

【公報種別】特許法第 17 条の 2 の規定による補正の掲載

【部門区分】第 6 部門第 1 区分

【発行日】平成30年2月8日(2018.2.8)

【公表番号】特表2016-537622(P2016-537622A)

【公表日】平成28年12月1日(2016.12.1)

【年通号数】公開・登録公報2016-066

【出願番号】特願2016-519371(P2016-519371)

【国際特許分類】

G 0 1 S 3/808 (2006.01)

H 0 4 R 1/40 (2006.01)

【F I】

G 0 1 S 3/808

H 0 4 R 1/40 3 2 0 Z

【誤訳訂正書】

【提出日】平成29年12月6日(2017.12.6)

【誤訳訂正 1】

【訂正対象書類名】明細書

【訂正対象項目名】全文

【訂正方法】変更

【訂正の内容】

【発明の詳細な説明】

【発明の名称】音源の位置を特定する方法、および当該方法を用いる人型ロボット

【技術分野】

【0 0 0 1】

本発明は、音源を空間的に定位する方法、および当該方法を利用および実装可能な人型ロボットに関する。

【背景技術】

【0 0 0 2】

音源の空間的な定位は、人型ロボット工学を含むがこれに限定されない多くのアプリケーションで必要である。

【0 0 0 3】

ロボットは、頭部、胴体、2本の腕、場合によっては2本の脚等、人間の何らかの属性、外観および特徴を有する場合、人型ロボットと見なすことができる。一般に、人型ロボットは、人の存在に気付いたならば、会話を交わすなどして、自身の言語を含め、可能な限り「自然に」人間とコミュニケーションすることが求められる。この目的を達成するために、音源を定位する能力は極めて有用であるか、または必須である。具体的には、このような能力により、音声が発せられている方向を人型ロボットが判定して、当該方向に自身の頭を向けることができる。当該音声の人から発せられている場合、ロボットが顔認識ソフトウェアパッケージを起動し、音声認識システムを最適に構成して、当該人の運動を「自身の視線」で追跡する等ができる。

【0 0 0 4】

複数の音源の空間位置を感知する複数の方法およびシステムが従来技術において知られている。これらの方法およびシステムは一般に、無指向性であるかまたはあまり指向的でない複数のマイクロホン、および前記マイクロホンにより捕捉された信号のデジタル処理に基づいている。

【0 0 0 5】

M . S . B r a n d s t e i n および D . B . W a r d 編「M i c r o p h o n e A r r a y s : S i g n a l P r o c e s s i n g T e c h n i q u e s a n d A

pplications」(Springer-Verlag, 2001, Berlin, Germany)に掲載されたJ. DiBiase他による論文「Robust localization in reverberant rooms」に、音源を定位する3通りの主要なアプローチが記述されている。

【0006】

第1のアプローチは、マイクロホンにより捕捉された信号の相関行列に基づくスペクトル推定技術を用いる。当該アプローチに基づく方法は、モデリング誤差の影響を受け易く、膨大な計算能力を要する傾向がある。これらは主に狭帯域信号に適している。

【0007】

第2のアプローチは、マイクロホン対により受信された音声信号間の時間シフトの推定に基づいている(「到達時間差」またはTDOA技術)。これらの推定をマイクロホンの位置に関する知識と共に用いて、交点が音源の位置を与える双曲線を計算する。時間シフトは特に、フィルタリングにより以前に「白色化」された信号間の相互相関の計算を利用するPHAT-GCC(「Phase Transform-Generalized Cross-Correlation(位相変換-一般化相互相関)により推定することができる。PHAT-GCC法は、Ch. H. KnappおよびG. C. Carterによる論文「The Generalized Correlation Method for Estimation of Time Delay」, IEEE Transaction on Acoustics, Speech and Signal Processing, Vol. ASSP-24, No. 4, August 1976 pp. 320-327に更に詳述されている。これらの方法は、計算負荷は軽い、複数の音源から生じた有相関ノイズに対して堅牢ではなく、「偽陽性」を示し易い。更に、PHAT-GCC法を除いて反響に対してあまり堅牢でない。

