総説

２チャネル信号に基づく到来音方向推定の計算モデル

永 田 仁 史*

工学的な方向推定処理における主要な従来法，および，本稿で新たに導入するチャネル間差信号に基づく一般化相互相関関数について説明し，簡単な性能比較によってこの関数を用いた方法の優位性を確かめる。さらに，従来の聴覚の音源定位の計算モデルについて述べた後，とくにチャネル間差信号に基づく一般化相互相関関数の音源定位モデルとしての予備的検討を行う。これにより，この方法を用いて両耳受聴音からの２次元の定位が可能であり，工学処理の中で高性能であるとされる multiple signal classification (MUSIC) 法と同等の性能であること，また，聴覚モデルに要求されるのと同様，両耳間レベル差のみによって方向推定可能であることを確かめる。聴覚の音源定位モデルとしての提案までにはまだ未検討の課題が多いが，実環境における聴覚の高い定位性能を模擬できるような音源定位の計算モデルとなる可能性がある。

１．はじめに

ヒトには２つの耳でとらえた音から音源の方向や距離を知覚する音源定位の能力があり，その背景メカニズムの解明を目指して音源定位処理の研究が行われている。音源定位のモデル化によるアプローチでは，聴覚の情報処理を人工的に模擬し，モデルから導かれる予測と実際の聴覚の特性との一致を調べる。模擬は信号処理によって行うことから，信号処理アプローチとも称されており，相互相関関数からニューラルネットワークに至る様々な方法による計算モデルが提案されている[5,9,10,30]。従来，これらのモデルにおいては，幅広い音響環境において高い音源定位性能を達成することではなく，単純化した条件において聴覚の特性を説明することに重点が置かれ，複数音源や反射音，背景雑音などの存在する一般的な条件で聴覚の示す高い性能を説明できるようなモデルはまだ提案されていない[5,30]。

一方，動物の研究，とくに昆虫の研究においては，聴覚が昆虫の重要な情報獲得手段であることから，聴覚と行動の関係をめぐる多様な研究がなされている[30,37]。刺激を与えて反応を観察する行動生理学的なアプローチがある一方で，生体の特徴を備えたロボットを作成し，行動を模擬することによって処理の実体に迫るアプローチもある。例えば，コオロギの生体模擬ロボット[37,pp.791-796]においては，聴覚の役割を担う内部処理としてニューラルネットワークによる音源追跡処理が用いられている。生物模擬ロボットは，信号処理アプローチの発展形と見ることもできる。

ところで，音源定位は，生物による到来音方向推定であるといえ，その信号処理としての手法は工学に求めることができる。工学では，音声や環境音などの可聴音よりも，むしろ，通信や探査に必要な電波や水中伝播音の到来方向推定において長い研究の歴史がある[16,29,35]。可聴音を対象とした方向推定に関しては，近年，ロボットや音声対話システムにおいて，発話者や雑音源などの位置情報の重要性が認識され，ロボット聴覚と称される研究の１テーマとして注目を集めている[14,28]。到来音方向推定には，電波や水中音の場合と同じ信号処理が用いられているが，一度に処理すべき帯域幅が広いことや，音声のような極端な非定常音も対象となることなど，可聴音特有の問題がいくつもある。試作されているロボットは，このような状況でも性能を維持するため多数のセンサを備え，方向推定には高分解能法とよばれる高性能な処理を用いていることが多い[3,14]。これに対し，最近では，聴覚心理学の知見を工学へ応用する動きが現れ[7,28,36]，先に述べたロボット聴覚の研究においても，少数のセンサを用いた手法が注目されるようになっている。

以上に述べたように，音源定位モデルの研究においては，様々な信号処理が用いられており，研究の進展に伴ってさらに進んだ工学手法が使用されてゆく可能性がある。また，同時に，ヒトや動物の聴覚研究の知見を工学処理に導入する流れもあり，今後も両分野相互のインタラクションによって研究が進展してゆくに違いない。そこで，本稿は，工学研究者の立場から音源定位モデルに関係の深い２チャネルの工学的方向推定処理をとりあげ，代表的といえる一般化相互相関関数 (generalized cross-correlation function (GCC)) [18] と高分解能法について説明する。GCCは相互相関関数の改良版であるが，元の相互相関関数は，従来から聴覚の音源定位の基本的な計算モデルとして用いられ，多くの検討がなされている。これに対し，高分解能法は工学的な面が強いためか聴覚処理として検討された例は見当たらないものの，有色性の非定常音源の場合や高レベルの雑音下など，より一般的な状況における性能が高い。さらに，本稿では，新たな手法として，筆者の最近報告した方向推定法[24-26]の中核処理部分であるチャネル間差信号に基づく GCC (差信号 GCC) を説明し，これらの手法の性能をシミュレーションによって比較する。次に，この方法の聴覚の音源定位モデルとしての可能性について検討するため，頭部伝達関数を用いた２次元の音源方向推定において，逐次的な複数音源推定処理を本提案法に適用し，高分解能法と性能を比較する。また，高域において聴覚がチャネル間の時間差情報を獲得できず，振幅情報のみによって推定処理していることに対応し，差信号 GCCにおいても振幅情報のみによる推定が可能であることを示す。

２．音源方向推定の工学的手法

図１に示すように，２個のセンサが間隔 r だけ離れて位

*Yoshifumi NAGATA，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4-3-5）

置し，方向θからの到来音をセンサで観測する。反射波はなく，音源は十分遠方に位置して到来信号は平面波とみなせるものとする。図中のOは便宜的に定めた信号観測の基準点であり，音源信号からセンサ信号までの伝達関数を考える代わりに，Oにおいて観測される信号とセンサ信号との間の伝達関数を考える。反射に関する対策は，屋内の音響処理では重要な問題であり，工学処理としては，信号の立ち上がり区間の検出に基づいた方法[34]や線形予測に基づいた方法[12]などが報告されており，一方，聴覚の音源定位モデルでは先行音効果をモデルのメカニズムに含めて説明する場合があるが，本稿の課題としては扱わない。また，センサは必ずしも離れた位置に置く必要はなく，同じ位置に指向性の向きを変えた複数のセンサを置く場合があり，このようなシステムを vector sensor array[19,27]とよぶ。vector sensor の処理は，センサ間の距離が小さい小動物の音源定位や，後に述べるように，高域におけるヒトの聴覚の音源定位にその枠組みを当てはめることができる。

