周波数領域ブラインド音源分離と周波数領域適応ビームフォーマの関係について*

◎荒木 章子†, 牧野 昭二†, 向井 良†, 猿渡 洋‡
(† 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所, ‡ 奈良先端科学技術大学院大学)

1. はじめに

ブラインド音源分離 (Blind Source Separation:BSS) は、観測された混合信号のみから音源信号を推定する手法であり、各音源信号同士の独立性の仮定に基づく手法をはじめ近年多くの手法が提案されている。

本稿では、周波数領域 BSS の枠組と周波数領域適応ビームフォーマ (adaptive beamformer:ABF) の枠組を統一的に論じる。ref マイクに目的音の漏れがあるノイズキャンセラ (NC) の枠組と BSS の関係については既に議論がなされており[1,2]、BSS による分離音の無相関化が、目的音が無い場合の NC の二乗誤差最小化と等価であることが示されている。本稿では、二次の統計量を用いる周波数領域 BSS が周波数領域 ABF と二乗誤差最小の意味で等価であることを理論的に示す[3]。これより、BSS が残響に弱い理由を説明できるものと考える。また ABF の性能が BSS の性能の上限を与えるであろうことも指摘する。

2. 周波数領域 ABF

本稿では、音源数 $N=2$(目的音 1、妨害音 1)、マイク数 $M=2$ の場合を考えるが、$N=M$ であれば一般性は失われない。また $\boldsymbol{S}(\omega,m)=[S_1(\omega,m),S_2(\omega,m)]^T$、$\boldsymbol{X}(\omega,m)=[X_1(\omega,m),X_2(\omega,m)]^T$、$\boldsymbol{Y}(\omega,m)=[Y_1(\omega,m),Y_2(\omega,m)]^T$ は、それぞれ音源、観測信号および分離 (出力) 信号のフーリエ変換である。$\boldsymbol{H}(\omega)$ は音源 i からマイク j への周波数応答 $H_{ji}(\omega)$ を要素とする (2×2) の混合行列である。

ABF では、目的音方向既知を仮定する。まず、目的音が S_1、妨害音が S_2 の場合を考える [Fig.1 (a)]。目的音 S_1=0 の時間に出力 Y_1

$$Y_1(\omega,m) = \boldsymbol{W}(\omega)\boldsymbol{X}(\omega,m) \quad (1)$$

を最小にするよう、各周波数で分離フィルタ $\boldsymbol{W}(\omega)$ を推定する。ここで $\boldsymbol{W}(\omega)$=$[W_{11}(\omega),W_{12}(\omega)]$ である。二乗誤差最小の規範によりエラー関数を次のように定義する。

$$\begin{aligned} J(\omega) &= E[Y_1^2(\omega,m)] \\ &= \boldsymbol{W}(\omega)E[\boldsymbol{X}(\omega,m)\boldsymbol{X}^*(\omega,m)]\boldsymbol{W}^*(\omega) \\ &= \boldsymbol{W}(\omega)\mathbf{R}(\omega)\boldsymbol{W}^*(\omega) \end{aligned} \quad (2)$$

ここで * は共役転置を、E は期待値を表す。以下、(ω) を省く。エラー関数の最小化

$$\frac{\partial J}{\partial \boldsymbol{W}} = 2\mathbf{R}\boldsymbol{W}^* = 0 \quad (3)$$

を、目的音に対して拘束条件 $(W_{11}H_{11}+W_{12}H_{21})S_1 = c_1S_1$ を付加して解くと

$$W_{11}H_{12} + W_{12}H_{22} = 0 \quad (4)$$

$$W_{11}H_{11} + W_{12}H_{21} = c_1 \quad (5)$$

*Relationship between frequency domain blind source separation and frequency domain adaptive beamformers, by S. Araki, S. Makino, R. Mukai (NTT Communication Science Laboratories, NTT Corporation) and H. Saruwatari (Nara Institute of Science and Technology).

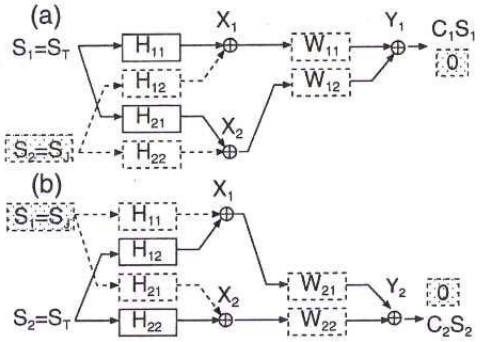

Fig. 1 Two sets of adaptive beamformers.