【0008】

第3のアプローチは、各種のマイクロホンにより捕捉されて可変な時間シフトが適用された信号を追加することにより向き付け可能な音響ビームを合成して、このように受信されたコンポジット信号の出力を最大化するビームの向きを識別するものである。当該アプローチに基づく方法は、膨大な計算能力を要する特定の変型例を除いて反響およびノイズに対してあまり堅牢でない傾向がある。

【0009】

上述の論文はより具体的に、向き付け可能な音響ビームの合成および一般化相関を位相変換と組み合わせる方法を記述している。当該方法は、SRP-PHAT(「Steered Response Power-PHASE Transform(被制御応答出力-位相変換)」)と表記する。PHAT-GCC法と比べてノイズに対してより堅牢であるが反響に影響され易い。

【先行技術文献】

【非特許文献】

【0010】

【非特許文献1】J. DiBiase et al. "Robust localization in reverberant rooms" in M. S. Brandstein and D. B. Ward "Microphone Arrays: Signal Processing Techniques and Applications", Springer-Verlag, 2001, Berlin, Germany

【非特許文献2】Ch. H. KnappおよびG. C. Carter "The Generalized Correlation Method for Estimation of Time Delay", IEEE Transaction on Acoustics, Speech and Signal Processing, Vol. ASSP-24, No. 4, August 1976 pp. 320-327

【発明の概要】

【発明が解決しようとする課題】

【 0 0 1 1 】

本発明は、公知の先行技術方法に比べてノイズおよび反響に対してより良好な耐性を有すると共に、人型ロボット等のオンボードシステムにおいてリアルタイム且つ他のタスクと並行して実行する際に計算負荷が十分に軽い音源定位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】

【 0 0 1 2 】

本発明によれば、当該目的は、上述の S R P - P H A T 法から派生した方法により、且つ当該派生的方法を実行する手段を含む人型ロボットにより実現される。

【 0 0 1 3 】

本発明の一主題は従って、音源定位方法であって、

a) 少なくとも 3 個のマイクロホンの配列により、定位したい音源からの音声信号を捕捉するステップと、

b) 前記配列から少なくとも 3 対のマイクロホンを選択して、前記各対について、捕捉した音声信号の一般化相互相関を計算するステップ、すなわち前記音声信号間の両耳間時間差と呼ばれる遅延の複数の値について前記計算を実行するステップと、

c) 前記一般化相互相関から、前記マイクロホンの各対における両耳間時間差のベクトルの関数として表される被制御応答出力を計算するステップと、

d) 前記被制御応答出力を最大化する両耳間時間差ベクトルを判定するステップと、

e) ステップ d) で判定された両耳間時間差ベクトルに依存する前記音源の定位方向を推定するステップとを含み、

- 前記ステップ c) および d) が、前記マイクロホンから無限遠点にある単一の音源から発せられた音声信号と整合するベクトルの第 1 の部分集合と、前記マイクロホンから無限遠点にある単一の音源から発せられた音声信号と整合しないベクトルの第 2 の部分集合とを含む集合を形成する複数の両耳間時間差ベクトルを考慮することにより実行され、

- 前記第 1 の部分集合の各ベクトルには対応する単一音源の定位方向が関連付けられ、前記第 2 の部分集合の各ベクトルには所定計量に従い最も近い前記第 1 の部分集合のベクトルに関連付けられた定位方向が関連付けられていて、

- 前記ステップ e) で推定された方向は、前記ステップ d) で判定された両耳間時間差ベクトルに関連付けられたものである。

【 0 0 1 4 】

上述の方法の有利な特徴によれば、

- 前記ステップ b) で計算された一般化相互相関は、以前に白色化された信号の相互相関に対応する G C C - P H A T 型であってよく、

- 前記ステップ b) において、前記配列のマイクロホンの全ての対が選択されてよく、

- 前記被制御応答出力は、前記マイクロホン対についての一般化相互相関の和により与えられるか、または比例してよく、

- 前記第 1 の部分集合は、C h a s l e s 条件と呼ばれる条件

$$i j + j k = i k$$

を自身の要素が満たす前記集合のベクトルを含んでいてよく、

ここに添え字 i j k は前記配列の各種のマイクロホンを表し、 $i j$ 、 $j k$ 、 $i k$ は各々が対 (i, j) 、 (j, k) 、 (i, k) における両耳間時間差を表し、

