

マイクロホンのペア分散配置によるブラインド音源分離と ポストフィルタによる性能向上の検討

小野 拓磨[†] 伊藤 信貴[†] 宮部 滋樹[†] 小野 順貴[†] 嵯峨山茂樹[†]

[†] 東京大学大学院情報理工学系研究科システム情報学専攻

〒 113-8656 東京都文京区本郷 7-3-1

E-mail: †{tono,ito,miyabe,onono,sagayama}@hil.t.u-tokyo.ac.jp

あらまし 本研究では、分散型マイクロホンアレーを用いた高品質な音源分離を目的として、マイクロホンの配置法と分離手法について論じる。分散型マイクロホンアレーとは、分散配置された複数の録音機器をマイクロホンアレーとして用いる枠組みであり、空間的に広い範囲に適用可能であること、ハードウェア構成が簡便であること、素子数の増加が容易なこと、といった利点を持つ。一方、録音信号が時間的に非同期であり、録音機器の位置情報が通常未知であるといった難しさも存在する。本稿では、独立成分分析に基づくブラインド音源分離を前提に、1)チャンネル間の相互相関に基づく時間同期化、2)マイクロホンペアの分散配置による効率的なパーミュテーション解法、3)拡散性雑音の無相関性に基づくポストフィルタ設計法、について議論し、合わせて実環境実験による検証結果を報告する。キーワード ブラインド音源分離, 独立成分分析, 分散型マイクロホンアレー, ポストフィルタ

Blind source separation with distributed microphone pairs and investigation of performance gain using post-filter

Takuma ONO[†], Nobutaka ITO[†], Shigeki MIYABE[†], Nobutaka ONO[†], and Shigeki SAGAYAMA[†]

[†] Graduate School of Information Science and Technology, The University of Tokyo

7-3-1 Hongo Bunkyo-ku Tokyo 113-8656, Japan

E-mail: †{tono,ito,miyabe,onono,sagayama}@hil.t.u-tokyo.ac.jp

Abstract In this paper, we discuss microphone location and the separation method using frequency-domain independent component analysis aiming to high-performance source separation with distributed microphone arrays. In the distributed microphone array framework, which has the advantages of its abilities to separate wide-spread sources, simplify hardware configurations and make it easy to increase devices, we discuss 1)the time alignment of each channel using cross correlation, 2)permutation correction with distributed microphone pairs and 3)designing of post-filter based on the uncorrelatedness of diffuse noises. The experimental results in real reverberant environment are also shown.

Key words blind source separation, independent component analysis, distributed microphone array, post-filter

1. はじめに

様々な雑音が重畳する実環境において目的とする音源信号のみを取得するための枠組みとして、マイクロホンアレーを用いた音源分離技術は大変有用な枠組みであり、これまでも活発な研究が行われてきた。従来のマイクロホンアレー信号処理においては、数個の近接して配置されたマイクロホンによりアレーが構成され、かつマルチチャンネルA/D変換器により時間同期した録音信号が取得されることが前提であったが、近年では、複数の録音機器を分散配置させ、録音信号を後で組織的に統合

することによりアレー信号処理を行う、分散型マイクロホンアレーという新しい枠組みが試みられてきている [1], [2]。

分散型マイクロホンアレーは、

- 広範囲にマイクロホンを配置するため、音源-マイクロホンの距離が小さくなりやすく、高SN比の入力を得ることが可能
 - 複数録音機器を用いることにより多素子数化が容易
 - 固定配置を必要とせずハードウェア構成が簡便
- といった魅力的な特長を有し、実環境における音源分離の適用範囲を格段に広げる可能性を有しているが、一方では、録音信

号同士が非同期、マイクロホン位置が未知、各録音機器間の感度のばらつき、などの特有の問題を解決する必要がある。

我々も分散型マイクロホンアレーの特長に着目し、これまでに、音源位置・マイクロホン位置・非同期録音信号の時空間的ブラインドアライメント[3]やブラインド音源分離(Blind Source Separation; BSS)への応用を進めてきた。特にブラインド音源分離への応用においては、周波数領域における独立成分分析(Independent Component Analysis)[4]におけるパーミュテーション問題[5],[6]に着目し、マイクロホンを近接したペアとして分散配置させることにより、マイクロホンの位置情報が未知で、かつペア間の時間同期がとれていなくても、効率的にパーミュテーション問題を解決する手法を提案してきた[7],[8]。

