

## 両耳受聴音からの音源方向推定の精度向上に関する一検討

## HRTF データへの適用

若竹 孝幸<sup>†</sup> 永田 仁史<sup>†</sup> 岩崎 聡<sup>††</sup> 針山 孝彦<sup>†††</sup> 藤岡 豊太<sup>†</sup>  
安倍 正人<sup>†</sup><sup>†</sup> 岩手大学工学部情報システム工学科 〒020-8551 岩手県盛岡市上田 4-3-1<sup>††</sup> 愛知医科大学耳鼻咽喉科 〒480-1195 愛知県愛知郡長久手町<sup>†††</sup> 浜松医科大学 〒431-3192 浜松市半田山 1-20-1

E-mail: †{h15j75,nagata,toy,abe}@cis.iwate-u.ac.jp

**あらまし** 両耳受聴音からの音源方向推定の精度向上のため、高分解能方向推定法におけるスペクトル成分の強調方法と重み付け Wiener 利得 [1] に基づく逐次的な推定法の二つの手法を提案する。高分解能法におけるスペクトル成分の強調は、周波数ごとの自己相関行列の固有値の比を用いるものであり、音源が 1 個の場合の推定精度を大幅に向上できる。一方、重み付き Wiener 利得に基づく逐次的な処理法は、音源方向推定に用いる空間スペクトル上の最大ピークだけを音源方向とみなし、最大ピークに寄与するスペクトル成分を抑圧するフィルタを用いて再び音源方向推定を行うものであり、複数の音源がある場合に適用できる。シミュレーションによりこれらの手法の性能を評価した結果、音源 1 個の場合、固有値の比で重み付けする高分解能法と重み付き Wiener 利得に基づく方法は同等の性能であったが、音源が複数の場合は、重み付き Wiener 利得に基づく逐次法のみが高い方向推定精度を達成できることがわかった。

**キーワード** HRTF, Wiener Gain, 2次元方向推定, 固有値, 逐次推定

## Performance improvement on DOA estimation from binaural sound

Takayuki WAKATAKE<sup>†</sup>, Yoshifumi NAGATA<sup>†</sup>, Satoshi IWASAKI<sup>††</sup>, Takahiko HARIYAMA<sup>†††</sup>,  
Toyota FUJIOKA<sup>†</sup>, and Masato ABE<sup>†</sup><sup>†</sup> Department of Computer and Information Science, Iwate University, Ueda4-3-1, Morioka, 020-8551,  
Japan<sup>††</sup> Aichi Medical University, Nagakute Aichi-pref. 480-1195, Japan<sup>†††</sup> Hamamatsu University School of Medicine, Handayama1-20-1, Hamamatsu, 480-1195, Japan

E-mail: †{h15j75,nagata,toy,abe}@cis.iwate-u.ac.jp

**Abstract** In order to improve estimation accuracy of directions of arrival(DOA) of sound sources, we propose a new weighting function for averaging multiple spectral components. We exploit a ratio between the eigen-values obtained from the auto correlation matrix at each frequency bin for an weighting function. This method improves the estimatuon accuracy in the cases where a single sound source exists. Furthermore, we propose a iterative method based on the weighted Wiener gain to coop with multiple source environment. At first, this method regards the maxmum peak in the spatial spectrum as a one source. Next, a filter to reduce the spectral components that contribute the maximum peak is applied and re-estimate the second source to detect the maximum peak in the residual spatial spectrum. While the methods using the eigen-ratio weighting function and the weighted Wiener gain were almost equivalent in a single source environment, the iterative method for the weighted Wiener gain only attained good performance for the cases where multiple sound sources exist.

**Key words** directional microphone, HRTF, Wiener gain, elevation angle estimation, two-dimensional DOA estimation

## 1. はじめに

筆者等は、回転対称に設置した2個の指向性マイクロホンの信号から重み付きウィナー利得(WWG) [1]に基づいて方位-仰角の2次元の音源方向を推定する手法 [2] を提案し、さらに、この方法において、マイクロホンの指向性を人間の頭部伝達関数(HRTF)に置き換えれば、HRTFによる変形を受けた信号からも2次元の方向推定が可能であることを示した [3], [4]。

