複合自己回帰系に基づく音響信号の生成モデル*

○亀岡弘和，柏野邦夫
(NTT コミュニケーション科学基礎研究所)

1 序論

観測データがある確率分布に従って生成されたものと見なせるとき，その確率分布の族を一般に統計モデルという。本稿では，音声などを対象とした様々なブラインド信号処理の問題への応用を念頭に置いた，音響信号の新しい統計モデルについて論じる。

ブラインド音源分離やブラインド残響除去では，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ことが目的となる。明らかにこれは不良設定問題であるため，音源に関する何らかの仮定をもとに音源らしさの規準を定め，これにより立てられる目的関数を最小 (最大) 化する最適化問題として定式化されることが多い。例えば，独立成分分析に基づくブラインド音源分離では，音源間の信号の統計的独立性を測るための一つの指標である信号の非 Gauss 性がしばしば音源らしさを表す規準として用いられる。このように，音源らしさの規準は，どのような信号を生成しやすいか，という何らかの傾向を表した統計モデルを仮定することにより定義できる。どのような統計モデルを仮定すべきかは，それが対象とする音源信号の実際のふるまいや現象にいかに即しているか，数理的技法が馴染みやすい最適化問題に帰着するか，を考慮して判断することが重要である。

今，M 個の音源の信号を M 個のマイクロホンで収音する状況を考え，これらの収音信号を短時間 Fourier 変換 (STFT) により展開した時間周波数成分を観測データと見なすことにする。ここで，l 番目のマイクロホンで観測される信号を $Y_{k,n}^{(l)}$(ただし，k, n は周波数，時刻に対応するインデックス) とし，$\mathbf{Y}_{k,n} = (Y_{k,n}^{(1)}, \cdots, Y_{k,n}^{(M)})^{\mathrm{T}}$ とする。M 個の音源成分を並べたベクトルを $\mathbf{S}_{k,n} = (S_{k,n}^{(1)}, \cdots, S_{k,n}^{(M)})^{\mathrm{T}}$ と表すと，$\mathbf{Y}_{k,n}$ と $\mathbf{S}_{k,n}$ は近似的に

$$\mathbf{Y}_{k,n} = \sum_{t=1}^{n_l} \mathbf{G}_{k,t} \mathbf{Y}_{k,n-t} + \hat{\mathbf{S}}_{k,n} \tag{1}$$

$$\hat{\mathbf{S}}_{k,n} = \mathbf{W}_k^{-1} \mathbf{S}_{k,n} \tag{2}$$

の関係によって結ばれる [1]。ただし，$\mathbf{G}_{k,t}$ は周波数ごとの残響除去フィルタのフィルタ係数行列 (n_l はフィルタ長)，$\mathbf{W}_k$ は周波数領域の瞬時混合に対する分離行列に対応する。さて，ここで，$S_{k,n}^{(m)}$ を確率変数とし，$S_{k,n}^{(m)}$ と $S_{k',n'}^{(m')}$ は $(k,n,m) \neq (k',n',m')$ のとき独立と仮定する。そして確率密度関数 $f_{S_{k,n}^{(m)}}(s_{k,n}^{(m)}; \theta^{(m)})$ を $S_{k,n}^{(m)}$ の統計モデルとしよう。ただし，$\theta^{(m)}$ は統計モデルのパラメータである。まず，$S_{k,n}^{(m)}$ と $S_{k',n'}^{(m')}$ $(m \neq m')$ の独立性より，$\mathbf{S}_{k,n}$ の確率密度関数は，$\theta = \{\theta^{(m)}\}_m$ とすると，

$$f_{\mathbf{S}_{k,n}}(\mathbf{s}_{k,n}; \theta) = \prod_m f_{S_{k,n}^{(m)}}(s_{k,n}^{(m)}; \theta^{(m)}) \tag{3}$$

である。次に，式 (2) より $\hat{\mathbf{S}}_{k,n}$ の確率密度関数は

$$f_{\hat{\mathbf{S}}_{k,n}}(\hat{\mathbf{s}}_{k,n}) = |\det \mathbf{W}_k| f_{\mathbf{S}_{k,n}}(\mathbf{W}_k \hat{\mathbf{s}}_{k,n}; \theta) \tag{4}$$

