社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

[特別講演] 統計モデルに基づく時変フィルタによる音源分離

小野　順貴† 　和泉　洋介† 　伊藤　信貴† 　嵯峨山茂樹†

† 東京大学大学院 情報理工学系研究科 システム情報学専攻
〒 113-8656 東京都文京区本郷 7-3-1
E-mail: †{onono,izumi,itou,sagayama}@hil.t.u-tokyo.ac.jp

あらまし　本稿では，1) マイクロホン数より音源数が多い場合の雑音環境下のブラインド音源分離，2) 拡散性雑音環境下での目的信号の分離，の 2 つの話題について，最近の著者らの取り組みを紹介する。ビームフォーミングや独立成分分析といったアレイ信号処理は，線形時不変フィルタによる分離系を構成する手法であるのに対し，提案する処理は時間周波数マスキングを含む時変フィルタになっている。この違いは，対象としている環境や観測モデルに起因する。本稿では，統計的観測モデルと提案法のアルゴリズムについて論じ，いくつかの基礎実験の結果を示す。
キーワード　時間周波数マスキング，時変フィルタ，最尤推定，EM アルゴリズム，等方性雑音

Source Separation by Time-Variant Filters Based on Stochastic Models

Nobutaka ONO†, Yosuke IZUMI†, Nobutaka ITO†, and Shigeki SAGAYAMA†

† Department of Information Physics and Computing, Graduate School of Information Science and Technology, The University of Tokyo
7-3-1 Hongo Bunkyo-ku Tokyo 113-8656, Japan
E-mail: †{onono,izumi,itou,sagayama}@hil.t.u-tokyo.ac.jp

Abstract　In this paper, we introduce our recent approach to two topics: 1) underdetermined ICA in noisy environment, 2) target source separation from surrounding diffuse noise. While the popular framework of array signal processing like adaptive beamformers or ICA gives us a linear time-invariant filter as a separation system, our proposed method is a time-variant filter including time-frequency masking. The difference between them is caused by observation models of interest. In this paper, we discuss the stochastic observation models and proposed algorithm, and show several experimental results.
Key words　time-frequency masking, time-variant filter, maximum likelihood estimation, EM algorithm, isotropic noise

1. はじめに

音源分離とは，1 つ，もしくは複数のマイクロホンにより観測された信号から，空間中に存在する個々の音源から生じた信号を推定する問題であり，音声認識をはじめとした音の認識の前処理，通信系や補聴器などにおけるターゲット音以外の干渉音抑圧，セキュリティや異常検出のためのモニタリング，音による周囲環境把握のためのロボット聴覚など，実環境の様々な分野で必要となる技術である。また，1) 音源数とマイクロホン数の大小関係，2) 混合系に対する先験情報 (音源方向，残響時間など) の有無，3) 背景雑音環境に対する先験情報 (パワースペクトル，確率密度関数) の有無，4) 音源信号に対する先験情報の有無，などによって多様な問題設定が生じ，様々な手法が研究されている。

本稿ではこうした中で，1) マイクロホン数より音源数が多い場合の雑音環境下のブラインド音源分離，2) 拡散性雑音環境下での目的信号の分離，の 2 つの話題について，最近の著者らの取り組みを紹介する。以下では音響信号は，自由空間に配置された音圧型マイクロホンで観測することを前提として議論を進める。

2. 音源分離問題のモデルと解法

2.1 時間周波数領域における一般的な観測モデル

音響信号処理で扱う系は，通常，時間遅れを含む畳み込み混合で表され，これを時間領域で扱うためには，非常に次数の大きな行列計算が必要となる。これに対し，音源から観測点までのインパルス応答長に対して十分に長い時間窓をもつ時間周波数分解 (短時間 Fourier 変換，wavelet 変換など) を用いると，

この畳み込みを近似的に乗算で表すことができるため，アレイ信号処理では時間周波数表現が好んで用いられる。

いま，M 本のマイクロホンアレイで観測された信号の時間周波数表現を $\boldsymbol{O}(\omega,t)=(O_1(\omega,t),\cdots,O_M(\omega,t)$ とすると，観測モデルは

$$\boldsymbol{O}(\omega,t)=A(\omega)\boldsymbol{S}(\omega,t)+\boldsymbol{N}(\omega,t) \tag{1}$$

のように表される。ここで $\boldsymbol{S}(\omega,t)=(S_1(\omega,t),\cdots,S_K(\omega,t))$ は音源信号の時間周波数表現であり，各音源から各マイクロホンへの伝達周波数特性を表す $A(\omega)$ は時不変であることを仮定している。また $\boldsymbol{N}(\omega,t)$ は，多数の方向から到来する背景雑音や，フレーム長を超える残響成分など，時不変な伝達特性として表現できない成分を表すものとする。

