

反射板を用いた2本のマイクによる2音源の2次元位置推定

中島 弘道*・河本 満**

伊藤 雅紀***・向井 利春*

A Two-dimensional Position Estimate of Two Sound Sources Using Two Microphones with Reflectors

Hiromichi NAKASHIMA*, Mitsuru KAWAMOTO**,
Masanori ITO*** and Toshiharu MUKAI*

Human beings and living things have the capability of identifying the directions of two or more sounds by a certain amount of correctness with only two ears. However it is difficult to give this capability to robots. Almost all the robots which have been proposed until now have three or more microphones in order to localize sound sources. In this paper, we propose a technique of estimating two kinds of directions, that is, vertical and horizontal directions, using a robot head consisted of two microphones, where the microphones of the robot head have reflectors working like the pinna.

Key Words: sound source separation, sound source localization, two or more sound sources, robot audition

1. ま え が き

われわれ人間は、五感の中でも主に視聴覚のセンサを用いている。よって、人間と接するロボットの開発において、これら2つのセンサは非常に重要であると考えられる。人間のもつすぐれた認識能力をロボットで実現するためにさまざまな研究がなされているが、人とロボットの自然なコミュニケーションを実現するための重要な機能であるロボットの聴覚機能の研究は視覚の研究に比べると少なく、ロボットにおける聴覚機能はほとんど有効に使われていないのが現状である。しかし、実際の生活環境においてロボットを使用する場合に、聴覚でなければならないこと、聴覚を使わなければならないことが数多く存在し、人間に近い機能を有する聴覚機能があれば、ロボットにもっと多様なタスクを行なわせることが可能になると考えられる。

人とロボットのコミュニケーションにおいて、人間からの呼びかけに反応して話しかけられた方向を向く音源定位機能は重要な役割をはたす。ロボットが人と会話する場合、ロボッ

トは話者のほうを向いて会話をするほうが望ましく、話者の方向の検出が必要である。また、実際にロボットが利用される生活環境においては、マイクで検出される音は複数の音源からの音が混ざり合った混合音となることが頻繁に起こると考えられる。そのような環境においてロボットを使用する場合には、ロボットはこのような混合音を聞き分け音源方向を見つける必要がある。

安藤らによって開発された SmartHead¹⁾は4本のマイクと2台のカメラを用いて低レベルな情報を抽出し、それらを統合することによって複数音源の定位を可能としている。また、中臺らは、アクティブな動作と視聴覚を統合することにより、2本のマイクで水平面上にある複数音源を定位・分離し、さらにこの分離フィルタをフロントエンド処理として利用した三話者同時発話の認識を実現している²⁾。さらに東芝の聞き分けロボットは6つのマイクで全周囲から音声を取り込み、人から話しかけられた方向と内容を認識出来る。これにより、複数の人による全方向からの呼び掛けに対して、それぞれに回答することが可能である。2本のマイクで上下左右方向の1つの音源に対して定位が可能なモデルには、小野ら³⁾の研究がある。このモデルでは、メンブクロウの耳の機能を模倣して、左右非対称に指向性マイクを配置している。左右対称の場合には、正中面内の音に対して、両耳間到達時間差 (Inter-aural Time Difference, 以下 ITD と略記) と両耳間音圧差 (Inter-aural Level Difference, 以下 ILD と略記) 共に常にゼロとなるが、非対称の場合には、正中面内の音に対しても、ITD および ILD が常にゼロとはならず、ITD と ILD の組合せによって方向を特定することが可能である。

人間や多くの生物は複数の音源を2つの耳のみで分離し、

* 理化学研究所ロボット感覚情報研究チーム
名古屋市守山区大字下志段味字穴ヶ洞 2271-130

** 産業技術総合研究所 つくば市梅園 1-1-1

*** 東北学院大学教養学部 仙台市泉区天神沢 2-1-1

* RIKEN BMC Research Center, 2271-130 Anagahora,
Shimoshidami, Moriyama-ku, Nagoya

** Advanced Industrial Science and Technology, 1-1-1
Umezono, Tsukuba

*** Faculty of Liberal Arts, Tohoku Gakuin University, 2-1-1
Tenjinzawa, Izumi-ku, Sendai

(Received August 29, 2008)

(Revised May 28, 2009)

その音源方向を見つけることができる。本研究では、このような処理を工学的に実現する手法を提案する。提案手法を用いることにより、2音源からの音を分離し、それぞれの音の上下左右の方向を推定することが可能となる。

2. 提案する2音源の定位法

通常、人の話す音声にはたくさんの無音部分が存在する。よって、複数の人間が同時に話していたとしても短い時間内であれば話をしている人間が一人だけとなる区間がいくらか存在すると考えられる。そこで、音声の中で1音源とみなせるような、一方の音が出ていないまたは音量が小さい区間を見つけ、その区間のみを注目すれば1音源での音源定位の処理方法を用いることが可能であると考えられる。具体的な処理の流れを以下に示す。