と表すことができ，さらに $S_{k,n}^{(m)}$ と $S_{k',n'}^{(m')}$ $(n \neq n')$ の独立性より，$\hat{\mathbf{S}}_k = (\hat{\mathbf{S}}_{k,1}^{\mathrm{T}}, \cdots, \hat{\mathbf{S}}_{k,N}^{\mathrm{T}})^{\mathrm{T}}$ と置くと，

$$f_{\hat{\mathbf{S}}_k}(\hat{\mathbf{s}}_k) = \prod_n f_{\hat{\mathbf{S}}_{k,n}}(\hat{\mathbf{s}}_{k,n}) \tag{5}$$

を得る。次に，$\mathbf{Y}_k = (\mathbf{Y}_{k,1}^{\mathrm{T}}, \cdots, \mathbf{Y}_{k,N}^{\mathrm{T}})^{\mathrm{T}}$ と置けば，式 (1) は $\mathbf{G}_k \mathbf{Y}_k = \hat{\mathbf{S}}_k$ と書けるため $\mathbf{Y}_k$ の確率密度関数は，

$$f_{\mathbf{Y}_k}(\mathbf{y}_k) = |\det \mathbf{G}_k| f_{\hat{\mathbf{S}}_k}(\mathbf{G}_k \mathbf{y}_k) \tag{6}$$

で与えられる。ただし，

$$\mathbf{G}_k = \begin{bmatrix} \mathbf{I} & & & & & O \\ -\mathbf{G}_{k,1} & \ddots & & & & \\ \vdots & \ddots & \ddots & & & \\ -\mathbf{G}_{k,n_l} & \cdots & -\mathbf{G}_{k,1} & \ddots & & \\ & \ddots & \vdots & \ddots & \ddots & \\ O & & -\mathbf{G}_{k,n_l} & \cdots & -\mathbf{G}_{k,1} & \mathbf{I} \end{bmatrix} \tag{7}$$

であり，故に $\det \mathbf{G}_k = 1$ である。最後に，$S_{k,n}^{(m)}$ と $S_{k',n'}^{(m')}$ $(k \neq k')$ の独立性より，$\mathbf{Y}_1, \cdots, \mathbf{Y}_K$ の同時密度関数は

$$f_{\mathbf{Y}_1,\cdots,\mathbf{Y}_K}(\mathbf{y}_1, \cdots, \mathbf{y}_K) = \prod_k f_{\mathbf{Y}_k}(\mathbf{y}_k) \tag{8}$$

と書かれる。以上から分かるように，式 (3)〜(8) より，統計モデル $f_{S_{k,n}^{(m)}}(s_{k,n}^{(m)}; \theta^{(m)})$ の具体形を決めれば，自動的に観測信号 $\mathbf{Y}_1, \cdots, \mathbf{Y}_K$ の同時密度関数が定義される。この同時密度関数は，未知パラメータ θ, $\{\mathbf{W}_k\}$, $\{\mathbf{G}_{k,t}\}$ の尤度関数に対応するものであり，[1] のように最尤推定の枠組でブラインド音源分離とブラインド残響除去を同時に行うための統一的な目的関数となる。

以上を踏まえ，本稿では，$f_{S_{k,n}^{(m)}}(s_{k,n}^{(m)}; \theta^{(m)})$ に該当する音響信号の新しい統計モデルを提案する。また，提案する統計モデルに基づき，本報告では，最もシンプルなケースである $\mathbf{G}_{k,1} = \cdots = \mathbf{G}_{k,n_l} = O$, $\mathbf{W}_k = \mathbf{I}$, $M = 1$ の場合の θ に関する最適化問題を解く基本アルゴリズムを導く。従って，以後は $^{(m)}$ の表記は無意味なので省略する。

---

*Statistical model for acoustic signals based on composite autoregressive system. by KAMEOKA Hirokazu (NTT Corporation), KASHINO Kunio (NTT Corporation)

2 音響信号の統計モデル

音響信号のスペクトルに見られる微細構造と包絡構造は相補的な量であり，前者は，周期性の有無の情報や周期性を有する場合にはその周期ないし基本周波数の情報を含み，後者は音素や音色に相当する情報を含んでいる。音声や楽音などの意思伝達や表現の媒体として使われる音響信号の多くは，微細構造と包絡構造を個別に時間変化させて多様なスペクトルを形成しながらこれら相補的な情報を時間に載せて人間に伝達している。またさらに，音声 (楽音) においては，声の高さ (音階) や音素 (音色) の範囲や種類は限られるため，限られた種類の微細構造ないし包絡構造が個別に明滅しながら時変スペクトルが構成されていると仮定できそうである。