2.2 線形時不変システムによる分離系

音源分離のアルゴリズムとしてよく用いられる分離系の 1 つは，j 番目の音源信号を

$$\hat{S}_j(\omega,t)=\boldsymbol{W}_j(\omega)^h\boldsymbol{O}(\omega,t) \tag{2}$$

により推定するものである。すなわち時間周波数領域において，観測信号の観測信号の複素数荷重和により音源信号を生成するものであり，荷重が時間に依らないことから，この分離系は線形時不変システムとなる。

この分離法の妥当性の 1 つは，式 (1) において，雑音項 $\boldsymbol{N}(\omega,t)$ が無視でき，かつ音源信号数 K に対しマイクロホン数 M が等しいかそれより大きい場合には，混合行列 $A(\omega)$ の擬似逆行列を用いることにより，

$$\boldsymbol{S}(\omega,t)=(A(\omega)^hA(\omega))^{-1}A(\omega)^h\boldsymbol{O}(\omega,t) \tag{3}$$

の関係が得られることにある。

よって分離系を式 (2) のような線形フィルタとし，重みベクトル $\boldsymbol{W}_j(\omega)$ を設計することが，アレイ信号処理における一つの標準的な音源分離手法であり，遅延和アレイよるビームフォーミングにおいては目的音源方向，最小分散ビームフォーマにおいては目的音源方向と観測信号の共分散行列によって，$\boldsymbol{W}_j(\omega)$ が設計される。近年大きく発展した独立成分分析は，音源方向の情報が未知であっても，観測信号の高次統計量を用いてこの推定を可能にする枠組みであると位置づけられる。

2.3 時変システムとしての時間周波数マスキング

式 (2) を拡張して，

$$\hat{S}_j(\omega,t)=\boldsymbol{W}_j(\omega,t)^h\boldsymbol{O}(\omega,t) \tag{4}$$

のように，重みベクトル $\boldsymbol{W}_j$ を t 方向にも変化を許すと，この分離系は線形時変システムとなる。時変システムという用語は，対象とする系の時間変化に追従するような適応的なシステムに対して用いられることが多いが，音源分離において用いられる場合は，観測点 (ω,t) 毎に分離系を切り分けるような目的で用いられることが多い。

このクラスに含まれる分離系でよく用いられるものの 1 つは時間周波数マスキングであり，その多くは，

$$\hat{S}_j(\omega,t)=m_j(\boldsymbol{O}(\omega,t))\boldsymbol{W}(\omega)^h\boldsymbol{O}(\omega,t) \tag{5}$$

のような形式で表される。ここで $m_j(\boldsymbol{O}(\omega,t))$ は，観測ベクトル $\boldsymbol{O}(\omega,t)$ に依存して 0 または 1 の離散値，もしくは 0 から 1 の間の実数値をとるスカラであり，時間周波数領域において，目的とする音源信号 S_j のみを通過させその他の音源信号を阻止する役割を果たすため，時間周波数マスクと呼ばれる。

時間周波数マスキングのアプローチは，特にスパースな音源信号に対して正当化される。もし，各音源信号の時間周波数表現 $S_j(\omega,t)$ が多くの領域でほぼ 0 であり，有意なエネルギーをもつ領域が互いにほとんど重ならないとすれば，

$$S_j(\omega,t)\simeq m_j(\omega,t)\boldsymbol{e}^h\boldsymbol{O}(\omega,t) \tag{6}$$

を満たすような理想的なマスク m_j が存在することになる [1]。ただし，$\boldsymbol{e}=(1\ 0\ \cdots\ 0)^t$ である。時間周波数マスキングは，近年では特に，このようなスパース性に基づき，マイクロホン数より多くの数の音源を扱う BSS の枠組みとして研究されている [2], [3]。また，式 (5) における線形時不変フィルタ部分である $\boldsymbol{W}(\omega)$ は，簡単には $\boldsymbol{W}(\omega)=Be$ のように選ばれるが，通常の遅延和アレイや，最小分散ビームフォーマなどが用いられることもあるし，ICA により $\boldsymbol{W}(\omega)$ が決定され，時間周波数マスキングが併用されることもある [4], [5]。

我々もこの時変フィルタの自由度を活かし，時間周波数マスク $m_j(\omega,t)$ による時間周波数毎のエネルギーのオンオフと，線形時不変フィルタ $\boldsymbol{W}(\omega)$ の組み合わせ，もしくは，時間周波数毎にフィルタが変化するような $\boldsymbol{W}(\omega,t)$ の設計により，線形時不変フィルタのみでは困難であるような環境での音源分離を改善することを目的に，研究を進めている。