本研究では、分散型マイクロホンアレーの分散配置と多素子化に基づき、音源分離性能を更に向上させる方法について検討する。一般には、背景雑音や後部残響のような成分は、ICAによる指向性制御だけでは抑制が難しい。そこで我々は、分散配置を積極的に利用し、サブアレーが分散配置されている場合には、これらの成分はサブアレー間では無相関とみなせるのではないかと、この着眼に基づき、サブアレー毎のICAとマルチチャネルWienerフィルタによる後段処理に基づく分離手法を検討する。特に本研究ではその典型的な場合として、サブアレーとしてはマイクロホンペアの分散配置、対象とする音源数は2の場合に絞り、1)非同期録音信号の同期化、2)マイクロホンペア配置を利用した周波数領域ICAのパーミュテーション解法について述べた後、3)分散配置を利用したマルチチャネルWienerフィルタの設計法、について論じ、実環境実験により、このような後段処理の有効性を検討する。

2. チャネル間の相互相関に基づく時間同期化

複数録音機器間が非同期の場合、機器間での録音開始時刻のずれを修正するようなキャリブレーションが必要である。フレーム長に対し録音開始時刻のずれが十分に小さくできれば録音開始時刻のずれを吸収した形で後段の推定が可能のため、正確な機器間での録音開始時刻のずれを推定する必要はない。ここでは、大まかなアライメントとして録音開始時刻のずれの近似値 T_{i1} を基準のチャンネルの観測信号 $x_1(t)$ との時間領域での相互相関関数 $R_{i1}(t_m)$ を用いて

$$T_{i1} = \underset{t_m}{\operatorname{argmax}} R_{i1}(t_m) \quad (1)$$

として求め、観測信号 $x_i(t)$ を T_{i1} だけ時間シフトした。

尚、ワイアレスシステム等で互いの機器間で正確な時間同期がとれている場合この操作は必要ない。また本研究では、サンプリング周波数のずれによる影響については無視できるほど小さいとして扱わない。以下では、全てのチャンネル間で観測の時点で正確な時間同期がとれている、または相互相関関数に基づくキャリブレーションがされたとする。

3. マイクロホンペア分散配置によるBSS

本手法は、図1のように複数チャンネル間の相関を計算する

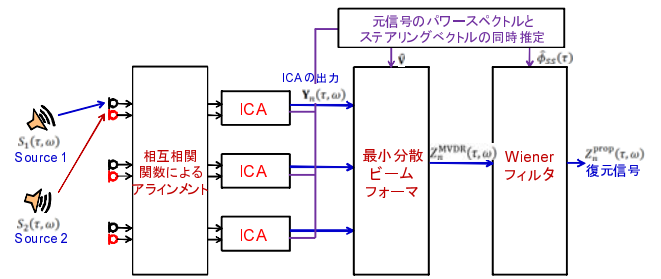


図1 提案法のフローチャート

為にマイクロホンペア毎でのICAの出力信号に対しモデル化を行い、ペア間での拡散性雑音の無相関性に基づく元信号のパワースペクトルとステアリングベクトル(音源から各チャンネルまでの周波数特性を並べたベクトル)の同時推定を行いマルチチャネルWienerフィルタを設計する。

マルチチャネルWienerフィルタの設計の為に、消し残りの残響成分から成る拡散性雑音を含んだペア毎でのICAの出力信号のみから、元信号のパワースペクトルとステアリングベクトルの両方を推定する必要がある。距離の離れたマイクロホンペア間の拡散性雑音の無相関性に着目すれば、ある音源についてのペア毎でのICAの出力信号のクロススペクトルの非対角成分は雑音の影響を無視できるので、クロススペクトルの非対角成分のみに着目した評価関数を設計することで、元信号のパワースペクトルとステアリングベクトルを同時推定可能である[9]。これにより得られたステアリングベクトルの推定値を用いて最小分散ビームフォーマを実装し、その出力に対し、元信号のパワースペクトルの推定値を用いたWienerフィルタをかけることで、より高品質な音源分離を実現する。

以下音源数を2とするが、音源数が3以上の場合でも音源数以上のマイクロホンをユニットとし、ユニットを分散させることで以下の枠組みは容易に拡張できる。まず、ペア毎でのICAの出力信号を求める。