先の報告 [4] においては、ランダムに設定した様々な音源方向に関し、どれだけの割合で検出が可能かを計算機シミュレーションにより調べた。この結果、S/N10dBで音源数が1の場合は、許容誤差を5度以上としたとき、提案法が90%以上と高い検出率が得られたが、音源数が2の場合は30%以下と大幅に低下した。また、指向性マイクロホンを用いた場合との比較においても大幅に低い性能であったことから、両耳受聴音の場合には複数音源環境下での性能に問題のあることがわかった。しかしながら、このとき、音源方向推定の指標として用いる空間スペクトルにおいて、最大ピークが2つの音源のいずれかに正しく対応している割合が90%以上と高いことが分った。

そこで、本報告では、2chの重み付きウィナー利得(WWG)に基づいた音源方向推定において、逐次的に複数の音源を推定する方法を検討することとした。この方法は、最初に得られた空間スペクトル上の最大ピークを1個の音源方向とし、次に、この最大ピークに寄与するスペクトル成分を抑圧するようなフィルタを掛け、フィルタリング後の残りの成分によって再度音源方向を推定する。音源成分の抑圧方法はいくつか考えられるが、ここでは、バイナリマスクのような選択処理とはせず、重み付きWiener利得の重み関数に統合する形で処理することとした。

一方、他の従来法においても、WWGと同様に方向推定に寄与するスペクトル成分の選択あるいは強調によって性能が向上するものと期待できる。そこで、高分解能法として一般的なMUSICと最小分散法(MV)に対し、入力信号の自己相関行列から得られる固有値の比に基づく方法を検討することとした。

本報告では、まず、音源が1個の場合について、高分解能法における成分強調について検討し、次に、音源が複数の場合の逐次的な方向推定について述べる。

## 2. 固有値の比に基づいた周波数成分の強調

2chの方向推定の場合、高分解能法により検出可能な音源数は1であるので、音源数がマイクロホン数以上となる状況が容易に起こりうる。しかし、音声のスペクトル成分が時間-周波数平面上に疎に分布していることから、各周波数成分の推定結果を統合すれば複数の音声の到来方向が推定できる。統合方法としては、例えば、入力信号の自己相関行列の固有値の和の値で重み付けする方法 [5] などがある。

MUSICの場合、まず、背景雑音の部分空間に相当する固有ベクトルを自己相関行列の固有ベクトルから選択する必要がある。一般には固有値の大きさに基づいてその選択を行う [6] [7] [8] が、マイクロホン数が少ない場合、背景雑音のみに相当する固

有値が存在しないか、判定が困難なことが多い。そこで、周波数成分ごとに小さい方の固有値に相当する固有ベクトルを雑音の固有ベクトルとみなすこととし、小さい方の固有値と大きい方の固有値の比に基づいて成分を重み付けすることを考える。小さい方の固有値が背景雑音ではなく、到来音の寄与によるものであるとしても、固有値の比が十分大きければ、その成分は1個の音源の影響が支配的な成分であると考えられる。従って、この重みによって音源方向推定に有利な成分を強調することができる。チャンネル数が3以上の場合も同様に、最小と最大の固有値の比を用いればよいと考えられる。

以上を考慮し、ここでは、MUSICとMVに関し、固有値の比により強調した空間スペクトルを次式のように計算することとした。

$$S_{\text{MUSIC-EW}}(d) = \sum_k r_{n,k} \log\left(\frac{1}{a(d)^* E_{n,k} E_{n,k}^* a(d)}\right), \quad (1)$$

$$S_{\text{MV-EW}}(d) = \sum_k r_{n,k} / (a(d)^* R_{n,k}^{-1} a(d)), \quad (2)$$

$$R_{n,k} = \begin{bmatrix} W_{xx,n,k} & W_{xy,n,k}^* \\ W_{xy,n,k} & W_{yy,n,k} \end{bmatrix}, \quad (3)$$