3. 最尤時間周波数マスキングによる雑音環境下 2ch BSS

3.1 スパース音源の観測モデル

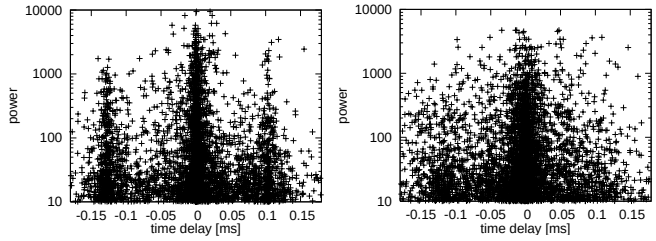
本節では，音源信号のスパース性の仮定に基づく BSS における時間周波数マスキングについて論じる。特に 2 チャンネルの場合に議論を絞るが，本手法は，より多チャンネルのアレイに対しても同様に適用可能である。

2ch のマイクロホンによる観測信号の時間周波数領域表現を $\boldsymbol{O}(\omega,t)=(M_L(\omega,t),M_R(\omega,t))^T$ とする。いま，複数存在する音源信号が時間周波数領域でスパースであり，各時間周波数成分にはいずれか 1 つの音源のみが寄与する，と仮定すると，観測モデル式 (1) は，

$$\boldsymbol{O}(\omega,t)=S_k(\omega,t)\boldsymbol{a}(\omega)_k+\boldsymbol{N}(\omega,t) \tag{7}$$

のように書き直せる。ただし，$S_k(\omega,t)$ は (τ,ω) に寄与する音源信号，$\boldsymbol{a}(\omega)_k=(1,\exp(j\omega\delta_k))^T$ は k 番目の音源から 2 個のマイクロホンまでの伝達周波数特性であり，ここでは平面波伝播を仮定し，2 個のマイクロホン間の時間差を δ_k としている。

もし $\boldsymbol{N}(\omega,t)=0$ であれば，観測ベクトル $\boldsymbol{O}$ は $\boldsymbol{a}_{k,\omega}$ に常に平行であり，観測信号 M_L と M_R の間の時間差は，真の時間差 δ_k に一致する。このため従来は，各時間周波数点 (t,ω) 毎に時間差を検出し，そのクラスタリングによって行う手法がとられ



(a) $T_R = 0$[ms] (b) $T_R = 170$[ms]

図 1　3 個の音源信号を 2 マイクロホンで観測した場合に各時間周波数点から検出された時間差分布の例

てきた。しかし，実環境においては $\boldsymbol{N}(\omega,t)$ の影響により，観測から得られる時間差は大きなばらつきを持つ (図 1 参照)。

我々のアプローチは，時間差領域でのばらつきをモデル化するのではなく，このばらつきの要因である $\boldsymbol{N}(\omega,t)$ に対して確率モデルを仮定し，最尤解を求めるというものである。これにより，1) 時間差のばらつきの周波数依存性のモデル化を容易にし，2) 雑音の共分散行列やパワースペクトルといった先験情報を利用しやすくする，などといった利点がある。また従来のアレイ信号処理との整合性もよい。

いま，$\boldsymbol{N}(\omega,t)$ が平均 0，共分散行列 $V(\omega)$ の複素ガウス分布に従うと仮定すると，推定されるべきパラメータは，

1)　各音源の時間差 δ_k

2)　雑音共分散行列 $V(\omega)$

3)　各時間周波数点 (ω,t) においてアクティブな音源信号 $S_{k(\omega,t)}(\omega,t)$

となる。また，以下では表記の簡単さのため，$\boldsymbol{O}(\omega,t)$ を $\boldsymbol{O}_{\omega,t}$ などと表記する。

各時間周波数点 (ω,t) に寄与する音源のインデックス $k(\omega,t)$ が既知であれば，ある (ω,t) において $\boldsymbol{O}_{\omega,t}$ が観測される尤度は，

$$p_k(\boldsymbol{O}_{\omega,t}|\ \delta_k, V_\omega, S_{k,\omega,t}) = \frac{1}{2\pi\sqrt{|V_\omega|}} \times \exp\left(-\frac{1}{2}(\boldsymbol{O}_{\omega,t} - S_{k,\omega,t}\boldsymbol{a}_{k,\omega})^h V_\omega^{-1}(\boldsymbol{O}_{\omega,t} - S_{k,\omega,t}\boldsymbol{a}_{k,\omega})\right) \tag{8}$$

のように与えられる。しかし実際には $k(\omega,t)$ は観測できない隠れ変数であるので，$k(\omega,t)$ に関して周辺化することにより，式 (7) の観測モデルに基づく $\boldsymbol{O}_{\omega,t}$ の対数尤度 L は，

$$L = \sum_{\omega,t} \log \sum_k \log r_k p_k(\boldsymbol{O}_{\omega,t}|\ \delta_k, V_\omega, S_{k,\omega,t}) \tag{9}$$

のように表される。ここで r_k は k 番目の音源がアクティブになる事前分布に相当するパラメータであり，$\sum_k r_k = 1$ を満たすものとする。

3.2　EM アルゴリズムによる最尤解の導出

L の最大化問題は隠れ変数 $k(\omega,t)$ を含む最尤化問題であり，これは EM アルゴリズムにより効率的に解くことができる [6], [7]。パラメータの更新式は下記のようにまとめられる。

$$S^{(i)}_{k,\omega,t} = \frac{{\boldsymbol{a}^{(i)}_{k,\omega}}^h V_\omega^{-1}\boldsymbol{O}_{\omega,t}}{{\boldsymbol{a}^{(i)}_{k,\omega}}^h V_\omega^{-1}\boldsymbol{a}^{(i)}_{k,\omega}}. \tag{10}$$

