

## 両耳受聴音からの音源方向推定処理に関する検討

小原 富太<sup>†</sup> 永田 仁史<sup>†</sup> 藤岡 豊太<sup>†</sup> 岩崎 聡<sup>††</sup> 針山 孝彦<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> 浜松医科大学耳鼻咽喉科学教室

<sup>†††</sup> 浜松医科大学総合人間科学学生物学教室

E-mail: <sup>†</sup>{h13j18,nagata,toy,abe}@cis.iwate-u.ac.jp

**あらまし** 2つのマイクロホン信号に基づく音源方向推定処理の両耳受聴音に対する有効性を検討するため、4体の頭部モデルの作成とそのHRTFを測定を行い、各モデルのHRTFを用いて音源方向推定精度を調べた。一般化相互相関関数、MUSICなどの一般的な音源方向推定法、および、筆者等が提案している重み付きウィナー利得[1]に関し、方位と仰角の2次元的な方向推定精度を調べた結果、とくに2音源からの音声の到来する環境において、少なくとも1つの音源方向を検出する性能において、本提案方法が大幅に他の方法よりも高精度であることが確かめられた。

**キーワード** 指向性マイクロホン, HRTF, Wiener Gain, 仰角方向推定, 2次元方向推定, GCC

## A study on DOA estimation from binaural sound

Tomita OBARA<sup>†</sup>, Yoshifumi NAGATA<sup>†</sup>, Toyota FUJIOKA<sup>†</sup>, Satoshi IWASAKI<sup>††</sup>, Takahiko HARIYAMA<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> Department of Otolaryngology, Hamamatsu University School of Medicine

<sup>†††</sup> Department of Biology, Hamamatsu University School of Medicine

E-mail: <sup>†</sup>{h13j18,nagata,toy,abe}@cis.iwate-u.ac.jp

**Abstract** We have made four artificial head models and measured these HRTFs to examine the effectiveness of a method to estimate the DOA estimation methods based on two microphone received signals. By using these HRTFs, we examined source detection rate. As a result of two dimensional DOA estimation of azimuth and elevation obtained from GCC, MUSIC and weighted Wiener gain which we have proposed [1], our method can estimate DOA more precise than other methods in the environment where two 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、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] に基づいて音源方向を推定する手法を提案してきた [2]。この方法におけるマイクロホンの配置は、例えば、図1に示す180°の回転対称の場合のように、2個のマイクロホンの仰角が角度差をもつようにすることが重要であり、チャンネル間で仰角方向には振幅差、方位角方向には時間差が生ずるようなマイクロホン配置をとることによって2次元の分解能を得ることができる。

一方、人の頭部伝達関数 (HRTF) は、周波数帯域によっては左右が非対称であり、仰角方向にメインローブがずれていることがある。この場合、2個の指向性マイクロホンの仰角をずらした場合に似た状況となるため、WWGに基づく2次元方向推定法が適用可能であると考えられる。そこで既報 [3] において、方向推定処理におけるマイクロホンの指向性を HRTF データに置き換えて行った方向推定のシミュレーション結果を報告した。このとき HRTF としては、名古屋大学のデータベースの中の「HRTF data (2)」[4] を用いた。この HRTF データの場合、MUSIC, GCC などの従来法と WWG とを音源方向の検出精

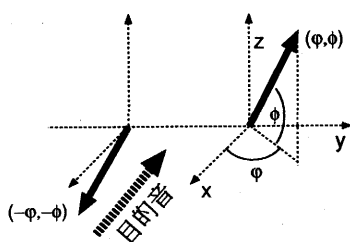


図1 マイクロホン配置

Fig.1 Microphone arrangement

度に関して比較した結果、音源が1個の場合に WWG の方が他の方法よりも高い検出精度が得られた。しかしながら、用いた HRTF は、ヘッドホンによる音源定位等を主目的としていることもあり、外耳入口にマイクロホンを設置して測定されている。音源方向推定を目的とした場合には、外耳道入口ではなく、鼓膜位置の受信信号の方が人間の処理する信号に近いと考えられる。そこで実際に頭部モデルを作成し、鼓膜位置に埋めこんだマイクロホンを用いて HRTF を測定した。頭部モデルの作成においては、耳骨の部分に X 線による CT スキャン、頭部の表面をレーザーによる 3 次元スキャンにより 3 次元データを測定し、これらに基づいて耳介をシリコンゴムで、他の部分を石膏で作成した。

