3—4—10

間隔の異なる複数のマイクペアによるブラインド音源分離 *

○澤田 宏, 荒木 章子, 向井 良, 牧野 昭二
(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)

1 はじめに

ブラインド音源分離 (BSS: Blind Source Separation) とは, 混合された複数の音を, 元の音や混合の過程を知ることなしに分離する技術である. 近年, 独立成分分析 (ICA: Independent Component Analysis) [1] に基づく手法が盛んに研究されている.

本稿では, マイク間隔が分離性能に与える影響に着目し, 低周波数に対しては間隔の広いマイクペアを用い, 高周波数に対しては間隔の狭いマイクペアを用いるブラインド音源分離手法を提案する. 議論の簡素化のため, 2 音源の場合を考える. ICA によって作られる分離フィルタの指向特性は, 多くの場合, 消したい音の方向に適応的に死角を向けるものとなる [2, 3]. そこでは, 2 つのマイクにおける目的音と妨害音の位相差の違いを利用している. しかし, もしマイク間隔が扱う音の最大周波数の半波長より長ければ, 空間的 aliasing により複数の方向に死角を作ることになり, 分離性能が劣化することがある. 一方, 波長の長い低音に着目すると, 十分な位相差を確保するためには長いマイク間隔が望まれる. このように, 扱う音の周波数に従って適切なマイク間隔が違ってくる.

2 システム構成

まず, 一般の 2 音源 2 マイクの BSS を定式化する. 音源 $s_p(t)$ が室内で残響も含めて混合され, マイク $x_q(t) = \sum_{p=1}^{2} \boldsymbol{h}_{qp} * s_p(t)$ で観測されたとする. ここで, $\boldsymbol{h}_{qp}$ は音源 p からマイク q へのインパルス応答, $*$ は畳み込みを示す. BSS の目的は, $s_p(t)$ や $\boldsymbol{h}_{qp}$ を知ることなしに, 分離のための FIR フィルタ係数 $\boldsymbol{w}_{rq}$ と分離信号 $y_r(t) = \sum_{q=1}^{2} \boldsymbol{w}_{rq} * x_q(t)$ を求めることにある. ICA を用いる手法では, 瞬時混合の問題に帰着させるために, 周波数領域で処理を行うことが多い. その場合は, マイク q での観測信号を短時間フーリエ変換 $X_q(f,m)$ し, 分離のための周波数応答 $\mathbf{W}(f)$ と分離信号 $\mathbf{Y}(f,m) = \mathbf{W}(f)\mathbf{X}(f,m)$ を求める.

本稿で提案する BSS システムは, 複数のマイクペア (x_1^n, x_2^n), $n = 1, 2, \ldots$ とそれに対応した分離系 n で構成される. 各分離系は, 混合された音の分離と周波数帯域選択の役目を果たす. 各分離系の出力 y_r^n は, 後段で統合 $y_r = \sum_n y_r^n$ され, システム全体が出力する分離信号となる. 図 1 に 2 つのマイクペアによるシステム例を示す. 第 1 のマイクペア (x_1^1, x_2^1) と分離系 1 は低音の分離に用いられ, 第 2 のマイクペア (x_1^2, x_2^2) と分離系 2 は高音の分離に用いられる. 本例のように, マイク x_2^1 と x_2^2 を共有することで, 3 つのマイクで 2 つのペアを構成することも可能である.

*Blind source separation using pairs of microphones with different distances, by Hiroshi Sawada, Shoko Araki, Ryo Mukai, Shoji Makino (NTT Communication Science Laboratories, NTT Corporation)