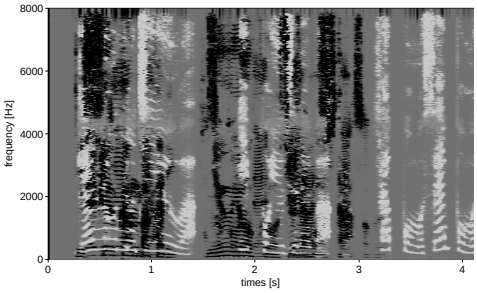


図 2　期待値マスクの例

$$m^{(i)}_{k,\omega,t} = \frac{r^{(i)}_k p_k(\boldsymbol{O}_{\omega,t}|\ \delta^{(i)}_k, V^{(i)}_\omega, S^{(i)}_{k,\omega,t})}{\sum_{k'} r^{(i)}_k p_k(\boldsymbol{O}_{\omega,t}|\ \delta^{(i)}_{k'}, V^{(i)}_\omega, S^{(i)}_{k,\omega,t})} \tag{11}$$

$$Q(\delta_k;\delta^{(i)}_k) = \sum_{t,\omega} m^{(i)}_{k,\omega,t} \log r^{(i)}_k p_k(\boldsymbol{O}_{\omega,t}|\ \delta^{(i)}_k, V^{(i)}_\omega, S^{(i)}_{k,\omega,t}) \tag{12}$$

$$r^{(i+1)}_k = \frac{\sum_{\omega,t} m^{(i)}_{k,\omega,t}}{\sum_{k',\omega,t} m^{(i)}_{k',\omega,t}} \tag{13}$$

$$\delta^{(i+1)}_k = \mathrm{argmax}_{\delta_k} Q(\delta_k;\delta^{(i)}_k). \tag{14}$$

$$V^{(i+1)}_\omega = \frac{1}{C}\sum_{\omega,t} m^{(i)}_{k,\omega,t} \times (\boldsymbol{O}_{\omega,t} - S^{(i)}_{k,\omega,t}\boldsymbol{a}^{(i)}_{k,\omega})(\boldsymbol{O}_{\omega,t} - S^{(i)}_{k,\omega,t}\boldsymbol{a}^{(i)}_{k,\omega})^h \tag{15}$$

ただし，i は反復の回数，C は全体の時間周波数点数である。式 (14) に関しては解析解が求まらないため，適当な数の離散化した δ_k に対して $Q(\delta_k;\ \delta^{(i)}_k)$ を全て計算し，最大値を選択することにより行なっている。

3.3　最小二乗誤差を与える期待値マスク

EM アルゴリズムによるパラメータ推定後，音源分離を行なう一つの方法は，各時間周波数点において $m_{k,\omega,t}$ (k 番目の音源がアクティブである確率) が最大の音源のみがアクティブであるとハードに決定し，アクティブでない音源は 0 と推定するものである。しかし，パラメータを推定した後では，二乗誤差最小という意味での最適推定値である期待値も求めることができる。k 番目の音源信号 $S_{k,\omega,t}$ の期待値は，

$$\begin{aligned} E[S_{k,\omega,t}] &= \sum_{k'} m_{k,\omega,t} E_{k'}[S_{k,\omega,t}] \\ &= m_{k,\omega,t} E_k[S_{k,\omega,t}] + \sum_{k'\neq k} m_{k',\omega,t} E_{k'}[S_{k,\omega,t}] \\ &= m_{\tau,\omega,k} \frac{\boldsymbol{a}^h_{k,\omega} V_\omega^{-1}\boldsymbol{O}_{\omega,t}}{\boldsymbol{a}^h_{k,\omega} V_\omega^{-1}\boldsymbol{a}_{k,\omega}} \end{aligned} \tag{16}$$

のように得られる。ただし，$E_{k'}[S_{k,\omega,t}]$ は，k' 番目の音源がアクティブな場合の $S_{k,\omega,t}$ の推定値であり，我々のモデルにおいては，各時間周波数点においてアクティブな音源は 1 つのみであるから，$k' \neq k$ であれば $E_{k'}[S_k] = 0$ となる。よって，分配関数 $m_{\tau,\omega,k}$ 自体が，期待値を与える連続値マスクとして働くことがわかる (図 2 参照)。