$$r_{n,k} = e_{L,n,k} / e_{S,n,k}. \quad (4)$$

ここで、 $n$ はフレーム番号、 $k$ は周波数成分の番号、 $d = (\theta, \varphi)$ は方位角 $\theta$ 、仰角 $\varphi$ により決まる方向、 $X_{n,k}$ 、 $Y_{n,k}$ は、各々、1ch目と2ch目の入力信号のFFTスペクトル、 $W_{xx,n,k}$ 、 $W_{yy,n,k}$ は、各々、1ch目と2ch目のパワースペクトル、 $W_{xy,n,k}$ は、1ch目と2ch目の信号間のクロススペクトル、 $R_{n,k}$ は入力信号の共分散行列である。また、 $E_{n,k}$ は $R_{n,k}$ の固有ベクトルのうち、小さい方の固有値に対応するもの、 $r_{n,k}$ は大きい方の固有値 $e_{L,n,k}$ と小さい方の固有値 $e_{S,n,k}$ の比である。 $a(d)$ はモードベクトル [9]、または、ステアリングベクトルであり、ここではHRTFに相当する。すなわち、 $H_{x,k}(d)$ 、 $H_{y,k}(d)$ を各々、左と右のHRTFのDFTとすると、 $a(d) = (H_{x,k}(d), H_{y,k}(d))$ であり、事前に測定して既知であるとする。なお、(\*)は複素共役を表す。

## 3. 重み付きWiener利得に基づく逐次的な複数音源方向推定

2chの指向性マイクロホンを用いた場合、音源の増加に伴う推定精度の低下は比較的緩やかであった [2] が、1章で述べたように、両耳受聴音からの推定では音源が2個の場合に既に非常に低い精度となっていた。しかしながら、WWGの空間スペクトルにおいては、最大ピークの方向が2個の音源のいずれかと一致する割合が高かった。そこで、この性質を利用して音源方向を一つずつ求めて行く逐次的方法を検討した。

まず、WWGによる空間スペクトルを次式に示す。ただし、フレーム番号 $n$ は省略した。

$$S_{\text{WWG}}(d) = \frac{\sum_k \text{Re}[G_{xy,k}(d)]\Psi_k(d)\Phi_k(d)}{\sum_k G_{zz,k}(d)\Psi_k(d)}. \quad (5)$$

ここで,

$$\begin{aligned} G_{xy,k}(d) &= \overline{X_k^* A_{x,k}^*(d) Y_k A_{y,k}(d)} \\ &= W_{xy,k} A_{x,k}^*(d) A_{y,k}(d), \end{aligned} \quad (6)$$

$$\begin{aligned} G_{zz,k}(d) &= \overline{[(X_k A_{x,k}(d) + Y_k A_{y,k}(d))/2]^2}, \\ &= W_{xx,k} |A_{x,k}(d)|^2/4 + W_{yy,k} |A_{y,k}(d)|^2/4 \\ &\quad + \text{Re}[W_{xy,k} A_{x,k}(d) A_{y,k}^*(d)]/2, \end{aligned} \quad (7)$$

$$\begin{aligned} G_{dd,k}(d) &= \overline{[X_k A_{x,k}(d) - Y_k A_{y,k}(d)]^2}, \\ &= W_{xx,k} |A_{x,k}(d)|^2 + W_{yy,k} |A_{y,k}(d)|^2 \\ &\quad - 2\text{Re}[W_{xy,k} A_{x,k}(d) A_{y,k}^*(d)], \end{aligned} \quad (8)$$

$$\Psi_k(d) = 1/G_{dd,k}^\beta(d), \quad (9)$$

$$\Phi_k(d) = \frac{|G_{xy,k}(d)| - \gamma G_{dd,k}(d)/\nu_k(d)}{|G_{xy,k}(d)|}, \quad (10)$$

であり,  $A_{x,k}(d)$ ,  $A_{y,k}(d)$  は, 各々, 左 ch と右 ch の HRTF の逆特性である.

$S_{\text{WWG}}(d)$  から得られた空間スペクトルにおいて, 最大ピークの方角を

$$d_1 = \arg[\max_d S_{\text{WWG}}(d)] \quad (11)$$

とおく. 前述したように, 最大ピークが真の音源のいずれかである割合が高いので, まず, このピークを第 1 番目の音源方向であるとみなす. 次に, このピークに寄与するスペクトル成分を抑圧してから再び空間スペクトルを求めれば, 埋もれていた音源のピークが浮かび上がってくるものと期待できる. 抑圧する方法としては様々なものが考えられるが, WWG においては, 2ch-SS に対応する重み関数  $\Phi_k(d)$  は方向  $d$  からの到来音パワーに関する信号対信号+雑音比に相当しているので, 方向  $d_1$  からの到来音を抑圧するには,  $\Phi_k(d_1)$  に逆比例するような重みを与えればよいことが分る.  $\Phi_k(d)$  が 0 から 1 の値をとることから, ここでは簡単に,