２－１．時間差検出と一般化相互相関関数

図１において，音速をｃ，基準点Oにおける観測信号に対する各チャネルの信号の遅延をτ_1，τ_2とする。また，チャネル間の遅延時間差を$\tau = \tau_1 - \tau_2$とすると，$\tau = r\sin\theta/c$であるため，時間差τから到来方向θが求まる。時間差検出には，次式の相互相関関数がよく使われる。

$$\gamma(m) = E\left[x_1(i+m)x_2(i)\right] \qquad (1)$$

ここで，mは，サンプリング周期Tを単位としたタイムラグ，$x_n(i)$ (n=1,2) は観測信号，E[ ]は期待値演算を表す。有限個のサンプルから上式を計算する方法はいくつかあるが，例えば，切り出したNサンプルの区間に対し，

$$\gamma(m) = \frac{1}{N}\sum_{i=0}^{i\le N-m-1} x_1(i+m)\cdot x_2(i) \qquad (2)$$

とするシンプルな計算法がよく用いられる。上式は，計算に用いられるサンプル数がタイムラグmに反比例するため，重み1/(m+1)が$\gamma(m)$にかかることになるが，値の信頼性から見てこの重みは妥当であるとされている。図１の場合，$\gamma(m)$は，チャネル間の時間差に相当するタイムラグの位置で最大となるため，関数上のピーク位置から時間差を得ることができる。しかし，音声のように有色性の強い信号の場合，ピークがブロードとなると同時に疑似ピークが出現して推定精度が低下する。そこで，次式のように，周波数領域において入力信号を白色化し，逆離散フーリエ変換によって相互相関関数を求める。これを一般化相互相関関数 (GCC) [18]とよぶ。

$$\gamma_{GCC}(m) = \mathrm{IDFT}\left[\overline{X_{1,k}X^*_{2,k}}\cdot\Psi_k\right] = \mathrm{IDFT}\left[Q_{12,k}\Psi_k\right]$$
$$= \frac{1}{L}\sum_k \overline{X_{1,k}X^*_{2,k}}\cdot\Psi_k\cdot e^{j2\pi km/L} \qquad (3)$$

ここで，kは離散化周波数，$X_{n,k}$ (n = 1,2) は$x_n(i)$の離散フーリエ変換 (DFT)，IDFT[ ]は逆離散フーリエ変換，Ｌは DFT 点数，Ψ_kは白色化のための重み関数，$Q_{12,k} = \overline{X_{1,k}X^*_{2,k}}$はチャネル間のクロススペクトルである。また，$\overline{[\ ]}$は時間平均化であり，処理対象区間の短時間 DFTから求めた計算値について，処理対象区間をずらしながら平均化する操作を意味する。重み関数Ψ_kに関しては，

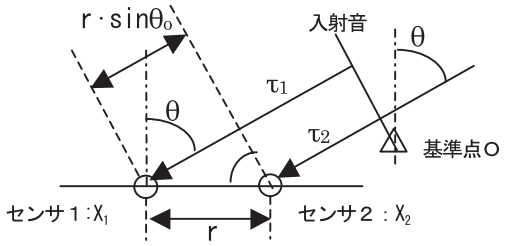

図１　センサと到来音の関係

$$\Psi_k = \frac{1}{|Q_{12,k}|} \qquad (4)$$

とし，全周波数成分の絶対値を強制的に１に揃える方法をPhase Transfer (PHAT)，チャネル間で無相関な雑音の影響を最尤推定法 (ML 法) に基づいて最小化するように重み関数Ψ_kを

$$\Psi_k = \frac{\Gamma_k}{1-\Gamma_k}\frac{1}{|Q_{12,k}|},\ \Gamma_k = \frac{|Q_{12,k}|^2}{Q_{11,k}Q_{22,k}} \qquad (5)$$

とする方法を最尤推定に基づくGCCとよぶ[18]。(５) 式で，$Q_{nn,k} = \overline{|X_{n,k}|^2}$ (n = 1,2) は，ｎチャネル目の信号のパワースペクトルである。

ここで，相互相関関数と GCCの例を図２に示す。図２は，図１のセンサ配置においてセンサ間隔rを１ｍとし，方位角$\theta = 5°$と$-30°$の２つの方向から異なる信号が到来している場合に得られた結果である。図２において，上の図には，信号が白色雑音の場合と音声の場合の相互相関関数を重ねて示してあり，いずれも信号の到来時間差に相当するタイムラグの位置にピークが現れているが，音声の場合のピークは幅が広いため時間分解能が低いことがわかる。また，図２における下の図は，同じ音声に関する普通の相

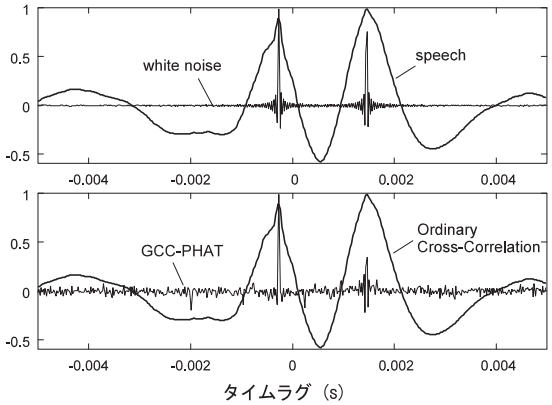

図２　相互相関関数と GCC　上は到来音が白色雑音の場合と音声の場合各々の相互相関関数，下は到来音が音声の場合の相互相関関数と一般化相互相関関数 (GCC-PHAT) である。到来する２つの信号の時間差に相当する時間遅れの位置にピークがある。音声の相互相関関数のピークはブロードであるため分解能が低いが，GCC-PHATにおいては分解能が改善されていることがわかる。

互相関関数とGCC-PHATの計算結果であり，白色化によって分解能が向上していることがわかる。

2－2．空間スペクトル

パワーなどの音源の属性と方向との関係を表したデータを角度スペクトル，または，空間スペクトルとよぶ。空間スペクトル上のピークをもって音源であると判断するため，真の音源方向に際立ったピークが出るような処理法が有利である。相互相関関数やGCCの場合，直接求まるものは時間差であり，方向推定性能が関数上のピーク選択方法によって影響を受け，他の方法との比較評価が難しくなるため，空間スペクトルに変換して扱うこととする。この場合，GCCの成分をそのタイムラグに相当する到来方向の上に移せばよいが，方向に関するサンプリング点はタイムラグのサンプリング点と通常一致しないため，時間領域の時間シフトが周波数領域の位相変化に対応することを用い，周波数領域において処理する。

まず，（3）式から，

$$\gamma_{GCC}(m+\nu) = \Sigma_k \overline{X_{1,k}X^*_{2,k}}\,\Psi_k \cdot e^{j2\pi k(m+\nu)/L} \qquad (6)$$