本章では，以上の考えに基づき，音響信号を I 種類のパワースペクトル密度 (以後，PSD) をもつ Gauss 性雑音の駆動信号と J 種類の全極型フィルタからなる複合系から生成されたものと捉えてモデル化する。

2.1 準備

まず，ある短時間信号 $\{x[t]\}$ を，P 次の自己回帰過程

$$x[t] = \sum_{p=1}^{P} a_p x[t-p] + \epsilon[t] \tag{9}$$

からの標本値系列と仮定する。ここで，$\epsilon[t]$ は，自己相関関数が $h[t]$ の定常な Gauss 性雑音とし，必ずしも白色雑音とは限らない点を強調しておく。さて，$x[1], \cdots, x[K]$ の離散 Fourier 変換 (以後，DFT) を $\mathbf{X} = (X_1, \cdots, X_K)^{\mathrm{T}} \in \mathbb{C}^K$ とすると，式 (9)，DFT の線形性，および，$\epsilon[t]$ の定常性と Gauss 性により，$\mathbf{X}$ は，平均が 0，分散共分散行列が $\mathbf{\Lambda} = \mathrm{diag}\big(\lambda_1, \cdots, \lambda_K\big)$ の多次元複素正規分布 $\mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{\Lambda})$ に従う。ただし，

$$\lambda_k = \frac{H_k}{|A(e^{\mathrm{j}2\pi k/K})|^2}, \tag{10}$$

$$A(z) = 1 - a_1 z^{-1} \cdots - a_P z^{-P} \tag{11}$$

である。$H_1, \cdots, H_K$ は $h[1], \cdots, h[K]$ の DFT，すなわち駆動信号 $\epsilon[t]$ の PSD であり，スペクトル微細構造に対応する。一方で，$1/|A(e^{\mathrm{j}2\pi k/K})|^2$ は全極型伝達関数の PSD であり，スペクトル包絡に対応する。

2.2 複合自己回帰系

ここで，上述のとおり，I 種類の駆動信号 PSD と J 種類の全極型フィルタにより構成される複合系により音響信号の STFT(フレーム内の信号の DFT) をモデル化する。この複合系は，異なる PSD をもつ合計 $I \times J$ 種類の要素信号をフレームごとに個別のゲインによってアクティベートし，それらを重畳したものを生成する機構をもつ。さて，i 番目の駆動信号 PSD と j 番目の全極型伝達関数をそれぞれ H_k^i, $1/A^j(e^{\mathrm{j}2\pi k/K})$ とし，$\{i, j\}$ 番目の要素信号の n 番目のフレームにおけるゲインを $U_n^{i,j}$ で表すと，2.1 節での議論のとおり，その STFT $\mathbf{X}_n^{i,j} = (X_{1,n}^{i,j}, \cdots, X_{K,n}^{i,j})^{\mathrm{T}} \in \mathbb{C}^K$ は，

$$\mathbf{X}_n^{i,j} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{\Lambda}_n^{i,j}) \tag{12}$$

のように分散共分散行列が $\mathbf{\Lambda}_n^{i,j} = \mathrm{diag}\big(\lambda_{1,n}^{i,j}, \cdots, \lambda_{K,n}^{i,j}\big)$ の多次元複素正規分布に従う。ただし，

$$\lambda_{k,n}^{i,j} = \frac{H_k^i U_n^{i,j}}{|A^j(e^{\mathrm{j}2\pi k/K})|^2} \tag{13}$$

$$A^j(z) = 1 - a_1^j z^{-1} \cdots - a_P^j z^{-P} \tag{14}$$

である。ここで，$\mathbf{X}_n^{i,j}$ と $\mathbf{X}_{n'}^{i',j'}$ は $(i,j) \neq (i',j')$ または $n \neq n'$ のとき独立と仮定すると，$\mathbf{X}_n^{1,1}, \cdots, \mathbf{X}_n^{I,J}$ の和信号の STFT $\mathbf{S}_n \in \mathbb{C}^K$ はやはり正規分布に従い，