$$1 - \Phi_k(d_1) \quad (12)$$

を抑圧フィルタとすることとし, この関数を WWG の重みの一部として新たな空間スペクトル

$$S_{\text{WWG}}^{(d_1)}(d) = \frac{\sum_k \text{Re}[G_{xy,k}(d)]\Psi_k(d)\Phi_k(d)[1 - \Phi_k(d_1)]}{\sum_k G_{zz,k}(d)\Psi_k(d)} \quad (13)$$

を計算する. 新たな空間スペクトルから得られた最大ピークを

$$d_2 = \arg[\max_d S_{\text{WWG}}^{(d_1)}(d)] \quad (14)$$

とし, 以降は, 仮定する音源数に達するまで同様の処理を繰り返す. 従って,  $i(\geq 2)$  番目の音源を推定するための空間スペクトルは,

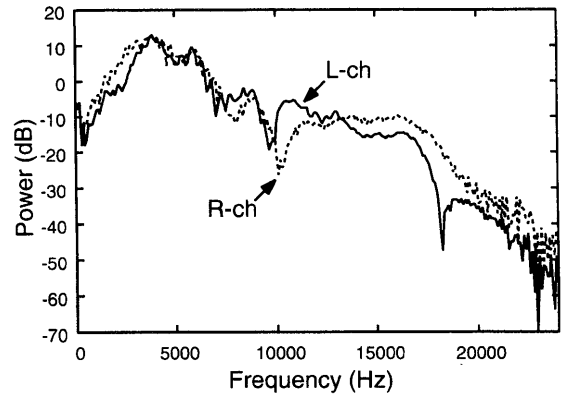


図 1 正面方向  $(\theta, \varphi) = (0, 0)$  の HRTF (モデル A)

Fig. 1 HRTF in the direction  $(\theta, \varphi) = (0, 0)$  obtained from model A

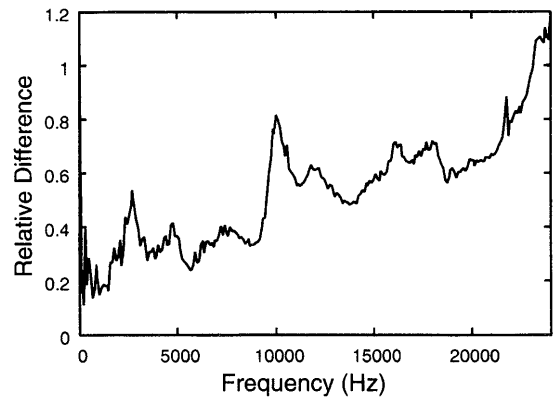


図 2 HRTF の空間分布全体の相対左右差

Fig. 2 Difference measure between L- and R-ch HRTF spatial distribution.

$$S_{\text{WWG}}^{(d_1, d_2, \dots, d_{i-1})}(d) = \frac{\sum_k \text{Re}[G_{xy,k}(d)]\Psi_k(d)\Phi_k(d) \prod_{j=1}^{i-1} [1 - \Phi_k(d_j)]}{\sum_k G_{zz,k}(d)\Psi_k(d)} \quad (15)$$

$$d_j = \arg[\max_d S_{\text{WWG}}^{(d_1, \dots, d_{j-1})}(d)], \quad (16)$$

と表すことができる.

#### 4. HRTF データ

音源方向推定性能の評価には, 先に本研究グループで測定した 4 組の HRTF を用いた [4]. これは, 本研究グループが作成した精密な 4 体の頭部モデルを使って測定したものであり, データ本体はサンプリング周波数 48kHz, 512 点のインパルス応答である. 測定範囲は, 方位角が  $-180^\circ \leq \theta \leq 180^\circ$ , 仰角が  $-60^\circ \leq \varphi \leq 90^\circ$  であり, 仰角の角度刻みは  $5^\circ$ , 方位角の刻みは仰角  $0^\circ$  付近が  $5^\circ$  で最も小さく, 仰角の絶対値が  $60^\circ$  以上においては徐々に方位角の刻みを大きくしてある. 例として, 図 1 にモデル A に関して測定した HRTF の正面方向  $(\theta, \varphi) = (0, 0)$  の振幅特性を示す. なお, 既報 [4] の HRTF の左右の非対称性を示すための相対左右差のデータに計算誤りがあったため, 図 2 に, モデル A に関して測定した HRTF の例と計算した相対左右差を改めて示す. この相対左右差は次式に

