

混合 DOA モデルに基づく多チャンネル複素 NMF による劣決定 BSS *

武田和馬 (筑波大), 亀岡弘和 (東大/NTT CS 研),
澤田宏, 荒木章子 (NTT CS 研), 宮部滋樹, 山田武志, 牧野昭二 (筑波大)

1 序論

ブラインド音源分離 (Blind Source Separation; BSS) とは, 音源の成分と音源からマイクロホンまでの伝達特性がともに未知のもとで, マイクロホン入力信号から音源成分を復元する技術である。音声信号を対象とした BSS は, ハンズフリーテレビ会議システムなど, 多くの応用が期待されている。

以下, まず BSS の問題を定式化する。\$M\$ 個の信号源から到来する音源信号を \$J\$ 個のマイクロホンで観測する場合を考え, \$j\$ 番目のマイクロホンで観測される信号の時間周波数成分を \$y_j(\omega, t)\$, \$m\$ 番目の音源信号の時間周波数成分を \$s_m(\omega, t)\$ とし, \$\mathbf{y}(\omega, t) = (y_1(\omega, t), \dots, y_J(\omega, t))^T \in \mathbb{C}^J\$, \$\mathbf{s}(\omega, t) = (s_1(\omega, t), \dots, s_M(\omega, t))^T \in \mathbb{C}^M\$ とする。ただし, \$1 \le \omega \le \Omega\$, \$1 \le t \le T\$ は, それぞれ周波数および時刻に対応するインデックスである。ここで, 時間周波数分析 (短時間 Fourier 変換など) の分析窓長に比べて, 信号源から観測点までのインパルス応答長が十分短いとすると, 観測信号モデルは近似的に

$$\mathbf{y}(\omega, t) = \mathbf{A}(\omega)\mathbf{s}(\omega, t) + \mathbf{n}(\omega, t) \quad (1)$$

のように瞬時混合の形で表すことができる。この近似は, 各狭帯域信号において畳み込み混合モデルを仮定していることに相当するため, 狭帯域近似という。\$m\$ 番目の信号源から \$j\$ 番目のマイクロホンまでの伝達周波数特性 \$a_{j,m}(\omega)\$ を要素にした行列 \$\mathbf{A}(\omega) = (a_{j,m}(\omega))_{J \times M} = (\mathbf{a}_1(\omega), \dots, \mathbf{a}_M(\omega)) \in \mathbb{C}^{J \times M}\$ を混合行列と呼び, 以下ではこれを時不変と仮定する。\$\mathbf{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。このモデルに基づく BSS は周波数領域 BSS と呼ばれ, 時間領域の畳み込み混合モデルに基づく BSS に比べて演算量の少ないアルゴリズムを構成できる点が特長である。

BSS では観測信号だけから未知の混合行列と音源成分を推定する必要があるため, 通常は音源に関して何らかの仮定を置き, これにより立てられる規準をもとに両未知変数を最適推定する問題として定式化される。例えば, 観測信号数が信号源数以上の場合には, 独立成分分析が有効な手法として知られ [1], 音源信号間の独立性を最大化するように分離行列 (混合行列の逆行列) を推定することが目的となる。しかし, 信号源の数が観測信号の数よりも多い劣決定の条件下では, たとえ混合過程が既知であったとしても式 (1) を満たす音源成分の解は無数に存在し, 独立性の規準ではこの中から一意に解が決められない。このため, 劣決定の条件では, 音源に関してスパース性などの独立性よりさらに強い仮定が必要となる [2-6]。また一方で, 仮に \$\omega\$ ごとに分離が達成され, \$M\$ 個の分離信号

\$s_1(\omega, t), \dots, s_M(\omega, t)\$ が得られたとしても, これら \$M\$ 個の成分がそれぞれどの信号源のものであるかは一意に決定できず, その対応づけのためには別の手が必要となる。この問題はパーミュテーション整合問題と呼ばれ, 周波数領域 BSS では通常 BSS の後段処理としてパーミュテーションの整合が行われ, 分離信号が最終的に得られる。パーミュテーション整合の方法として, これまで音源到来方向 [8], 帯域間の振幅相関 [9], それらの組み合わせと調波性 [10], デルタ振幅成分の帯域間の同期性 [11] などを利用した様々な解法が検討されている。