- 前記第 2 の部分集合は、前記第 1 の部分集合に属していない前記集合のベクトルおよび当該ベクトルから前記第 1 の部分集合のベクトルまでの距離を含み、当該距離は前記所定計量に従い決定され、正規化後は所定の閾値未満であり、

- 前記所定計量はユークリッド計量であってよく、

- 前記ステップ e) は、1 個の定位方向を前記集合の各値に関連付ける参照テーブルに基づいて実行されてよい。

【 0 0 1 5 】

本発明の別の主題は、

- 少なくとも 3 個のマイクロホンの配列と、

- 前記マイクロホンにより捕捉された入力音声信号として受信すべく前記マイクロホンに接続されたプロセッサであって先行請求項のいずれか 1 項に記載の方法を実行すべくプログラミングまたは構成されたプロセッサを含む人型ロボットである。

【0016】

有利な特徴として、前記マイクロホンは、前記ロボットの頭部の上面に配置されている。

【0017】

本発明の他の特徴、詳細事項、および利点は、例示的な添付図面を参照しながら以下の記述を精査することでより明らかになる。

【図面の簡単な説明】

【0018】

【図 1】本発明の一実施形態による方法の実装を示すブロック図である。

【図 2】本発明の一実施形態による人型ロボットの頭部におけるマイクロホンの配置を示す。

【図 3 A】本発明の一実施形態による方法の性能を示すグラフである。

【図 3 B】比較として従来技術による方法の性能を示すグラフである。

【図 4 A】本発明の一実施形態による方法の性能を示すグラフである。

【図 4 B】比較として従来技術による方法の性能を示すグラフである。

【図 5 A】本発明の一実施形態による方法の性能を示すグラフである。

【図 5 B】比較として従来技術による方法の性能を示すグラフである。

【図 6 A】本発明の一実施形態による方法の性能を示すグラフである。

【図 6 B】比較として従来技術による方法の性能を示すグラフである。

【図 7 A】本発明の一実施形態による方法の性能を示すグラフである。

【図 7 B】比較として従来技術による方法の性能を示すグラフである。

【発明を実施するための形態】

【0019】

図 1 は、マイクロホン M 1 ~ M 4 およびプロセッサ P R のネットワークを含むシステムによる本発明の一実施形態による方法の実装を極めて模式的に示す。

【0020】

同図のケースでは、ネットワークは 4 個のマイクロホンを含んでいるが、この数はより少なくても（少なくとも 3 ）またはより大きくてもよく、更に、マイクロホンが整列して示されているが、任意の空間配置が可能である。好適には、マスキングのケース、すなわちマイクロホンと、定位したい音源の定位可能な空間領域との間に障害物が入るケースが最小になる配置が選択される。

【0021】

実際のシステムはまた、マイクロホンから発せられた信号を調整してこれらの増幅、フィルタリング、サンプリング、およびアナログの / デジタル変換を保証する回路も含んでいる。当該回路は、図を簡潔且つ読み易くすべく省略されている。従って以下では、マイクロホン M 1 ~ M 4 がデジタルフォーマットで信号を配信するものとする。

【0022】

プロセッサ P R は、以下の動作を行うべくプログラミングまたは構成されている。

【0023】

1. 各種のマイクロホンから発せられた信号の一般化相互相関、特に G C C - P H A T 型の一般化相互相関の計算。

【0024】

$s_i(t)$ 、 $s_j(t)$ をマイクロホン M_i 、 M_j ($i, j = 1 - 4 : i \neq j$) から発せられた信号とする。当該信号の一般化相互相関 R_{s_i, s_j} は、事前フィルタリング後の信号間の相互相関として定義する。これは時間を単位とする変数の関数であって、2 個のマイクロホンから発せられた信号間の時間シフト（両耳間時間差）を表す。有利な特