- (1) 2本のマイクで取得した音声の分離を行なう。
- (2) 分離後の音声信号を有音区間と無音区間に分ける。
- (3) 上の結果から音源が1つのみとなる区間を求める。
- (4) 分離前の音声信号から音源が1つのみの部分を抽出する。
- (5) 抽出した部分の音声信号を用いて定位を行なう。

2.1 音源分離

音源分離には、時間領域におけるブラインド分離法を用いる⁴⁾。本研究では、Fig. 1に示す入力信号が $x_j(t)$ (マイクによる観測信号)、出力信号が $y_i(t)$ の分離フィルタ $W_{ij}(z) = \sum_{k=0}^M w_{ij}(k)z^{-k}$ を用いて、もとの信号を分離して取り出すことを考える。このとき、出力信号 $y_i(t)$ は次式で与えられる。

$$y_i(t) = x_i(t-L) + \sum_{j=1}^J \sum_{k=0}^M w_{ij}(k)x_j(t-k), \quad (1)$$

$i = 1, 2, \dots, I$ 。ここで I はマイクの本数、 J は音源数である。Fig. 1では、 $J = 2, I = 2$ の場合を示している。また、 M はフィルタのタップ数を表わし、係数 L は、 $0 < L < M$ を満足する正の定数である。分離は、出力信号 $y_i(t)$ を使って、 $W_{ij}(z)$ の係数 $w_{ij}(k)$ を以下の式で更新することによって実現する。

$$w_{ij}^{new}(k) = w_{ij}^{old}(k) - \alpha \frac{y_i(t-L)y_j(t-k)}{\phi_i(t)}, \quad (2)$$

ただし、 $w_{ij}^{old}(k)$ は更新前、 $w_{ij}^{new}(k)$ は更新後の $w_{ij}(k)$ の値を表わしている。 α は学習係数で1より小さい正定数である。 $\phi_i(t)$ は、以下の式で求まる $y_i(t-L)^2$ の移動平均値である。

$$\phi_i(t) = \gamma\phi_i(t-1) + (1-\gamma)y_i(t-L)^2 \quad (3)$$

ただし、 γ は、1より小さい正数である(3章の実験では、 $\gamma = 0.9$ とした)。(2)式を繰り返し計算することによって、分離は実現されるが、詳しい分離の原理に関しては、文献4)を参照のこと。分離後、さらに、分離の精度を上げるために、分離した信号 $y_i(t)$ に対して、バイナリマスク⁵⁾をかける。以上の処理によって、得られた信号を本研究では分離した信号とし

て取り扱う。

2.2 単一音源部分の抽出法

N 個の音源信号をそれぞれ $s_1(t), \dots, s_N(t)$ 、左右のマイクによる観測信号を $x_l(t), x_r(t)$ 、分離した結果得られた N 個の信号を $y_1(t), \dots, y_N(t)$ とする。これらの分離信号のそれぞれに対して、一定時間(T)以上連続して閾値(I_1, \dots, I_N)未満となる部分を無音区間、それ以外の部分を有音区間とする。Fig. 2に2音源の場合の分離結果を示す。図中で網掛けとなっている区間が無音区間と判断された領域である。ここで、有音区間と無音区間を表わす関数 $b_1(t), \dots, b_N(t)$ は以下のように定義される。

$$b_i(t) = \begin{cases} 0, & t \text{ が無音区間} \\ 1, & t \text{ が有音区間} \end{cases} \quad i = 1, \dots, N \quad (4)$$

この関数から、音源が1つのみとなる区間は以下のように表わせる。

$$\begin{cases} m^1(t) = b_1(t)\bar{b}_2(t)\bar{b}_3(t)\cdots\bar{b}_N(t) \\ m^2(t) = \bar{b}_1(t)b_2(t)\bar{b}_3(t)\cdots\bar{b}_N(t) \\ \vdots \\ m^N(t) = \bar{b}_1(t)\bar{b}_2(t)\bar{b}_3(t)\cdots b_N(t) \end{cases} \quad (5)$$

ここで、 $\bar{b}_i(t)$ は $b_i(t)$ の値が0のとき1、1のときに0に反転させる関数と定義する。すなわち、 $\bar{b}_i(t)$ は音がある部分を表わす。この関数 $m^i(t)$ は、分離結果 $y_i(t)$ の音源のみが存在する時間 t において値が1となる関数である。この関数をFig. 2の分離結果に対して作成した結果がFig. 3である。図中の上の図が分離結果 $y_1(t)$ の音のみが出ている区間 $m^1(t)$ であり、下の図が分離結果 $y_2(t)$ の音のみが出ている区間 $m^2(t)$ を表わしている。この関数 $m^i(t)$ をマスクとして用いると、観測信号 $x_l(t), x_r(t)$ にマスクを掛けることにより、 i 番目の分離音のみとなる部分が抽出される。

$$\begin{cases} x_l^i(t) = m^i(t)x_l(t) \\ x_r^i(t) = m^i(t)x_r(t) \end{cases} \quad (6)$$