3.4　シミュレーションによる残響環境下での評価実験

EM アルゴリズムによる音源分離実験を，図 3 のように 3 つ

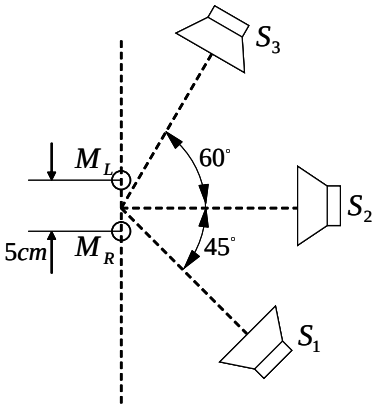


図 3　シミュレーションにおけるマイクロフォンと音源の位置関係

表 1　音源定位結果の比較 (時間差 [μs])

手法	s_1	s_2	s_3
従来手法	-51	-42	10
提案手法	-80	0	90
提案手法 (拡散音場)	-90	0	121
真の位置	-88	0	103

表 2　音源分離性能結果の比較 ([dB])

手法	s_1	s_2	s_3
従来手法	6.0	-11.4	2.2
提案手法	7.2	3.9	7.8
提案手法 (拡散音場)	7.4	5.3	7.4

表 3　σ^2 の推定値と残響時間の関係

残響時間 [ms]	0	90	170	270	370
σ^2	0.12	0.14	0.17	0.21	0.25

の音源および 2 つのマイクロフォンを配置し，鏡像法 [8] による残響シミュレーションを行った．分離性能の評価には分離の前後での元音声に対する S/N 比の改善値を用いた．音声データは研究用連続音声データベース (©板橋秀一 [日本音響学会／編]) を使用した．サンプリング周期 16kHz，フレーム長は 1024 点, シフトは 512 点，窓関数を Hamming 窓として，観測信号を短時間 Fourier 変換して時間周波数表現を得た．比較対象とした従来法は Yilmaz らの手法 [1] に基づき，時間差のヒストグラムのピークを検出することで音源定位を行い，次にこれに基づいて最尤推定で音源分離マスクを設計するものである．残響時間 376ms の場合の音源分離・定位結果を表 1 ,2 に示す。また，本実験では雑音共分散行列 V を対角行列 I，もしくは拡散音場モデル [9], [10] から導かれる行列に固定し，そのパワー σ^2 のみを推定している。表 3 に，残響時間を変えたときの σ_2 の推定結果を示している。

これらの結果から，提案法は従来法よりも高い分離性能をもっていることが確認できる．さらに残響時間が増加するにしたがい雑音の分散 σ^2 の推定値も増加し，雑音環境の推定も同時に行えていることがわかる．

4.　結晶型アレイを用いた等方的雑音抑圧

4.1　等方的雑音場の問題設定

本節では，カクテルパーティ，駅や空港などの雑踏などの多数話者による背景雑音，室内の残響のような拡散性雑音の除去を目的とした処理について論じる。目的音源以外の音源は背景雑音のみとすると，観測モデルは，

$$\boldsymbol{O}(\omega,t) = S(\omega,t)\boldsymbol{a}(\omega) + \boldsymbol{N}(\omega,t) \tag{17}$$

と表される。

いま，観測点の周囲に多数の独立な雑音源が存在するような拡散性雑音場の理想的なモデルとして，

1)　雑音パワースペクトルが観測位置に依らない

2)　雑音の空間的クロススペクトルが 2 点の観測位置の方向に依らない

という性質を満たす等方性雑音場を考える。この拡散性雑音の共分散行列 $E[\boldsymbol{N}(\omega,t)\boldsymbol{N}(\omega,t)^h] = V_N(\omega,t)$ を考えると，上記 1) の仮定より，V_N の対角成分は等しい値をとり，また 2) の仮定より，アレイが距離の等しいマイクロホン対を複数含めば，それらに対応する非対角成分は等しい値をとる。例えば正方形アレイに対し，周を回る順にマイクロホンに番号をつければ，

$$E[\boldsymbol{N}\boldsymbol{N}^h] = \boldsymbol{V} = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_2 \\ \alpha_2 & \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_3 & \alpha_2 & \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 & \alpha_2 & \alpha_1 \end{pmatrix} \tag{18}$$

のように，全成分が 3 つのパラメータ α_i $(i = 1, 2, 3)$ のみで表わされる。ただし，これらは (ω, t) に依存することに注意する。

4.2　等方的雑音を無相関化するアレイ配置

式 (18) のような行列は一般にフルランクで零空間をもたないため，波形領域での雑音抑圧性能には限界がある。そこで我々は式 (17) のパワースペクトル表現に着目した。目的信号 S と雑音 $\boldsymbol{N}$ が無相関であれば，

$$E[\boldsymbol{O}\boldsymbol{O}^h] = V_O = \Phi_{SS}\boldsymbol{a}\boldsymbol{a}^h + V_N \tag{19}$$

と表される。ただし Φ_{SS} は目的信号のパワースペクトルである。また以下では簡単のため，(ω, t) は省略して表記する。

V_N はエルミート行列なので，あるユニタリ行列 P により

$$P^h V_O P = \Phi_{SS}(P^h\boldsymbol{a})(P^h\boldsymbol{a})^h + \Lambda \tag{20}$$