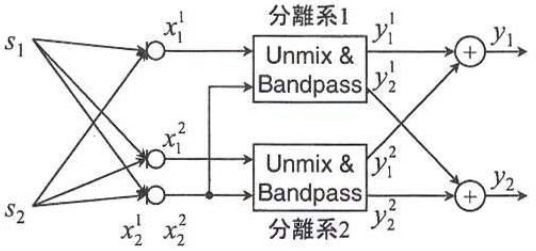

図 1: システム構成

本システムの処理としては, まず, 各分離系がどの周波数帯域を担当するかを決める. これは, 3 章で述べる基準により, マイクペアの間隔と推定された音源の方向から決定される. その後, 各分離系では, 担当する周波数帯域に渡って, ICA を用いて分離のための周波数応答 $\mathbf{W}(f)$ を求める. 我々が用いる ICA の詳細は文献 [4] を参照されたい. 次に, 担当部分以外の周波数を阻止するために, その周波数応答を 0 とする. 最後に, 全体の周波数応答に対して逆 DFT を適用することで, FIR フィルタの係数 $\boldsymbol{w}_{rq}$ を求める. ただしこのままでは, 遮断周波数付近で急峻な特性となり現実的な長さの FIR フィルタでは誤差が大きくなるため, 得られた FIR フィルタに Hanning 窓を掛ける. これにより, ICA で求めた周波数応答 $\mathbf{W}(f)$ が近傍の周波数により平均化されることになるが, 元々短時間フーリエ変換 $X_q(f,m)$ の際にも Hanning 窓を掛けて平均化していたので, 大きな問題にはならないと考える.

3 マイク間隔に対する適切な周波数の範囲

本章では, MUSIC 法など何らかの手法で音源方向を推定できた際に, 個々のマイク間隔を持つ分離系がどの周波数までを担当するかについて議論する. 音源 s_p の到来方向を $0° \leq \theta_p \leq 180°$, マイク q の位置を d_q とする. 混合系のインパルス応答を直接音のみで近似し, さらに平面波を仮定すると, 混合系の周波数応答は $H_{qp}(f) = e^{j2\pi f c^{-1} d_q \cos\theta_p}$ と表現できる (c は音速). 座標系を $d_1 = 0$, $d_2 = d$ と設定して行列形式で書くと

$$\mathbf{H}(f) = \begin{bmatrix} 1 & 1 \\ e^{j2\pi f c^{-1} d \cos\theta_1} & e^{j2\pi f c^{-1} d \cos\theta_2} \end{bmatrix}$$

となる. これを分離系の周波数応答 $\mathbf{W}(f)$ に掛けて $\mathbf{W}(f)\mathbf{H}(f)$ を求めると, 音源 $[s_1, s_2]^T$ から分離信号 $[y_1, y_2]^T$ の周波数応答となる:

$$\begin{bmatrix} W_{11}+W_{12}e^{j2\pi f c^{-1} d \cos\theta_1} & W_{11}+W_{12}e^{j2\pi f c^{-1} d \cos\theta_2} \\ W_{21}+W_{22}e^{j2\pi f c^{-1} d \cos\theta_1} & W_{21}+W_{22}e^{j2\pi f c^{-1} d \cos\theta_2} \end{bmatrix}$$

本行列の各要素は, 音源方向 θ_p に従って変化するため, 指向特性と呼ばれる. 図 2 に r 行に対応する指向特性の

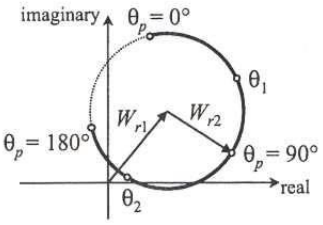

図 2: 指向特性の複素平面上での解釈

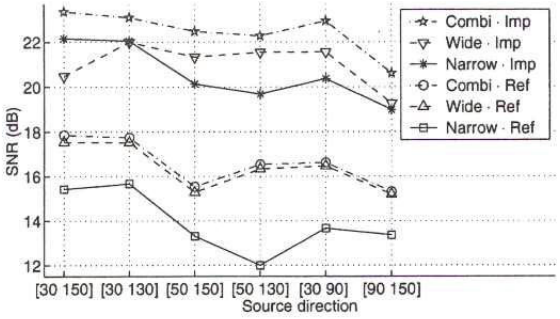

図 3: 分離性能

複素平面上での解釈を示す．方向 θ_p の変化により $\cos\theta_p$ が 1 から -1 まで動くことで，中心 W_{r1}，半径 W_{r2} の円周上を特性が動く．これより，空間的 aliasing を起さない条件は $2\pi f c^{-1} d < \pi$ であり，周波数 $f < c/(2d)$ の範囲を扱えることがわかる．しかし $\mathbf{W}(f)$ が性能の良い分離フィルタとなるためには，より強い条件が望まれる．すなわち，θ_1 でゲイン特性が十分に大きく θ_2 でゲイン特性が十分に小さいこと，さらに θ_1 と θ_2 の間でゲイン特性が極大にならないことである．そのためには $2\pi f c^{-1} d \cos\theta_1 - 2\pi f c^{-1} d \cos\theta_2 \leq \pi$ が条件となり，扱える周波数は $f \leq c/[2d(\cos\theta_1 - \cos\theta_2)]$ となる．