より計算した。

$$E_{LR}(k) = \frac{\sum_d ||H_{x,k}(d)|^2 - |H_{y,k}(d')|^2|}{\sum_d (|H_{x,k}(d)|^2 + |H_{y,k}(d')|^2)/2} \quad (17)$$

ここで、 $d' = d(-\theta, \varphi)$ である。用いた HRTF においては、マイクロホン特性とスピーカ特性の補正は行っていない。後述するように、本シミュレーションでは評価に用いた周波数帯域を 260Hz から 4kHz としたが、この帯域においては、マイクロホンとスピーカはいずれも平坦な特性であるため、これらの特性の音源方向推定への影響は無視できるものと考えられる。

## 5. 音源方向推定シミュレーション

固有値の比に基づく成分強調と WWG に基づく逐次推定法の有効性を調べるため、シミュレーションにより音源方向推定精度について検討した。

### 5.1 音源データとシミュレーション条件

到来音の信号として用いるため、ラジオのインタビュー番組から男性の音声を約 5 分間録音し、パワーに基づいて無音区間を削除した。背景雑音に関しては、マイクロホンを鼓膜位置に設置した 1 体の頭部モデルを使い、多数の計算機のある実験室の計算機雑音をバイノーラル録音した。

評価に用いた帯域は 260Hz から 4kHz とし、雑音と無音区間を削除した音声の両方をこの帯域の帯域通過フィルタに通した。方向推定シミュレーションは、S/N、音源数、および、HRTF のモデルを変えるごとに各々 1000 回の試行を行うこととし、1 秒の区間を帯域通過後の音声から試行ごとにランダムに切り出して音源信号とした。S/N については、各試行において切り出した 1 秒の音源信号が正面  $(\theta, \varphi) = (0, 0)$  から到来したと仮定したときの受信点における S/N が所望の値になるように振幅を調整して重畳した。複数の音源を仮定する場合は、各音源からの信号が背景雑音に対し同じ S/N となるようにした。

HRTF は、先に述べたように、本研究グループで作成した 4 体の頭部モデルのものをを用い、実験結果は 4 体の平均値で表した。仮定する音源方向は、HRTF の測定方向 [4] のいずれかに対応する方向とし、複数音源を仮定する場合は、音源方向の間隔が  $10^\circ$  以上となるように制限した。サンプリング周波数は 48kHz、FFT 長は 2048 点、窓関数はハニング窓、フレーム周期は 1024 点、 $W_{xx}, W_{yy}, W_{xy}$  等の推定のための平均化フレーム数は 3 とした。

評価に用いた音源方向推定法は、WWG、MUSIC、MV、相互相関関数、および、一般化相互相関関数 (GCC) であり、WWG に基づいた逐次法を WWG-ITR、また、MUSIC、MV を固有値の比で強調した方法を、各々、MUSIC-EW、MV-EW と呼ぶことにした。用いた GCC は、雑音が無相関な場合の最尤推定に基づく方法 [10] であり、これを GCC-ML とし、普通の相互相関関数を CC とよぶことにした。また、WWG のパラメータは、 $(\beta, \gamma) = (0.8, 3)$  とした。

### 5.2 音源方向推定精度の評価

音源方向をランダムに多数設定し、空間スペクトルからピー

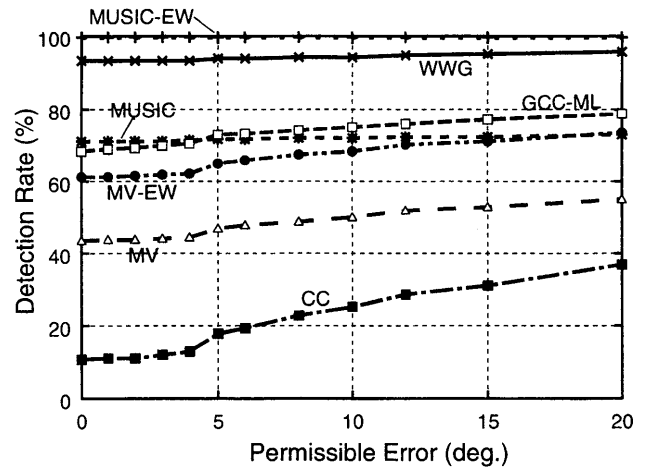


図 3 音源方向検出率 (音源 1 個, S/N=10dB)

