音源の W-DO 性を仮定した多チャンネル複素 NMF による劣決定 BSS*

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

1 序 論

ブラインド音源分離 (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, K 個の信号源から到来する音源信号を M 個のマイクロホンで観測する場合を考え，BSS の問題を定式化する。m 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 $y_m(\omega,t)$, k 番目の音源信号の時間周波数成分を $s_k(\omega,t)$ とし，$\boldsymbol{y}(\omega,t)=(y_1(\omega,t),\cdots,y_M(\omega,t))^{\mathrm{T}}\in\mathbb{C}^M$，$\boldsymbol{s}(\omega,t)=(s_1(\omega,t),\cdots,s_K(\omega,t))^{\mathrm{T}}\in\mathbb{C}^K$ 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，観測モデルは近似的に

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

のように瞬時混合の形で表すことができる。信号源 k からマイクロホン m までの伝達周波数特性 $a_{m,k}(\omega)$ を要素にした行列 $\boldsymbol{A}(\omega)=(a_{m,k}(\omega))_{M\times K}=(\boldsymbol{a}_1(\omega),\cdots,\boldsymbol{a}_K(\omega))\in\mathbb{C}^{M\times K}$ を混合行列と呼び，以下ではこれを時不変と仮定する。$\boldsymbol{n}(\omega,t)$ 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。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 ω ごとに分離が達成され，K 個の成分 $s_1(\omega,t),\cdots,s_K(\omega,t)$ が得られたとしても，これら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かりが必要となる。この問題はパーミュテーション問題と呼ばれ，これまで音源到来方向 [7]，帯域間の振幅相関 [8]，それらの組み合わせと調波性 [9]，デルタ振幅成分の帯域間の同期性 [10] などを利用した様々な解法が検討されていることから分かるとおり，周波数領域 BSS における重要課題の一つである。

以上のように，劣決定の周波数領域 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が，以上の解の非一意性は，(1) 音源の時間周波数成分のスパース性を仮定した観測 (混合過程) モデル，(2) 振幅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もつ音源モデル，を統合化することで効果的に解消できる可能性がある。これを実現するアプローチとして，これまで我々は多チャンネル複素 NMF と呼ぶ劣決定 BSS の解法の枠組を提案し，この思想に基づきいくつかの手法を検討してきた [11,12]。また，近年，同様の考え方に基づき，NMF(Non-negative Matrix Factorization) を多チャンネル化した BSS の手法が提案されている [13]。本稿では，この枠組の一環として，音源の混合過程に関し音源成分間の W-Disjoint Orthogonality (W-DO 性) [2](スパース性の一種) を仮定した多チャンネル複素 NMF を提案する。

2 多チャンネル複素 NMF

2.1 音源信号モデル

以下，式 (1) の観測モデルを想定し，$\boldsymbol{n}(\omega,t)$ を時間，周波数および空間的に無相関な Gauss 性雑音

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

とする。また，$\boldsymbol{A}(\omega)$ と $\boldsymbol{s}(\omega,t)$ の間のスケールおよび符号の任意性を除くため，

$$\|\boldsymbol{a}_k(\omega)\|_2^2=1,\ a_{1,k}(\omega)\ge0,\ 1\le k\le K \tag{3}$$

を仮定する。ここで，k 番目の音源の成分を表す $s_k(\omega,t)$ に関して次の仮定を立てる。

仮定 1 (振幅成分の時間変化のコヒーレント性)**.** 各音源信号を，振幅成分が周波数間でコヒーレントに時間変化する要素信号を I 個加算したものと見なす。

---

*Underdetermined Blind Source Separation with Multichannel Complex NMF Assuming W-Disjoint Orthogonality of Sources. by Kazuma TAKEDA (University of Tsukuba), Hirokazu KAMEOKA, Hiroshi SAWADA, Shoko ARAKI (NTT Corporation), Takeshi YAMADA, Shoji MAKINO (University of Tsukuba)

ある音源 k の i 番目の要素信号の周波数成分比を $H_k^i(\omega)$ 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 $U_k^i(t)$, $\phi_k^i(\omega,t)$ とすると，仮定 1 より，要素信号は

$$x_k^i(\omega,t) = H_k^i(\omega)U_k^i(t)e^{\mathrm{j}\phi_k^i(\omega,t)} \tag{4}$$