実際の抽出結果をFig. 4, 5に示す。この抽出された部分の信号を用いて音源定位の処理を行なう。

2.3 単一音源の定位法

音源定位とは、音源が発する音からその音源物体の位置を判定することである。人間や多くの生物は、音の方向のある程度の正確さで識別することができる。この識別には、ITD、

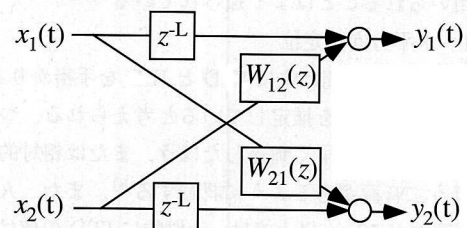


Fig. 1 Separation filter

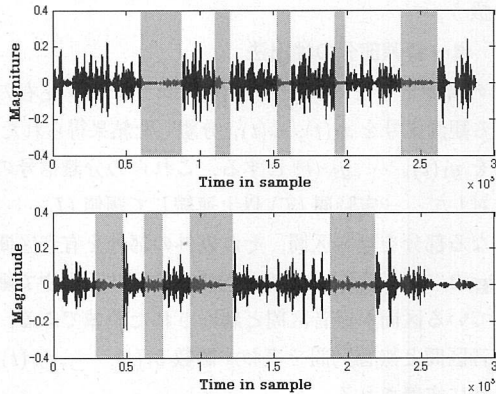


Fig. 2 silent section of separation result

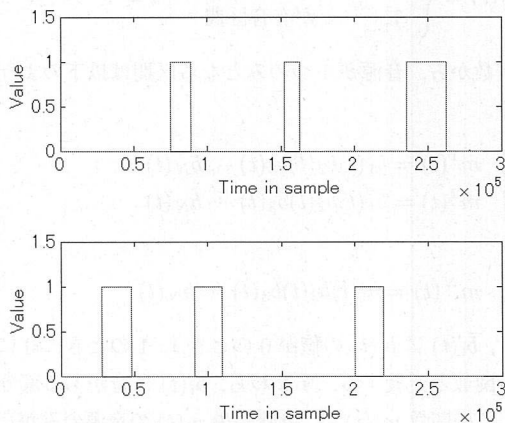


Fig. 3 One sound source section

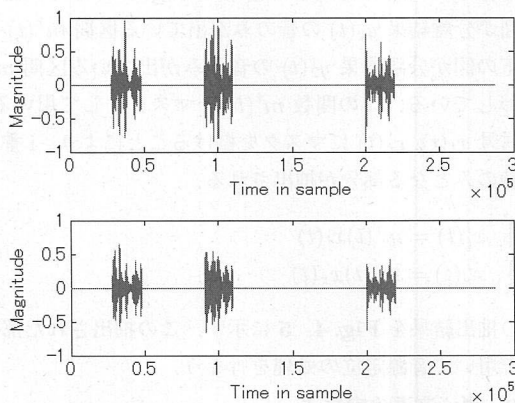


Fig. 4 Extract only the sound of the 1st sound source ($x_1^1(t), x_2^1(t)$)

ILD, 耳介によるスペクトルパターンの変化などが手掛かりとして用いられることはよく知られている^{6),7)}.

2.3.1 水平方向の定位

心理実験から、人間は主に ITD と ILD を手掛かりとして、水平面内の音源方向を推定していると考えられる。つまり人間は、相対的に音が早く到達したほう、または相対的に音が大きいほうに音源があるように認識する⁸⁾。また、人間の場合、音源距離が 50cm 以上では、近似的に ITD の値は音源方向に対して、ほぼ一定となる⁹⁾。人間では、時間差も音圧差

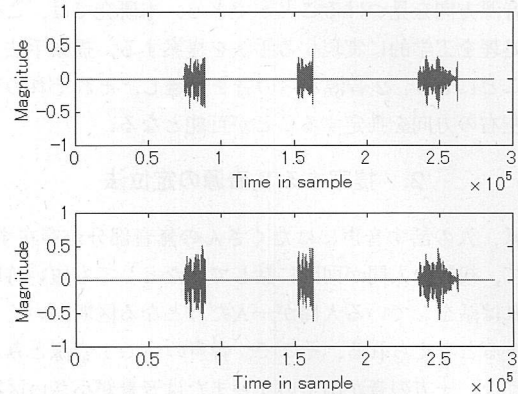


Fig. 5 Extract only the sound of the 2nd sound source ($x_1^2(t), x_2^2(t)$)

も、音源の左右の位置を識別する手掛かりとして用いられているが、メンクロウの場合には、音圧差は上下の音源を識別するために用いられている¹⁰⁾。このようなメンクロウの音源定位機構を模倣した、音源定位システムも提案されている¹¹⁾。