徴として、一般化相互相関は、周波数領域において計算される。

【数 1】

$$R_{si,sj}(\tau) = \int_{-\infty}^{+\infty} \Phi_i(\omega) S_i(\omega) \Phi_j^*(\omega) S_j^*(\omega) e^{i\omega\tau} d\omega \quad (1)$$

ここに、 $S_{i,j}(\omega)$ は信号 $s_{i,j}(t)$ のフーリエ変換、 $*$ は共役複素数演算子、 $\Phi_i(\omega)$ はフィルタの伝達関数である。GCC-PHAT一般化相互相関の場合、信号を「白色化する」、すなわち位相情報だけを保存すべく自身の全てのスペクトル成分の強度を等しくするフィルタが選択される。

【数 2】

$$\Phi_{i,PHAT} = \frac{1}{|S_i(\omega)|} \quad (2)$$

【0025】

無論、本方法の実用的な実施形態において、フーリエ変換は例えば高速フーリエ変換 (FFT) アルゴリズムにより計算される離散フーリエ変換であり、積分項は有限個の項の和で代替される。

【0026】

従って図 1 に示すように、マイクロホン $M_1 \sim M_4$ から発せられた信号は周波数領域 (例えば高速フーリエ変換すなわち「FFT」アルゴリズムを用いて離散フーリエ変換を実行する FT ブロック) に変換され、主にノイズを含むまたはノイズだけを含むスペクトル成分が白色化フィルタにより増幅されることを防止 (スペクトル閾値化ブロック SCSS) すべく、強度が所定の値未満であるスペクトル成分を抑制するために閾値化され、次いで前記白色化フィルタ PHAT によりフィルタリングされるが、他の種類のフィルタを用いてもよく、Ch. H. Knapp および G. C. Carter による上述の論文を参照されたい。次に、フィルタリングされた信号は対毎に乗算され、それらの積は時間領域に再変換される (FT⁻¹ ブロックは、特に高速フーリエ変換すなわち「FFT」アルゴリズムを用いて逆離散フーリエ変換を実行する)。従って、各種のマイクロホンにより捕捉された音声信号の一般化相互相関が得られ、対毎に考慮される。マイクロホンの全ての対 (N をマイクロホンの個数とすれば対の数は $N(N-1)/2$) からの信号を扱うように選択されているが、選択する対の数を 3 個に限定できるが、 $3 \sim N(N-1)/2$ の間の任意の個数の対を選択してもよい。

【0027】

周波数領域への、および周波数領域からの変換は相互相関の計算に必須ではないが、極めて有利である。

【0028】

計算は別々に行われるため、各々の一般化相互相関は両耳間時間差の値の離散的集合についてのみ計算される。

【0029】

2. 被制御応答出力 (SRP) の計算。当該計算は、単に各種の一般化相互相関を加算することにより行われる。

$$SRP(\omega) = \sum_{i,j} s_i s_j R_{s_i, s_j}(\omega) \quad (3)$$

ここに、和は、マイクロホンの各対に関連付けられた信号 s_i, s_j の M 個の対に対して計算される ($3 \leq M \leq N(N-1)/2$)。ベクトル変数 ω は、M 個の成分を有し、その各々が 1 個の前記対における両耳間時間差に対応する。より一般的には、SRP は一般化相互相関の一次結合として定義できる。

【0030】

全ての値、従って値 s_i, s_j の全ての組合せは「物理的に可能」ではない。具体的には、マイクロホンに到達する音波が平面であると考えられる程度にマイクロホンのネットワークから十分に離れた単一の音源を考慮する場合、前記音源の定位方向は 2 個の

両耳間時間差により完全に識別される。換言すれば、ベクトル の 2 個の成分は、他の成分の値を明確に定義する（少なくとも理論上）。

【 0 0 3 1 】

3. 被制御応答出力の最大化

音源の定位方向は、ベクトル変数 に対して関数 $S R P ()$ を最大化することにより識別される。当該最適化課題を解決するために多くの数値方法を用いてよい。勾配アルゴリズムおよび「総当たり」法の非限定的な例により行うことができる。