である。上式のνは，サンプリング周期Tの実数倍で表された遅延であるとする。すなわち，$\nu = \tau/T$である。ここで，$m = 0$とおき，$\nu_n = \tau_n/T$ $(n = 1,2)$，$\nu = \nu_1 - \nu_2$を使うと，

$$\gamma_{GCC}(\nu) = \Sigma_k \overline{X_{1,k}X^*_{2,k}}\,\Psi_k \cdot e^{j2\pi k(\nu_1-\nu_2)/L} = \Sigma_k \overline{X_{1,k}e^{j2\pi k\nu_1/L} \cdot \left[X_{2,k}e^{j2\pi k\nu_2/L}\right]^*}\,\Psi_k \qquad (7)$$

であるので，チャネルごとに基準点に対する遅延分だけ補正したクロススペクトルを用いて任意の遅延時間に対応したGCCの値を計算できることになる。したがって，ある注目する方向をθとすると，θからν_1，ν_2が決まるので，これに応じて（7）式から$\gamma_{GCC}(\nu = r\sin\theta/cT)$を計算すれば空間スペクトルを求められる。$\theta$は，look-directionとよばれる。（7）式の値は一般に複素数であるが，θが音源方向と一致した場合には実数となってその方向で極大となるので，実部だけを考えれば十分である。

以上では，基準点からセンサへの伝達関数として遅延だけを考えたが，センサが指向性をもつ場合などは振幅の変化も考慮する必要がある。そこで，基準点からセンサへの振幅も含めた伝達関数を$A_{n,k}(\theta)$ $(n = 1,2)$とすると，上述のスペクトル補正値はその逆数$A^{-1}_{n,k}(\theta)$であり，これを用いてGCCの空間スペクトルの一般形が次式のように表わされる。

$$S_{GCC}(\theta) = \overline{\Sigma_k \mathrm{Re}\left[G_{12,k}(\theta)\right]\Psi_k(\theta)} \qquad (8)$$

$$G_{12,k}(\theta) = \overline{X_{1,k}A^{-1}_{1,k}(\theta) \cdot X^*_{2,k}\left[A^{-1}_{2,k}(\theta)\right]^*} = Q_{12,k} \cdot A^{-1}_{1,k}(\theta)\left[A^{-1}_{2,k}(\theta)\right]^* \qquad (9)$$

ここで，Re[ ]は 複素数の実部をとる操作，$G_{12,k}(\theta)$は，方向θに対応して補正されたクロススペクトルである。また，$A_{n,k}(\theta)$ $(n = 1,2)$をベクトルにまとめた

$$\boldsymbol{\alpha}_k(\theta) = \{A_{1,k}(\theta), A_{2,k}(\theta)\}^T \qquad (10)$$

をステアリングベクトル，または，モードベクトルと呼ぶ。例えば，基準点を1ch目のセンサ位置とした場合，2ch目は1ch目に対して$r\sin\theta/c$だけ先行しているので，ステアリングベクトルは次式となる。

$$\boldsymbol{\alpha}_k(\theta) = \left\{1, e^{j2\pi\frac{r\sin\theta}{cT}\cdot\frac{k}{L}}\right\}^T \qquad (11)$$

2－3．高分解能法

次に，高分解能法の中でよく使用される2つの方法を説明する。

a）最小分散法（minimum variance method（MV））[8,16]

高分解能法では，次式のように，入力信号をチャネルごとにフィルタリングし，その和を出力する処理を仮定する。

$$Y_k = h^*_{1,k}X_{1,k} + h^*_{2,k}X_{2,k} = \boldsymbol{h}^H_k\boldsymbol{X}_k \qquad (12)$$

ここで，$\boldsymbol{h}_k = \{h_{1,k}, h_{2,k}\}^T$はフィルタ係数ベクトル，$\boldsymbol{X}_k = \{X_{1,k}, X_{2,k}\}^T$は入力信号ベクトル，$[\ ]^H$は転置して複素共役をとる操作である。このとき，look-directionθからの到来音を変化させずに出力信号のパワーを最小化するようなフィルタ$\boldsymbol{h}^{opt}_k$をθごとに求め，同時に$\boldsymbol{h}^{opt}_k$を使った場合の出力パワーを空間スペクトルとする方法が最小分散法である。出力の平均パワーは，

$$\overline{\left[Y_k(\theta)\right]^2} = \overline{\boldsymbol{h}^H_k(\theta)\boldsymbol{X}_k\boldsymbol{X}^H_k\boldsymbol{h}_k(\theta)} = \boldsymbol{h}^H_k(\theta)R_k\boldsymbol{h}_k(\theta) \qquad (13)$$

となる。ここで，$R_k = \overline{\boldsymbol{X}_k\boldsymbol{X}^H_k}$は入力信号の相関行列である。方向$\theta$から音が到来したとして，基準点における到来音スペクトルを$X_{o,k}$とすると，$\boldsymbol{X}_k = \boldsymbol{\alpha}_k(\theta)X_{o,k}$であるので，$X_{o,k}$を変化させずにフィルタリングして出力する条件は，$Y_k = \boldsymbol{h}^H_k(\theta)\,\boldsymbol{\alpha}_k(\theta)X_{o,k} = X_{o,k}$，すなわち，

$$\boldsymbol{h}^H_k(\theta)\boldsymbol{\alpha}_k(\theta) = 1 \qquad (14)$$

である。この拘束条件の下で（13）式を最小化する問題をラグランジュの未定係数法を用いて解くことにより，最適なフィルタは，

$$\boldsymbol{h}^{opt}_k(\theta) = \frac{R^{-1}_k\boldsymbol{\alpha}_k(\theta)}{\boldsymbol{\alpha}^H_k(\theta)R^{-1}_k\boldsymbol{\alpha}_k(\theta)} \qquad (15)$$

と求められる。これを再度（13）式の$\boldsymbol{h}_k(\theta)$に代入すると，MV法の空間スペクトルが次式のように求められる。

$$S_{MV,k}(\theta) = \frac{1}{\boldsymbol{\alpha}^H_k(\theta)R^{-1}_k\boldsymbol{\alpha}_k(\theta)} \qquad (16)$$

b）MUSIC（multiple signal classification[16,33]）法

MUSIC法においては，MV法における出力パワー最小化の拘束条件である（14）式が

$$\boldsymbol{h}^H_k(\theta)\boldsymbol{h}_k(\theta) = 1 \qquad (17)$$

に変わる。これは，フィルタの大きさを1に保つことを意味する。この結果，拘束条件付き最小化は，

$$R_k \boldsymbol{h}_k(\theta) = \lambda \boldsymbol{h}_k(\theta) \quad (18)$$

の固有値問題となり，これを解いて得られる R_k の固有ベクトル $\boldsymbol{e}_k$ が求めるフィルタ係数である。$\boldsymbol{e}_k$ はチャネル数個求められるが，そのうちで最小の固有値に対応するものを $\boldsymbol{e}_{min,k}$ とし，

$$S_{MN,k}(\theta) = 1/|\boldsymbol{e}^H_{min,k} \boldsymbol{\alpha}_k(\theta)|^2 \quad (19)$$