本研究では、水平方向を推定する手掛かりとして、ITD の情報のみを利用することにした。これは、ILD の情報は雑音に弱く正確な方向推定に用いることは難しいのに比べ、ITD は検出が比較的容易であり、マイク間距離に比べて音源までの距離が十分大きい場合、原理的に ITD の値のみで音源方向が一意に決まり、ITD 情報のみで音源方向の推定が可能であるからである。

ITD の値は、左右の音データの相互相関係数値が最大値となる標本点の差 Δt とサンプリングレート f_s から求められ $\Delta t/f_s$ である。この ITD 値から音源方向の推定を行なう。

Fig. 6 上の到達距離差 l は ITD 値と音速 c からつぎの式で求められる。

$$l = \frac{c\Delta t}{f_s} \quad (7)$$

この到達距離差 l から音源とマイク間の距離が十分遠い場合の音源方向 θ は、左右のマイク間距離を d とすると次式のよう求められる。

$$\theta = \arccos(l/d) \quad (8)$$

2.3.2 垂直方向の定位

人間は、音の上下方向や前後方向の判断には、耳介などによる単耳手掛かりを使っている。耳介は、到達音に対して音源方向に依存したスペクトル上の変化を与えることによって、音源方向を推定する手掛かりを作っている。音源が広い範囲の周波数成分を多く含むほど、広い範囲にわたってスペクトルの変化が起こるので、音源方向に関しての情報が多く得られ、音源方向の推定精度が高くなると考えられる。逆に、純音では情報がほとんど得られないため、定位することが困難となる。Hebrank らの心理実験によると、人間は 4k-15kHz の比較的高い周波数領域の音声情報を定位手掛かりとして用

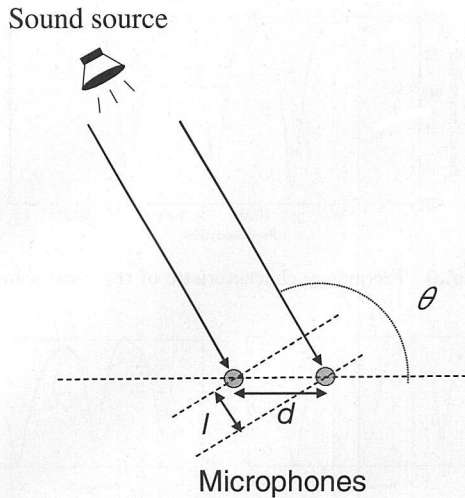


Fig. 6 Estimate the direction of the sound source on the horizontal plane

いていると考えられる¹²⁾。また Shaw らは、人工外耳を用いた測定により、スペクトルに音源方向に依存した複数の窪みが存在することを示した¹³⁾。さらに Hofman らは、ベイズ統計を用いたスペクトルを再構築することによって、スペクトルの特徴と音源方向の関連性を明らかにした¹⁴⁾。小野らは、スペクトル形状が音源方向に依存して変化する特徴を利用し、反射板を用いた単耳音源定位システムを実現した¹⁵⁾。このシステムでは、スペクトル波形上で極小値をとる周波数の対数が、音源方向に対して線形関係となるように反射板を設計することにより、広帯域雑音の音源方向を定位可能である。

われわれは、上下方向（鉛直面方向）の音源定位動作を自己組織的に学習するロボットシステムの構築を行ない、実験によってシステムが定位能力を学習可能であることを示した¹⁶⁾。このシステムでの上下方向の定位手掛かりには、小野らの設計した反射板によって生じる振幅特性の周波数軸上での極小値を用いた。小野らのシステムでは、一部の周波数帯のみを使用し、キャリアレーションすることにより正しい定位を可能としている。これに対し、本システムでは、より広い周波数帯を使用し、自己組織特徴マップによる学習によって定位能力を獲得している。より具体的には、小野らは1つの極小値周波数のみを含む周波数帯において、極小値周波数の対数と音源方向が線形になるように設計しているのに対し、本システムでは、複数の極小値周波数を含むような広い周波数帯を使用し、極小値周波数と音源方向の関係を学習によって獲得する。

複数の極小値を扱うために、本システムでは特徴ベクトルを使用した。特徴ベクトル I は、音源位置 $\pm 60^\circ$ の範囲を 20° 刻みで取り込んだデータから、各音源位置において Fig. 7 のように極小点周波数の位置から、作成する。図の上段が振幅特性、下段が特徴ベクトルである。極小値が N 個ある場合、極小値を取る周波数を、 f_1, f_2, \dots, f_N 、使用する周波数範囲を、 $f_{\min} \sim f_{\max}$ 、特徴ベクトルの次元を M とした場合の i 番目

