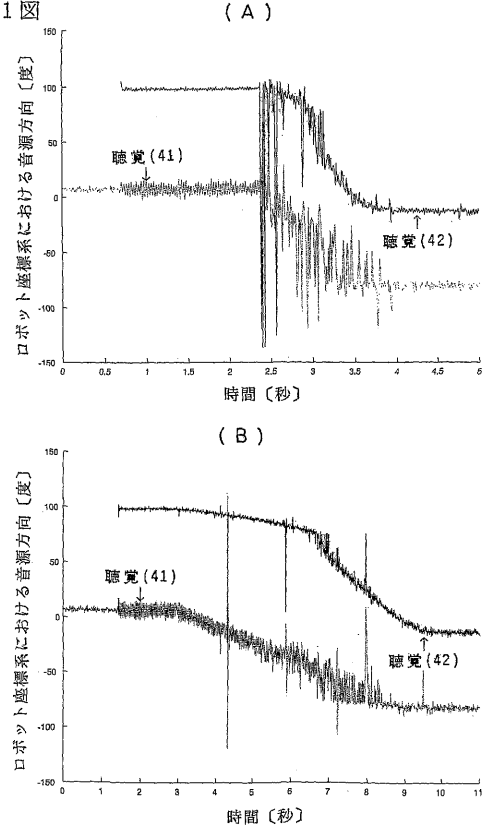
【図9】  
第9図



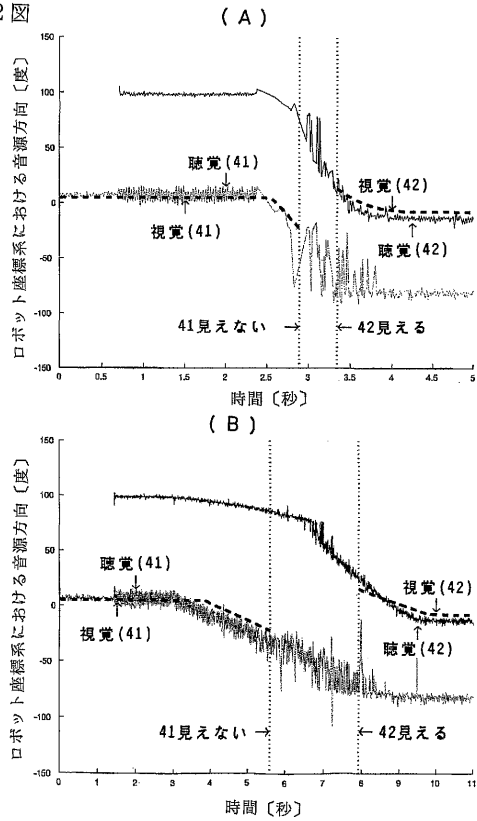
【図10】  
第10図



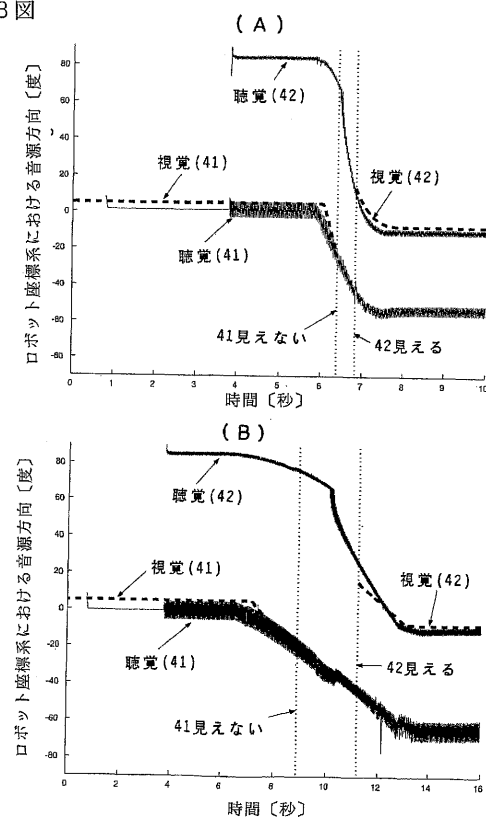
【図11】  
第11図



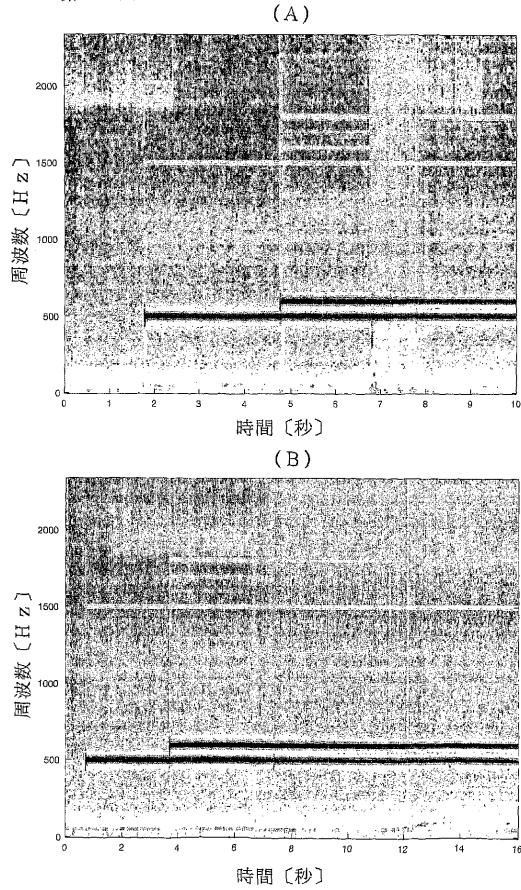
【図12】  
第12図