以上のように, 劣決定条件の周波数領域 BSS は, 周波数ごとの分離問題の劣決定性とパーミュテーション問題の任意性を同時に有することになる。これまで我々は, 複素 NMF と呼ぶモノラル信号分離手法 [12] において仮定した信号モデルを多チャンネルに拡張し, 周波数ごとの信号分離とパーミュテーション整合を同時解決することを目的とした, 多チャンネル複素 NMF と呼ぶ枠組を提案した。そして, このコンセプトに基づきいくつかの手法を検討してきた [13-16]。一方, 近年, 非負値行列因子分解 (Non-negative Matrix Factorization; NMF) を多チャンネル化した BSS の手法が提案され, 周波数ごとの信号分離とパーミュテーション整合を同時解決できる手法として注目されている [17, 18]。多チャンネル複素 NMF や多チャンネル NMF では, 同一音源の信号の周波数成分はコヒーレントに時間変化する, という仮定に基づき観測信号がモデル化されるが, 本稿では, この仮定に加え, 同一音源の信号は単一方向から平面波として到来する, という仮定を組み込んだ統合モデルを提案する。

2 多チャンネル複素 NMF

まず, 多チャンネル複素 NMF の基本モデルについて述べる。観測信号に関し, 以下を仮定する。

仮定 1. 観測信号は, 振幅成分が周波数間でコヒーレントに時間変化する, \$I\$ 個の要素信号 \$s_i(\omega, t)\$ からなる。

ここで, 要素信号とは, 各音源信号をより細かく分解した成分を指す。よって, 各要素信号はいずれかの音源に属することになる。\$i\$ 番目の要素信号の周波数成分比を \$H_i(\omega)\$ とし, これに対応する時刻 \$t\$ におけるゲインおよび位相スペクトルをそれぞれ \$U_i(t)\$, \$\phi_i(\omega, t)\$ とすると, 仮定 1 より, 要素信号は

$$s_i(\omega, t) = H_i(\omega)U_i(t)e^{j\phi_i(\omega, t)} \quad (2)$$

と表される。ただし,

$$\forall i: \sum_{\omega} H_i(\omega) = 1 \quad (3)$$

*Underdetermined blind source separation with multichannel complex NMF incorporating DOA mixture model. by Kazuma TAKEDA (University of Tsukuba), Hirokazu KAMEOKA (University of Tokyo/NTT Corporation), Hiroshi SAWADA, Shoko ARAKI (NTT Corporation), Shigeki Miyabe, Takeshi YAMADA, Shoji MAKINO (University of Tsukuba)

とする。周波数成分比 $H_i(\omega)$ が時不変であることから, $s_i(\omega, t)$ の振幅成分が周波数間でコヒーレントに時間変化するような拘束を与えていることになる。以上より, i 番目の要素信号と j 番目のマイクロホンとの間の伝達周波数特性を $a_{j,i}(\omega)$ とし, $\mathbf{a}_i(\omega) = (a_{1,i}(\omega), \dots, a_{J,i}(\omega))^T$ とすると, 観測信号 $\mathbf{y}(\omega, t)$ は,

$$\mathbf{y}(\omega, t) = \sum_{i=1}^I \mathbf{a}_i(\omega) s_i(\omega, t) + \mathbf{n}(\omega, t) \quad (4)$$

と表される。ここで, $\mathbf{a}_i(\omega)$ と $s_i(\omega, t)$ のスケールの任意性を除くため, $a_{1,i}(\omega) = 1$ とする。雑音成分 $\mathbf{n}(\omega, t)$ は, 平均が $\mathbf{0}$, 共分散が $\sigma^2(\omega)\mathbf{I}$ の複素正規分布 $\mathcal{N}_{\mathbb{C}}(0, \sigma^2(\omega)\mathbf{I})$ に従うと仮定する。

$$\mathbf{n}(\omega, t) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma^2(\omega)\mathbf{I}) \quad (5)$$

また, [12] と同様, ゲイン関数 $U_i(t)$ をスパースにする目的で事前分布として一般化正規分布を仮定する。

$$U_i(t) \sim \mathcal{GN}(0, v, \gamma) \quad (6)$$

ただし, $x \sim \mathcal{GN}(0, v, \gamma) \propto e^{-|x|^\gamma/v^\gamma}$ である。

観測信号を以上のような要素信号の和で表す考え方は, 複素 NMF [12] における観測信号のモデル化に基づくため, 式 (2) の要素信号モデルに基づく BSS 方式を多チャンネル複素 NMF と呼ぶ。