$$\mathbf{S}_n = \sum_i \sum_j \mathbf{X}_n^{i,j} \sim \mathcal{N}_{\mathbb{C}}\big(\mathbf{0}, \mathbf{\Phi}_n\big) \tag{15}$$

$$\mathbf{\Phi}_n = \sum_i \sum_j \mathbf{\Lambda}_n^{i,j} \tag{16}$$

が言える。式 (16) は，正規分布する独立な確率変数の和は各々の正規分布の分散の和を分散とする正規分布に従うことを意味する。複合系のパラメータをまとめて

$$\theta = \bigcup_{k,n,i,j} \{H_k^i, a_p^j, U_n^{i,j}\} \tag{17}$$

とすると，以上より，$S_{k,n}$ の確率密度関数が具体的に

$$f_{S_{k,n}}(s_{k,n}; \theta) = \frac{1}{\pi \phi_{k,n}} \exp\left(-\frac{|s_{k,n}|^2}{\phi_{k,n}}\right) \tag{18}$$

と定義されたことになる。ここで，$\mathbf{\Phi}_n$ の対角要素の $\phi_{k,n}$ は，複合系から生成される不規則信号の PSD であり，

$$\phi_{k,n} = \sum_i \sum_j \frac{H_k^i U_n^{i,j}}{|A^j(e^{\mathrm{j}2\pi k/K})|^2} \tag{19}$$

で与えられる。以上より，式 (18) が複合自己回帰系に基づく音響信号の統計モデルを与える。

3 最適化アルゴリズム

3.1 問題設定

ここでは，1 章で示した問題において，$\mathbf{G}_{k,1} = \cdots = \mathbf{G}_{k,n_l} = O$, $\mathbf{W}_k = \mathbf{I}$, $M = 1$ の場合，すなわち，2 章で仮定した統計モデル $f_{S_{k,n}}(s_{k,n}; \theta)$ の実現値が直接観測されるケースを考える。従って，観測信号 $y_{1,1}, \cdots, y_{K,N}$ が与えられたとき，

$$\log \prod_{k,n} f_{S_{k,n}}(y_{k,n}; \theta) + \log f_\theta(\theta) \tag{20}$$

を最大化する θ を求める最大事後確率推定 (MAP) 問題を考える。式 (18) より，$\log f_{S_{k,n}}(y_{k,n}; \theta)$ は観測パワー

スペクトル $|y_{k,n}|^2$ と PSD モデル $\phi_{k,n}$ との板倉斎藤距離と定数項を除いて等しい。一方，第二項の事前確率は，同時にアクティベートできる要素信号の個数に関し，観測データに依らない何らかの傾向を統計モデルに持たせる用途に用いる。ここでは，逆ガンマ分布

$$f_\theta(\theta) \propto \prod_{i,j,n} \frac{1}{U_n^{i,j\,\alpha}} \exp\left(-\frac{\beta}{U_n^{i,j}}\right) \tag{21}$$

を仮定する。この事前確率はアクティベーションをスパース化する効果をもち，その効果は α が大きいほど高くなる。従って，$f_{S_{k,n}}(s_{k,n};\theta)$ を単一音声の統計モデルとして扱いたい場面では α を大きく取れば良く，逆に，音楽のように複数の音が同時に発音しうる音響信号の統計モデルとして扱いたい場面では小さく取れば良い。

3.2　EM アルゴリズム

以上の MAP 推定の局所最適解は，要素信号 $\hat{\mathbf{Y}}_{k,n} = (X_{k,n}^{1,1},\cdots,X_{k,n}^{I,J})^{\mathrm{T}}$ を完全データと見なして EM アルゴリズム [2] により解くことができる。

$X_{k,n}^{i,j}$ と $X_{k',n'}^{i',j'}$ は $(k,n,i,j) \neq (k',n',i',j')$ のとき独立であるため，完全データ $\hat{\mathbf{Y}} = (\hat{\mathbf{Y}}_{1,1}{}^{\mathrm{T}},\cdots,\hat{\mathbf{Y}}_{K,N}{}^{\mathrm{T}})^{\mathrm{T}}$ の対数尤度は具体的に

$$\begin{aligned}\log f_{\hat{\mathbf{Y}}}(\hat{\mathbf{y}};\theta) &= -\sum_{k,n}\left[\log\det\pi\mathbf{\Lambda}_{k,n} + \hat{\mathbf{y}}_{k,n}^{\mathrm{H}}\mathbf{\Lambda}_{k,n}^{-1}\hat{\mathbf{y}}_{k,n}\right] \\ &= -\sum_{k,n}\left[\log\det\pi\mathbf{\Lambda}_{k,n} + \mathrm{tr}\big(\mathbf{\Lambda}_{k,n}^{-1}\hat{\mathbf{y}}_{k,n}\hat{\mathbf{y}}_{k,n}^{\mathrm{H}}\big)\right]\end{aligned}$$