【 0 0 3 2 】

この最大化は、上で述べた意味で「物理的に可能な」、すなわち「無限」の距離にある単一の音源から発せられた音声信号と整合する の値だけを考慮することにより、制約の下で実行できる。このアプローチは特に、J. H. D i B i a s e 他による上述の論文から公知であるが、実際にはあまり堅牢でないことが分かっている。具体的には、ノイズ、マスキング効果（障害物が音源と 1 個以上のマイクロホンとの間に入り込む）、および特に反響効果に起因して、 $S R P ()$ の最大絶対値が の「不可能な」値に対応することが多い。

【 0 0 3 3 】

本発明によれば、音源の空間的な定位を見つけるアルゴリズムの堅牢性を向上させるべく、 $S P R ()$ の最大値の探索は、 の「可能な」値だけでなく、理論的に受容できない値、すなわち「無限の」距離にある信号源から発せられた音声信号と整合しない値に対しても実行される。より正確には、図 1 に示すように、関数 $S R P ()$ は、2 個の部分集合からなる集合 E を形成するベクトル の値を考慮することにより最大化される（関数ブロック「M A X」）。

- マイクロホンのネットワークから無限遠点にある単一の音源から発せられた音声信号と整合するベクトル の第 1 の部分集合 E 1。数学的観点から、これらのベクトルは「C h a s l e s」条件と呼ばれる条件を満たす成分を有している。

$$i j + j k = i k \quad (4)$$

ここに添え字 $i j k$ は、前記配列の各種マイクロホンを表し、 $i j$ 、 $j k$ 、および $i k$ は各々が対 (i, j) 、 (j, k) 、および (i, k) における両耳間時間差を表している。これらの条件は、ベクトル が無限遠点にある単一の音源から発せられた音声信号と整合するための必要条件であるが、十分条件ではない。

- マイクロホンのネットワークから無限遠点にある単一の音源から発せられた音声信号と整合しないベクトル の第 2 の部分集合 E 2。特に、当該第 2 の部分集合は、前記第 1 の部分集合に属していない前記集合のベクトルおよび当該ベクトルから前記第 1 の部分集合のベクトルまでの距離を含み、当該距離は所定の（特にユークリッド）計量に従い決定され、正規化（各ベクトルをノルムで除算）後は所定の閾値未満である。当該閾値は本方法を調整するものであって、実験的に決定されてよい。

【 0 0 3 4 】

以下の手順を用いて集合 E に属するベクトル を選択して部分集合 E 1 および E 2 に振り分ける。

【 0 0 3 5 】

最初に、音源の方向を指す単位ベクトルすなわち x と表記するベクトルを両耳間時間差のベクトルに関連付けるシステムを行列形式で記述することを推奨する。

$$= A x \quad (5)$$

ここに A は $M \times 3$ 行列である（M はベクトル の成分の個数であることを想起されたい）。行列 A は正方行列でないため、逆行列を直接求めることはできない。マイクロホンが同一平面上にない場合、行列 A の階数は 3 である。従って、3 個の線形独立な行を選択して逆行列演算可能な正方行列

【数 3】

\tilde{A}

を構築することが可能である。行列 A の前記独立成分に対応する の 3 個の成分を取ることにより得られる三次元ベクトルを

【数 4】

$\tilde{\tau}$

と表記する。従って、

【数 5】

$$\tilde{\tau} = \tilde{A}x \quad (6)$$

従って、

【数 6】

$$x = \tilde{A}^{-1}\tilde{\tau} \quad (7)$$

となる。

【0036】

次に、(またはその各成分のモジュラスに受容可能な最大値が与えられるため)