頭部モデルは、4 人に対応する 4 体を作成した。本報では、この 4 体のモデルの HRTF を用いて、シミュレーションにより、2 次元の音源方向推定精度について検討する。

まず、測定した HRTF データの指向性の左右差について検討し、次に、MUSIC、GCC などの従来方向推定法と提案法とを用いて音源方向推定のシミュレーションを行い、音源方向の検出精度に関して比較する。

## 2. 2ch 指向性マイクロホンによる音源方向推定

### 2.1 WWG に基づいた音源方向推定

本音源方向推定法におけるマイクロホン配置の例を図 1 に示す。図中の二つの太い矢印は指向性マイクロホンの方向を表し、 $180^\circ$  の回転対称の位置にある。マイクロホン素子の位置は、ベクトルの始点である。この場合、方向を方位角と仰角の組  $(\theta, \varphi)$  で表すと、1 番目のマイクロホンの方向は  $(\theta, \varphi)$ 、2 番目の方向は  $(-\theta, -\varphi)$  となる。

WWG は、観測信号に対し自動利得制御 (AGC) によって目的信号を近似しようとする際の利得として、次式の重み付き最小二乗法を解いて得られる [2]。

$$J_n = \sum_k \Psi_{n,k} |\alpha_n Z_{n,k} - S_{n,k}|^2 \quad (1)$$

ここで、 $n$  はフレーム番号、 $k$  は周波数成分の番号、 $Z_{n,k}$  は 2ch のマイクロホン信号  $X_{n,k}, Y_{n,k}$  の加算平均値  $(X_{n,k} + Y_{n,k})/2$ 、 $\alpha_n$  は利得、 $S_{n,k}$  は目的信号、 $\Psi_{n,k}$  は重み関数である。

解  $\alpha_n$  は、背景雑音チャンネル間で無相関の場合は

$$\alpha_n = \frac{\sum_k \Psi_{n,k} S_{n,k}^2}{\sum_k \Psi_{n,k} Z_{n,k}^2} \approx \frac{\text{Re}(\sum_k \Psi_{n,k} G_{xy,n,k})}{\sum_k \Psi_{n,k} G_{zz,n,k}} \quad (2)$$

と近似できる。ここで、 $\text{Re}()$  は実部をとる操作、 $G_{xy,n,k}$  は、チャンネル間のクロススペクトル、 $G_{zz,n,k}$  は、両チャンネルの加算平均信号のパワースペクトルである。さらに、相関のある雑音の場合にも対処するため、2ch スペクトルサブトラクション (2ch-SS) [6] に基づいた雑音抑圧の重み関数  $\Phi_{n,k}$  を導入し、 $\alpha_n$  の代わりに

$$\rho(\Psi, \Phi)_n = \frac{\text{Re}(\sum_k \Psi_{n,k} G_{xy,n,k} \Phi_{n,k})}{\sum_k \Psi_{n,k} G_{zz,n,k}} \quad (3)$$

を用いる。  $\Psi$  としては、雑音を近似的に白色化し、周波数軸に沿った平均化による雑音相殺効果を高めるため、次式のように、差信号パワーの逆数を用いる。

$$\Psi_{n,k} = (1/G_{dd,n,k})^\beta \quad (4)$$

$$G_{dd,n,k} = |X_{n,k} - Y_{n,k}|^2 \quad (5)$$

ここで、指数  $\beta$  は、白色化の程度を調節するために導入した指数である。また、 $\Phi$  は、

$$\Phi_{n,k,\gamma} = \frac{|G_{xy,n,k}| - \gamma G_{dd,n,k}/\nu_{n,k}}{|G_{xy,n,k}|} \quad (6)$$