と書ける。上式に対し，$Y_{k,n} = y_{k,n}$ および $\theta = \theta'$ のときの条件付期待値を取り，$\log f_\theta(\theta)$ を加えると，Q 関数

$$\begin{aligned}Q(\theta,\theta') = \log f_\theta(\theta) - \sum_{k,n}\Big[&\log\det\pi\mathbf{\Lambda}_{k,n} \\ &+ \mathrm{tr}\big(\mathbf{\Lambda}_{k,n}^{-1}\mathbb{E}\big[\hat{\mathbf{y}}_{k,n}\hat{\mathbf{y}}_{k,n}^{\mathrm{H}}|Y_{k,n}=y_{k,n};\theta'\big]\big)\Big]\end{aligned} \tag{22}$$

を得る。ここで，不完全データ (観測データ)$Y_{k,n}$ と完全データ $\hat{\mathbf{Y}}_{k,n}$ に $Y_{k,n} = \mathbf{C}\hat{\mathbf{Y}}_{k,n}$(ただし，$\mathbf{C} = [1,\cdots,1]$) なる関係があるため，

$$\begin{aligned}&\mathbb{E}\big[\hat{\mathbf{y}}_{k,n}\hat{\mathbf{y}}_{k,n}^{\mathrm{H}}|Y_{k,n}=y_{k,n};\theta'\big] = \\ &\mathbf{\Lambda}'_{k,n} - \mathbf{\Lambda}'_{k,n}\mathbf{C}^{\mathrm{T}}\big(\mathbf{C}\mathbf{\Lambda}'_{k,n}\mathbf{C}^{\mathrm{T}}\big)^{-1}\mathbf{C}\mathbf{\Lambda}'_{k,n} + \\ &|y_{k,n}|^2\mathbf{\Lambda}'_{k,n}\mathbf{C}^{\mathrm{T}}\big(\mathbf{C}\mathbf{\Lambda}'_{k,n}\mathbf{C}^{\mathrm{T}}\big)^{-1}\big(\mathbf{C}\mathbf{\Lambda}'_{k,n}\mathbf{C}^{\mathrm{T}}\big)^{-1}\mathbf{C}\mathbf{\Lambda}'_{k,n}\end{aligned} \tag{23}$$

である。ただし，$\mathbf{\Lambda}'_{k,n}$ は $\mathbf{\Lambda}_{k,n}$ に $\theta = \theta'$ を代入したものを表す。θ に依らない項をまとめて c とすると上式は

$$\begin{aligned}Q(\theta,\theta') = &-\sum_{k,n}\sum_{i,j}\left[\log H_k^i U_n^{i,j} + \frac{\Psi_{k,n}^{i,j}|A^j(e^{\mathrm{j}2\pi k/K})|^2}{H_k^i U_n^{i,j}}\right] \\ &-\sum_n\sum_{i,j}\left[\alpha\log U_n^{i,j} + \frac{\beta}{U_n^{i,j}}\right] + c\end{aligned} \tag{24}$$

と書き直せる。ただし，

$$\Psi_{k,n}^{i,j} = \lambda'^{i,j}_{k,n}\left[1 + \frac{\lambda'^{i,j}_{k,n}\big(|y_{k,n}|^2 - \phi'_{k,n}\big)}{\phi'^2_{k,n}}\right] \tag{25}$$

は各要素信号の PSD 推定値であり，E ステップでは θ' の更新に伴いこの値が更新されることになる。

以上より，各パラメータの M ステップ更新則が導ける。Q 関数を H_k^i と $U_n^{i,j}$ に関して偏微分して 0 と置くと，

$$H_k^i = \frac{1}{NJ}\sum_n\sum_j \frac{\Psi_{k,n}^{i,j}|A^j(e^{\mathrm{j}2\pi k/K})|^2}{U_n^{i,j}} \tag{26}$$

$$U_n^{i,j} = \frac{1}{K+\alpha}\left[\beta + \sum_k \frac{\Psi_{k,n}^{i,j}|A^j(e^{\mathrm{j}2\pi k/K})|^2}{H_k^i}\right] \tag{27}$$