【数 7】

\mathbb{R}^M

内の球体または立方体に対応する、ベクトル の離散的な起点集合 E ' (これらの成分の離散的な値しか考慮しないため) について考える。

【0037】

E ' の各ベクトル に対して、三次元 C h a s l e s 関係に関するベクトルの部分空間への射影

【数 8】

$\bar{\tau}$

を計算する。差

【数 9】

$$d_1 = |\tau - \bar{\tau}|$$

が所定の閾値 を上回る場合、当該ベクトルは廃棄される。さもなければ、

【数 10】

$\bar{\tau}$

からベクトル

【数 11】

$\tilde{\tau}$

(すなわち、3 個の線形独立な成分のベクトル) を抽出し、

【数 12】

$$d_2 = \left| \frac{\sqrt{\|\tau\|^2 - d_1^2}}{\|x\|} - 1 \right| \quad (8)$$

で与えるように距離 d_2 を計算する。

【0038】

式 (8) は、ベクトル

【数 1 3】

 $\tilde{\tau}$

を方向を変更せずに値 d_2 だけ短くすることにより、

【数 1 4】

 $\tilde{\tau}$

がユニタリノルムを有することを示す。

【0 0 3 9】

【数 1 5】

$$d = \sqrt{d_1^2 + d_2^2}$$

が閾値 未満である場合、ベクトル は従って集合 E の一部 ($d = 0$ ならば部分集合 E 1 の、さもなければ部分集合 E 2 の、但しこの違いは意図的ではなく、集合 E の全てのベクトルが同様に扱われる点に注意されたい) を形成する。

【0 0 4 0】

式 7 により、前記ベクトルの各々を音源の定位方向に関連付けることができる。

【0 0 4 1】

これらの計算は極めて高負荷であるため、一度だけ実行して結果を参照テーブル (下記参照) に保存することが有利である。

【0 0 4 2】

4. 音源の定位方向の識別

S R P () を最大化するベクトル E ($m_{a x}$ と表記) を識別した後、これを音源の定位方向に関連付けることが必要である。 $m_{a x}$ E 1 である場合、両耳間時間差のベクトルは唯一の定位方向と整合するため、これは問題ではない。 $m_{a x}$ E 2 である場合、前記所定計量に従う $m_{a x}$ に最も近い E 1 に属するベクトルと整合する定位方向が選択される。

【0 0 4 3】

有利な特徴として、集合 E の各ベクトルは 1 回だけ定位方向に関連付けられる。集合の各ベクトルおよび関連付けられた定位方向はプロセッサのメモリにロードされたファイルに保存されて上述のように構成された参照テーブルを形成する。従って、音源の定位方向は、単にテーブルから読み出すことにより $m_{a x}$ の値から決定される。

【0 0 4 4】

有利な特徴として、参照テーブルを含むファイルは、以下のように編成されている (6 個の成分を有するベクトル を考慮する) 。

【0 0 4 5】

【表 1】

添え字	ベクトル τ	(方位角、仰角)	[隣接]
0	(0,6,6,6,0)	(0°, 30°)	{2,3}
1	(0,0,0,0,0)	(0°, 90°)	{}
2	(1,6,6,6,0)	(3°, 28°)	{0,3}
3	(1,7,6,6,0)	(5°, 27°)	{0,2}

【0 0 4 6】

ファイルには添え字が明示的に記述されておらず、順序通りである。「隣接」欄は、偏差が 1 以下 (または他の所定の値以下) であるベクトルの添え字を含んでいる。当該欄は、S R P を最大化する勾配アルゴリズムの実行を容易にすべく有用である。

【0 0 4 7】

プロセッサが起動したならば、参照テーブルを含むファイルの内容をメモリにロードす

る。音声を検知される都度、参照テーブルに保存されている全てのベクトル について S R P () 値を計算し、次いでこれらの値の最大値を求め、対応するベクトル を識別して、当該ベクトルに関連付けられた方位角および仰角の値を参照テーブルから読み出す。一変型例として、プロセッサは、複数定位 (多数の音源の同時定位) を実行すべく S R P () の $N > 1$ 個の最も高い局所最大値を探索することができるが、当該アプローチはあまり堅牢でないことが分かっている。