Fig. 3 Source detection rate vs. permissible error. (single sound source)

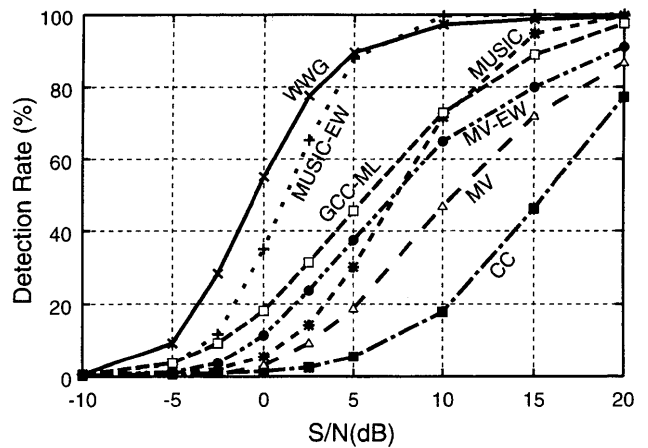


図 4 音源方向検出率 (音源 1 個, 許容誤差  $5^\circ$  のとき)

Fig. 4 Source detection rate vs. S/N. (single sound source)

ク検出によって得られた音源方向が設定方向と一致するか否かを調べた。音源方向は、 $-180^\circ \leq \theta \leq 180^\circ, -55^\circ \leq \varphi \leq 85^\circ$  の範囲とした。方向推定のための空間スペクトルは、 $-180^\circ \leq \theta \leq 180^\circ, -60^\circ \leq \varphi \leq 90^\circ$  の範囲で  $5^\circ$  間隔の格子点上の値を計算した。この空間スペクトルから推定精度を求める際は、音源数を既知とし、検出した最大ピークと正しい設定方向との間の誤差をどこまで許容するかを定める許容誤差を決め、許容誤差範囲内であれば正しく検出できたものとした。音源が複数の場合は、すべての音源において推定値と対応する正解との間の誤差が許容誤差内であれば正解とした。正しい結果の得られた回数の総数に対する比を検出率とした。

### 5.3 音源が 1 個の場合

まず、音源数を 1 とし、S/N=10dB のときに得られた音源方向検出率を図 3 に示す。図は、WWG、MUSIC、MUSIC-EW、MV、MV-EW、CC、GCC-ML に関する結果であり、許容誤差の関数として表してある。

図を見ると、例えば許容誤差  $10^\circ$  のとき、MUSIC と MV は検出率 72% と 50% であったが、固有値の比により強調した方

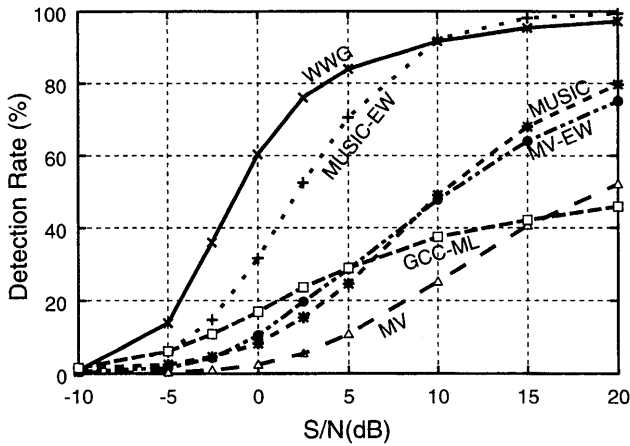


図 5 最大ピークの正解率 (音源 2 個)

Fig. 5 Maximum peak correct rate vs. S/N. (two sound sources)