が得られる。同様に，$a_1^j,\cdots,a_P^j$ に関して偏微分して 0 と置き，連立させると，Yule-Walker 方程式

$$r_p^j = \sum_{q=1}^P a_q^j r_{p-q}^j \ (p = 1,\cdots,P) \tag{28}$$

を得る。ただし，r_p^j は，

$$r_p^j = \sum_k\left[\sum_n\sum_i \frac{\Psi_{k,n}^{i,j}}{H_k^i U_n^{i,j}}\right] e^{p\mathrm{j}2\pi k/K} \tag{29}$$

である。以上より，$a_1^j,\cdots,a_P^j$ の更新値は

$$\begin{bmatrix} a_1^j \\ \vdots \\ a_P^j \end{bmatrix} = \begin{bmatrix} r_0^j & \cdots & r_{1-P}^j \\ \vdots & \ddots & \vdots \\ r_{P-1}^j & \cdots & r_0^j \end{bmatrix}^{-1} \begin{bmatrix} r_1^j \\ \vdots \\ r_P^j \end{bmatrix} \tag{30}$$

を解くことで得られる。これは，Levinson-Durbin アルゴリズムにより高速に計算できる。

4　提案法の主要な特徴

1. $P = 0$, $J = 1$, $f_\theta(\theta) = \text{const.}$ において，$P = 0 \Rightarrow |A^j(e^{\mathrm{j}2\pi k/K})|^2 = 1$ より式 (26), (27) は

$$H_k^i = \frac{1}{N}\sum_n \frac{\Psi_{k,n}^{i,1}}{U_n^{i,1}},\ \ U_n^{i,1} = \frac{1}{K}\sum_k \frac{\Psi_{k,n}^{i,1}}{H_k^i} \tag{31}$$

   と書ける。これは，板倉斎藤距離最小化規範の非負値行列因子分解 (NMF) の乗法更新則 [3] と等価である。これよりアルゴリズム面で提案法は NMF と関係が深いことが示唆される。

2. 白色雑音を駆動とする通常の自己回帰モデルは，基本周波数が高い音声には適合しないことが知られるが，これは駆動信号の白色性の仮定からの乖離が顕著になるためである。一方，提案モデルでは，駆動信号の白色性は仮定せず，その自己相関関数 (または PSD) 自体が推定すべきパラメータとなる。

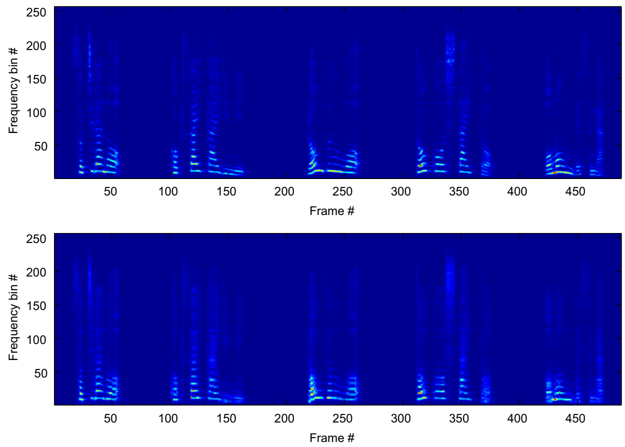

図 1　観測信号 (上段) と推定した PSD モデル (下段)

3. [1] で用いられる時変全極モデルは自己回帰パラメータがフレームごとに自由度をもつが，提案モデルでは全フレームにわたって高々J 組の自己回帰パラメータしか自由度をもたない。また，駆動信号の PSD パラメータの自由度も全フレームにわたって高々I である。これらの拘束が，系と入力の両特性が未知のもとで系を同定する手がかりを与える。

4. 式 (26) は，E ステップで更新された要素信号の PSD 推定値 $\Psi^{i,1}_{k,n}, \cdots, \Psi^{i,J}_{k,n}$ を前ステップで更新された全極型伝達関数の PSD で除算し，スペクトル包絡をフラットにしたものを，スケールを等化してから平均化する操作となっており，このことから $\Psi^{i,1}_{k,1}, \cdots, \Psi^{i,J}_{k,N}$ に含まれる共通の微細構造を抽出しようとする働きが理解される。