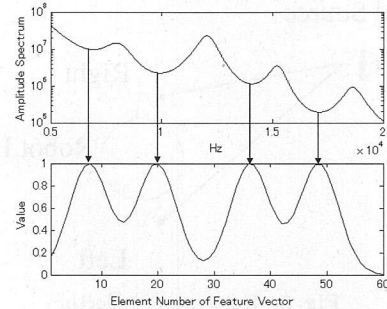


Fig. 7 A creation method of a feature vector

の要素 $I_i (i = 1, \dots, M)$ を、

$$I_i = \sum_{j=1}^N \exp\left(-\frac{(i-p_j)^2}{\sigma^2}\right) \quad (9)$$

$$p_j = \frac{f_j - f_{\min}}{f_{\max} - f_{\min}} M \quad (10)$$

とする。ここでは、 $\sigma = 4$ 、 $f_{\min} = 5000$ 、 $f_{\max} = 20000$ 、 $M = 60$ とした。この特徴ベクトルを入力として Kohonen の自己組織化学習によって、特徴マップを作成し、その特徴マップのテーブルを用いてテーブルルックアップ法により、入力データと音源位置の対応関係が学習される。このような手法を用いることにより、単一の情報のみを使用するのに比べて環境の変化に対してロバストな能力が獲得可能であり、また計算量も少なく処理時間が短く実時間で処理が可能である。本研究では、このわれわれの提案した手法を用いて上下方向の音源位置を推定する。

2.3.3 任意の音源への対応

この反射板を用いた上下定位の手法は音源に白色雑音を使用することを前提とした方法であるため、音声などの場合には上下方向を正しく定位することができない。これは音源が白色雑音などの振幅特性がフラットなもの以外の場合には、反射板の振幅特性を得るために音源の振幅特性が必要となるためである。

われわれはこの問題を解決するために、定位には使用しないもう一方のマイク情報を利用する手法を提案する。

Fig. 8 はマイクと反射板の配置を上から見たものである。反射板の形状から、音源に近い側のマイクに入る音 ($X_1(\omega)$) は反射板の影響を受けるが、音源から遠い側のマイクに入る音 ($X_2(\omega)$) は反射板の影響を受けない。ここで、音源の伝達特性を $S(\omega)$ 、音源から左右のマイクまでの伝達特性をそれぞれ $H_1(\omega)$ 、 $H_2(\omega)$ 、反射板の伝達特性を $R(\omega)$ 、マイクによる観測信号の伝達特性をそれぞれ $X_1(\omega)$ 、 $X_2(\omega)$ とすると、これらの関係はつぎのように表わされる。

$$X_1(\omega) = R(\omega)H_1(\omega)S(\omega) \quad (11)$$

$$X_2(\omega) = H_2(\omega)S(\omega) \quad (12)$$

この式から反射板の振幅特性 $|R(\omega)|$ は、

$$|R(\omega)| = \frac{|H_2(\omega)||X_1(\omega)|}{|H_1(\omega)||X_2(\omega)|} \quad (13)$$

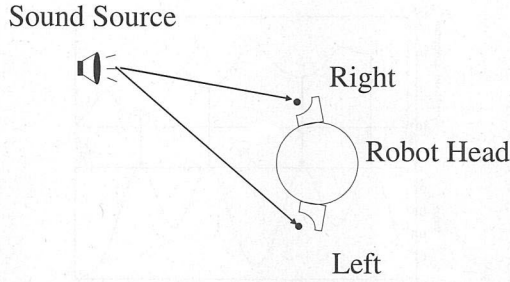


Fig. 8 Influence of reflector

となる。両耳への振幅特性の違いは無視できると仮定する。すなわち、音源-マイク間の振幅特性 $|H_1(\omega)|, |H_2(\omega)|$ を $|H_1(\omega)| = |H_2(\omega)|$ とすると。

$$|R(\omega)| = \frac{|X_1(\omega)|}{|X_2(\omega)|} \tag{14}$$

となり、左右のマイクによる観測信号のみから反射板の振幅特性を $|R(\omega)|$ を計算することが可能となる。

この提案手法の検証のために実験を行なった。Fig. 9 に示すような、周波数軸上に3つの窪みのある音源を用いて、左右のマイクで音を取り込んだ。Fig. 10(a) は音源に近い側のマイクにおける振幅特性 ($|X_1(\omega)|$)、Fig. 10(b) は音源に遠い側のマイクにおける振幅特性 ($|X_2(\omega)|$) である。また Fig. 10(c) が音源に白色雑音を用いた場合における音源-マイク間の振幅特性である。反響や伝搬による減衰などのない理想的な環境では、この振幅特性が反射板の振幅特性となる。そして、Fig. 10(d) が計算により得られた振幅特性 ($|R(\omega)|$) となる。Fig. 10(c) と Fig. 10(d) を比較すると、形状はかなり異なるが、窪みの位置は近いことがわかる。白色雑音によって作成される特徴マップ¹⁶⁾と比較するために音源の高さを -60 度 \sim 60 度まで変化させ特徴マップを作成した。その結果を Fig. 11 に示す。左の図が白色雑音によって作成された特徴マップ、右の図が Fig. 9 で示した音から作成された特徴マップであり、白色雑音のものに近いものになっていることがわかる。このことから、この提案手法を用いることによって音源が白色雑音の場合とほぼ同等の定位が可能であると考えられる。よって、この手法を用いて音声の上下方向推定を行なうこととした。