の連立方程式となり、この解の $\boldsymbol{W}$ を用いて一組の ABF が得られる。

同様に目的音が S_2、妨害音が S_1 の場合の ABF [Fig.1 (b)] も考え、二組の ABF をまとめて表すと、各周波数で次の関係を得る。

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix} \quad (6)$$

3. 周波数領域 BSS

周波数領域 BSS では、各周波数において入力信号が互いに独立であるという仮定と観測信号

$$\boldsymbol{X}(\omega,m) = \boldsymbol{H}(\omega)\boldsymbol{S}(\omega,m) \quad (7)$$

のみを用いて分離を行う。BSS のブロック図を Fig.2 に示す。

各周波数において $Y_1(\omega,m)$ と $Y_2(\omega,m)$ が互いに独立となるよう、逆混合行列 $\boldsymbol{W}(\omega)$ を推定する。

$$\boldsymbol{Y}(\omega,m) = \boldsymbol{W}(\omega)\boldsymbol{X}(\omega,m) \quad (8)$$

逆混合行列を求めるにはいくつかの手法があるが、ここでは、二次の統計量 (SOS) を用いる[2,4]。

非定常な音源信号 $S_1(\omega,m)$ と $S_2(\omega,m)$ が、平均 0、無相関であると仮定する。逆混合行列 $\boldsymbol{W}(\omega)$ は、出力の共分散行列 $\boldsymbol{R}_Y(\omega,k)$ の非対角要素が全てのブロック k で同時に 0 となるよう決定される。

$$\begin{aligned} \boldsymbol{R}_Y(\omega,k) &= \boldsymbol{W}(\omega)\boldsymbol{R}_X(\omega,k)\boldsymbol{W}^*(\omega) \\ &= \boldsymbol{W}(\omega)\boldsymbol{H}(\omega)\boldsymbol{\Lambda}_s(\omega,k)\boldsymbol{H}^*(\omega)\boldsymbol{W}^*(\omega) \\ &\rightarrow \boldsymbol{\Lambda}_c(\omega,k) \end{aligned} \quad (9)$$

ここで $\boldsymbol{\Lambda}_s(\omega,k)$ は $\boldsymbol{S}(\omega)$ の共分散行列であり対角行列である。また $\boldsymbol{R}_X(\omega,k)$ は $\boldsymbol{X}(\omega)$ のブロック k での共分散行列、$\boldsymbol{\Lambda}_c(\omega,k)$ は任意の対角行列である。

$\boldsymbol{R}_Y(\omega,k)$ の対角化は、次の最小化問題となる。

$$\arg\min_{W(\omega)} \sum_k ||\text{off-diag}\boldsymbol{W}(\omega)\boldsymbol{R}_X(\omega,k)\boldsymbol{W}^*(\omega)||^2 \quad (10)$$

$$s.t., \quad \sum_k \text{diag}||\boldsymbol{W}(\omega)\boldsymbol{R}_X(\omega,k)\boldsymbol{W}^*(\omega)||^2 \neq 0$$

ここで $||\boldsymbol{A}||^2$ は、$\boldsymbol{A}$ のフロベニウスノルムを表す。

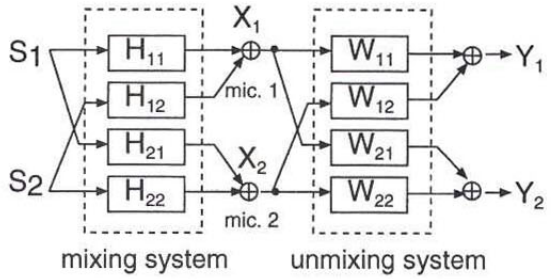

Fig. 2 BSS system configuration.