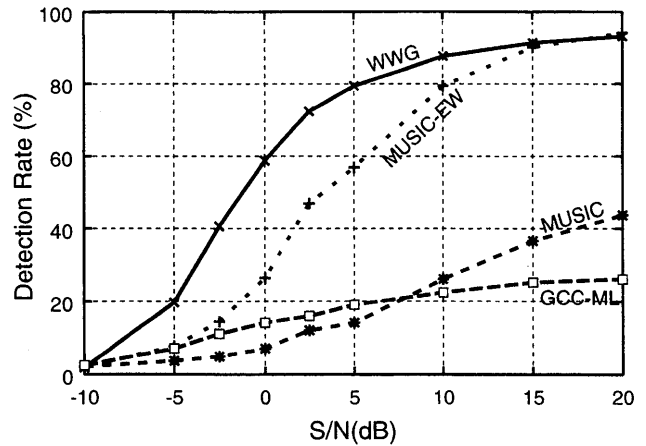


図 7 最大ピークの正解率 (音源 3 個)

Fig. 7 Maximum peak correct rate vs. S/N. (three sound sources)

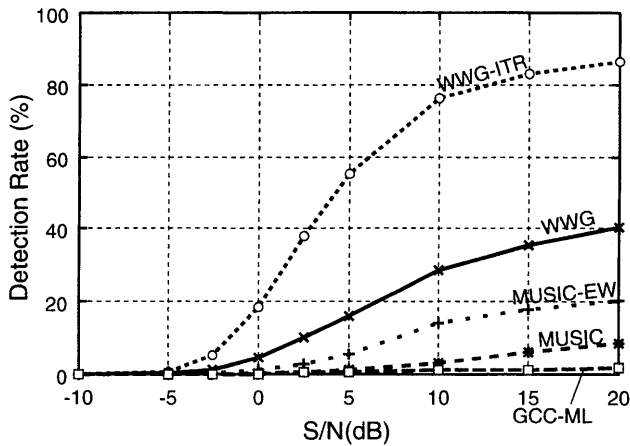


図 6 音源方向検出率 (音源 2 個, 許容誤差 5° のとき)

Fig. 6 Source detection rate vs. S/N. (two sound sources)

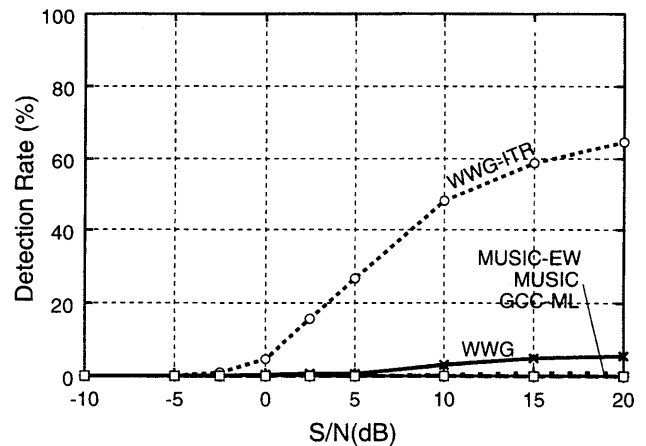


図 8 音源方向検出率 (音源 3 個, 許容誤差 5° のとき)

Fig. 8 Source detection rate vs. S/N. (three sound sources)

法の MUSIC-EW と MV-EW では、各々、99% と 68% に向上したことがわかる。とくに、MUSIC-EW の性能が高くなり、WWG を若干上回る結果となった。一方、CC は検出率 30%、GCC-ML は 75% 程度であった。

次に、S/N を変えながら検出率を求めた結果を図 4 に示す。これは、許容誤差が 5° のときの結果である。図を見ると、S/N が -5dB から 5dB の間で検出率が大きく変化しており、WWG と MUSIC-EW が同程度で最も性能が高いことがわかる。

#### 5.4 音源が 2 個の場合

次に、音源が 2 個の場合の結果について、まず、許容誤差を 5° としたときに空間スペクトル上の最大ピークが正解である率を S/N の関数として図 5 に示す。図を見ると、WWG、MUSIC-EW の正解率は、S/N 10dB 以上ではいずれも 90% 以上であり、最大ピークの正解率に関しては、両者が同程度の性能であり、且つ、他の方法よりも大幅に高い性能であることがわかる。MV と MV-EW、および、CC は性能が低く、あまり特徴がないため以降の実験では表示を省略する。