によって空間スペクトルを計算する方法を最少ノルム法という。分母はフィルタリングの出力パワーであるが，その逆数の演算は，音源方向において鋭い正のピークが生じるようにするための便宜的な処理であるといえる。

R_k の固有値は，通常，背景雑音に対応するものと，到来信号に対応するものとに分けることができ，雑音と信号に対応する固有ベクトルの張る空間を，各々，雑音部分空間，信号部分空間とよぶ。MUSIC 法は，$\boldsymbol{e}_k$ のうち，雑音部分空間に属する複数の固有ベクトルを用いて上式の分母を平均化する方法である。しかし，2ch システムの場合は，最小の固有値に対応した固有ベクトル 1 個を用いることができるだけなので，最小ノルム法と同じ処理となる。ただし，ステアリングベクトルの大きさを補正した次式が用いられる。

$$S_{MUSIC,k}(\theta) = \boldsymbol{\alpha}^H_k(\theta) \boldsymbol{\alpha}_k(\theta) / |\boldsymbol{e}^H_{min,k} \boldsymbol{\alpha}_k(\theta)|^2 \quad (20)$$

MV 法と MUSIC 法は，似たような条件から導かれた方法であるが，一般に MUSICの方が高性能であるとされている。

高分解能法におけるパワー最小化は，look-direction 以外からの到来音にフィルタの指向性の谷を向けることによってなされている。(12) 式からわかるように，フィルタリングは入力信号ベクトルとステアリングベクトルの内積であるので，指向性の谷を到来方向に向けるフィルタは，到来音方向のステアリングベクトルと直交するようなフィルタであることがわかる。また，N個の到来音があるときには，対応するN個のステアリングベクトルすべてに同時に直交するフィルタを求める必要があるが，そのようなフィルタは，一般に，ベクトルの次元数がNより大きい場合でないと存在しない。すなわち，音源数よりセンサ数の方が多くなければならない。

2－4．差信号 GCC

筆者が提案した重み付き Wiener 利得[24,25] に基づく音源方向推定法は，観測したクロススペクトルに対し，スペクトル減算[4,17] による雑音低減と差信号パワーに基づく分解能向上処理を組み合わせ，さらに正規化して Wiener 利得を求める処理であり，複数の要素が関係するため働きが分かりにくい。そこで本稿では，重み付き Wiener 利得においてスペクトル減算と正規化処理を除き，単純化した残りの部分からなる処理を差信号 GCCとよび，これについて検討する。差信号 GCCの空間スペクトルは次式によって計算する。

$$S_{DIF}(\theta) = \overline{\sum_k \mathrm{Re}\left[G_{12,k}(\theta)\right] / G_{dd,k}(\theta)} \quad (21)$$

ここで，

$$G_{dd,k}(\theta) = \overline{|X_{1,k} A^{-1}_{1,k}(\theta) - X_{2,k} A^{-1}_{2,k}(\theta)|^2} \quad (22)$$

は，両チャネルの観測信号を補正して求めた基準点信号の差のパワーである。この関数は，look-direction θ と到来音方向とが一致した場合に小さな値となるため，その逆数 $1/G_{dd,k}(\theta)$ は到来方向において非線形に増大し，分解能向上に大きく寄与する。クロススペクトルを除いたこの関数単独の働きについても明らかにするため，次節の性能比較において，

$$S_{PSI,k}(\theta) = \overline{\sum_k 1/G_{dd,k}(\theta)} \quad (23)$$

に関しても検討することとする。

2－5．工学的手法の性能

上述した方向推定法について，性能比較のシミュレーションを行った。図 1 のように，無指向性のセンサを 2 個 15㎝間隔で置き，$-90° \leq \theta \leq 90°$ の範囲から音が到来するものとした。性能比較は，この条件でランダムに音源方向を変えながら方向推定の試行を多数回行い，各方向推定法の推定結果が正解と一致する率を調べることによって行った。音源数は 1 個から 3 個で既知とし，空間スペクトルにおいて，大きさ順に音源数個のピークを検出して正解と比べた。このとき，ピークと対応する正解との誤差がすべての音源に関して 5° 以内であればその試行は成功したものとし，成功回数／試行回数を検出率とした。試行回数は500とした。音源は男性の音声とし，音声の内容は音源と試行ごとに変えた。また，各試行の信号長は 2 秒とし，背景雑音はチャネル間で無相関な白色雑音とした。音源が複数の場合は，音源の平均パワーが等しくなるように振幅を調整した。 サンプリング周波数は48kHz，計算に用いた周波数帯域は260Hz～4kHzである。周波数分析は2,048点の高速フーリエ変換をフレーム周期1,024点で行い，窓関数にはハニング窓を用いた。

前節で述べたように，センサが 2 個の場合，高分解能法が推定可能な音源数は 1 個である。しかし，音源が音声の場合は，強いパワーの成分が時間―周波数上に疎に分布するスパース性のため，1 個の音源の成分を主成分とする周波数が多く存在する。このため，各周波数においては音源 1 個を仮定して推定した後，複数の周波数に亘る空間スペクトルの平均化によって複数の音源方向が推定できる。このとき，複数の音源の成分を含んだ周波数成分は推定精度低下の原因となるので，このような成分を除外するため，高分解能法については固有値の比に基づいて周波数成分の選択を行った[22]。この成分選択法は，音源定位モデルにおいて Fallerらが用いた両耳間コヒーレンスに基づく方法[11] と等価である。結果を図 3 に示す。

図 3 を見ると，通常の相互相関関数と従来の 2 つの GCC の PHATと MLは，音源が 1 個の場合の低 S/N 時と音源が 2 個と 3 個の場合に大幅に性能が低下し，一方で，MUSIC，MV，および，「DIF」で示した曲線の差信号 GCC は，S/N 低下と音源数増加に伴う性能低下が緩やかであることがわかる。 このシミュレーションでは，差信号 GCC が一貫して性能が最も高かった。なお，「PSI」で示した差信号パワーの逆数による方法 ((23) 式) の性能は，高分解能法と従来の GCCの中間程度であった。この結果から，差