【 0 0 4 8 】

図 1 を用いて動作を示したプロセッサ P R は、集合 E および参照テーブル L U T を保存すべく 1 個以上のメモリに接続され且つ適切にプログラムされたマイクロプロセッサを含んでいてよい。当該マイクロプロセッサは、音源の定位専用であっても、または他の目的を達成するものであってもよい。必要に応じて、ポータブルまたはデスクトップコンピュータあるいは極めて単純なロボットの唯一のプロセッサであってもよい。同様に、集合 E および参照テーブル L U T は、専用のメモリ装置または中央メモリに保存されていてもよい。プロセッサはまた、任意選択的にプログラム可能であり得る専用の論理回路を含んでいてよい。

【 0 0 4 9 】

図 2 に、本発明の実施に適した人型ロボットの頭部 T R H を示し、前記頭部の上面に配置された 4 個のマイクロホン $M_1 \sim M_4$ のネットワークを含んでいる。当該構成により、音源がロボットの頭部よりも高い位置にある場合にマスキング効果を回避できる。これは特に、大抵の人間よりもサイズが小さい、例えば 5 0 ~ 1 5 0 c m の人型ロボット側で人間話者の検知および定位を行う場合に推奨される。2 個の隣接するマイクロホン間の距離は、例示的に 3 ~ 3 0 c m の間である。

【 0 0 5 0 】

本発明の方法は、図 2 によれば、出願人企業が製作した「N a o」ロボットにマイクロホンを備えることにより試験された。ロボットおよび音源は通常の部屋に配置され、ロボットの前方、右側、左側、前方右側 (前後軸に対して - 4 5 ° の角度) および前方左側 (前後軸に対して + 4 5 ° の角度) に音源が配置された状態で「正常な」反響を示した。音源は、音声信号を発する音声再生設備であった。定位は、1 0 2 4 個のサンプルの計算ウィンドウを各々考慮することにより行われた。複数回の試行を反復して、ロボットにより判定された定位方向 (方位角および迎角により識別された) を集計してヒストグラムを形成した。

【 0 0 5 1 】

図 3 A / 3 B は、ロボットの前方に定位された音源に対応している (理論方位角 : 0 °) 。

【 0 0 5 2 】

図 4 A / 4 B は、ロボットの左側に定位された音源に対応している (名目方位角 : 9 0 °) 。

【 0 0 5 3 】

図 5 A / 5 B は、ロボットの前方左側に定位された音源に対応している (名目方位角 : 4 5 °) 。

【 0 0 5 4 】

図 6 A / 6 B は、ロボットの右側に定位された音源に対応している (名目方位角 : - 9 0 °) 。

【 0 0 5 5 】

図 7 A / 7 B は、ロボットの前方右側に定位された音源に対応している (名目方位角 : - 4 5 °) 。

【 0 0 5 6 】

名目仰角は測定されなかった。主に関心対象であるロボットアプリケーションでは特に、仰角は方位角ほど重要でない。

【 0 0 5 7 】

図「A」は上述したような本発明の方法に関する。図「B」は従来のTDOA法により得られたものであり、比較として与える。当該参照方法において、最大正規化SRP値が所定の閾値未満である場合、定位は失敗したものと考えられる。より一般的には、本発明による方法においても、当該正規化最大値は、定位の信頼度のインジケータと見なすことができる。正規化は次式で与えられ、

【数 16】

$$\lambda = \sum_{s_i s_{ij}} \sqrt{W(s_i)W(s_j)}$$

ここに $W(s_i)$ 、 $W(s_j)$ はマイクロホン i 、 j から発せられた白色化信号のエネルギーである。

【0058】

本発明の場合（図「A」）、大多数の試行が音源の方位角の数度以内の満足すべき推定に至ったのに対し、参照方法（図「B」）は失敗率が極めて高いことを示す点に注意されたい。基本的に関心対象であるロボットアプリケーションは高い精度を要求せず（数度の定位誤差は影響しない）、むしろ高い堅牢性と相対的な計算の簡便性を合わせて要求することを強調すべきである。