と表すことができる。ただし，

$$\sum_{\omega} H_k^i(\omega) = 1 \tag{5}$$

とする。周波数成分比 $H_k^i(\omega)$ が時不変であることから，$x_k^i(\omega,t)$ 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自動的に与えていることになっている。これを用いて音源信号モデルを

$$s_k(\omega,t) = \sum_{i=1}^{I_k} x_k^i(\omega,t) + \epsilon_k(\omega,t) \tag{6}$$

と表すこととする。ここで，$\epsilon_k(\omega,t)$ は音源ごとのモデル化誤差である。音源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 [14] における観測信号のモデル化に基づくため，仮定 1 により立てられる音源モデルに基づく BSS を多チャンネル複素 NMF と呼ぶ。

次に，音源の混合過程に関して次の仮定を立てる。

仮定 2 (音源の時間周波数成分のスパース性)**.** 各時間周波数において単一音源の成分だけがアクティブとなる。

以上の仮定は，[2] では W-Disjoint Orthogonality (W-DO) と呼ばれており，音声などのスパースな音源を対象とした劣決定 BSS に有効であることが知られ，広く用いられている。この仮定により，各時間周波数でアクティブな音源インデックスを $\hat{k}(\omega,t)$ とすると，式 (1) は

$$\begin{aligned}\boldsymbol{y}(\omega,t) &= \boldsymbol{a}_{\hat{k}(\omega,t)}(\omega)s_{\hat{k}(\omega,t)}(\omega,t) + \boldsymbol{n}(\omega,t) &(7)\\ &= \boldsymbol{A}(\omega)\boldsymbol{Z}_{\hat{k}(\omega,t)}\boldsymbol{s}(\omega,t) + \boldsymbol{n}(\omega,t) &(8)\end{aligned}$$

と書き直すことができる。$\boldsymbol{Z}_{\hat{k}(\omega,t)}$ は n 行 m 列の成分が $\delta_{n,\hat{k}(\omega,t)} \cdot \delta_{m,\hat{k}(\omega,t)}$ (ただし，$\delta_{n,m}$ は Kronecker デルタ) の $K \times K$ 行列である。

2.2 観測信号の統計モデル化

以上の仮定に基づき観測信号の確率密度関数を導く。音源 k のモデル化誤差を，分散が ν_k^2 の白色 Gauss 性雑音

$$\epsilon_k(\omega,t) \sim \mathcal{N}_{\mathbb{C}}(0,\nu_k^2) \tag{9}$$

と仮定すると，音源成分 $\boldsymbol{s}(\omega,t)$ は

$$\boldsymbol{s}(\omega,t) \sim \mathcal{N}_{\mathbb{C}}\big(\boldsymbol{\mu}(\omega,t),\boldsymbol{\Lambda}\big) \tag{10}$$

に従う。ただし，

$$\boldsymbol{\mu}(\omega,t) := \big(\mu_1(\omega,t),\cdots,\mu_K(\omega,t)\big)^{\mathrm{T}} \tag{11}$$

$$\mu_k(\omega,t) := \sum_{i=1}^{I_k} H_k^i(\omega)U_k^i(t)e^{\mathrm{j}\phi_k^i(\omega,t)} \tag{12}$$

$$\boldsymbol{\Lambda} := \mathrm{diag}\big(\nu_1^2,\cdots,\nu_K^2\big) \tag{13}$$

である。式 (2), (8), (10) より，未知パラメータ

$$\begin{aligned}\theta &:= \{\boldsymbol{a}_k(\omega),\pi_k,\sigma^2(\omega),\nu_k^2\}_{1\le k\le K,1\le\omega\le\Omega} &(14)\\ \pi_k &:= \{H_k^i(\omega),U_k^i(t),\phi_k^i(\omega,t)\}_{1\le i\le I_k,1\le\omega\le\Omega,1\le t\le T}\end{aligned}$$

および $\hat{k}(\omega,t)$ が既知の下での $\boldsymbol{y}(\omega,t)$ の確率密度関数は，

$$\begin{aligned}p(\boldsymbol{y}(\omega,t)|\hat{k}(\omega,t),\theta) = \mathcal{N}_{\mathbb{C}}\big(&\boldsymbol{A}(\omega)\boldsymbol{Z}_{\hat{k}(\omega,t)}\boldsymbol{\mu}(\omega,t),\\ &\boldsymbol{A}(\omega)\boldsymbol{Z}_{\hat{k}(\omega,t)}\boldsymbol{\Lambda}\boldsymbol{Z}_{\hat{k}(\omega,t)}\boldsymbol{A}(\omega)^{\mathrm{H}} + \sigma^2(\omega)\boldsymbol{I}\big) \quad (15)\end{aligned}$$