3.1 マイクロホンペア毎での分離行列の更新

後段のマルチチャネルWienerフィルタ設計に必要な複数出力を得る為の、 L 組のマイクロホンペアごとのICAによる分離行列の更新と復元信号の推定を行う。観測信号のベクトル列 $\mathbf{x}_l(t) = [x_{l1}(t), x_{l2}(t)]^T$ ($l = 1, \dots, L$) は

$$\mathbf{x}_l(t) = \sum_{u=0}^{\infty} \begin{bmatrix} a_{11}^l(u) & a_{12}^l(u) \\ a_{21}^l(u) & a_{22}^l(u) \end{bmatrix} \mathbf{s}(t-u) \quad (l = 1, \dots, L) \quad (2)$$

と表せる。ここで、 $\mathbf{s}(t) = [s_1(t), s_2(t)]^T$ は元信号のベクトル列、 $a_{mn}^l(t)$ は n 番目 ($n = 1, 2$) の音源から l 組目のペア内の m 番目 ($m = 1, 2$) のマイクロホンへのインパルス応答である。音声の短時間での定常性を仮定し、短時間フーリエ変換すれば

$$\mathbf{X}_l(\tau, \omega) \simeq \mathbf{A}_l(\omega) \mathbf{S}(\tau, \omega) \quad (l = 1, \dots, L) \quad (3)$$

と表せる。ここで τ, ω はフレーム、周波数のインデックス、 $\mathbf{A}_l(\omega)$ は l 組目のペアでの混合行列である。周波数領域ICAにより、混合過程が未知の条件で観測信号から元信号を復元した信号

$$\begin{aligned} \mathbf{Y}_l(\tau, \omega) &= [Y_{l1}(\tau), Y_{l2}(\tau)]^T \\ &= W_l(\omega) \mathbf{X}_l(\tau, \omega) \quad (l = 1, \dots, L) \end{aligned} \quad (4)$$

を推定する。ここで $W_l(\omega)$ は l 組目のペアでの分離行列である。分離行列 $W_l(\omega)$ の更新には補助関数法に基づく手法 [10] を用いた。 $W_l(\omega) = W = [\mathbf{w}_1, \mathbf{w}_2]^T$, $\mathbf{X}_l(\tau, \omega) = \mathbf{X}(\tau)$ と書き直せばこの手法では、自然勾配法による手法 [11] 同様に Infomax タイプの評価関数

$$J_{\text{ICA}} = \sum_{n=1}^2 E [G(\mathbf{w}_n^H \mathbf{X}(\tau))] - \log |\det W| \quad (5)$$

を最小にすることを考える。補助関数法を用いた更新則は

$$V_n = E \left[\frac{G'(r_n(\tau))}{r_n(\tau)} \mathbf{X}(\tau) \mathbf{X}^H(\tau) \right] \quad (6)$$

$$P_1 = V_1 \mathbf{w}_2 \quad (7)$$

$$P_2 = V_2 \mathbf{w}_1 \quad (8)$$

$$\mathbf{w}_n \leftarrow \mathbf{w}_n - P_n (P_n^H P_n)^{-1} P_n^H \mathbf{w}_n \quad (9)$$

$$\mathbf{w}_n \leftarrow \mathbf{w}_n / \sqrt{\mathbf{w}_n^H V_n \mathbf{w}_n} \quad (10)$$

となる。ただし $r_n(\tau) = |\mathbf{w}_n^H \mathbf{X}(\tau)|$ であり、 $G(r)$ はコントラスト関数と呼ばれる非線形関数で、文献 [4] 等で用いられる $G(r) = \log \cosh r$ を用いた。

このように、各マイクロホンペア、各周波数ごとに独立に分離行列が更新されるが、独立成分分析は観測信号のみから元信号と分離行列の両方を推定するので、解の中にスケールとパーミュテーションの任意性が残る。対角行列 $B_l(\omega)$ と置換行列 $P_l(\omega)$ が推定できれば

$$\hat{S}_l(\tau, \omega) = B_l(\omega) P_l(\omega) W_l(\omega) \mathbf{X}_l(\tau, \omega) \quad (l = 1, \dots, L) \quad (11)$$

により元信号の推定値を得る。

3.2 パーミュテーション問題の解決

マイクロホンペア毎での ICA により、ペア間と周波数間でそれぞれ独立に分離行列を更新するので、双方のパーミュテーション解決が必要である。この時、各マイクロホンペアでそれぞれ同一の音源を分離するので、各ペアで推定した分離行列は相互にパーミュテーション解決の手掛かりになる。従って、パーミュテーション解決をペア毎で独立に行うより、全てのペアで同時に行う方が妥当である。全てのペアで同時にパーミュテーション解決する為に、先に同一の周波数内でのペア間のパーミュテーションを揃える必要がある。従って、先にペア間では振幅包絡の類似性に基づくクラスタリングを用い、周波数間ではペア内到来時間差を並べたベクトルのクラスタリングを用いることで以下では双方のパーミュテーション問題を順番に解決する。