で計算され、 $\nu_{n,k}$  はクロススペクトル中の雑音成分パワーを差信号のパワーから推定するための補正係数であり、雑音区間の差信号のパワーを用いて更新する。しかしながら、音源方向推定の場合は、到来を仮定する方向ごとに雑音区間の決定を行わなければならない。演算量と記憶量が著しく増大するため、音源方向推定処理においては、 $\nu_{n,k} = 1$  で一定とおくこととした。 $\gamma$  はサブトラクションの強さを決める定数である。

式 (3) の WWG は、広帯域の場合おおよそ  $0 \leq \rho \leq 1$  の値をとり、チャンネル間で同位相、同振幅の信号が到来したときに大きくなる。そこで、仮想的な到来方向ごとの WWG を計算し、到来範囲にわたる空間分布を求めてそのピーク位置から音源方向を推定する。仮想的な到来方向に関する計算は、仮想方向からの到来音成分が同位相、同振幅となるようなマイクロホン信号の位相と振幅を補正することにより行う。

仮想方向を方位角と仰角の組  $d = (\theta, \varphi)$  で表し、 $d$  に対応した  $X_k, Y_k$  に関する補正係数を  $A_{x,k}(d), A_{y,k}(d)$  とすると、補正後の受信信号のスペクトルは、

$$X'_k = X_k A_{x,k}(d) \quad (7)$$

$$Y'_k = Y_k A_{y,k}(d) \quad (8)$$

である。なお、以降は簡単化のため、フレーム番号  $n$  の添字を省略する。式 (3) から (6) の各式は、 $G_{xy}, G_{zz}, G_{dd}$  に依存しているので、これらを次式の補正スペクトル  $G'_{xy}, G'_{zz}, G'_{dd}$  に入れ替えて方向  $d$  に関する WWG である  $\rho(\Psi, \Phi, d)$  を求めることにより、仮想方向ごとの平均化処理を省略できる。

$$\begin{aligned} G'_{xy,k} &= \overline{X'_k Y'^*_k} \\ &= G_{xy,k} A_{x,k}(d) A_{y,k}^*(d) \end{aligned} \quad (9)$$

$$\begin{aligned} G'_{zz,k} &= \overline{[(X'_k + Y'_k)/2]^2} \\ &= G_{xx,k} |A_{x,k}(d)|^2/4 + G_{yy,k} |A_{y,k}(d)|^2/4 \\ &\quad + \text{Re}[G_{xy,k} A_{x,k}(d) A_{y,k}^*(d)]/2 \end{aligned} \quad (10)$$

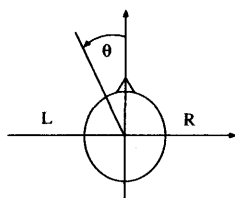


図2 方位角の設定

Fig. 2 Arrangement of the azimuthal angle

$$\begin{aligned}
 G'_{dd,k} &= |X'_k - Y'_k|^2 \\
 &= G_{xx,k}|A_{x,k}(d)|^2 + G_{yy,k}|A_{y,k}(d)|^2 \\
 &\quad - 2\text{Re}[G_{xy,k}A_{x,k}(d)A_{y,k}^*(d)]
 \end{aligned} \quad (11)$$

ここで、(\*)は複素共役、 $G_{xx,k}, G_{yy,k}$ は各々、 $X_k, Y_k$ のパワースペクトルである。また、仮想方向 $d$ に関する補正係数は次式により計算する。

$$\begin{aligned}
 A_{x,k}(d) &= 1/D_x(d) \\
 A_{y,k}(d) &= \exp(j2\pi\tau_d f_s k/N)/D_y(d)
 \end{aligned} \quad (12)$$

ここで、 $\tau_d$ は1番目のマイクロホンを基準にしたときに方向 $d$ からの到来音に関して2番目のマイクロホンに生じる伝搬遅延、 $D_x(d), D_y(d)$ は、 $d$ 方向に関するマイクロホンの相対感度である。

