

カーネルビームフォーマの性能評価*

◎宮部滋樹 (東大), ジュアング・ビンファン (ジョージア工科大),
猿渡洋, 鹿野清宏 (奈良先端大)

1 はじめに

本発表では多チャネルスペクトル減算による非線形ビームフォーマの指向特性の評価を行う。2つの線形ビームフォーマとスペクトル減算を組み合わせた相補型ビームフォーマ [1] は、線形ビームフォーマよりも強力な指向性制御が可能である。しかし、相補型ビームフォーマを最適化するためには、目的関数が二次関数の場合でも収束が不安定な繰り返し演算を用いる必要があり、適応に問題を抱えていた。前回の発表では、相補型ビームフォーマが二次形式型の再生核（カーネル）を用いて記述することができ、このカーネルにより導入される再生核ヒルベルト空間上でユークリッド空間と同様の最適化が定式化できることを示した [2]。またその応用例として、カーネル独立成分分析による劣決定ブラインド音源分離 [3] を紹介した。本稿では、相補型ビームフォーマと等価にならない3素子以上のマイクロホンアレーを用いた場合に二次カーネルビームフォーマが死角を構成するメカニズムとその指向特性について議論し、また教師あり適応を掛けた場合の性能を評価する。

2 線形ビームフォーマ

本論文の議論は全て周波数領域にて行う。複素振幅 s の目的音源と $N - 1$ 個の点雑音源 d_n , $n = 2, \dots, N$ を K 素子のマイクロホンアレーで観測したとすると、マイクロホンのアレーの k 番目のマイクロホン素子での観測信号を x_k , $k = 1, \dots, K$ は以下のように表される。

$$x_k = a_{k1}s + \sum_{n=2}^N a_{kn}d_n \quad (1)$$

ここで a_{kn} は n 番目の音源から k 番目のマイクロホン素子までの伝達関数である。線形ビームフォーマの出力は各素子の観測信号をフィルタリングして足し合わせたもので、その出力信号 y は以下のように表される。

$$y = \mathbf{g}\mathbf{x}, \quad \mathbf{g} = [w_1, \dots, w_K] \quad (2)$$

$$\mathbf{x} = [x_1, \dots, x_K]^T = \mathbf{a}_1s + \sum_{n=2}^N \mathbf{a}_n d_n \quad (3)$$

$$\mathbf{a}_n = [a_{1n}, \dots, a_{Kn}]^T \quad \text{for } n = 1, \dots, N \quad (4)$$

ここで $\{\cdot\}^T$ は転置を表し、 w_k は k 番目のマイクロホンにおける観測信号のためのフィルタ係数である。ビームフォーマ出力が $y = s$ のように目的音源の推定となるのは、 \mathbf{g} が全ての雑音の伝達関数に直交 ($\mathbf{g}\mathbf{a}_n = 0$ for $n = 2, \dots, N$) し、かつ音源の伝達関数に対する利得が 1 となる場合 ($\mathbf{g}\mathbf{a}_1 = 1$) である。このような \mathbf{g} が存在するためには、目的信号の伝達関数ベクトル \mathbf{a}_1 が他のどの伝達関数ベクトル \mathbf{a}_n とも方向が同じにならず、かつ \mathbf{a}_n , $n = 1, \dots, N - 1$ が張る空間が $K - 1$ 次元以下である必要がある。この条件は、全ての伝達関数ベクトルの方向が異なると仮定すると $K > N$ となる。

3 相補型ビームフォーマ

相補型ビームフォーマの処理の流れを Fig. 1 に示す。相補型ビームフォーマは、2つの線形ビームフォーマ \mathbf{g}_1 と \mathbf{g}_2 の和の出力 $y^{(p)}$ と差の出力 $y^{(r)}$ の間でパワースペクトル減