で与えられる。また，式 (8) より，$\boldsymbol{s}(\omega,t)$ も既知の下では $\boldsymbol{y}(\omega,t)$ の確率密度関数は，

$$\begin{aligned}&p(\boldsymbol{y}(\omega,t)|\boldsymbol{s}(\omega,t),\hat{k}(\omega,t),\theta)\\ &\qquad= \mathcal{N}_{\mathbb{C}}\big(\boldsymbol{A}(\omega)\boldsymbol{Z}_{\hat{k}(\omega,t)}\boldsymbol{s}(\omega,t),\sigma^2(\omega)\boldsymbol{I}\big) \quad (16)\end{aligned}$$

で与えられる。

3 EM アルゴリズムによる $\boldsymbol{\theta}$ の最尤推定

以下に，観測信号 $y := \{\boldsymbol{y}(\omega,t)\}_{1\le\omega\le\Omega,1\le t\le T}$ が与えられたもとで最尤の θ を推定するアルゴリズムを導く。y を不完全データと見なし，y および音源成分 $s = \{\boldsymbol{s}(\omega,t)\}_{1\le\omega\le\Omega,1\le t\le T}$ とアクティブな音源インデックス $\hat{k} = \{\hat{k}(\omega,t)\}_{1\le\omega\le\Omega,1\le t\le T}$ を合わせて完全データとして扱うと，$\log p(y|\theta)$ を θ に関して最大化する問題は EM アルゴリズムによる不完全データ問題に帰着できる。このとき Q 関数は，完全データ $c := \{y,s,\hat{k}\}$ に対する θ の対数尤度 $\log p(c|\theta)$ に対し，観測データ y と $\theta = \theta'$ が与えられた下での $s,\hat{k}$ に関する条件つき期待値

$$\begin{aligned}Q(\theta,\theta') &= \big\langle \log p(c|\theta)\big\rangle_{p(s,\hat{k}|y,\theta=\theta')} &(17)\\ &= \Big\langle\big\langle \log p(c|\theta)\big\rangle_{p(s|\hat{k},y,\theta=\theta')}\Big\rangle_{p(\hat{k}|y,\theta=\theta')} &(18)\\ &\overset{\theta}{=} \Big\langle\big\langle \log P(y|s,\hat{k},\theta)\big\rangle_{p(s|\hat{k},y,\theta=\theta')}\Big\rangle_{p(\hat{k}|y,\theta=\theta')}\\ &\quad+ \Big\langle\big\langle \log P(s|\theta)\big\rangle_{p(s|\hat{k},y,\theta=\theta')}\Big\rangle_{p(\hat{k}|y,\theta=\theta')} &(19)\end{aligned}$$

と定義される。ただし，$\langle f(\xi)\rangle_{p(\xi)}$ は分布 $p(\xi)$ に従う ξ に関する $f(\xi)$ の期待値を，$\overset{\xi}{=}$ は ξ に関係する項のみの等号を表す。式 (19) の各項の具体形を以下に示す。まず，$\langle\langle\log p(y|s,\hat{k},\theta)\rangle_{p(s|\hat{k},y,\theta=\theta')}\rangle_{p(\hat{k}|y,\theta=\theta')}$ は，

$$\begin{aligned}&\Big\langle\big\langle \log p(y|s,\hat{k},\theta)\big\rangle_{p(s|\hat{k},y,\theta=\theta')}\Big\rangle_{p(\hat{k}|y,\theta=\theta')}\\ &\quad\overset{\theta}{=} -TM\sum_{\omega}\log\sigma^2(\omega)\\ &\qquad- \sum_{\omega,t}\sum_{\hat{k}(\omega,t)=1}^{K} m_{\hat{k}(\omega,t)}\frac{\mathrm{tr}\big(\boldsymbol{F}^y_{\hat{k}(\omega,t)}(\omega,t)\big)}{\sigma^2(\omega)} \quad (20)\end{aligned}$$