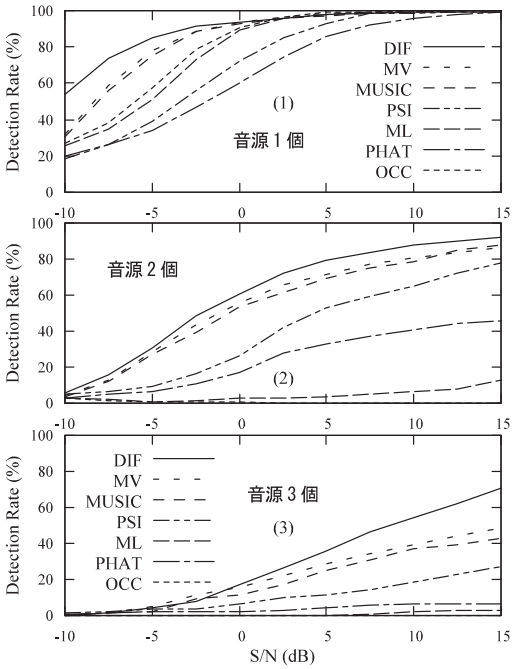

図３　音源方向推定手法の性能比較結果（音源：音声） 曲線名のDIFは差信号スペクトルに基づくGCC，PSIは差信号パワーの逆数，MUSICはMUSIC法，MVは最小分散法，MLは最尤推定に基づくGCC，PHATはPHAT-GCC，OCCは通常の相互相関関数を表す。音源が複数になると、MUSIC，MV，WWGによる方法の優位性が高くなる。

信号パワーの逆数は単独では性能が低いが，GCCの重み関数として用いることによってそのGCCの性能が大幅に高まることがわかる。なお，音源が白色雑音の場合は，どの方法の性能も普通の相互相関関数と大差なく，S/N = − 5 dB 以上で検出率はほぼ一定値となり，音源１個のときに98%以上，音源２個のときに約65%，音源３個のときに約25%となった。

３．聴覚の音源定位モデルに基づく方法

３−１．頭部伝達関数と２次元方向推定

ここまでは，方位角θだけの一次元の方向推定に限定してきたが，聴覚には，方位角だけではなく，方位と仰角の２次元の定位，さらには，距離感覚も含めた３次元の定位能力があるとされている。また，聴覚の場合は，自由空間における受音の場合と違い，頭や耳介などによる回折，反射の影響を受けて到来音の到達時間と振幅が変化する。この変化は到来方向の複雑な関数であり，頭部伝達関数（head-related transfer function（HRTF））と呼ばれている。距離感覚についてはここでは扱わず，伝達系としてHRTFを想定した方位θ―仰角ϕの２次元の方向推定について考えることとする。この場合，方向は，図４に示すように，極座標を用いて d =（θ，ϕ）によって表すものとし，頭の正面方向を（0°，0°）とおくことにする。

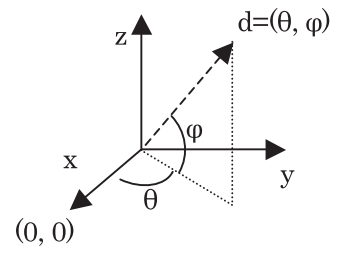

図４　極座標による２次元方向の表現

３−２．定位キューと２次元モデル

観測信号中の音源定位をもたらす情報を音源定位キューという。音源定位キューと音源定位モデルについては長い研究の歴史があり，詳細については，例えば，Popperらによる成書[30]に系統的にまとめられているので，ここでは，これを参考にして２次元の音源定位に関してのみ簡単に述べる。

音源定位キューとして広く認められているのは，両耳信号間の到来時間差（interaural time difference（ITD））とパワーレベル差（interaural level difference（ILD））を指す両耳間差キュー，および，観測信号のパワースペクトルにおける特徴を指すスペクトルキューである。ITDは，高い周波数帯域において，位相のあいまい性と聴覚の符号化の時間分解能の問題から精度が低下するため，包絡信号の時間差が高域のITDとして使われることもあるとされる。ただし，一定振幅の正弦波を重畳した信号のように，包絡信号が直流となる場合はITDは求められないことになる。また，duplex theory[30, pp.286]によると，２〜３kHzを境に，ITDは低域側で，ILDは高域側で優位になるとされている。一方，スペクトルキューは，HRTFの形状を反映した到来音の音質変化であり，仰角の定位において重要視されている。しかし，スペクトル上の微細な構造は定位に無関係であるとされ，何が実質的なキューなのかについて様々な仮説がある。現在のところは，HRTF上で仰角に対応して変化するいくつかの大きなノッチとピークの位置がキューとして有力視されている[13]。しかし，ノッチやピークは，信号スペクトルが既知であるか，あるいは，キューの存在する帯域の信号スペクトルが平坦で滑らかであるかなどの条件がないとキューとしての抽出が難しいという問題もある[30, pp. 297]。

２次元定位のモデルとしては，ITD，ILDとスペクトルキューの組み合せ[21,23]が提案されている一方で，スペクトルキューが存在しないとされる３kHz以下の帯域だけでも正中面から離れるに従って仰角の音源定位精度が高くなること[2]や，正中面においてもILDが存在することが示されており[31,32]，両耳間差キューだけを用いる２次元の音源方向推定モデルも報告されている[6,15,20]。例えば，Martin[20]の方法は，HRTFから学習した両耳間差キューをテンプレートとし，観測された両耳間差キューとの照合を尤度関数により行う。HRTFを用いた工学的な２次元推定については，筆者らも性能比較の結果を報告しており[25]，本稿では，この方法に基づいた２次元の定位に注目することとする。

３−３．音源定位の信号処理

音源定位モデルの中で，２章に述べた工学的な方向推定処理と関係が深いのは，Jeffressの相互相関モデルとDurlachの等化打消説（Equalization-Cancellation Theory）[5,9,30]で

ある。相互相関関数に関しては，聴覚処理モデルでは聴覚心理学の知見に基づいた様々な修正が行われているが，時間差を求めるための処理であることには変わりがない。また，等化打消説の処理は，両耳信号間の時間差と振幅差を補正して，両耳信号が最も一致するときの方向を求める。２－３節で述べた工学的な高分解能法は，look-direction以外からの到来音の最小化が処理の要であり，等化打消説と同じ意味の処理を行っているが，音源らしさの指標である空間スペクトルをフィルタの出力パワーやその逆数として与えるところが重要であり，これによって複数帯域にわたる平均化などの演算が容易になる。差信号GCCも，等化打消説と同じ意味の処理を行うが，高分解能法と同様，空間スペクトルをチャネル間差信号パワーの逆数として評価するところが重要である。なお，高分解能法の場合，計算に必要な固有値展開や逆行列演算は聴覚の処理として想定しにくいが，差信号GCCの場合，必要な演算は，相互相関とチャネル間差信号パワーの逆数であるので，演算は高分解能法よりだいぶ簡単であり，聴覚の計算モデルの候補としては有利であるといえる。