と対角化される。ただし $\Lambda = P^h V_N P$ である。理論的にはこの誤差の直交化により $P^h V_O P$ の非対角成分からは雑音が除かれることになる。また $\boldsymbol{a}$ も既知の量であるから，式 (20) に基づき $P^h V_O P$ の非対角成分から目的音源信号のパワースペクトル Φ_{SS} を推定できないか，というのが我々の着眼点である。

一般に V_N を直交化する行列 P は V_N に依存し，V_N を得るためには雑音のみを観測する必要があるため，通常は前節で述べたような手法を行なうことはできない。しかし我々は，等方的な雑音場においては行列 V_N がセンサ配置によって式 (18) のような特別な構造をもつことに着目し，V_N の各成分の値に依

らず，この構造だけで行列 P が定まるセンサ配置があることを発見した。すなわち，以下の定理が成り立つ。

［定理 1］ マイクロホンが，1) 正多角形，2) 長方形，3) 正多面体，4) 直方体，5) 正多角柱，の頂点に配置されているならば，等方的雑音場の共分散行列は，その行列の要素に依らず，ある定数行列によって対角化できる。ただし，正多角柱は上面と下面が中心周りに任意の回転角をもってねじれていてもよい。

紙面のスペース上，定理の詳細は述べることができないが，鍵となるのは，巡回行列が DFT 行列 Z_n で対角化される [11] という点にある。ここで DFT 行列とは，

$$Z_n = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & \zeta & \cdots & \zeta^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \zeta^{n-1} & \cdots & \zeta^{(n-1)(n-1)} \end{pmatrix}, \tag{21}$$

$$\zeta = e^{-j2\pi/n}. \tag{22}$$

のように定義される行列である。

正多角形配置の場合には，周を回る順にマイクロホンに番号をつければ，等方性雑音場における共分散行列は式 (18) に示したように巡回行列になる。よって式 (18) は，α_i に依らず，$P = Z_4$ によって対角化される。

正多角柱，正六面体，正八面体，配置の場合には，マイクロホンに対する適切な番号付けによって (例を図 4)，等方性雑音場における共分散行列は $\begin{pmatrix} A & B \\ B & A \end{pmatrix}$ のようなブロック巡回型で，かつ A，B も巡回行列となるため，$P = Z_n \otimes Z_2$ によって対角化される。ただし $\otimes$ は直積を表わす。

長方形，直方体，正十二面体，正二十面体の場合にも，それぞれ固有の定行列 P が存在するが，やや形が複雑なのでここでは省略する。(より詳細は [13] を参照のこと。) また，構造だけで P が定まるようなアレイ配置が，前述の配置で全てであるかどうかはまだ不明であり [15]，検討中の課題の 1 つである。

4.3 ノイズフリークロススペクトルに基づくパワースペクトルの最尤推定

P によって変換された観測信号の共分散行列 $\hat{V}_O = P^h V_O P$ の (i,j) 成分を Φ_{ij} と表す。理論的にはこの非対角成分はノイズフリーであるが，実際には，期待値を有限時間区間における平均により算出することに起因する誤差，及び雑音場の等方性からのずれに起因するモデル化誤差が含まれるため，これを ε_{ij} と表すと Φ_{ij} は

$$\Phi_{ij} = a_{ij}\Phi_{SS} + \varepsilon_{ij} \tag{23}$$

と表される。ここで，a_{ij} は式 (20) 中の $\boldsymbol{a}\boldsymbol{a}^h$ の (i,j) 成分であり，例えば，

$$a_{12} = -(j\sin\omega\tau_1 + \sin\omega\tau_2)(\cos\omega\tau_1 + \cos\omega\tau_2) \tag{24}$$

$$\tau_1 = \frac{D\cos\theta}{c}, \quad \tau_2 = \frac{D\sin\theta}{c} \tag{25}$$

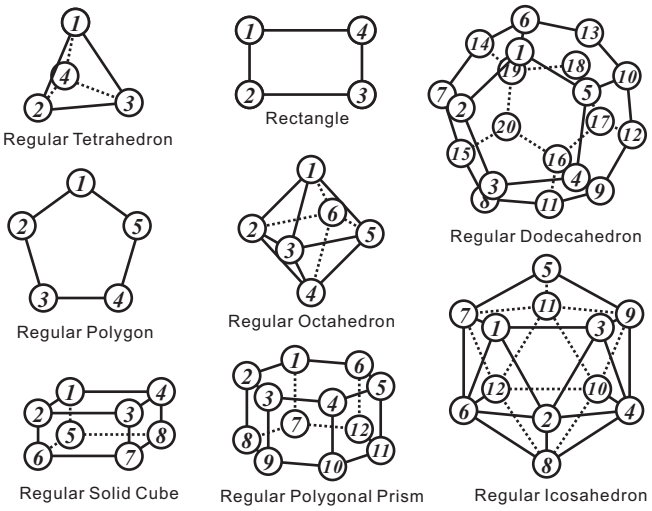


図 4　等方的雑音場を直交化する結晶型センサ配置

などとなる。ただし、D はマイクロフォンと原点の距離，c は音速，θ は音源方向である。

モデル化誤差 ε_{ij} の大きさも未知であり，また成分 ij によって異なると考えられるので，ここではこれら分散の異なるガウス分布に従うと仮定し，これらの分散も推定すべきパラメータとして最尤法を適用する。具体的には定数項を除いた対数尤度関数