の形に書ける。ただし，

$$\begin{aligned}\boldsymbol{F}^y_{\hat{k}(\omega,t)}(\omega,t) :=& \boldsymbol{y}(\omega,t)\boldsymbol{y}(\omega,t)^{\mathrm{H}} &(21)\\ &-\boldsymbol{A}(\omega)\boldsymbol{Z}_{\hat{k}(\omega,t)}\bar{\boldsymbol{s}}_{\hat{k}(\omega,t)}(\omega,t)\boldsymbol{y}(\omega,t)^{\mathrm{H}}\\ &-\boldsymbol{Z}_{\hat{k}(\omega,t)}\boldsymbol{A}(\omega)^{\mathrm{H}}\boldsymbol{y}(\omega,t)\bar{\boldsymbol{s}}_{\hat{k}(\omega,t)}(\omega,t)^{\mathrm{H}}\\ &+\boldsymbol{Z}_{\hat{k}(\omega,t)}\boldsymbol{A}(\omega)^{\mathrm{H}}\boldsymbol{A}(\omega)\boldsymbol{Z}_{\hat{k}(\omega,t)}\boldsymbol{R}_{\hat{k}(\omega,t)}(\omega,t)\end{aligned}$$

であり，

$$\bar{\boldsymbol{s}}_{\hat{k}(\omega,t)}(\omega,t) := \boldsymbol{\mu}'(\omega,t) + \boldsymbol{G}_{\hat{k}(\omega,t)}(\omega)\boldsymbol{y}(\omega,t) \tag{22}$$

$$\boldsymbol{R}_{\hat{k}(\omega,t)}(\omega,t) := \bar{\boldsymbol{s}}_{\hat{k}(\omega,t)}(\omega,t)\bar{\boldsymbol{s}}_{\hat{k}(\omega,t)}(\omega,t)^{\mathrm{H}} + \boldsymbol{\Lambda}' - \boldsymbol{G}_{\hat{k}(\omega,t)}(\omega)\boldsymbol{A}'(\omega)\boldsymbol{Z}_{\hat{k}(\omega,t)}\boldsymbol{\Lambda}' \tag{23}$$

$$\boldsymbol{G}_{\hat{k}(\omega,t)}(\omega) := \boldsymbol{\Lambda}'\boldsymbol{Z}_{\hat{k}(\omega,t)}\boldsymbol{A}'(\omega)^{\mathrm{H}} \big(\boldsymbol{A}'(\omega)\boldsymbol{Z}_{\hat{k}(\omega,t)}\boldsymbol{\Lambda}'\boldsymbol{Z}_{\hat{k}(\omega,t)}\boldsymbol{A}'(\omega)^{\mathrm{H}} + \sigma'^2(\omega)\boldsymbol{I}\big)^{-1} \tag{24}$$

である。$\bar{\boldsymbol{s}}_{\hat{k}(\omega,t)}(\omega,t)$ と $\boldsymbol{R}_{\hat{k}(\omega,t)}(\omega,t)$ はそれぞれ $\langle \boldsymbol{s}(\omega,t)\rangle_{p(s|y,\hat{k},\theta=\theta')}$ と $\langle \boldsymbol{s}(\omega,t)\boldsymbol{s}(\omega,t)^{\mathrm{H}}\rangle_{p(s|y,\hat{k},\theta=\theta')}$ に対応した量であり，$\boldsymbol{G}_{\hat{k}(\omega,t)}(\omega)$ は Wiener フィルタである。また，$m_{\hat{k}(\omega,t)}$ は，

$$m_{\hat{k}(\omega,t)} := p(\hat{k}(\omega,t)|\boldsymbol{y}(\omega,t),\theta=\theta') \tag{25}$$

を簡略表記したものであり，(ω,t) において $\hat{k}$ 番目の音源がアクティブである確率を表す。なお，ξ' は $\theta=\theta'$ のときの ξ の値を表すものとする。

同様に，$\langle\langle \log P(s|\theta)\rangle_{p(s|\hat{k},y,\theta=\theta')}\rangle_{p(\hat{k}|y,\theta=\theta')}$ は，

$$\begin{aligned}&\big\langle\big\langle \log P(s|\theta)\big\rangle_{p(s|\hat{k},y,\theta=\theta')}\big\rangle_{p(\hat{k}|y,\theta=\theta')}\\ &\overset{\theta}{=} -T\sum_{\omega}\log|\boldsymbol{\Lambda}|\\ &\quad -\sum_{\omega,t}\sum_{\hat{k}(\omega,t)=1}^{K} m_{\hat{k}(\omega,t)}\mathrm{tr}\big(\boldsymbol{\Lambda}^{-1}\boldsymbol{F}^{s}_{\hat{k}(\omega,t)}(\omega,t)\big)\end{aligned} \tag{26}$$