４．差信号GCCの２次元音源定位性能

本章では，差信号GCCの聴覚処理モデルへの適用可能性を検討する。聴覚処理モデルとして提案するには，まず，受聴音を分析する聴覚末梢系を信号処理によって模擬し，評価する必要があるが，ここでは予備的検討として，工学処理の枠組みで評価する。音声処理の場合，聴覚末梢系の処理によって一般的な工学的手法よりも良い結果が期待できることがある[1)]との考えもあり，工学的枠組における評価も無視できないものと考えている。

４－１．逐次法による複数音源推定

両耳受聴音からの方向推定においては，測定したHRTFを(10)式のステアリングベクトルとして用いることによって工学的な処理が適用可能となる。すなわち，左右のHRTFを各々，$H_{1,k}(d)$，$H_{2,k}(d)$，$(d = (\theta, \phi))$とすると，ステアリングベクトルとして，

$$\boldsymbol{\alpha}_k(d) = \{H_{1,k}(d),\ H_{2,k}(d)\}^T \qquad (24)$$

を用いればよい。しかし，両耳受聴音からの方向推定においては，空間スペクトルの形状が複雑になり，真の音源方向以外に疑似ピークが複数出現し、マイクロホンを自由空間に置いた場合よりも方向推定が困難な状況となる。これは，HRTFが複雑な形状をもち，異なった複数の方向が近い伝達関数をもつ場合があるためである。２個の音声が到来している場合に差信号GCCによって計算した２次元の空間スペクトルの例を図５に示す。図５の（ａ）は(21)式を方向$d = (\theta, \phi)$に関する式に拡張し，２次元の空間走査を行って得られた空間スペクトルであり，図中の矢印の方向に音源ＡとＢがある。計算に用いた周波数帯域は，260Hzから４kHzである。図５（ａ）においては，到来時間差の等しくなる矢状面に沿って多数のピークが出現し，その中に音源のピークがある。また，音源Ｂのピークは音源Ａの近傍にある疑似ピークよりも低いため，大きさ順のピーク検出では見落とすことになる。このように，複数の

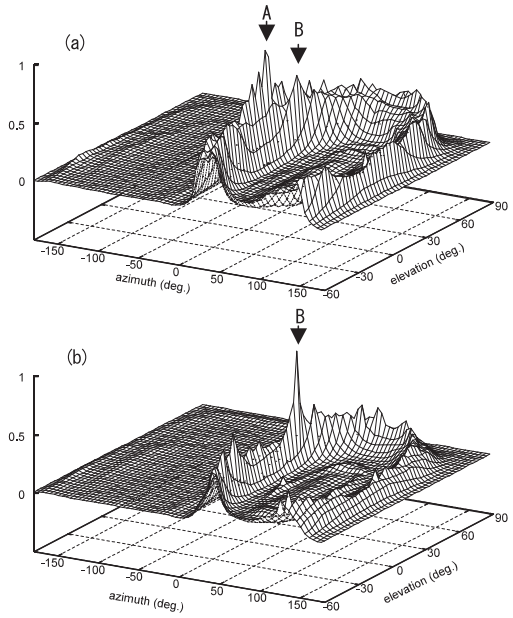

図５　両耳受聴音からの差信号に基づいたGCCの２次元空間スペクトル　（ａ）は初期スペクトル，（ｂ）は音源Ａのスペクトル減衰後のスペクトルである。初期スペクトルには疑似ピークが多数出現するため，音源Ｂの検出は困難であるが，音源Ａのスペクトル減衰後は（ｂ）のように最大ピークとなって検出可能になる。

音源の存在する環境では，最初の空間スペクトルから複数個の音源方向を推定することは難しい問題となる。この問題に対して，工学処理では，周波数ごとに得られるピーク位置をクラスタリングによってまとめる方法[22)]があり，一方，音源定位モデルにおいては，汎用モデル[5,30)]に従い，周波数ごとに得られたITDとILDをベクトルとしてパタン照合を適用するなどの方法があるが，２次元の場合，狭帯域の空間スペクトル上には音源方向の疑似候補が多数生じるため，扱いは容易ではない。これに対し，この問題は以下に述べる逐次法によって解決が可能である[25,26)]。

この方法は，(21)式の初期スペクトルを出発点とし，その最大ピークを第一の音源であるとみなし，このピークを構成する周波数成分を減じるような重み関数を求めて周波数ごとに(21)式のクロススペクトルに乗じ，再度，空間スペクトルを計算して最大ピークを２番目の音源とするものである。得られた空間スペクトルに対して同様に反復を重ね，反復終了の基準に達するまで計算を続ける。この計算を行うと，音源Ａに起因する複数のピークは反復ごとに小さくなるのに対し，音源Ｂが音源Ａと異なる周波数スペクトルを持つ場合は，音源Ｂのピークの高さはそれほど変化しないため，それまで小さかった音源Ｂのピークが最大ピークとなる。図４（ｂ）はこの処理によって得られた空間スペクトルであり，音源Ｂのピークが最大となっていることがわかる。この方法によって複数の音源方向が求められる。ただし，同じ時間・周波数スペクトルをもつ音源が複数ある場合には適用できない。

ここで，この手法を差信号GCCと高分解能法に適用して音源方向推定性能を調べた結果を図６に示す。評価方法とデータ分析の条件は２－５節と同じであり，音源の存在範

囲は，$-180^\circ \leq \theta \leq 180^\circ$，$-60^\circ \leq \phi \leq 90^\circ$とした。HRTFは，以前に作成した４人の頭部モデルの測定データ[25]を用いた。図６からわかるように，音源が１個と２個の場合，差信号GCCの性能はMUSICと同等であり，S/Nが高い場合は音源が３個までは十分推定可能であることがわかる。音源が３個のときは MUSIC 法の方が若干性能が高いが，これは，音源数が多いほどスペクトルのスパース性が低下し，MUSICの場合は複数音源の寄与した成分を除く成分選択の効果が出ているためである。しかし，簡単な処理である差信号 GCCが代表的な工学的手法の性能とあまり変わらないことは注目すべきことであると考えられる。なお，MV法は，音源が２個以上の場合に大幅に性能が低下したため図への表示は省いた。