$$L(\Phi_{SS},\boldsymbol{\sigma}) = \sum_{i\neq j}^{K} -\log\sigma_{ij} - \frac{|\Phi_{ij} - a_{ij}\Phi_{SS}|^2}{2\sigma_{ij}^2} \tag{26}$$

をパワースペクトル Φ_{SS}，分散 σ_{ij} に関して最大化する。ここで，K は非対角成分の総数を表す (等価な成分，つまり対称成分は除く). これは非線形であるが，対数尤度関数の各変数での偏微分を 0 とおくことにより求まる以下の更新式

$$\sigma_{ij}^{2\,(t+1)} = |\Phi_{ij} - a_{ij}\Phi_{SS}^{(t)}|^2,\ \Phi_{SS}^{(t+1)} = \frac{\sum_{i\neq j}^{K} \frac{a_{ij}\Phi_{ij}}{\sigma_{ij}^{2\,(t+1)}}}{\sum_{i\neq j}^{K} \frac{|a_{ij}|^2}{\sigma_{ij}^{2\,(t+1)}}} \tag{27}$$

を反復的に適用することで Φ_{SS} が求まる。以上が，パワースペクトル推定の概要である。図 5 に，本手法によるパワースペクトル推定結果の例を示す [12], [13]。他の手法に比べ，低い入力 SN 比条件でもスペクトル歪みが改善されていることがわかる。

4.4 等方的雑音抑圧の音源分離への応用

我々は，前述のパワースペクトル領域での等方的雑音抑圧を音源分離へ応用することを試みている [14]。具体的には，1) Delay-and-Sum (DS) 法によりビームフォーミングを行ない，2) 振幅スペクトルを前述の手法により推定されたパワースペクトル $\hat{\Phi}_{SS}$ の平方根 $\sqrt{\hat{\Phi}_{SS}}$ により置き換える，というものである。但し，もし負の推定値が得られた場合には，確からしくない推定値として 0 で置き換える。以上の処理では，振幅スペクトルの推定部が時間周波数マスキングのように働く。ビームフォーミングには最小分散ビームフォーマを利用する方法も考

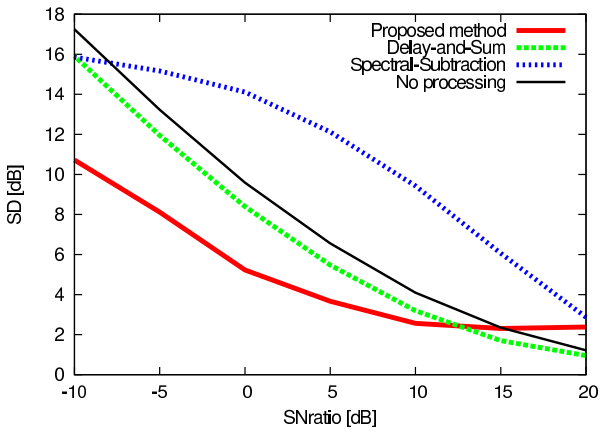


図 5　非定常雑音 (音声) 環境下におけるスペクトル歪み SD の比較 [12], [13]

えられ，今後の検討課題である。

4.5　等方的雑音場による音源分離のシミュレーション実験と結果

本節では，等方的雑音場において，提案法により既知の方向から到来する目的信号の波形を分離するシミュレーションの結果を述べる。比較のため，DS 法，MV 法 (minimum variance beamforming) [16] による結果も示す。

xy 平面内，x 軸からの偏角 $\frac{360^\circ}{64}j$ $(j = 0, \cdots, 63)$ の 64 方向から到来する白色雑音または相異なる音声の平面波により，等方的雑音場をシミュレートした。目的信号としては，xy 平面内，x 軸からの偏角 60° の方向から到来する音声の平面波を加えた。音声は ATR の音声データベース B セットの連続音声をサンプリング周波数 16kHz に変換したものを用いた。使用したアレイは，半径 0.1 m，各マイクロフォンの座標が $\big(0.1\cos\big(90^\circ i\big), 0.1\sin\big(90^\circ i\big)\big)$[m] $(i = 0, 1, 2, 3)$ の正方形アレイである。分析条件はフレーム長 2^8 点，フレームシフト 2^4 点，窓関数は Hamming 窓とした。以上の条件の下，白色雑音，音声雑音のそれぞれの場合について，観測信号の SN 比を 5.0dB，0.0dB，−5.0dB と変えて，提案法，DS 法，MV 法の各手法により目的信号を波形分離した。尚，提案法におけるパワースペクトル推定では，雑音相関行列を対角化する行列として DFT 行列を用い，2^4 フレーム毎に推定を行い，最尤法の反復計算回数は 15 回とした。また，MV 法において，観測信号の相関行列は全観測区間に亙る平均により算出し，その逆行列の算出の際には正則化を行い，正則化パラメータは結果が最良となるように調節した。