## 2.2 HRTF データへの適用

上記の手法における補正係数 $A_{x,k}(d), A_{y,k}(d)$ をHRTFの逆特性に置き換えることにより、HRTFの場合の音源方向推定処理が行える。HRTFの振幅特性にディップがある場合は、逆特性が不安定となる可能性があるが、ここでは、単純に周波数領域で逆数を計算した。すなわち、HRTF用に修正した補正係数は、

$$\begin{aligned}
 A'_{x,k}(d) &= H_{x,k}^*(d)/|H_{x,k}(d)|^2 \\
 A'_{y,k}(d) &= H_{y,k}^*(d)/|H_{y,k}(d)|^2
 \end{aligned} \quad (13)$$

である。ここで、 $H_{x,k}(d), H_{y,k}(d)$ は、各々、左右の複素HRTF伝達関数である。

今回測定したデータの測定間隔は、仰角は5度、方位角は仰角によって表1の角度で変化させた。これは仰角が $\pm 90^\circ$ 方向に近づくにつれ方位角の測定間隔が狭まるので、多くのデータを必要としないためである。このデータから任意の方向の指向性を得るためデータの補間を行った。補間の方法は種々あるが、ここでは簡単のため、周波数成分ごとにFFT複素スペクトルを線形補間した。また、図2に示すように、被測定者から見て左方向に正の方位角をとるものとした。HRTF測定の際のテスト信号はサンプリング周波数48kHzで8192点のTSP[5]を用い、測定したHRTFのインパルス応答の長さは512点、以降の周波数分析に用いたFFTの点数は512点とした。

仰角	方位角の刻み
+90°	回転なし
+85°, +80°	30°
+75°, +70°	15°
+65°, ±60°	10°
-55° ~ +55°	5°

表1 方位角測定の刻み

Table 1 Azimuth step in the measure unit

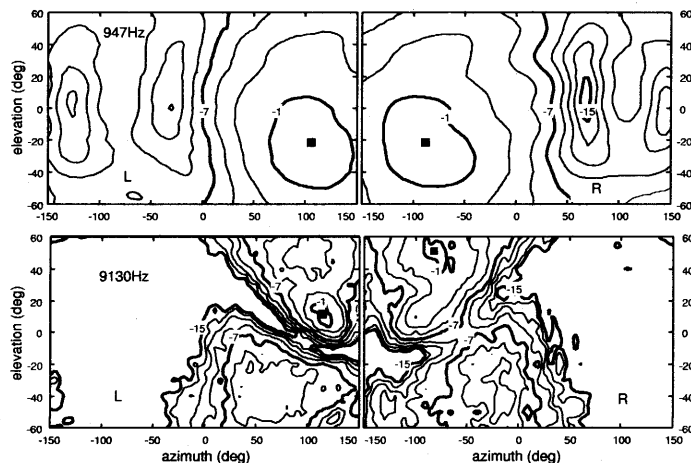


図3 HRTFの振幅の分布

Fig. 3 Spatial distribution of HRTF amplitude

## 3. HRTF 測定データに関する検討

測定したHRTFにおける左右の非対称性を確認するため、HRTFの振幅の分布を計算した。モデルCのHRTFにおける周波数が947Hz、9130Hzの場合の計算結果を図3に示す。図3は、左右各々の最大振幅で正規化してあり、■は振幅最大値を示す方向である。また、最大から-15dB未満の領域の表示は省いてある。

周波数が947Hzの場合、メインローブ付近の左右の感度分布は対称に近いが、9130Hzの場合、メインローブの方向が仰角方向に左右で40度ほどずれていることがわかる。モデルCの場合、全体としては、5kHzより低い周波数では左右は対称で差は小さいが、これより周波数が高くなるほどずれが大きくなっていく。

メインローブの方向の周波数による変化を調べるため、左右各々の感度分布において最大値を示す仰角方向を周波数ごとに求めた。仰角について求めた結果を図4に示す。図4からわかるように、感度が最大となる方向は、4kHz付近から左右間でずれ始め、周波数が高くなるほどずれが大きくなっていくことがわかる。7.5kHz、8kHzあるいはそれ以上の周波数で最大方向が大きく振れるように見えるのは、メインローブがHRTFの測定範囲からはみ出すためである。