4 実験結果および考察

提案した手法の有効性を示すため，残響下で混合された音声を分離する実験を行った．混合音声は，ASJ 研究用音声コーパスから選んだ 8 秒の音声データに，RWCP 実環境音声・音響データベースから選んだ残響時間 300ms のインパルス応答を畳み込んで作成した．図 3 に分離性能を示す．横軸は 2 音源の方向を示し，縦軸は SNR を示している．"Wide"，"Narrow" ではマイク幅がそれぞれ，141.5mm，28.3mm であり，"Combi" では提案手法により双方を利用した．"Ref" は音声信号自身で計測した SNR を示す．音声信号には低音域にパワーが集中しているため，全帯域の効果も見るためにインパルスでも SNR を計測した ("Imp")．

角度が [50 150] の状況を詳しく見る．周波数毎に "Wide" の SNR から "Narrow" の SNR を引いたものを図 4 に示す．"Combi" の場合には，推定された 2 音源の角度が 53° と 138° であったため，893Hz (図中の縦線) より低い音に対して広いマイクを用い，それより高い音に対して狭いマイクを用いた．この縦線よりも左側では正の値が多く，逆に右側では負の値が多いことから，"Combi" の効果が理解できる．

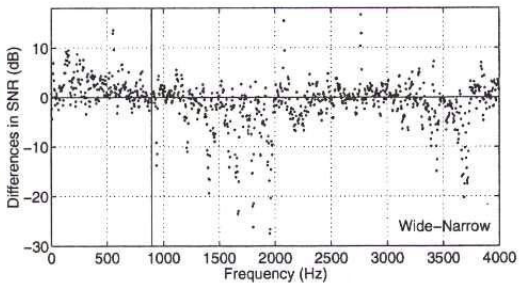

図 4: 周波数毎の分離性能の違い

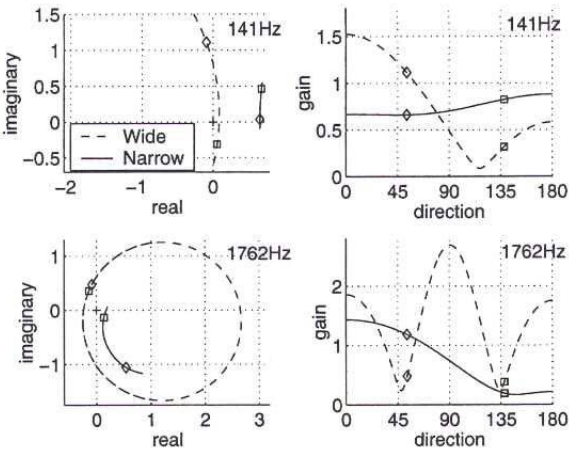

図 5: 指向特性

次に指向特性を見る．図 5 の上半分は $f = 141$Hz，下半分は $f = 1762$Hz のものであり，左側が複素平面上で示したもの，右側がゲイン特性のみを示したものである．◇と□は推定された 2 音源の方向を示す．低周波数で "Narrow" を用いると，妨害音方向に死角を形成しにくいという状況が伺える．一方，高周波数で "Wide" を用いると空間的 aliasing が起こり，2 音源の方向の指向特性を大きく異なるものにすることができない．

5 おわりに

低周波数には間隔の広いマイク，高周波数には間隔の狭いマイクが有効であることが，BSS においても成立することが分かった．提案手法では，複数のマイクペアを用いてその有効性を活かし，単一のマイクペアよりも高い分離性能を得た．

参考文献

[1] T. W. Lee, *Independent component analysis - Theory and applications*, Kluwer academic publishers, 1998.

[2] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in *Proc. ICASSP2000*, June 2000, pp. 3140–3143.

[3] S. Araki, S. Makino, R. Mukai, and H. Saruwatari, "Equivalence between frequency domain blind source separation and frequency domain adaptive null beamformers," in *Proc. Eurospeech2001*, Sept. 2001, pp. 2591–2594.

[4] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Polar coordinate based nonlinear function for frequency-domain blind source separation," in *Proc. ICASSP 2002*, accepted.