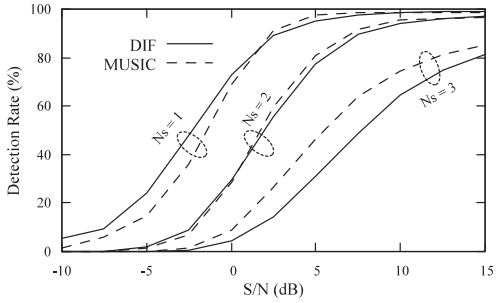

図６　両耳受聴音からの２次元的な音源方向推定結果　MUSIC 法と差信号 GCCによる音源方向検出率を示す。Nsは音源数である。音源が３個のときは MUSICの性能が若干上回っているが，音源が１個と２個のときは同等の性能であることがわかる。

４－２．振幅情報のみによる音源方向推定

３－１節でも述べたように，聴覚処理においては，高域はILDが優位であり，また，包絡信号の時間差は検出できない場合もあるため，聴覚の定位モデルは高域では振幅情報のみによって方向を推定できなければならないことになる。振幅情報だけを使う工学的な方向推定処理は，２章でも述べたように，vector sensor array とよばれており，同じ位置に指向性の向きの異なるセンサを置き，推定自体はMUSIC などの通常の方法で行う。センサ間隔が実際は０でない場合でも０とおき，入力信号の DFTの位相を無視して絶対値だけを用いることによって Vector Sensorと等価な処理にできる。この場合，ステアリングベクトルや相関行列は実数になる。例えば，差信号 GCCにおいては，（９）式の $G_{12,k}(\theta)$ において θ を $d = (\theta, \phi)$ をと置き，さらに

$$G_{12,k}^{VSA}(d) = \overline{|X_{1,k}|\,|A_{1,k}^{-1}(d)| \cdot |X^{*}_{2,k}|\,|A_{2,k}^{-1}(d)^{*}|}$$
$$= \overline{|X_{1,k}|\,|X_{2,k}|} \cdot |A_{1,k}^{-1}(d)|\,|A_{2,k}^{-1}(d)| \qquad (25)$$

とする。$G_{dd,k}$，$G_{zz,k}$も同様であり，入力信号 $X_{n,k}$とステアリングベクトルの要素 $A_{n,k}(d)$ を各々その絶対値に入れ替えればよい。

そこで，この処理の有効性を確認するため，差信号 GCCにおいて振幅情報のみを用いた方向推定性能を評価した。まず，図７に帯域が３kHz～3.5kHzの場合の差信号 GCCの空間スペクトルを示す。音源は１個の音声である。帯域幅が狭いため，スペクトル上には大きな疑似ピークが多数出現しているが，最大ピークは真の音源に対応しており，方向推定可能であることがわかる。

次に，振幅情報のみを用いた場合の２次元的な推定の場合と正中面上の仰角のみの推定の場合の結果を，各々，図８と図９に示す。図８を見ると，位相も使う場合の結果（図３の曲線「DIF」）に比べると音源数とS/Nの低下に伴う性

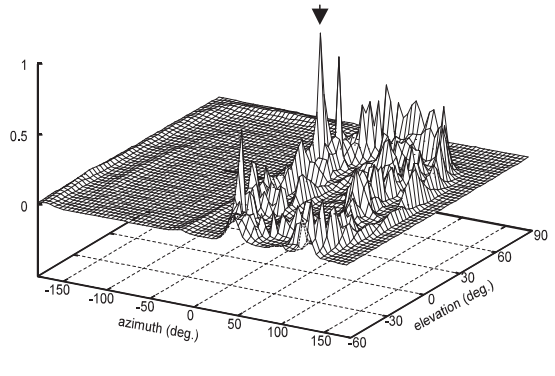

図７　差信号に基づいた GCCの振幅情報のみを用いた場合の２次元空間スペクトル　音源は矢印の位置に1個あり，計算に用いた周波数範囲は３kHz～3.5kHz，S/N=20dB の場合である。スペクトル上には大きな疑似ピークが多数出現しているが，最大ピークは真の音源に対応している。

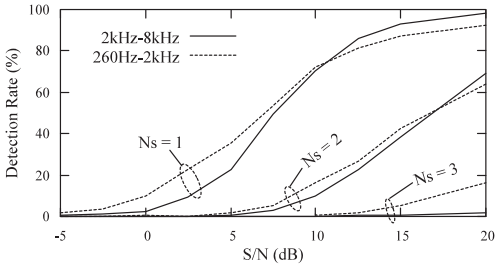

図８　差信号 GCCによる振幅情報のみからの２次元音源方向検出率　Nsは音源数である。音源は音声であり，計算に用いた帯域が２kHz～８kHzの場合と260Hz～２kHzの場合についての結果である。S/Nは帯域ごとに調整してあり，２つの帯域による性能差はあまりない。

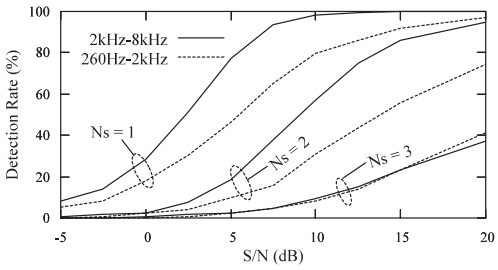

図９　差信号 GCCによる振幅情報のみからの正中面上の音源方向検出率　Nsは音源数である。音源は音声であり，計算に用いた帯域が２kHz～８kHzの場合と260Hz～２kHzの場合についての結果である。２kHz～８kHzの場合の方が高い性能である。

能低下が大きいが，高S/Nのときは，音源２個程度までは推定可能であることがわかる。また，図９からは，正中面においても同様に音源２個程度までは推定可能であることがわかる。また，正中面上の推定の場合，用いる周波数帯域は，260Hz～２kHzの低域よりも２kHz～８kHzの高い方の帯域の場合の方が性能がだいぶ高いことがわかる。この結果は，聴覚の精度よりも高いと思われるが，これは，測定により得られたHRTFをステアリングベクトルとして用いており，正確な伝達系の詳細を既知として処理していることになるためである。ヒトが自身のHRTFについてどこまで詳細な情報を持っているかが問題であるが，このような事前情報の量を減ずることによって聴覚の精度に近づくものと思われる。