さらに、HRTF全体の変化を捕らえるため、HRTFのパワー分布の左右の相対的な差を次式により計算した。

$$E_{LR}(k) = \frac{\sum_d ||H_L(d, k)|^2 - |H_R(d, k)|^2|}{\sum_d (|H_L(d, k)|^2 + |H_R(d, k)|^2)/2} \quad (14)$$

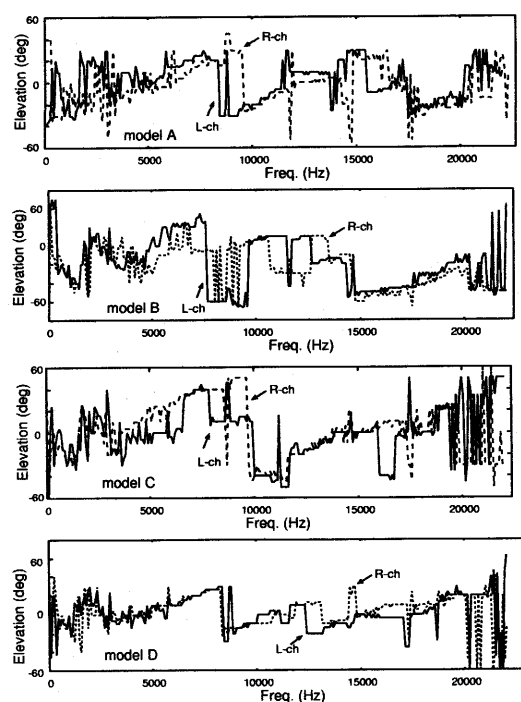


図4 感度最大方向の仰角

Fig. 4 Elevation angle exhibiting maximum amplitude in the HRTF

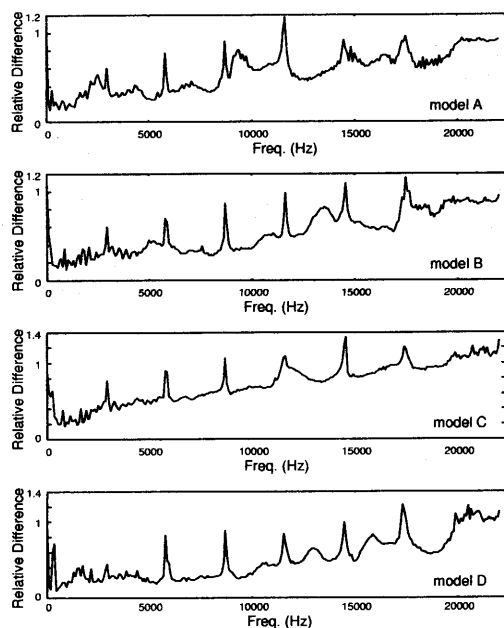


図5 HRTF の分布全体の相対左右差

Fig. 5 Difference measure between L- and R-ch HRTF spatial distribution

この結果を図5に示す。図5からわかるように、左右差は周波数が高くなるにつれ大きくなっており、音声帯域においては個人差があることがわかる。この結果は、先に示した感度最大の方向の傾向とおおよそ一致する。これらの結果から、本手法で仰角方向の分解能を得るには、なるべく高い周波数帯域を用いるのがよいと考えられるが、音声の場合、4~5kHz以上はパワーが小さく、S/Nが低いため、高域の情報だけを使うのは問

題がある。ここでは簡単のため、1kHzから6kHzの帯域を方向推定に用いることとした。

#### 4. 音源方向推定シミュレーション

##### 4.1 音源データとシミュレーション条件

音源信号としては、男性アナウンサーの約1秒の発声を用いた。発声内容は、「メジャーリーグ情報です」と「松井選手は結局」であり、音源を1個としたときは前者を、2個としたときには両方を用いた。背景雑音はチャンネル間で無相関な白色ノイズとし、両音源のS/Nは背景雑音に対し20dBとした。