3 混合 DOA モデル

次に, 以下を仮定する。

仮定 2. 要素信号は単一方向から平面波到来する

ここまで各要素信号の伝達周波数特性を周波数 ω ごとの独立な変数であるかのように扱っていたが, もし上記の仮定が成立するならば, 例えばマイクロホン数が 2 の場合, 伝達周波数特性の各 ω 間の関係は, 到来方向 (Direction-of-Arrival; DOA) θ の関数として

$$\ell_\theta(\omega) = \begin{bmatrix} 1 \\ e^{j\omega d \cos \theta/c} \end{bmatrix} \quad (7)$$

と陽に表される。ただし, $0 \leq \theta < 2\pi$, d をマイクロホンの間隔 (m), c を音速 (m/s) とする。実際には残響や狭帯域近似の影響で, $\mathbf{a}_i(\omega)$ は上記の理論式からは逸脱する。そこで, $\mathbf{a}_i(\omega)$ は, 到来方向 θ_i が既知のとき, $\ell_{\theta_i}(\omega)$ を中心とした複素正規分布より生成されたものと仮定する。到来方向 θ_i は実際には観測することができないため, これを潜在変数と見なすことにすると, $\mathbf{a}_i(\omega)$ の生成モデルは, DOA を潜在変数とした混合モデルとなる。これを多チャンネル複素 NMF に基づく観測信号の生成モデルに組み込み, 観測信号が与えられた下で全体の生成モデルのパラメータ推論を行うことで音源分離とパーミュテーション整合を同時に達成しようというのが提案法の要点である。

まず, DOA を K 個の候補値 $\Theta_1, \dots, \Theta_K$ からなる離散的な変数とする。例えば, 180 度を K 等分した角度の集合, $\Theta_k = (k-1)\pi/K$ とする。これらの DOA の候補値の中から一つの値が選択され, 各要素信号の DOA が決定される, というプロセスをモデル化した。このプロセスは以下のように記述できる。

$$z_i | \pi \sim \text{Discrete}(\pi_1, \dots, \pi_K) \quad (8)$$

$$\theta_i = \Theta_{z_i} \quad (9)$$

ただし, $\pi = (\pi_1, \dots, \pi_K)$ である。 $z_i \in \{1, \dots, K\}$ は i 番目の要素信号にどの DOA 候補値が割り当てられるかを表すインジケータ変数であり, 上式はこれが離散分布 (各確率値が π_1, \dots, π_K) から生成されることを意味している。このプロセスにより要素信号の DOA が決定され, 伝達周波数特性 $\mathbf{a}_i(\omega)$ は,

$$a_{j,i}(\omega) | z_i \sim \mathcal{N}_{\mathbb{C}}([\ell_{\theta_i}(\omega)]_j, \nu^2(\omega)), \quad (j \geq 2) \quad (10)$$

により生成される。ただし, $[\cdot]_j$ はベクトルの j 番目の要素を表す。

ここで, 音源が空間的にスパースに点在するなら, 要素信号の到来方向はスパースとなるはずである。従って, z_i が従う離散分布をできるだけスパースにしたい。すなわち, $1, \dots, K$ のうち, 一部の少数を除くほとんどのものにほぼ 0 の確率値が割り当てられているような離散分布にしたい, ということである。このためにはスパースな π を生成しやすいような事前分布を設定すれば良い。そこで,

$$\pi \sim \text{Dirichlet}(\alpha_0/K, \dots, \alpha_0/K) \quad (11)$$

を仮定する。ただし, $\text{Dirichlet}(\alpha_0/K, \dots, \alpha_0/K)$ は Dirichlet 分布であり, $0 < \alpha_0 < K$ のときスパースな π を生成しやすくなる。

以上の, 混合モデルに基づく $\mathbf{a}_i(\omega)$ の生成モデルを「混合 DOA モデル」と呼ぶ。

4 最大事後確率推定

以上の生成モデルに基づき, 観測信号 $Y := \{\mathbf{y}(\omega, t)\}_{1 \leq \omega \leq \Omega, 1 \leq t \leq T}$ が与えられたもとの事後確率が最大となる未知パラメータを推定するアルゴリズムを導く。観測信号 Y が与えられた下でのパラメータ

$$a := \{\mathbf{a}_i(\omega)\}_{1 \leq i \leq I, 1 \leq \omega \leq \Omega} \quad (12)$$

$$H := \{H_i(\omega)\}_{1 \leq i \leq I, 1 \leq \omega \leq \Omega} \quad (13)$$

$$U := \{U_i(t)\}_{1 \leq i \leq I, 1 \leq t \leq T} \quad (14)$$

$$\phi := \{\phi_i(\omega, t)\}_{1 \leq i \leq I, 1 \leq \omega \leq \Omega, 1 \leq t \leq T} \quad (15)$$