５．おわりに

本稿では，方向推定処理の性能比較により，提案する差信号GCCが工学的手法の中でも高性能であるとされるMUSIC法と同等の性能を持ち，さらに，HRTFの影響を受けた両耳信号から複数の音源方向を２次元的に推定することが可能であることを確認した。また，振幅だけの処理によっても推定可能であることを確認できた。本稿における検討は工学処理の枠組み内で行った予備的なものであり，聴覚処理モデルとしての本格的な検討は今後の課題であるが，可能性を示唆することはできたと思われる。この手法において，振幅情報のみからの２次元的な方向推定能力は，頭部伝達関数の高域における両耳間レベル差の空間的な分布が周波数間で不揃いであることから生じており，聴覚がこのようなばらつきを利用しているのかどうかも興味深い問題である。本稿で説明した推定方法は，現実的な環境にも対処できるような高性能な定位モデルを意図しているものの，聴覚モデルは聴覚の特性の説明が目的であって，高性能な処理は目指すところではない[5,pp.89)]，との指摘もある。しかし，実環境における高い性能も聴覚の音源定位の１つの特性であり，その模擬は音源定位モデルの研究からはずれているわけではないと考えている。

今後，音源定位モデルの研究を大きく発展させるためには，工学者と生物学者など既存の学問分野を広くまたがる新たな枠組みと協調が必要だろう。

文献

1) 赤木正人 : 電子情報通信学会誌 , Vol. 77, No. 9, 948-956 (1994)
2) Algazi V. R.: J. Acoust. Soc. Am., 109 (3), 1110-1122 (2001)
3) 浅野太 : 日本音響学会誌 , 63巻 , 1号 , 41-46 (2007)
4) Ball S.F.: IEEE Trans. Acoust., Speech, Signal Proces., Vol.ASSP-27, No.2, 113-120 (1979)
5) Blauert J.: Communication Acoustics, Springer, 75-103 (2005)
6) Breebaart J.: J. Acoust. Soc. Am., 110 (2), 1074-1088 (2001)
7) Bregman A.S.: Auditory Scene Analysis,MIT Press (1990)
8) Capon J.:Proc. IEEE, 57, 1408-1418 (1969)
9) Colburn H.S. and Durlach N.I.: Models of binaural interaction in Handbook of Perception Vol. IV edited by E Carterette and M. Friedman, Academic Press, 365-515 (1978)
10) 電子通信学会 : 聴覚と音声 , 電子通信学会 , 195-210 (1980)
11) Faller C. and Merimaa J.: J. Acoust. Soc. Am., 116 (5), 3075-3089 (2004)
12) Gaubitch N.D., Naylor P.A., Ward D.B.: IWAENC2003, 99-102 (2003)
13) Iida K., Itoh M., Itagaki A., and Morimoto M.: Applied Acoustics, 68, 835-850 (2007)
14) 石井カルロス寿憲 , シャトッ・オリビエ , 石黒浩 , 萩田紀博 : 人工知能学会研究会資料 , SIG-Challenge-A802-5, 27-32 (2008)
15) Jin C.: J. Acoust. Soc. Am., 108 (3), 1215-1235 (2000)
16) 菊間信良 : アレーアンテナによる適応信号処理 , 科学技術出版社 (1998)
17) Kim H.Y, Asano F., Suzuki Y., and Sone T.: IEICE Trans. Fundamentals, Vol.E97-A, 2151-2158 (1996)
18) Knapp C.H. and Carter G.C.: IEEE Trans. Acoust., Speech, Signal Proces., Vol.ASSP-24, No.4, 320-327 (1976)
19) Lockwood M.E. and Jones D L.: J. Acoust. Soc. Am. 119 (1), 608-619 (2006)
20) Martin K.D.: Applications of Signal Processing to Audio and Acoustics, 1995, IEEE ASSP Workshop on, 96-99 (1995)
21) Middlebrooks J.C.: J. Acoust. Soc. Am., 92 (5), 2607-2624 (1992)
22) Mohan S., Lockwood M.E., Kramer M.L., and Jones D.L.: J. Acoust. Soc. Am., 123 (4), 2136-2147 (2008)
23) Morimoto M., Iida K. and Ito M.: Acoust, Sci. & tech. 24, 5, 267-275 (2003)
24) Nagata Y., Fujioka T., and Abe M.: IEEE Trans. Audio, Speech, and Language Proces., Vol. 15, No. 2, 416-429, (2007)
25) Nagata Y., Iwasaki S., Hariyama T., Fujioka T., Obara T.,Wakatake T. , and Abe M.: IEEE Trans. Audio, Speech, and Language Process., Vol. 17, No. 1, 52-65, (2009)
26) 永田仁史 , 岩崎聡 , 針山孝彦 , 堀口弘子 , 藤岡豊太 , 安倍正人 : 電子情報通信学会論文誌 A, Vol. J92-A, No. 11, 864-873, (2009)
27) Nehorai A. and Paldi E.: IEEE Trans. On Signal Proces., Vol. 42, No.9, 2481-2491 (1994)
28) 奥乃博 : 情報処理 , Vol.49, No.1, 15-23 (2008)
29) Pillai S.U.:Array Signal Processing, Springer Verlag (1989)
30) Popper A.N. and Fay R.R. : Sound Source Localization, Springer, (2005)
31) Searle C.L., Braida L.D., Cuddy D.R., and Davis M.F.: J. Acoust. Soc. Am., Vo.57, No.2, 448-455 (1975)
32) Searle C.L., Braida L.D., Davis M.F., and Colburn H.S.: J. Acoust. Soc. Am., Vo.60, No.5, 1165-1175 (1976)
33) Schmidt R.O.: IEEE Trans. Antennas Propagat., Vol.

AP-34, No. 3, 276-280 (1986)
34) 鈴木敬 , 金田豊 : 日本音響学会誌 , Vol.65, No.10, 513-522 (2009)
35) Urick R.J.: 水中音響の原理 , 共立出版 (1978)
36) Wang D. and Brown G.J.: Computational Auditory Scene Analysis, IEEE Press (2006)
37) Webb B.: 昆虫ミメティクス (下澤盾夫 , 針山孝彦 監修), NTS, 791-796 (2008)

Abstract

Computational models for estimating directions of arrivals of sound sources using two-channel signal.

Yoshifumi NAGATA, Department of Electrical Engineering and Computer Science, Faculty of Engineering, Iwate Univ., Morioka 020-8551

I describe some engineering methods that have been used frequently for estimating directions of arrivals (DOAs) of sound sources first. I further present a new DOA estimation method based on the generalized cross correlation function (GCC) weighted by the inverse of the difference spectrum between the channels. I evaluate the performances of the above methods assuming that two omni directional microphones are placed in free space. Next, I investigate the availability of the proposed GCC as a binaural sound localization model. The results of the investigation demonstrate that the performances of the proposed method in two-dimensional DOA estimation from binaural sound are comparable to those of the high resolution method Multiple Signal Classification (MUSIC) method which is considered to be a central method in engineering field. Moreover, the validity as binaural localization model is demonstrated by the performance in case where the phase information is omitted to simulate the auditory characteristics at the high frequency band.