の形に書くことができ，$\boldsymbol{F}^{s}_{\hat{k}(\omega,t)}(\omega,t)$ は，

$$\begin{aligned}\boldsymbol{F}^{s}_{\hat{k}(\omega,t)}(\omega,t) := &\boldsymbol{R}_{\hat{k}(\omega,t)}(\omega,t)\\ &-\boldsymbol{\mu}(\omega,t)\bar{\boldsymbol{s}}_{\hat{k}(\omega,t)}(\omega,t)^{\mathrm{H}}\\ &-\bar{\boldsymbol{s}}_{\hat{k}(\omega,t)}(\omega,t)\boldsymbol{\mu}(\omega,t)^{\mathrm{H}}\\ &+\boldsymbol{\mu}(\omega,t)\boldsymbol{\mu}(\omega,t)^{\mathrm{H}}\end{aligned} \tag{27}$$

で与えられる。以上より，Q 関数の中で θ に関係する項は式 (20) と式 (26) の和で表される。

3.1 E ステップ

E ステップは，直前の M ステップで更新された θ を θ' に代入し，θ' に関係する変数に反映させるステップであるため，E ステップでは具体的には，$\bar{\boldsymbol{s}}_{\hat{k}(\omega,t)}(\omega,t)$ と $\boldsymbol{R}_{\hat{k}(\omega,t)}(\omega,t)$ と $m_{\hat{k}(\omega,t)}$ が式 (22), (23) と式 (25) により算出される。式 (25) はより具体的には，$p(\hat{k}(\omega,t)|\theta)=1/K$ とすると，

$$\begin{aligned}m_{\hat{k}(\omega,t)} :=& p(\hat{k}(\omega,t)|\boldsymbol{y}(\omega,t),\theta=\theta')\\ =&\frac{p(\boldsymbol{y}(\omega,t)|\hat{k}(\omega,t),\theta=\theta')}{\sum_{k'=1}^{K}p(\boldsymbol{y}(\omega,t)|\hat{k}(\omega,t)=k',\theta=\theta')}\end{aligned} \tag{28}$$

と書けるため，式 (15) の分布を用いて算出できる。

3.2 M ステップ

M ステップでは，Q 関数が最大化，または減少しないことが保証されるように θ を更新すれば良い。まず，$\boldsymbol{A}(\omega)$ 以外を固定した下で Q 関数を最大化する $\boldsymbol{A}(\omega)$ は，

$$\boldsymbol{A}(\omega) = \boldsymbol{\Gamma}(\omega)\boldsymbol{\Sigma}(\omega)^{-1} \tag{29}$$

$$\boldsymbol{\Gamma}(\omega) := \sum_{t}\boldsymbol{y}(\omega,t)\sum_{\hat{k}} m_{\hat{k}}\bar{\boldsymbol{s}}_{\hat{k}}(\omega,t)^{\mathrm{H}}\boldsymbol{Z}_{\hat{k}} \tag{30}$$

$$\boldsymbol{\Sigma}(\omega) := \sum_{t}\sum_{\hat{k}} m_{\hat{k}}\boldsymbol{Z}_{\hat{k}}\boldsymbol{R}_{\hat{k}}(\omega,t)\boldsymbol{Z}_{\hat{k}} \tag{31}$$

で与えられる。なお，上記に従って $\boldsymbol{A}(\omega)$ の更新を行う場合，必ずしも式 (3) を満たさないため，更新後に適切な正規化を行う必要がある。次に，音源パラメータ π_k については，式 (26) より Q 関数の中で π_k に関係する項が，

$$Q(\theta,\theta') \overset{\pi_k}{=} \frac{1}{\nu_k^2}\sum_{\omega,t}\big|[\hat{\boldsymbol{s}}(\omega,t)]_k - \mu_k(\omega,t)\big|^2 \tag{32}$$

$$\hat{\boldsymbol{s}}(\omega,t) := \sum_{\hat{k}(\omega,t)=1}^{K} m_{\hat{k}(\omega,t)}\bar{\boldsymbol{s}}_{\hat{k}(\omega,t)}(\omega,t) \tag{33}$$