分析に関しては、FFT長は512点、窓関数はハニング窓、フレーム周期は256点、スペクトル推定のための平均化フレーム数は3(180ms)とした。また、比較に用いた音源方向推定法はWWG, MUSIC, GCCであり、WWGの重み関数 $(\beta, \gamma)$ は $(0.5, 3)$ を用いた。また、GCCに関しては、普通の相互相関数の場合をGCC-0、振幅を正規化して位相だけの情報にした場合をGCC-1、無相関な雑音を最小化する重み関数の場合をGCC-2[7]と呼ぶことにし、これらを用いた。他によく用いられる方向推定法として最小分散法(MV)があるが、予備実験ではMVによる空間スペクトルは、音源方向にピークはできるものの、無関係な方向に著しく大きなピークが複数現れてしまい、方向推定が困難であったため、その記述は省略した。以上の従来法による空間スペクトルは、本提案と同様に、周波数領域において、式(12)の補正ベクトルを用いて計算した。

##### 4.2 音源方向推定精度

次に、音源方向推定精度について検討するため、音源数が1の場合についてその音源方向を多数設定し、空間スペクトルからピー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、HRTFを測定した方向のうちのいずれかと一致するように、 $-180^\circ \leq \theta \leq 180^\circ$ ,  $-55^\circ \leq \varphi \leq 85^\circ$ の範囲で設定した。方向推定のための空間スペクトルは $-60^\circ \leq \theta \leq 90^\circ$ ,  $-180^\circ \leq \varphi \leq 180^\circ$ 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、検出した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。正しい結果の得られた回数の総数に対する比を検出精度とし、音源数を1で既知とした場合の結果を、図6に示すように許容誤差の関数として表した。図は、WWG, GCC-0, GCC-1, GCC-2, および、MUSICに関する結果である。

図からわかるとおり、WWGが最も精度が高く、GCC-0とGCC-1はいずれも低い検出精度だった。またGCC-2はMUSICとGCC-1の中間の検出精度だった。MUSIC以外の方法では、許容角度を大きくするにつれて検出精度が上がるが、MUSICはモデルB以外は70%~90%の一定値となっている。これは、MUSICの場合、正しい方向の周囲に虚像が少ないことを表している一方、離れた方向に大きな虚像が出現する傾向があるためである。