の事後確率は,

$$P(a, H, U, \phi, \pi | Y) \propto \sum_z p(a, H, U, \phi, z, \pi | Y) \propto p(Y | a, H, U, \phi) p(\pi) p(U) \sum_z p(a | z) p(z | \pi) \quad (16)$$

で与えられ, 右辺の対数を $L(a, H, U, \phi, \pi)$ とすると,

$$L(a, H, U, \phi, \pi) = \log p(Y | a, H, U, \phi) + \log p(\pi) + \log p(U) + \log \sum_z p(a | z) p(z | \pi) \quad (17)$$

となる。ただし, $z := \{z_i\}_{1 \leq i \leq I}$ である。ここで, 式 (4), (5), (6), (8), (10) より,

$$p(Y | a, H, U, \phi) \propto \prod_{\omega, t} e^{-\|\mathbf{y}(\omega, t) - \sum_i \mathbf{a}_i(\omega) s_i(\omega, t)\|_2^2 / \sigma^2(\omega)} p(a | z) \propto \prod_{j \neq 1} \prod_{i, \omega} e^{-|a_{j,i}(\omega) - l_{j,z_i}(\omega)|^2 / \nu^2(\omega)}$$

$$\begin{aligned}
p(z|\pi) &= \prod_i \pi_{z_i} \\
p(\pi) &\propto \prod_k \pi_k^{\alpha_0/K-1} \\
p(U) &\propto \prod_i \prod_t e^{-2\rho|U_i(t)|^\gamma} \quad (18)
\end{aligned}$$

である。表記の簡略化のため $l_{j,k}(\omega) := [\ell_{\Theta_k}(\omega)]_j$, $\rho := 1/(2\nu^\gamma)$ とした。以下、目的関数 $L(a, H, U, \phi)$ を a, H, U, ϕ に関して最大化する最適化問題を考える。

以上の最適化問題の大域最適解は解析的に得ることは難しいが、[12] と同様、局所最適解を効率的に探索する反復アルゴリズムを補助関数法に基づき導くことができる。補助関数法は、最大化したい目的関数の下界関数（補助関数と呼ぶ）を設計し、その下界関数を反復的に増加させていくことで間接的に目的関数を増加させていく最適化方法であるが、これを適用するためには目的関数の下界関数を設計する必要がある。

まず、 $\log p(Y|a, H, U, \phi)$ の下界関数は以下の不等式

$$\begin{aligned}
&\log p(Y|a, H, U, \phi) \\
&\stackrel{c}{=} - \sum_{\omega, t} \left\| \mathbf{y}(\omega, t) - \sum_i \mathbf{a}_i(\omega) s_i(\omega, t) \right\|_2^2 / \sigma^2(\omega) \\
&\geq - \sum_{\omega, t} \frac{1}{\sigma^2(\omega)} \sum_{j, i} \frac{1}{\beta_{j,i}(\omega, t)} \\
&\quad \left| x_{j,i}(\omega, t) - a_{j,i}(\omega)^{1(j \neq 1)} s_i(\omega, t) \right|^2 \quad (19)
\end{aligned}$$

により設計できる [12]。 $\stackrel{c}{=}$ は定数項を除いた項の等号を表す。 $\mathbf{1}(\cdot)$ は括弧内の条件を満たす場合に 1 を返し、満たさない場合に 0 を返す指示関数である。また、 $\beta_{j,i}(\omega, t)$ は、 $\sum_i \beta_{j,i}(\omega, t) = 1$, $\beta_{j,i}(\omega, t) > 0$ を満たす範囲で任意に定めても良いパラメータである。 $x_{j,i}(\omega, t)$ は $\sum_i x_{j,i}(\omega, t) = y_j(\omega, t)$ を満たす補助変数であり、

$$\begin{aligned}
x_{j,i}(\omega, t) &= a_{j,i}(\omega)^{1(j \neq 1)} s_i(\omega, t) \\
&+ \beta_{j,i}(\omega, t) \left(y_j(\omega, t) - \sum_i a_{j,i}(\omega)^{1(j \neq 1)} s_i(\omega, t) \right) \quad (20)
\end{aligned}$$