表 1 に結果を示す。(a)，(b) は，それぞれ雑音として白色雑音，音声を用いた場合の結果である。提案法が他の 2 つの手法より高い SN 比を示している。

表 1 提案法，DS 法，MV 法による推定波形の SN 比 [dB]

	white noise			voice		
observed	5.0	0.0	−5.0	5.0	0.0	−5.0
proposed	**11.7**	**8.7**	**4.6**	**9.6**	**5.8**	**1.3**
delay-and-sum	9.4	4.8	−0.2	7.0	2.1	−3.0
minimum variance	9.5	5.1	0.5	9.1	5.1	0.3

5.　ま　と　め

本稿では，時間周波数マスキングを含む時変フィルタによる音源分離として，著者らの取り組みを紹介した。時変フィルタは，時不変フィルタより広いクラスの音源分離を可能にする枠組みと考えられるが，一方，分離音にはミュージカルノイズなど聴感上問題となる雑音が発生する場合もあり，この対処が今後の課題の一つと考えている。

文　　献

[1] O. Yilmaz and S. Rickard: “Blind Separation of Speech Mixtures via Time-Frequency Masking,” IEEE Transaction on Signal Processing, Vol. 52, No. 7, pp 1830-1847, 2004.

[2] S. Araki, H. Sawada, R. Mukai, and S. Makino, “Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors, “ Signal Processing, vol. 87, pp. 1833-1847, Feb. 2007.

[3] M. Mandel, D. Ellis and T. Jebara, “An EM algorithm for localizing multiple sound sources in reverberant environments,” Proc. Neural Info. Proc. Sys., 2006.

[4] H. Sawada, S. Araki, R. Mukai, and S. Makino, “Blind extraction of dominant target sources using ICA and time-frequency masking, “ IEEE Trans. Audio, Speech and Language Processing, vol. 14, no. 6, pp. 2165-2173, Nov. 2006.

[5] Y. Mori, H. Saruwatari, T. Takatani, S. Ukai, K. Shikano, T. Hiekata, and T. Morita, “Real-Time Implementation of Two-Stage Blind Source Separation Combining SIMO-ICA and Binary Masking,” Proc. IWAENC2005, pp.229 − 232, Sep. 2005.

[6] 和泉, 小野, 嵯峨山, “EM アルゴリズムを用いた音声スパース性に基づく 2ch BSS,” 日本音響学会春季研究発表会講演集, pp.555-556, 3 月, 2007.

[7] Y. Izumi, N. Ono, and S. Sagayama, “Sparseness-based 2ch BSS using the EM algorithm in reverberant environment,” Proc. WASPAA, Oct, 2007. (to appear in)

[8] J. B. Allen and A. Berkley, “Image method for efficiently simulating small room acoustics,” J. Acoust. Soc. Am, JASA, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[9] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelmann and M. C. Thompson, “Measurement of correlation coefficients in reverberant sound fields,” JASA, Vol. 27, No. 6, pp. 1072-1077, 1955.

[10] I. A. McCowan, H. Bourlard, “Microphone Array Post-Filter Based on Noise Field Coherence,” IEEE Trans. on Speech and Audio Processing, Vol. 11, No. 6, pp. 709-716, 2003.

[11] G. Golub and C. Van Loan, Matrix Computations, Johns Hopkins University Press, 1996.

[12] 清水, 松本, 小野, 嵯峨山, “等方的雑音場を直交化するアレイ信号処理の理論とパワースペクトル推定への応用,” 日本音響学会春季研究発表会講演集, pp.569-570, 3 月, 2007.

[13] H. Shimizu, N. Ono, K. Matsumoto, and S. Sagayama, “Isotropic Noise Suppression on Power Spectrum Domain by Symmetric Microphone Array,” Proc. WASPAA, Oct, 2007. (to appear in)

[14] 伊藤, 小野, 嵯峨山, “等方的雑音場における結晶型アレイを用いた非線形ビームフォーマ,” 日本音響学会秋季研究発表会講演集, 9 月, 2007. (掲載予定)

[15] 小野, 河野, 清水, 伊藤, 嵯峨山, “等方的雑音場直交化のためのアレイ配置に関する群論的検討,” 日本音響学会秋季研究発表会講演集, 9 月, 2007. (掲載予定)

[16] D. H. Johnson *et al.*, Array signal processing, Prentice Hall, 1993.