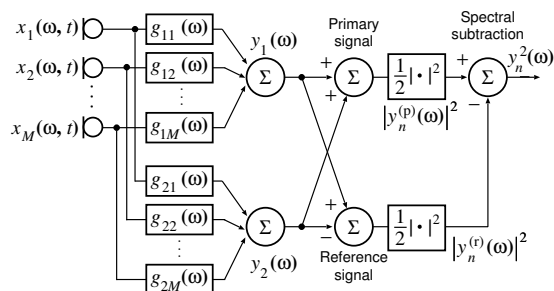


Fig. 1 相補型ビームフォーマ

算をとった信号 y^2 により目的音源のパワー $|s|^2$ の推定するものである。

$$y^2 = \frac{1}{2} |y^{(p)}|^2 - \frac{1}{2} |y^{(r)}|^2 = \mathbf{x}^H [\mathbf{g}_1^H \mathbf{g}_2 + \mathbf{g}_2^H \mathbf{g}_1] \mathbf{x} \quad (5)$$

$$y^{(p)} = (\mathbf{g}_1 + \mathbf{g}_2) \mathbf{x}, \quad y^{(r)} = (\mathbf{g}_1 - \mathbf{g}_2) \mathbf{x} \quad (6)$$

この手法では、2つのビームフォーマ \mathbf{g}_1 , \mathbf{g}_2 の両方が \mathbf{a}_0 に対する利得を 1 とし、かつ全ての雑音に関する伝達関数 \mathbf{a}_d , $d = 1, \dots, D$ が \mathbf{g}_1 か \mathbf{g}_2 のどちらかに直交する場合において、目的音源方向へのゲインが 1, 雑音方向へのゲインが 0 となる。したがって、それぞれのビームフォーマが別々の方位に死角を構成した場合に最大 $2(K - 1)$ 個の死角を構成できる。ただし、このスペクトル減算は位相を考慮することができず、各音源からの信号の混合が振幅ではなくパワーについて線形的に重ね合わさるという仮定を置いているため、実際には複数方位から到来する音波に対してこのような指向特性を実現できない。この誤差を緩和するため、信号を推定する際には引く側のパワーを増幅するオーバーサブトラクションと、推定パワーが負となった場合に 0 で置き換えるフロアリングが併用される。

相補型ビームフォーマの原典では2つのビームフォーマを交互に更新する繰り返し最適化演算が提案されているが、最小化する関数が \mathbf{g}_1 と \mathbf{g}_2 の積を含んだ形になっているため、勾配が干渉しあって収束が不安定になるという問題がある。

4 二次カーネルビームフォーマ

4.1 定式化

式 (5) は、以下の二次形式で書き換えられる。

$$y^2 = \mathbf{x}^H \mathbf{W} \mathbf{x} = \mathbf{x}^H [\mathbf{g}_1^H \mathbf{g}_2 + \mathbf{g}_2^H \mathbf{g}_1] \mathbf{x} \quad (7)$$

そのため、 y^2 の最適化は \mathbf{g}_1 と \mathbf{g}_2 ではなく \mathbf{G} を最適化と等化となる。このエルミート行列 \mathbf{W} は何らかの複数の K 次元縦ベクトル \mathbf{v}_i の重み付け和で表される。

$$\mathbf{W} = \sum_{i=1}^K \alpha_i \mathbf{v}_i \mathbf{v}_i^H \quad (8)$$

この形で表されるエルミート行列 \mathbf{W} は以下のような二次カーネル $K(\mathbf{v}_i, \mathbf{v}_j)$ を内積として導入されるヒルベルト空間上の元と等価である。

$$K(\mathbf{v}_i, \mathbf{v}_j) = \mathbf{v}_j^H \mathbf{v}_i \mathbf{v}_i^H \mathbf{v}_j \quad (9)$$

*Performance evaluation of kernel beamformer, by Shigeki Miyabe (the University of Tokyo), Biing-Hwang Juang (Georgia Institute of Technology), Hiroshi Saruwatari, and Kiyohiro Shikano (Nara Insitute of Science and Technology).