のとき上不等式の等号は成立する。

次に、 $\log \sum_z p(a|z)p(z|\pi)$ の下界関数は、負の対数関数が凸関数であることを利用し、Jensen の不等式

$$\begin{aligned}
&\log \sum_z p(a|z)p(z|\pi) \quad (21) \\
&= \sum_i \log \sum_k \pi_k \prod_{j \neq 1} \prod_\omega e^{-|a_{j,i}(\omega) - l_{j,k}(\omega)|^2 / \nu^2(\omega)} \\
&\quad \pi_k \prod_{j \neq 1} \prod_\omega e^{-|a_{j,i}(\omega) - l_{j,k}(\omega)|^2 / \nu^2(\omega)} \\
&\geq \sum_{i,k} \lambda_{i,k} \log \frac{\pi_k}{\lambda_{i,k}}
\end{aligned}$$

により設計できる。ここで、 $\lambda_{i,k}$ は、 i 番目の要素信号がそれぞれの方向からどのくらいの確率で到来しているかを意味する補助変数であり、 $\sum_k \lambda_{i,k} = 1$, $\lambda_{i,k} > 0$ を満たす。また、 $\lambda_{i,k}$ が

$$\lambda_{i,k} = \frac{\pi_k \prod_{j \neq 1} \prod_\omega e^{-|a_{j,i}(\omega) - l_{j,k}(\omega)|^2 / \nu^2(\omega)}}{\sum_n \pi_n \prod_{j \neq 1} \prod_\omega e^{-|a_{j,i}(\omega) - l_{j,n}(\omega)|^2 / \nu^2(\omega)}} \quad (22)$$

のとき上不等式の等号は成立する。

最後に、 $\log p(U)$ の下界関数を、 γ の値に応じて以下のいずれかの不等式

$$\begin{aligned}
\log p(U) &\stackrel{c}{=} -2\rho \sum_{i,t} |U_i(t)|^\gamma \geq -2\rho \sum_{i,t} f_\gamma[U_i(t)] \quad (23) \\
f_\gamma[U_i(t)] &= \begin{cases} \frac{\gamma|V_i(t)|^{\gamma-2} U_i(t)^2 + |V_i(t)|^\gamma - \frac{\gamma|V_i(t)|^\gamma}{2} & (1 < \gamma \leq 2) \\ \gamma|V_i(t)|^{\gamma-1} (U_i(t) - V_i(t)) + |V_i(t)|^\gamma & (0 < \gamma \leq 1) \end{cases}
\end{aligned}$$

を用いて設計する。いずれも、補助変数 $V_i(t)$ が

$$V_i(t) = U_i(t) \quad (24)$$

のとき等号成立する。

以上をまとめると、

$$\begin{aligned}
L^+(a, H, U, \phi, \pi, x, \lambda, V) \quad (25) \\
&= - \sum_{\omega, t} \frac{1}{\sigma^2(\omega)} \sum_{j, i} \frac{1}{\beta_{j,i}(\omega, t)} \\
&\quad \left| x_{j,i}(\omega, t) - a_{j,i}(\omega)^{1(j \neq 1)} s_i(\omega, t) \right|^2 \\
&+ \sum_{i,k} \lambda_{i,k} \left(\log \pi_k - \log \lambda_{i,k} \right. \\
&\quad \left. - \sum_{j \neq 1} \sum_\omega |a_{j,i}(\omega) - l_{j,k}(\omega)|^2 / \nu^2(\omega) \right) \\
&+ (\alpha_0/K - 1) \sum_k \log \pi_k - 2\rho \sum_{i,t} f_\gamma[U_i(t)] + C
\end{aligned}$$