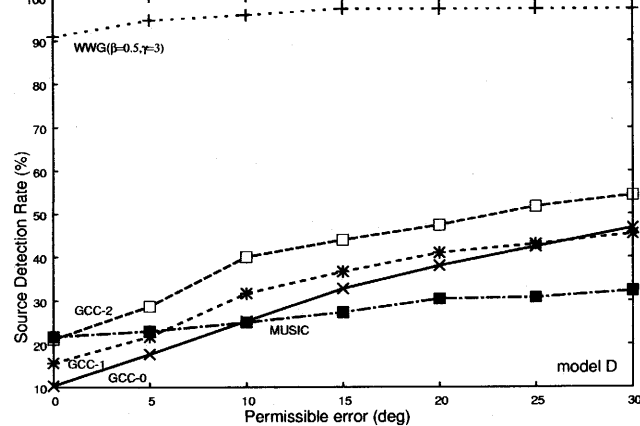
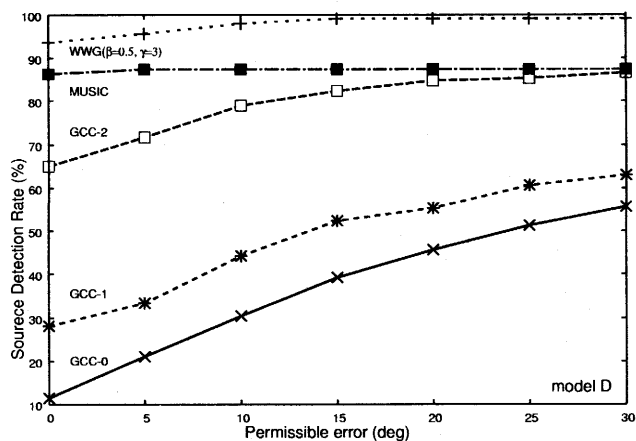
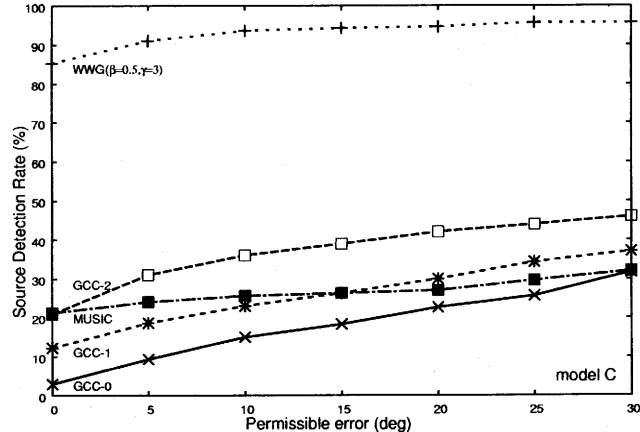
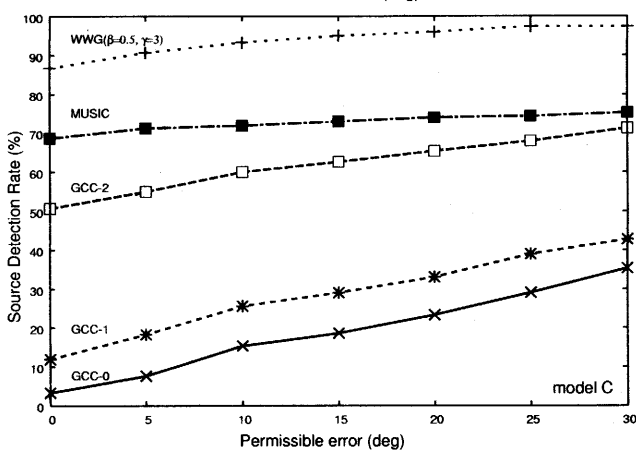
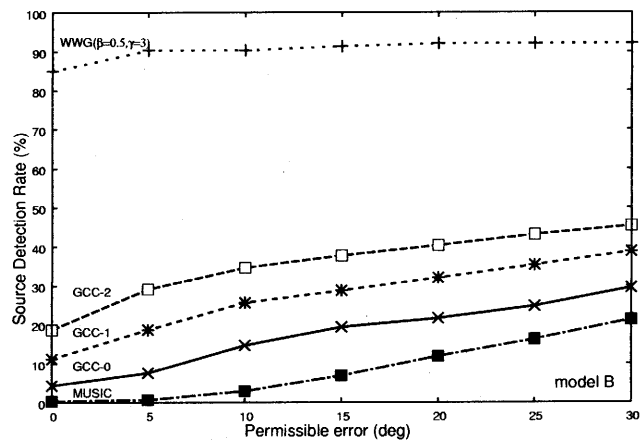
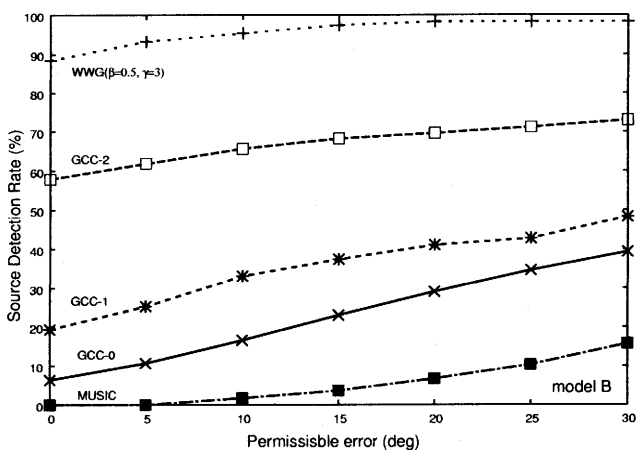
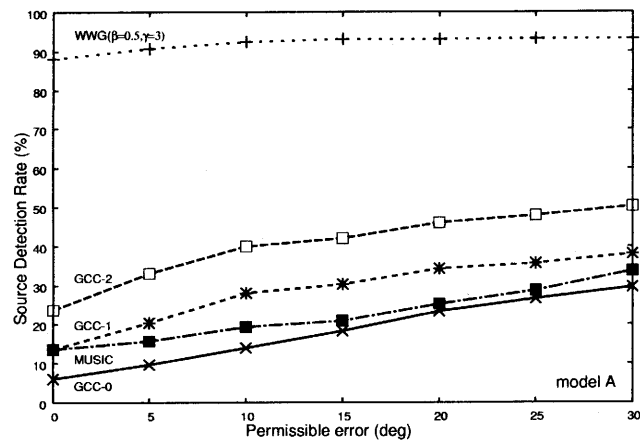
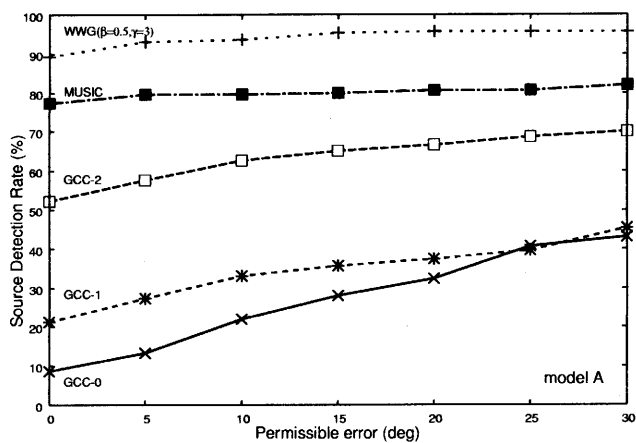