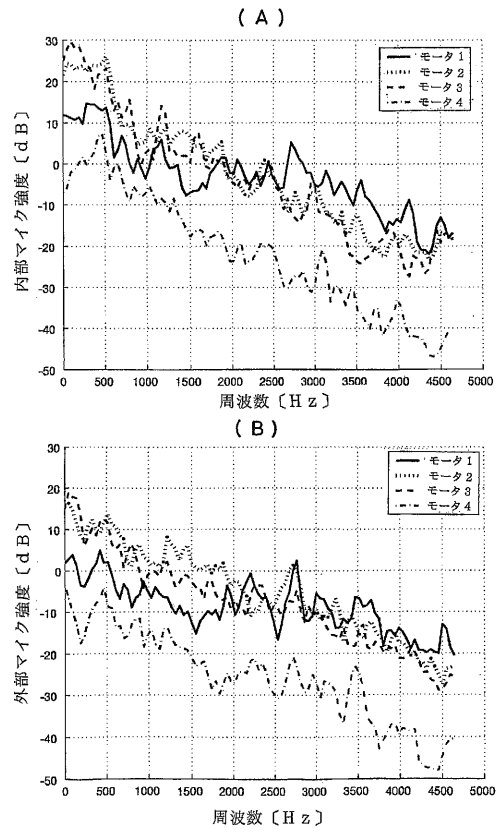
【図13】  
第13図



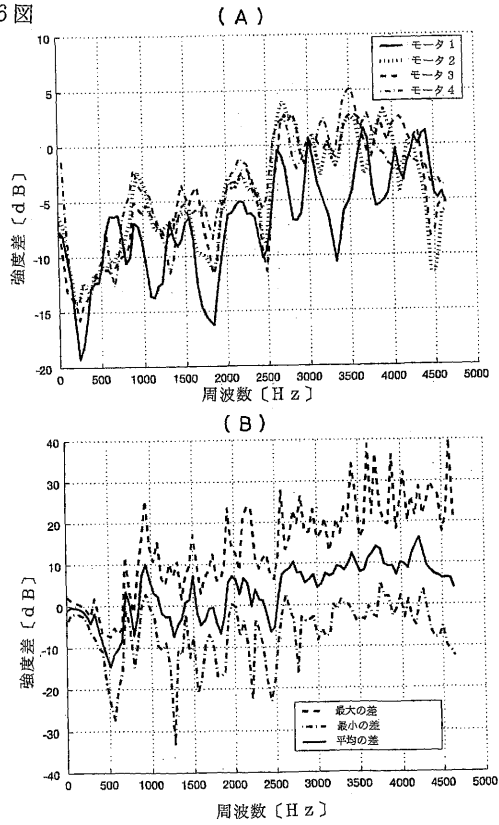
【図14】  
第14図



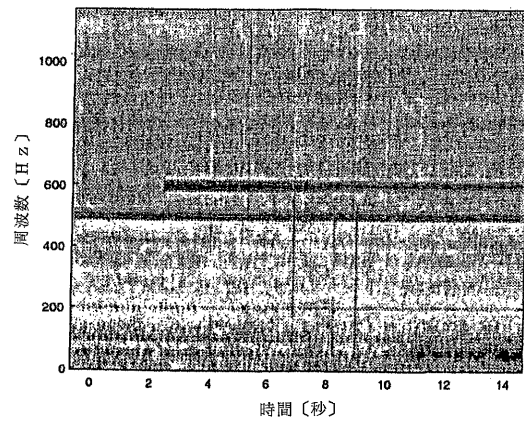
【図15】  
第15図



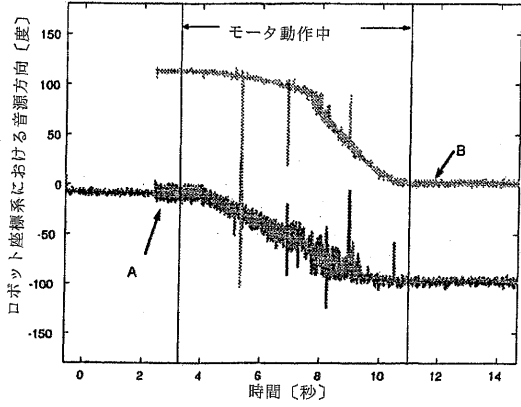
【図16】  