5. 式 (29) は，E ステップで更新された要素信号の PSD 推定値 $\Psi^{1,j}_{k,n}, \cdots, \Psi^{I,j}_{k,n}$ を $H^1_k, \cdots, H^I_k$ で除算し，微細構造をフラットにしたものを平均化した PSD を逆 Fourier 変換する操作となっている。式 (30) は，この結果求まる自己相関関数をもつ仮想的な信号に対し，従来の全極型の最尤スペクトル法を適用していることになる。

5　動作実験

ATR 音声データベースの音声データ (女性話者) を用いて提案法の基本動作の確認を行った。$y_{k,n}$ は STFT(標本化周波数 16kHz，　フレーム長 64ms，フレーム周期 32ms，Hanning 窓) により計算した。

図 2 は，図 1 上段の観測信号に対して $I = 10$, $J = 5$ の条件のもとで推定した駆動信号 PSD と全極型フィルタを示したものである。また，図 1 下段は推定した PSD モデル $\phi_{k,n}$ を図示したものである。図 2 を見ると，H^i_k の推定値には調波構造らしい特徴的な構造が表れていることが確認できる。提案する統計モデルがどの程度音声信号に合致したものとなっているかを確認するため，さ

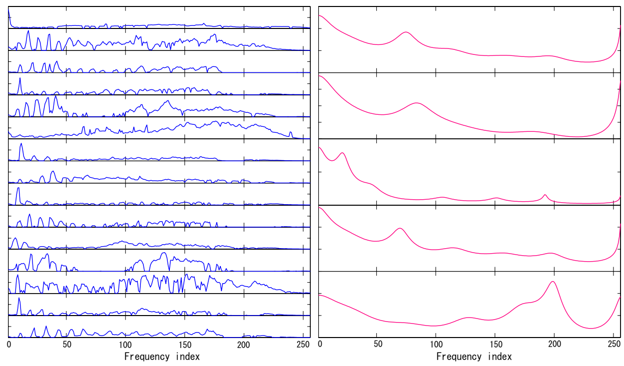

図 2　推定した駆動信号 PSD(左) と全極型フィルタ (右)

表 1　I, J の各条件でのモデル化誤差

$I \setminus J$	1	3	5
5	3.07	3.54	4.91
10	3.78	4.63	5.81
15	5.18	6.32	6.65

まざまな I と J の条件のもとで観測信号のパワースペクトルと推定した PSD モデルとの間の一致度を信号対雑音比 (SNR)

$$\mathrm{SNR(dB)} = 10 \log_{10} \frac{\sum_{k,n} |y_{k,n}|^2}{\sum_{k,n} \left| |y_{k,n}| - \sqrt{\phi_{k,n}} \right|^2}$$

により測定した。表 1 はこの結果を示したものである。

6　まとめ

本稿では，音声などを対象とした様々なブラインド信号処理の問題への応用を念頭に置き，I 種類の駆動信号の PSD と J 種類の全極型フィルタによって構成される複合系 (複合自己回帰系) に基づく音響信号の新しい統計モデルを提案した。この統計モデルの実現値が直接観測できる状況において，観測データが与えられたもとでのモデルパラメータの最大事後確率推定行う基本アルゴリズムを EM アルゴリズムにより導いた。このアルゴリズムは，板倉齋藤距離を規準とした非負値行列因子分解と同形の駆動信号 PSD の更新則と，Yule-Walker 方程式に基づく自己回帰パラメータの更新則からなる。

今後は，提案モデルをブラインド音源分離，ブラインド残響除去の問題に本格適用していく予定である。

参考文献

[1] 吉岡, 中谷, 三好, ブラインド音源分離と残響除去の統合のための一手法, 音講論 (秋), 703–704, 2008.

[2] M. Feder, E. Weinstein, Parameter estimation of superimposed signals using the EM algorithm, IEEE Trans. ASSP, **36**(4), pp. 477–489, 1988.

[3] C. Févotte, N. Bertin, J.-L. Durrieu, Nonnegative matrix factorization with the Itakura-Saito divergence with application to music analysis, Tech. Rep. TELECOM ParisTech 2008D006, 2008.