次に、音源が 2 個の場合の検出率について、許容誤差を 5° としたときの結果を S/N の関数として図 6 に示す。この実験

においては、WWG に基づく逐次推定 WWG-ITR の結果も合わせて示す。図を見ると、WWG は、S/N が 10dB のときの検出率が 28% であるのに対し、WWG-ITR では、77% と大幅に向上しているのがわかる。一方、MUSIC-EW を含む他の方法は、20% 以下の低い検出率となった。

#### 5.5 音源が 3 個の場合

次に、音源が 3 個の場合の結果について、まず、空間スペクトル上の最大ピークが正解である率を図 7 に示す。図から、WWG、MUSIC-EW の正解率はやはり高く、S/N 10dB 以上では約 80% 以上と、音源 2 個の場合と比べてときの正解率低下は小さいことがわかる。

次に、音源が 3 個の場合の検出率について、許容誤差を 5° としたときの結果を S/N の関数として図 8 に示す。図を見ると、WWG-ITR だけが低い検出率となっており、S/N が 10dB のとき、WWG の検出率が 3% であるのに対し、WWG-ITR では、約 50% と大幅に向上しているのがわかる。その他の方法の検出率はほとんど 0% であった。

## 6. ま と め

両耳受音信号からの音源方向推定における精度向上のための二つの方法を提案した。一つは、従来の高分解能法において音源方向推定に有利なスペクトル成分を強調するための固有値の比に基づいた重み付けであり、二つめは、重み付き Wiener 利得に基づく方法における逐次的な複数音源の推定法である。4体の頭部モデルについて測定した HRTF を用い、シミュレーションによりこれらの方法の性能を評価したところ、音源が1個の場合には提案した強調法によって MUSIC の性能が大きく改善され、許容誤差  $5^\circ$ 、 $S/N=10\text{dB}$  のときに音源方向検出率が 72% から 99% へと向上した。音源が複数の場合は、固有値の比による強調の効果は見られたものの検出率は低く、音源 2 個、許容誤差  $5^\circ$ 、 $S/N=10\text{dB}$  のときに検出率は 15% 以下であった。これに対し、複数音源の場合でも、重み付き Wiener 利得に基づく逐次的な推定法により、音源 2 個、許容誤差  $5^\circ$ 、 $S/N=10\text{dB}$  のときに、検出率は 28% から 77% へと大幅に向上した。また、音源が 3 個の場合も同様な結果が得られた。これらの結果より、提案法の有効性が確かめられた。高分解能法に関しても、逐次的な手法が可能であると考えられるので、今後検討してゆく予定である。また、逐次法では、音源数の推定が重要であるので、これについても検討する予定である。

## 文 献

- [1] Y. Nagata, F. Fujioka, and M. Abe, "Speech enhancement based on auto gain control," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, pp. 177–190, 1 2006.
- [2] Y. Nagata, F. Fujioka, and M. Abe, "Two-dimensional doa estimation of sound sources based on weighted wiener gain exploiting two directional microphones," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 15, pp. 416–429, Feb. 2007.
- [3] 永田仁史, 藤岡豊太, 安倍正人, "重み付きウィナー利得に基づく音源方向推定に関する検討," 電子情報通信学会技術報告, no. EA2005-43, pp. 19–24, 2005-8.
- [4] 小原富太, 永田仁史, 藤岡豊太, 岩崎聡, 針山孝彦, 安倍正人, "両耳受聴音からの音源方向推定処理に関する検討," 電子情報通信学会技術報告, no. EA2006-106, pp. 13–18, 2007-1.
- [5] 浅野太, "ロボットにおける音源位置推定," 日本音響学会誌, vol. 63, pp. 41–46, Jan. 2007.
- [6] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas and Propagat.*, vol. AP-34, pp. 276–280, Mar. 1986.
- [7] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech and Audio Processing*, vol. 3, pp. 251–266, July 1995.
- [8] F. Asano, S. Hayamizu, T. Yamada, and S. Nakamura, "Speech enhancement based on the subspace method," *IEEE Trans. Speech and Audio Processing*, vol. 8, pp. 497–507, Sep. 2000.
- [9] 菊間信良, ed., アレーアンテナによる適応信号処理. 科学技術出版, 1998.
- [10] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans, Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 320–327, Aug. 1976.