3. 定位実験

3.1 実験環境

反射板に無指向性マイクを取り付けたものを左右の耳としてロボットに搭載した (Fig. 12)。2本のマイク間距離は 32cm とした。また、これらのマイクの床からの高さは 1.35m とした。実験環境は、740cm \times 1400cm \times 270cm の会議室で、残響時間は 0.55 秒である。マイクの音はサウンドボードを通して量子化ビット数 16 ビット、サンプリング周波数 48kHz の PCM データに変換される。スピーカは Fig. 13 に示すようにマイクを搭載したロボットから 1.5m の距離の円周上に配置した。実験に使用する音声は、男性と女性の声を使用した。

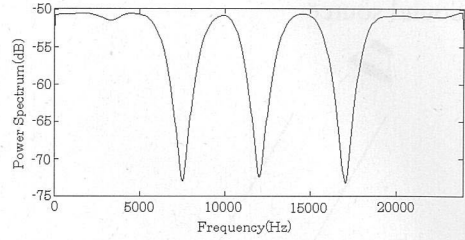


Fig. 9 Frequency characteristic of the used sound

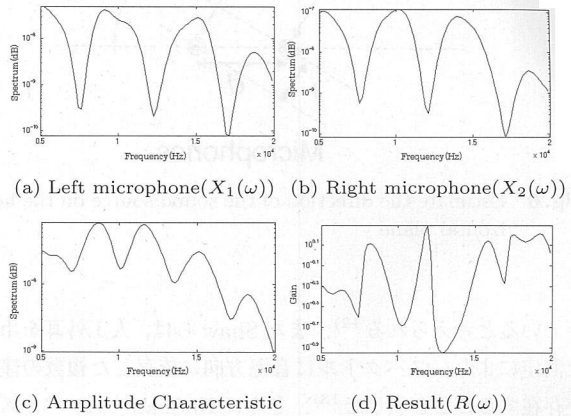


Fig. 10 Frequency characteristic

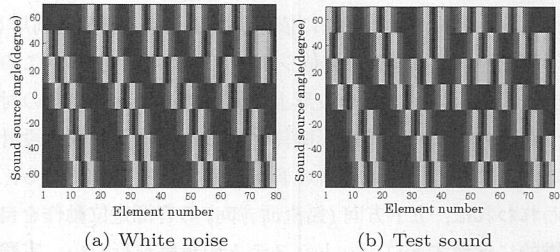


Fig. 11 Created feature vector

3.2 実験方法

Fig. 13 に示すように、1つ目のスピーカ (Spk1) は水平方向の角度 30 度、高さ 1.35m で固定とし、2つ目のスピーカ (Spk2) は水平方向の角度 45, 60, 80, 120 度、垂直方向の角度 $-10, 0, 10$ 度の計 12 カ所を移動させた。音声は 1つ目のスピーカから女性の声、2つ目のスピーカから男性の声が出力される。2つの混合音を 2本のマイクによって約 17 秒間録音し、左右のマイクにより録音された PCM データを用いて音源定位処理を行なう。無音部分を抽出するための閾値 (I_{m1}, I_{m2}) は分離結果から適切に設定した。左右方向は (8) 式を用いて得られた角度を音源方向とする。また、上下方向は仰角を $-10, 0, 10$ 度の 3つの高さで判別を行なった。この判別処理は、あらかじめ白色雑音を用いて作成した各音源方向に対する特徴ベクトルのテンプレートと計測された特徴ベクトルとの 2乗誤差の大きさから行なう。今回の実験において音源の高さを 10度刻みとしたのは、ロボットに搭載されているカメラの視野角がおおよそ ± 10 度であることからである。分離に関しては、録音した混合音を使って、オフライン



Fig. 12 Head robot

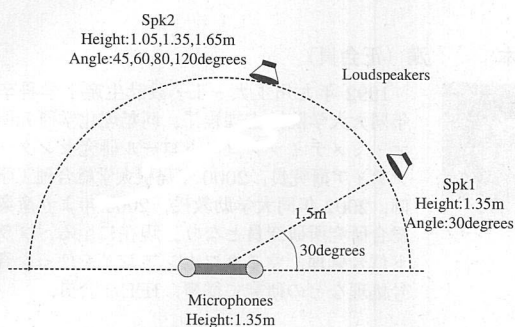


Fig. 13 Arrangement of microphones and loudspeakers

で (2) 式を使って分離を行なった. タップ数は $M=800$, 係数 L は 20 に設定した. 学習係数 α は 0.000002. 繰り返す計算の回数は, 10 回とした.