図 6 音源方向推定精度 (音源数 1)  
Fig. 6 Source detection rate(one source)

図 7 最大ピークの音源方向推定精度 (音源数 2)  
Fig. 7 Muximum peak detection rate(two source)

音源数を2とした場合は、空間スペクトル上の最大ピークが2つの音源のいずれか一方として正しく検出された場合について計算した。その結果を図7に示す、これより、2音源中の1音源についてはWWGが他の検出法より大幅に高い精度で検出できることがわかった。なお、設定した音源を2つとも定めた許容誤差範囲内であれば正しく検出できたものとした場合には、他の検出法が数%程度の検出率であったのに比べ、WWGは高い精度ではあったものの10~20%程度の検出精度に止まった。

## 5. ま と め

4個の頭部モデルにおいてHRTFを測定し、重み付きWiener利得に基づいた2次元の方向推定法をこのHRTFデータに当てはめて適用し、音源方向の推定精度について検討した。既提案の重み付きWiener Gainを用いる方法をGCCおよびMUSICと比較した結果、音源が1個の場合は、WWGとMUSICの性能が高く、その左右差によって正中面における仰角分解能が得られることが確認できた。一方、音源が2個の場合、MUSICおよびGCCは2つの音源のうち、いずれも推定できなくなるのに対し、提案法は音源2個のときでも少なくとも1つの音源を90%以上の検出率で推定できることがわかった。

## 謝 辞

本研究を行うにあたり、耳骨のCT測定に協力して頂いた浜松医科大学放射線医学教室・磯田治夫先生に感謝する。なお本研究は、科研費基盤研究(c)(課題番号17500104)の補助金を受けた。

## 文 献

- [1] Yoshifumi Nagata, Toyota Fujioka and Masato Abe, "Speech Enhancement Based on Auto Gain Control", IEEE Trans. on Audio, Speech, and Language Processing, Vol. 14, No. 1, pp.177-190 (2006.1)
- [2] Yoshifumi Nagata, Toyota Fujioka and Masato Abe, "Two-Dimensional DOA Estimation of Sound Sources Based on Weighted Wiener Gain Exploiting Two Directional Microphones", IEEE Trans. on Audio, Speech, and Language Processing, (in publication) 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, pp19-24, 2005-08
- [4] <http://www.itakura.nuee.nagoya-u.ac.jp/HRTF>.
- [5] <http://tosa.mri.co.jp/sounddb/tsp/index.htm>
- [6] H. Y. Kim, F. Asano, Y. Suzuki, and T. Sone, "Speech Enhancement Based on Short-Time Spectral Amplitude Estimation with Two-Channel Beamformer", IEICE Transactions on Fundamentals, Vol. E79-A, No. 12, pp. 2151-2158, Dec. (1996)
- [7] 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, No. 4, pp.320-327, Aug. (1976)