は $L(a, H, U, \phi, \pi)$ の補助関数の要件を満たす。ただし、 C はパラメータに依らない定数項である。

以下、この補助関数 L^+ をもとに $1 < \gamma \leq 2$ の場合のパラメータの更新則を導く。 H に関しては、厳密には式 (3) の条件を考慮に入れる必要があるが、ここでは簡単のため、無制約で L^+ を最大化する $H_i(\omega)$ を求めたのちに式 (3) を満たすように規格化する方法をとることとする。スペースの都合上、インデックス ω, t を下付き添字で表記することになると、 L^+ を最大にするパラメータは

$$a_{j,i,\omega} = \quad (26)$$

$$\begin{aligned}
H_{i,\omega} &= \frac{H_{i,\omega} \sum_t \frac{x_{j,i,\omega,t} U_{i,t} e^{-j\phi_{i,\omega,t}}}{\beta_{j,i,\omega,t}} + \frac{\sigma_\omega^2}{\nu_\omega^2} \sum_k \lambda_{i,k} l_{j,k,\omega}}{H_{i,\omega}^2 \sum_t \frac{U_{i,t}^2}{\beta_{j,i,\omega,t}} + \frac{\sigma_\omega^2}{\nu_\omega^2}} \\
H_{i,\omega} &= \frac{\sum_t U_{i,t} \sum_j \frac{\text{Re}(x_{j,i,\omega,t}^* a_{j,i,\omega}^{1(j \neq 1)}) e^{j\phi_{i,\omega,t}}}{\beta_{j,i,\omega,t}}}{\sum_t U_{i,t}^2 \sum_j \frac{|a_{j,i,\omega}^{1(j \neq 1)}|^2}{\beta_{j,i,\omega,t}}} \quad (27)
\end{aligned}$$

$$\begin{aligned}
U_{i,t} &= \frac{\sum_\omega \frac{H_{i,\omega}}{\sigma_\omega^2} \sum_j \frac{\text{Re}(x_{j,i,\omega,t}^* a_{j,i,\omega}^{1(j \neq 1)}) e^{j\phi_{i,\omega,t}}}{\beta_{j,i,\omega,t}}}{\sum_\omega \frac{H_{i,\omega}^2}{\sigma_\omega^2} \sum_j \frac{|a_{j,i,\omega}^{1(j \neq 1)}|^2}{\beta_{j,i,\omega,t}} + \rho\gamma|V_{i,t}|^{\gamma-2}} \quad (28)
\end{aligned}$$

$$e^{j\phi_{i,\omega,t}} = \frac{\sum_j x_{j,i,\omega,t} a_{j,i,\omega}^{*1(j \neq 1)} / \beta_{i,j,\omega,t}}{\left| \sum_j x_{j,i,\omega,t} a_{j,i,\omega}^{*1(j \neq 1)} / \beta_{i,j,\omega,t} \right|} \quad (29)$$

$$\pi_k = \frac{\sum_i \lambda_{i,k} + \alpha_0 / K - 1}{\sum_n \sum_m \lambda_{m,n} + \alpha_0 - K} \quad (30)$$

与えられる。 $\sigma^2(\omega)$ と $\nu^2(\omega)$ の更新則も同様に解析的に得ることができるが、以下の実装ではいずれも固定値とした。補助関数法の原理より、以上の更新則によるパラメータの更新と式 (20), (22), (24) による補助変数の更新を交互に繰り返すことで $L(a, H, U, \phi, \pi)$ を単調に増加させていくことができる。

上述の反復アルゴリズムにより得られたパラメータと補助変数を用い、以下の手順により分離信号を得る。

1. (DOA 推定) π_k の値より、各音源がどの角度から到来しているらしいかを決定する。
2. (要素信号のグルーピング) 前段の推定結果と $\lambda_{i,k}$ の値から、各要素信号がどの角度から到来していて、どの音源に属しているかを決定する。
3. (分離信号の獲得) 前段のグルーピング結果をもとに、各音源の信号を、補助変数 $x_{j,i}(\omega, t)$ を足し合わせることで構成する。

5 動作実験

本章では提案法の音源分離性能および DOA 推定性能の検証結果について述べる。

音源信号は、ピアノ、サクスの 2 つの音響信号とし、マイクロホン間の距離を 2.15[cm]、残響時間を 130[ms] として測定したインパルス応答を畳込み混合によって人工的に混合した 5[s] の信号を観測信号とした。また、観測信号の時間周波数成分は短時間 Fourier 変換 (標準化周波数 16k[Hz], フレーム長 128[ms], オーバーラップ 64[ms], Hanning 窓) により計算した。提案法の条件については、要素信号数 $I = 15$ 、観測雑音の分散 $\sigma^2(\omega) = 1$ 、ステアリングベクトルの分散 $\nu^2(\omega) = 5f/f_N$ 、角度の分割数 $K = 90$ とした。これらのパラメータは、実験に基づき決定した。補助関数法の繰り返し数は 300 とし、各パラメータ φ の初期値は乱数で設定した。分離信号の精度は、他音声の消し残りによる歪みを表す歪み具合を表す Source to Interference Ratio (SIR) [19] により評価した。単位は dB であり、この数値が高いほど性能が良い。