またその最適化は与えられた \mathbf{v}_k のもとで重み付け係数 α_k の最適化と単純化される。エルミート行列 \mathbf{W} を構成するためのベクトル \mathbf{v}_i にフレーム分析した観測信号の系列を用いることにより、 \mathbf{W} を適応することができる。式 (5) のパワースペクトル減算は、 \mathbf{g} に置いては固有値が正と負の両方の値を持ち、固有値が正の固有ベクトルによる線形ビームフォーマのパワーの和から固有値が正の固有ベクトルによる線形ビームフォーマのパワーの和が減算されると捉えることができる。

4.2 指向特性と次元数

このように最適化される \mathbf{W} は、ランクが 2 の場合において、特に $K = 2$ の場合において必ず相補型ビームフォーマと等価である。しかし、実際に得られる \mathbf{W} はフルランクである場合が多く、 $K > 2$ の場合における挙動の解析はまだなされていなかった。以下では \mathbf{W} のランクが 3 以上の場合の二次カーネルビームフォーマの挙動について議論する。

通常のビームフォーマで死角を向けることのできる方位の数は、観測信号の構成するユークリッド空間の次元数、つまりマイクロホンの素子数 K により上限が決められ、 $K-1$ 個である。それに対し、二次カーネルが導入するヒルベルト空間は、複素数の場合、 K 行 K 列エルミート行列の要素数 K^2 が次元数となり、例えば伝達関数が複素正規分布に基づくモデル化すれば、 K^2 個の K 次元伝達関数とその全空間を張る。しかし、実際には直線状アレーなどの伝達関数ベクトルの形状によって以下の表のように次元数は減少し、少ない素子数では相補型ビームフォーマと同程度の次元数に落ちる。しかしながら、次に説明する過減算によりこの限界をある程度緩和することができる。

Table 1 基底、音場モデルと次元数の関係

基底の形状	次元数	音場モデル
不規則	K^2	指向性素子アレー
実数	$K(K+1)/2$	瞬時混合
不等間隔位相差	$K^2 - K + 1$	不等間隔無指向性アレー
等間隔位相差	$K(K-1)/2$	等間隔直線状アレー

4.3 オーバーサブトラクション

3 節で述べた相補型ビームフォーマと同様に、二次カーネルビームフォーマは志向性を持った波の重ね合わせをパワー領域での線形重ね合わせとモデル化しているため、誤差を防ぐためにオーバーサブトラクションが必要である。そこで、正負両方の固有値を持つ対称行列 \mathbf{W} を引く側と引かれる側に分解する。

対称行列 \mathbf{W} を、目的方位に対する利得が一定の K 個のビームフォーマにより以下のように表されていると仮定する。

$$\mathbf{W} = \sum_{k=1}^K \sum_{l \neq k}^K \mathbf{g}_k \mathbf{g}_l^H \quad (10)$$

$$\mathbf{g}_k \mathbf{a}_0 = \frac{1}{\sqrt{K(K-1)}} \quad (11)$$

これにより、 \mathbf{W} の目的方位への利得 1 と正負両方の固有値が表現され、またそのフィルタリングは以下のようなスペクトル減算と捉えられる。

$$\mathbf{y}^2 = \mathbf{x}^H [\mathbf{h}^H \mathbf{h} - \mathbf{H}] \mathbf{x} \quad (12)$$

$$\mathbf{h} = \sum_{k=1}^K \mathbf{g}_k = \sqrt{\frac{K}{K-1}} \mathbf{W} \mathbf{a}_0 \quad (13)$$

$$\mathbf{H} = \left[\sum_{k=1}^K \mathbf{g}_k \right]^H \sum_{k=1}^K \mathbf{g}_k \quad (14)$$

最後に減算の重みによる利得変化の補正と \mathbf{G} の \mathbf{a}_0 に対する利得の 1 からの誤差を考慮に入れて、引かれる側のフィ

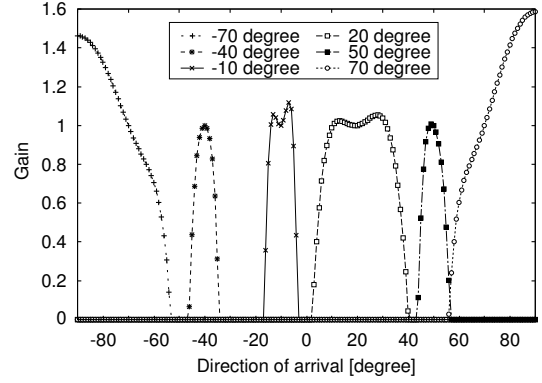


Fig. 2 方位 $-70^\circ, -40^\circ, -10^\circ, 20^\circ, 50^\circ, 70^\circ$ に対する二次カーネルビームフォーマ [2] の 250 Hz の指向特性。素子間隔 2.15 cm, 3 素子の直線型アレー, 減算係数 1.2。減算係数をさらに大きくするとより急峻な指向特性になる。