4. 周波数領域BSSと適応ビームフォーマの関係

SOSによるBSSのフレームワークでは、式(10)のように、システムの出力 $\boldsymbol{Y}$ の共分散行列

$$E\begin{bmatrix} Y_1Y_1^* & Y_1Y_2^* \\ Y_2Y_1^* & Y_2Y_2^* \end{bmatrix} \tag{11}$$

を、全てのブロックにおいて対角化する $\boldsymbol{W}$ が逆混合行列となる[式(9)参照]。

今、$\boldsymbol{H}$ と $\boldsymbol{W}$ のカスケード系を

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \tag{12}$$

と置き、式(11)の非対角要素の二乗を書き下すと、

$$\begin{aligned}&(E[Y_1Y_2^*])^2 \\ &= \left\{ad^*E[S_1S_2^*] + bc^*E[S_2S_1^*] + (ac^*E[S_1^2] + bd^*E[S_2^2])\right\}^2 \\ &= 0\end{aligned} \tag{13}$$

となる。式(13)において、S_1 と S_2 は無相関を仮定しているので、第一項と第二項は0になる。よって、SOSのBSSは式(13)の第3項を0にするよう進み、次の2通りの解が得られる。

CASE 1: $a = c_1, c = 0, b = 0, d = c_2$

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix} \tag{14}$$

これはABFの式(6)と同じである。

CASE 2: $a = 0, c = c_1, b = c_2, d = 0$

$$\begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} = \begin{bmatrix} 0 & c_2 \\ c_1 & 0 \end{bmatrix} \tag{15}$$

この式はpermutation解に相当する。

よって、BSSのエラー関数(非対角要素)の最小化はABFの二乗誤差最小化と等価であることが示された。

すなわちBSSでは、妨害音、目的音ともに存在する時間に適応できる2組のABFを形成することが分かる。更に、信号間の無相関の仮定が成り立たない場合、式(13)の第1,2項が値を持ちバイアスノイズとなることが分かる。

5. 考察

2マイクのマイクロホンアレイの支配的な動作は妨害音に1つの死角を向ける動作である。BSSにより得られた $\boldsymbol{W}$ による指向特性をFig.3に示す。BSSにより指向特性が得られること、及び、残響時間が

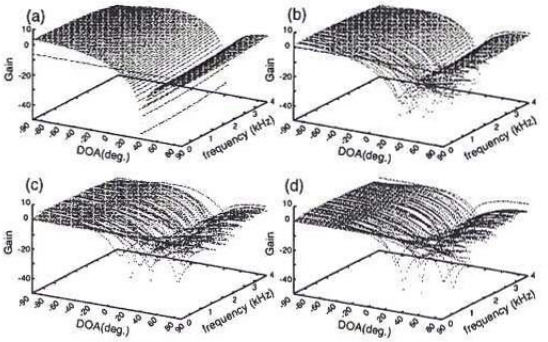

Fig. 3 指向特性 (a) 死角型ビームフォーマ,(b) BSS (残響時間 T_R=0 ms),(c) BSS (T_R=150 ms),(d) BSS (T_R=300 ms).

長い場合でも比較的鋭い指向特性が得られていることが分かる。BSSで残響音をある程度は消せることが分かっているが[5]、残響に対応するだけの長い分離フィルタの推定が難しいことも分かっている[6]。これより、様々な方向からの残響を完全には消せないことが、BSSが残響に弱い理由の一つであると考える。

また信号間の無相関の仮定が完全に成りたたない場合、BSSでは式(13)の第1,2項がバイアスノイズとして働くことから、ABFがBSSの性能の上限を与えるものと考える。

6. まとめ

本稿では、二次の統計量を用いる周波数領域BSSが周波数領域ABFと二乗誤差最小の意味で等価であることを理論的に示した。これより、BSSが残響に弱い理由を説明できるものと考える。また、ABFの性能がBSSの性能の上限を与えると考える。

参考文献

[1] S. V. Gerven and D. V. Compernolle, "Signal separation by symmetric adaptive decorrelation: stability, convergence, and uniqueness," *IEEE Trans. Speech Audio Processing*, vol. 43, no. 7, pp. 1602-1612, July 1995.

[2] E. Weinstein, M. Feder, and A. V. Oppenheim, "Multi-channel signal separation by decorrelation," *IEEE Trans. Speech Audio Processing*, vol. 1, no. 4, pp. 405-413, Oct. 1993.

[3] S. Araki, S. Makino, R. Mukai, and H. Saruwatari, "Equivalence between frequency domain blind source separation and frequency domain adaptive null beamformers," *Proc. Eurospeech2001*, Sept. 2001.

[4] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 3, pp. 320-327, May 2000.

[5] R. Mukai, S. Araki, and S. Makino, "Separation and dereverberation performance of frequency domain blind source separation for speech in a reverberant environment," *Proc. Eurospeech2001*, Sept. 2001.

[6] S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolutive mixture of speech," *Proc. ICASSP2001*, May 2001.