動作実験の結果、ピアノの SIR は分離前が 2.3[dB] だったのに対し、分離後は 11.4[dB] まで向上し、サクスの SIR は分離前が -2.0[dB] だったのに対し、分離後は 25.8[dB] まで向上した。また、Fig. 1 に混合音の作成に利用したステアリングベクトルの 2 チャンネル間の位相差と、推定ステアリングベクトルのチャンネル間の位相差 $\arg(a_{2,1}(\omega))$ を示す。Fig. 1 より、各音源の到来方向が概ね正しく推定できていることが分かる。

6 まとめ

本稿では、劣決定 BSS における、劣決定性およびパーミュテーション問題に起因する解の非一意性を同時に解消する手法として、混合 DOA モデルに基づく

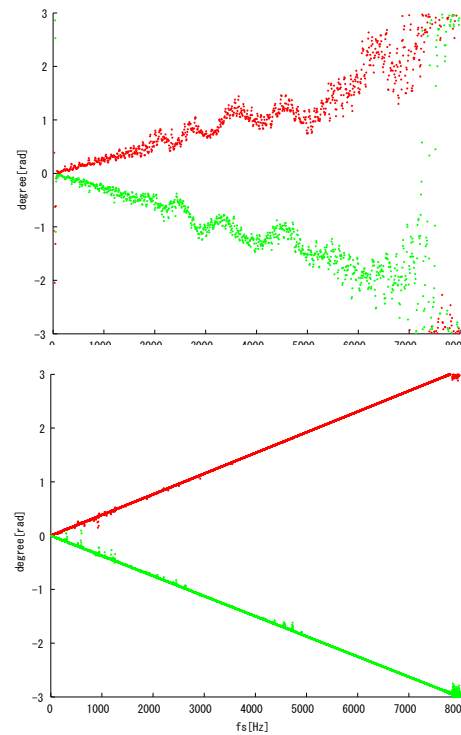


Fig. 1 正解チャンネル間位相差 (上) と推定チャンネル間位相差 $\arg(a_{2,i}(\omega))$ (下)

多チャンネル複素 NMF を提案した。音源分離実験を通して、各音源の伝達周波数特性を概ね正しく推定できたことから、パーミュテーション整合が適切に行われていることを確認した。

参考文献

- [1] 澤田他, 信学誌, Vol. 91, No. 4, pp. 292–296, 2008.
- [2] Yilmaz & Rickard, IEEE Trans. SP, **52**(7), pp. 1830–1847, 2004.
- [3] Mandel et al., Adv. Neural Inf. Proc. Sys., 2006.
- [4] Araki et al., Signal Process., **87**, pp. 1833–1847, 2007.
- [5] Mori et al., Proc. IWAENC'05, pp. 229–232, 2005.
- [6] 和泉他, 音講論 (春)'07, 2-1-5, pp. 555–556, 2007.
- [7] 小笠原基 音講論 (春)'09, 2-9-17, pp. 695–698, 2009
- [8] Kurita et al., Proc. ICASSP'00, pp. 3140–3143, 2000.
- [9] Murata et al., Neurocomputing, **41**(1–4), pp. 1–24, 2001.
- [10] Sawada et al., IEEE Trans. SAP, **12**(5), pp. 530–538, 2004.
- [11] 小野, 音講論 (秋)'10, 2-10-7, pp. 581–582, 2010.
- [12] 亀岡他, 音講論 (秋)'08, 2-8-13, pp. 657–660, 2008.
- [13] 北野他, 音講論 (秋)'09, 2-4-15, pp. 645–646, 2009.
- [14] Sawada et al., Proc. ICASSP'11, pp. 229–232, 2011.
- [15] Sawada et al., Proc. WASPAA'11, pp. 153–156, 2011.
- [16] 武田他, 音講論 (秋)'11, 1-Q-19, pp. 801–804, 2011.
- [17] Ozerov & Févotte, IEEE Trans. ASLP, **18**(3), pp. 550–563, 2010.
- [18] Sawada et al., Proc. ICASSP'12, to appear.
- [19] Vincent et al., Proc. ICA'07, pp. 552–559, 2007.