ルタ行列 $\mathbf{W}^{(p)}$ と引く側のフィルタ行列 $\mathbf{W}^{(r)}$ は以下のように求められる。

$$\mathbf{W}^{(p)} = \frac{\mathbf{G} \mathbf{r} \mathbf{r}^H \mathbf{G}^H}{\mathbf{r}^H \mathbf{G} \mathbf{r}} \quad (15)$$

$$\mathbf{W}^{(r)} = \frac{\mathbf{G} \mathbf{r} \mathbf{r}^H \mathbf{G}^H}{\mathbf{r}^H \mathbf{G} \mathbf{r}} - \mathbf{G} \quad (16)$$

Fig. 2 に 3 素子の直線型アレーによる 6 方位の二次カーネルビームフォーマ [2] の指向特性を示す。この条件ではヒルベルト空間の次元は 5 次元であるが、オーバーサブトラクションにより排他的な志向性制御が実現できている。

5 実験

残響時間 $T_{60} = 200$ ms の部屋で、無指向性マイクロホンの素子間隔 2.15 cm の直線状アレーを用いて評価を行った。音源方位はマイクロホンアレーから 1 m の距離に正面を 0° とした $-90^\circ - 90^\circ$ にほぼ等間隔に並んでおり、全ての音源は同じエネルギーで観測される。二次カーネルビームフォーマの設計はカーネル最小分散法を用い、比較手法として最小分散型ビームフォーマ (MVDR) を用いた。量手法ともに目的音源の存在しない 2 秒間で適応され、フィルタ長は 2048 点, 16 kHz, 提案手法の減算係数は 2 とした。

Tables 2, 3 にそれぞれ MVDR と提案手法の分離性能 noise reduction ratio (NRR) [dB] と ケプストラム距離 (CD) [dB] による歪評価を示す。MVDR は $K < N$ の場合に適応が不安定となるが、提案手法は音源の数が多くてもある程度の品質を保つことに成功している。

Table 2 MVDR の分離性能 (NRR) と歪 (CD, 括弧内)

	$K = 2$	$K = 3$	$K = 4$
$N = 2$	12.02 (3.84)	34.29 (6.77)	33.25 (7.26)
$N = 3$	1.35 (4.69)	13.25 (3.47)	25.48 (6.47)
$N = 4$	0.80 (5.65)	6.91 (4.28)	16.59 (5.90)
$N = 5$	1.45 (5.69)	2.67 (5.29)	9.07 (6.19)
$N = 6$	-0.93 (5.98)	2.76 (5.45)	1.14 (7.62)

Table 3 提案法の分離性能 (NRR) と歪 (CD, 括弧内)

	$K = 2$	$K = 3$	$K = 4$
$N = 2$	15.46 (5.23)	22.91 (6.43)	21.52 (6.79)
$N = 3$	8.64 (4.63)	16.48 (4.87)	18.89 (6.10)
$N = 4$	5.13 (5.92)	13.05 (5.04)	16.50 (6.74)
$N = 5$	4.65 (5.90)	7.02 (5.49)	9.45 (6.42)
$N = 6$	4.38 (6.15)	6.88 (5.67)	10.27 (8.00)

参考文献

- [1] Saruwatari *et al.*, IEICE Trans. Fundam., E83-A(5), 866–876, 2000.
- [2] 宮部他, 音講論 (春), 651–654, 2009.
- [3] Miyabe *et al.*, Proc. ICASSP, 1641–1644, 2009.
- [4] Miyabe *et al.*, Proc. IWAENC, #9084, 2008.