3.3 実験結果

それぞれのスピーカ配置に対する水平方向の推定結果を **Table 1** に示す. 表から 2 つの音源方向がほぼ正しく推定できていることがわかる. 正面付近に比べ両端での精度が悪い. これはサンプリング周波数を 48kHz とした場合, 正面付近では ITD の値が 1 サンプル違った場合に音源方向は 2 度程度しか変化しないが, 両端付近では ITD の値が 1 サンプル違った場合に音源方向がおよそ 15 度も変化してしまうためである. また, 垂直方向の推定結果を **Table 2** に示す. こちらも正しく推定できていることがわかる.

スピーカが水平方向に 60 度の位置にある場合の垂直方向の角度 $-10, 0, 10$ 度の 3 つに対する特徴ベクトルおよびテンプレートを **Fig. 14** に示す. グラフは図の上から垂直方向の角度 $-10, 0, 10$ 度の 3 つの高さに対応している. テンプレートには **Fig. 11** で示した白色雑音によって作成された特徴マップを使用した. 図から 3 つのどの角度においてもテンプレートと特徴ベクトルが完全には一致していないことがわかる. しかし, 3 つの山の位置すなわち周波数軸上の 3 つの極小値の位置のうち, 少なくとも 1 つはほぼテンプレートの理想値と一致していることがわかる. このように複数の極小値周波数を用いることによって, 誤認識を減らすことができていると考えられる.

Table 1 Estimation results on the horizontal plane

Elevation of speakers (Spk1, Spk2)	Direction of speakers (degree)							
	Spk1		Spk2		Spk1		Spk2	
	30	45	30	60	30	80	30	120
(10, -10)	35.5	50.7	35.5	64.6	35.5	79.6	35.5	111.2
(10, 0)	35.5	49.0	35.5	64.6	35.5	76.9	35.5	118.4
(10, 10)	35.5	43.6	35.5	60.2	35.5	79.6	35.5	118.4

Table 2 Estimation results on the vertical plane

Elevation of speakers (Spk1, Spk2)	Direction of speakers (degree)							
	Spk1		Spk2		Spk1		Spk2	
	30	45	30	60	30	80	30	120
(10, -10)	10	-10	10	-10	10	-10	10	-10
(10, 0)	10	0	10	0	10	0	10	0
(10, 10)	10	10	10	10	10	10	10	10

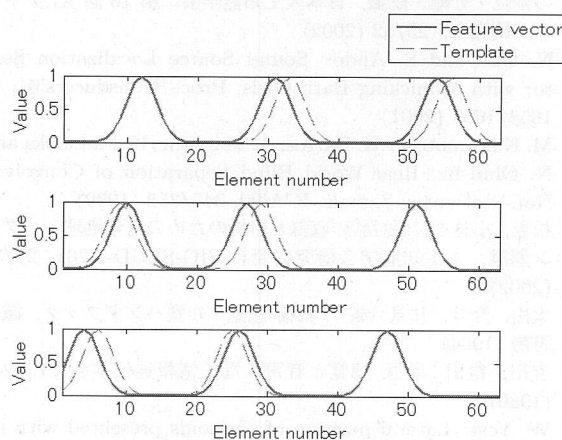


Fig. 14 Feature vector and feature map to each sound source direction

4. まとめ

本稿では, 2 つの異なる位置から発せられる音声の音源方向を推定する手法を提案し, 実験によりその有効性を示した. 本手法を用いることにより, 2 本のマイクのみで 2 つの音源の水平方向および垂直方向の位置を推定することが可能となった. 現在のシステムでは分離に ICA を使っているため, 2 本のマイクでは 2 音源しか取り扱うことはできないが, 分離に ICA 以外のものを使用し, 2 本のマイクで複数音源の分離を行なえば, 分離部分の変更のみで複数音源への拡張が可能になると考えられる.

今回提案手法を評価するために行なった実験では, スピーカを上下に 10 度刻みで 3 カ所の高さに配置して音源方向の識別を行なったが, 白色雑音を用いた場合にはより高い精度で識別が可能であるので, 今後音声を用いた場合においても, 精度を高めたい. また, 本手法では反射板を経由せず直接マイクに到達する側の音を音源の音とみなし, 音源の音の特徴を打ち消すことにより任意の音に対して上下方向の定位を可能にした. しかし, 現在使用している反射板の配置では正面付近の音源に対しては正しく動作しない. これは, 正面付近から来る音は左右共に反射版を経由しないかまたは左右共に反射を経由するかのどちらかになってしまうためである. 今後この問題に対する解決方法を検討したい. さらに, 実験に

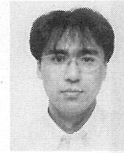
において使用した分離信号に対する閾値は、現在分離結果から適切な値を設定しているが、これについても今後分離信号自身から生成する手法を検討したいと考えている。また、分離フィルタ W の逆特性から伝達特性 R を計算することも可能であり、この方法で求めた R を使った音源定位法も検討していきたいと考えている。