と等しいため，この関数を大きくするように π_k を更新すれば良い。ただし，$[\cdot]_k$ はベクトルの k 番目の要素を表す。上述の式は，複素 NMF の目的関数 [14] と同形であるため，[14] で提案した反復アルゴリズムと同様の方法で π_k を更新することで，$Q(\theta,\theta')$ を π_k に関して局所最大化することができる。導出の詳細は [14] に委ねるが，

$$H_k^i(\omega) \leftarrow H_k^i(\omega)\frac{\sum_t U_k^i(t)|[\hat{\boldsymbol{s}}(\omega,t)]_k|}{\sum_t U_k^i(t)\sum_j H_k^j(\omega)U_k^j(t)} \tag{34}$$

$$U_k^i(t) \leftarrow U_k^i(t)\frac{\sum_\omega H_k^i(\omega)|[\hat{\boldsymbol{s}}(\omega,t)]_k|}{\sum_\omega H_k^i(\omega)\sum_j H_k^j(\omega)U_k^j(t)} \tag{35}$$

$$\phi_k^i(\omega,t) \leftarrow \arg([\hat{\boldsymbol{s}}(\omega,t)]_k) \tag{36}$$

のように繰り返し更新すれば式 (32) は局所最大化される。$\sigma^2(\omega)$ および ν_k^2 の更新則も，$Q(\theta,\theta')$ の偏微分を 0 と置くことで容易に得ることができるが，以下の実装ではこれらはいずれも固定値とした。

3.3 分離信号の獲得方法

上述の EM アルゴリズムにより得られた θ を用いて算出される $\langle\boldsymbol{s}(\omega,t)\rangle_{p(\boldsymbol{s}(\omega,t)|\boldsymbol{y}(\omega,t),\theta)}$ を分離信号として出力することにする。実は，式 (33) がこれに対応している。このことは，

$$\begin{aligned}&\langle\boldsymbol{s}(\omega,t)\rangle_{p(\boldsymbol{s}(\omega,t)|\boldsymbol{y}(\omega,t),\theta)}\\ &\quad= \big\langle\langle\boldsymbol{s}(\omega,t)\rangle_{p(\boldsymbol{s}(\omega,t)|\boldsymbol{y}(\omega,t),\hat{k}(\omega,t),\theta)}\big\rangle_{p(\hat{k}(\omega,t)|\boldsymbol{y}(\omega,t),\theta)}\\ &\quad= \langle\bar{\boldsymbol{s}}_{\hat{k}(\omega,t)}(\omega,t)\rangle_{p(\hat{k}(\omega,t)|\boldsymbol{y}(\omega,t),\theta)}\\ &\quad= \sum_{\hat{k}(\omega,t)=1}^{K} m_{\hat{k}(\omega,t)}\bar{\boldsymbol{s}}_{\hat{k}(\omega,t)}(\omega,t) = \hat{\boldsymbol{s}}(\omega,t)\end{aligned} \tag{37}$$

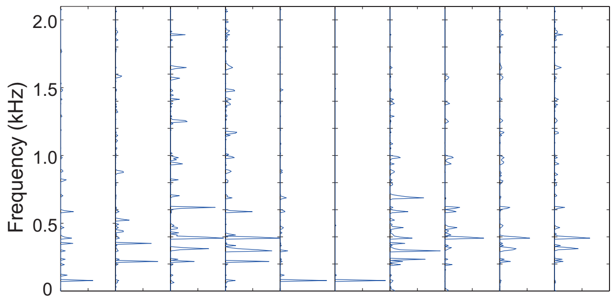

図 1　音源 1 (ピアノ音に対応) の基底スペクトル $H_1^i(\omega)$ $(H_1^1(\omega), \cdots, H_1^{10}(\omega)$ を横に並べて表示)

表 1　各音源の分離前/後の SIR[dB]

	分離前	分離後
音源 1: ピアノ	-6.8	17.7
音源 2: ボーカル	-6.1	15.4
音源 3: ドラム	2.9	14.0

により確認できる。以上より，分離信号の時間周波数成分 $\langle \boldsymbol{s}(\omega,t)\rangle_{p(\boldsymbol{s}(\omega,t)|\boldsymbol{y}(\omega,t),\theta)}$ は E ステップで算出される。

4　動作実験

本章では提案法の音源分離性能の検証結果について述べる。提案法は，周波数領域 BSS における劣決定性と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解消することを目指したものであるが，本実験ではトータルな性能評価に向けた予備検討として，劣決定性のみに対する提案法の効果を簡易に調べる目的で混合行列 $\boldsymbol{A}(\omega)$ を既知とした。