第16図



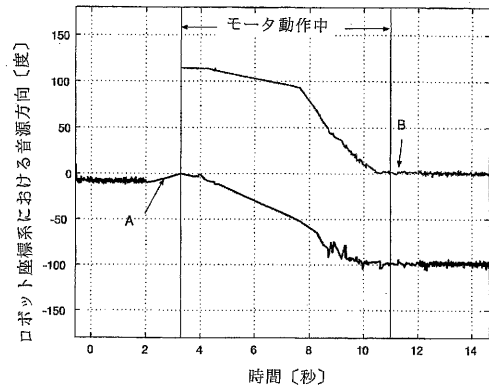
【図17】  
第17図



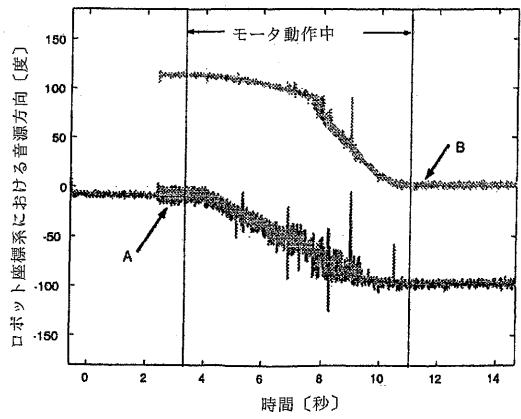
【図18】  
第18図



【図20】  
第20図



【図19】  
第19図



フロントページの続き

(58)調査した分野(Int.Cl. , DB名)

G10L 15/00-15/28、21/02

B25J 13/00