【誤訳訂正2】

【訂正対象書類名】特許請求の範囲

【訂正対象項目名】全文

【訂正方法】変更

【訂正の内容】

【特許請求の範囲】

【請求項1】

音源定位方法であって、

a) 少なくとも3個のマイクロホン（M1、M2、M3、M4）の配列により、定位したい音源からの音声信号を捕捉するステップと、

b) 前記配列から少なくとも3対のマイクロホンを選択して、前記各対について、捕捉した音声信号の一般化相互相関を計算するステップ、すなわち前記音声信号間の両耳間時間差と呼ばれる遅延の複数の値について前記計算を実行するステップと、

c) 前記マイクロホンの各対における両耳間時間差のベクトルの関数として表される前記一般化相互相関から被制御応答出力を計算するステップと、

d) 前記被制御応答出力を最大化する両耳間時間差ベクトルを判定するステップと、

e) ステップd) で判定された両耳間時間差ベクトルに依存する前記音源の定位方向を推定するステップとを含み、

- 前記ステップc) およびd) が、前記マイクロホンから無限遠点にある単一の音源から発せられた音声信号と整合するベクトルの第1の部分集合（E1）と、前記マイクロホンから無限遠点にある単一の音源から発せられた音声信号と整合しないベクトルの第2の部分集合（E2）とを含む集合（E）を形成する複数の両耳間時間差ベクトルを考慮することにより実行され、

- 前記ステップe) で推定された方向は、前記ステップd) で判定された両耳間時間差ベクトルが前記第1の部分集合に含まれる場合、対応する前記単一音源の定位方向に関連付けられ、前記第2の部分集合に含まれる場合、所定計量に従い最も近い前記第1の部分集合のベクトルに関連付けられた定位方向に関連付けられたものであることを特徴とする方法。

【請求項2】

前記ステップb) において計算された一般化相互相関が、以前に白色化された信号の相互相関に対応するGCC-PHAT型である、請求項1に記載の方法。

【請求項3】

前記ステップ b) において、前記配列の前記マイクロホンの全ての対が選択されている、請求項 1 ~ 2 のいずれか 1 項に記載の方法。

【請求項 4】

前記被制御応答出力が、前記マイクロホン対についての一般化相互相関の和により与えられるか、または比例している、請求項 1 ~ 3 のいずれか 1 項に記載の方法。

【請求項 5】

前記第 1 の部分集合が、C h a s l e s 条件と呼ばれる条件

$$i_j + j_k = i_k$$

を自身の要素が満たす前記集合のベクトルを含み、

ここに添え字 i j k は前記配列の各種のマイクロホンを表し、 i_j 、 j_k 、 i_k は各々が対 (i, j) 、 (j, k) 、 (i, k) における両耳間時間差を表す、請求項 1 ~ 4 のいずれか 1 項に記載の方法。

【請求項 6】

前記第 2 の部分集合が、前記第 1 の部分集合に属していない前記集合のベクトルおよび当該ベクトルから前記第 1 の部分集合のベクトルまでの距離を含み、当該距離は前記所定計量に従い決定され、正規化後は所定の閾値未満にある、請求項 1 ~ 5 のいずれか 1 項に記載の方法。

【請求項 7】

前記所定計量がユークリッド計量である、請求項 1 ~ 6 のいずれか 1 項に記載の方法。

【請求項 8】

前記ステップ e) が、1 個の定位方向を前記集合の各値に関連付ける参照テーブル (L U T) に基づいて実行される、請求項 1 ~ 7 のいずれか 1 項に記載の方法。

【請求項 9】

- 少なくとも 3 個のマイクロホンの配列 (M 1、M 2、M 3、M 4) と、
- 前記マイクロホンにより捕捉された入力音声信号として受信すべく前記マイクロホンに接続されたプロセッサ (P R) であって請求項 1 ~ 8のいずれか 1 項に記載の方法を実行すべくプログラミングまたは構成されたプロセッサを含む人型ロボット。

【請求項 10】

前記マイクロホンが、前記ロボットの頭部 (T R H) の上面に配置されている、請求項 9 に記載の人型ロボット。