対象とする信号は，ピアノ，ボーカル，ドラムの 3 つの音源と，マイク間の距離を 1m，残響時間を 100ms 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 10 秒の楽音信号とした。また，観測信号の時間周波数成分は短時間 Fourier 変換 (標本化周波数 16kHz，フレーム長 128ms，オーバーラップ 64ms，Hanning 窓) により計算した。提案法の条件については，EM アルゴリズムの繰り返し数を 50 とし，各パラメータ θ の初期値を混合行列 $\boldsymbol{A}(\omega)$ のみ正解を与え，その他の値は全て乱数で設定した。基底スペクトル数は $I_1 = \cdots = I_K = 10$ とした。また，分離信号は式 (36) により定義される推定値とした。獲得した分離信号は客観評価によって測定した。なお，客観評価値として，他音声の消し残りによる歪みを表す歪み具合を表す SIR(Source to Interference Ratio) [15] を用いた。単位は dB であり，この歪み尺度の数値が高いほど性能が良い。

客観評価の値を表 1 に示す。ここでは，分離前の信号と各音源信号との SIR，提案手法による分離後の信号と各音源信号との SIR の値を比較した。また，図 1 に提案法で求まったピアノ音の基底スペクトル $H_k^i(\omega)$，図 2 に各基底スペクトルのゲイン $U_k^i(t)$ の例を示す。

提案手法によって混合信号を各々の音源信号に分離できたかを考察する。表 1 を見ると分かる様に，SIR の値は提案手法によって大幅に上昇している。また，図 1, 2 からは，倍音構造をもつ基底スペクトルが，時間変化しながら混ざり合うことでピアノの音が表現されているのが読み取れ，仮定 1 の通りに音源がモデル化されていることが分かる。現段階では，混合行列 $A(\omega)$ が既知という条件下でなら十分な音源分離性能が得られることが確認できた。

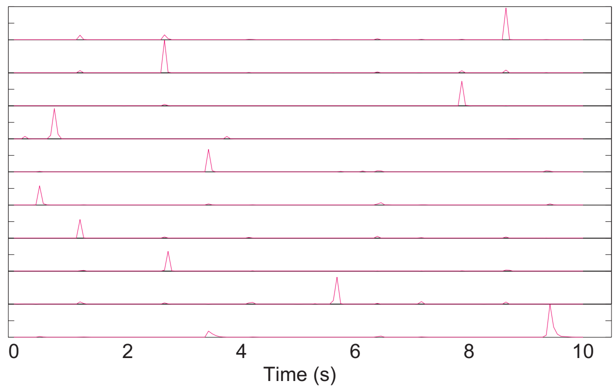

図 2　音源 1 (ピアノ音に対応) のアクティベーション $U_1^i(t)$ $(U_1^1(t), \cdots, U_1^{10}(t)$ を縦に並べて表示)

5　まとめ

本稿では，劣決定ブラインド音源分離における，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，音源の W-DO 性を仮定した多チャンネル複素 NMF を提案した。提案法による音源分離実験を通して，混合行列が既知であるなら高い精度で音源分離できることを確認した。今後の課題として，EM アルゴリズムにおけるパラメータの適切な初期値の選定方法が挙げられる。特に混合行列の初期値選定方法は最重要課題になると考えられる。また，提案法の優位性を示すために他の従来法との性能比較実験を行う予定である。

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] Kurita et al., Proc. ICASSP'00, pp. 3140–3143, 2000.
[8] Murata et al., Neurocomputing, **41**(1–4), pp. 1–24, 2001.
[9] Sawada et al., IEEE Trans. SAP, **12**(5), pp. 530–538, 2004.
[10] 小野, 音講論 (秋)'10, 2-10-7, pp. 581–582, 2010.
[11] 北野他, 音講論 (秋)'09, 2-4-15, pp. 645–646, 2009.
[12] Sawada et al., submitted to ICASSP'11 in 2010.
[13] Ozerov & Févotte, IEEE Trans. ASLP, **18**(3), pp. 550–563, 2010.
[14] 亀岡他, 音講論 (秋)'08, 2-8-13, pp. 657–660, 2008.
[15] Vincent et al., Proc. ICA'07, pp. 552–559, 2007.