参考文献

- 1) 安藤, 篠田, 小川, 光山: 時空間勾配法に基づく3次元音源定位センサシステム, 計測自動制御学会論文集, **29-5**, 520/528 (1993)
- 2) 中臺, 奥乃, 北野: アクティブオーディションによる複数音源の定位・分離・認識, 日本人工知能学会 第16回 AI チャレンジ研究会, 25/32 (2002)
- 3) N. Ono and S. Ando: Sound Source Localization Sensor with Mimicking Barn Owls, Proc. Transducers'01, **2**, 1654/1657 (2001)
- 4) M. Kawamoto, A.K. Barros, A. Mansour, K. Matsuoka and N. Ohnishi: Real World Blind Separation of Convolved Non-Stationary Signals, ICA'99, 347/352 (1999)
- 5) 松坂, 小林: 対面対話の収録と解析のための信号処理とパターン認識, 人工知能学会研究会資料 SIG-SLUD-A201, 27/32 (2002)
- 6) 大山, 今井, 和気 (編): 新編 感覚・知覚ハンドブック, 誠信書房 (1994)
- 7) 吉田, 亀田: 新版 聴覚と音声, 電子情報通信学会, 73/240 (1980)
- 8) W. Yost: Lateral position of sinusoids presented with interaural intensive and temporal differences, J. Acoust. Soc. Am., **70-2**, 397/409 (1981)
- 9) D.S. Brungart and W.M. Rabinowitz: Auditory localization of nearby sources. Head-related transfer functions, J. Acoust. Soc. Am., **106-3**, 1465/1479 (1999)
- 10) 小西正一: フクロウの音源定位の脳機構, 科学, **60-1**, 18/28 (1990)
- 11) L. Natale, G. Metta and G. Sandini: Development of auditory-evoked reflexes: Visuo-acoustic cues integration in a binocular head, Robotics and Autonomous Systems, **39**, 87/106 (2002)
- 12) J. Hebrank and D. Wright: Spectral cues used in the localization of sound sources on the median plane, J. Acoust. Soc. Am., **56-6**, 1829/1834 (1974)
- 13) E.A.G. Shaw and R. Teranishi: Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Point Source, J. Acoust. Soc. Am., **44-1**, 240/249 (1968)
- 14) P.M. Hofman and A.J.V. Opstal: Bayesian reconstruction for sound localization cues from responses to random spectra, Biological Cybernetics, **86**, 305/316 (2002)
- 15) N. Ono, Y. Zaitzu, T. Nomiya, A. Kimachi and S. Ando: Biomimicry Sound Source Localization with Fishbone, T.IEE, **121-E-6**, 313/319 (2001)
- 16) 中島, 向井, 大西: スペクトルの特徴マップを用いた上下方向音源定位学習システム, 電子情報通信学会論文誌, **J87-DII-11**, 2034/2044 (2004)

[著者紹介]

中島 弘道



1993年東北大学工学部電気情報系卒。95年同大学院通信工学専攻修士課程修了。同年沖電気工業株式会社勤務。97年名大大学院工学研究科情報工学専攻後期課程満了。同年理化学研究所リサーチアソシエイト。2004年より同研究所研究員。2007年より同研究所・ロボット感覚情報研究チームの研究員。神経回路, 学習, 音源定位の研究に従事。博士(工学)。2002年度日本神経回路学会研究賞, 電子情報通信学会, 情報処理学会, 日本音響学会各会員。

河本 満 (正会員)



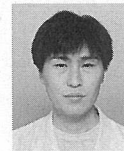
1992年九州工大・工・設計生産工学科卒, 97年同大学院博士課程了, 同年理化学研究所バイオ・ミメティックコントロール研究センターフロンティア研究員。2000年島根大学総合理工学部講師, 2002年同大学助教授, 2005年より産業技術総合研究所研究員となり, 現在に至る。ブラインド信号処理, マイクロホンアレイを使った音響信号処理などの研究に従事。IEEE 会員。

伊藤 雅紀



2001年名大・工・電気電子・情報工卒。2003年同大学院工学研究科情報工学専攻博士前期課程修了。2006年同大学情報科学研究科博士課程後期課程修了。2005年より日本学術振興会特別研究員。現在, 東北学院大学教養学部情報科学研究員。音源分離に興味をもつ。電子情報通信学会会員。

向井 利春 (正会員)



1995年東京大学大学院工学系研究科計数工学専攻博士課程修了。同年理化学研究所研究員。2000年より Laboratoire de Neurobiologie (フランス) にてポストドク。2001年より理化学研究所・生物型感覚統合センサー研究チーム, 2007年より同研究所・ロボット感覚情報研究チームのチームリーダーとなり現在に至る。生体情報処理の工学的応用, センサ情報処理, ロボティクスなどの研究に従事。博士(工学)。99年度計測自動制御学会論文賞, 2002年度日本神経回路学会研究賞, 2005年計測自動制御学会 SI 部門奨励賞など受賞。電子情報通信学会, 日本ロボット学会, IEEE などの会員。