3.2.1 ペア間パーミュテーション問題の解決

同一の周波数において、各マイクロホンペアでの出力信号はスケールの任意性を除けば理想的には同一となるため、各ペア間での出力信号の振幅包絡の類似性を調べることで、ペア間でのパーミュテーションを予め揃えることができる。 $\mathbf{Y}_l(\tau, \omega) = [Y_{l1}(\tau, \omega), Y_{l2}(\tau, \omega)]^T$ について、同一周波数での振幅包絡の類似性の尺度として相関係数 $r(Y_{ln}(\omega), Y_{l'n'}(\omega))$ を

$$\begin{aligned} & \text{Cov}(Y_{ln}(\omega), Y_{l'n'}(\omega)) \\ &= E [Y_{ln}(\tau, \omega) Y_{l'n'}(\tau, \omega)] - E [Y_{ln}(\tau, \omega)] E [Y_{l'n'}(\tau, \omega)] \quad (12) \\ & r(Y_{ln}(\omega), Y_{l'n'}(\omega)) \\ &= \frac{\text{Cov}(Y_{ln}(\omega), Y_{l'n'}(\omega))}{\sqrt{\text{Cov}(Y_{ln}(\omega), Y_{ln}(\omega)) \text{Cov}(Y_{l'n'}(\omega), Y_{l'n'}(\omega))}} \quad (13) \end{aligned}$$

と定義できる [12]。ただし $E[\cdot]$ は τ について期待値をとる操作である。同一周波数において、ペア間で出力音源を相関係数の和が最も大きくなるように並び替えることで、ペア間パーミュテーション問題を解決できる。以下同一周波数ではペア間でパーミュテーションが揃っているとする。

3.2.2 周波数間パーミュテーション問題の解決

距離の近いマイクロホンをペアとしたことで、ペア間での到来時間差を広帯域において空間的エイリアシングの影響を受けることなく推定することが可能で、これら理想的には周波数に依らない時間差をクラスタリングすることで、周波数間パーミュテーション問題を効率的に解くことができる。スケールとパーミュテーションの任意性を含んだ混合行列の推定値

$$\hat{A}_l(\omega) = W_l^{-1}(\omega) \quad (14)$$

を用いれば、ペア内での各音源の到来時間差 $d_n^l(\omega)$ は

$$d_n^l(\omega) = \{\arg(\hat{A}_{1n}^l(\omega)) - \arg(\hat{A}_{2n}^l(\omega))\} / \omega \quad (n = 1, 2) \quad (15)$$

と求まる [6]。ここで $\hat{A}_{mn}^l(\omega)$ は $\hat{A}_l(\omega)$ の (m, n) 要素である。直接波の到来時間差は周波数によらず、ほとんどの帯域で直接波がドミナントになるため、 ω 番目の周波数での時間差 $d_n^l(\omega)$ をマイクロホンペア数だけ並べた L 次元ベクトル $\mathbf{d}_{\omega n} = [d_n^1(\omega), \dots, d_n^L(\omega)]^T$ の推定値は、理想的には周波数に依らない。この時間差ベクトル $\mathbf{d}_{\omega n}$ をクラスタリングして直接波の到来時間差を推定することによりパーミュテーションを推定することができる。ここで、到来時間差の推定はパワーが大きいものほど信頼性が高いため、周波数帯域・マイクロホンペア毎にパワーに応じた信頼度を持たせる。ここでは各マイクロホンペアの各周波数での振幅のフレーム平均を、Minimal Distortion Principle [13] によりスケールの任意性を解消した出力信号

$$W_l(\omega) \leftarrow \text{diag}(W_l^{-1}(\omega)) W_l(\omega) \quad (16)$$

$$\mathbf{Y}_l(\tau, \omega) = W_l(\omega) \mathbf{X}_l(\tau, \omega) \quad (n = 1, 2) \quad (17)$$

を用いて求め、重みを

$$\gamma_{\omega n} = [E[|Y_{1n}(\tau, \omega)|], \dots, E[|Y_{Ln}(\tau, \omega)|]]^T \quad (18)$$

と定義した。この重みを信頼度として、重み付き距離を最小化する k -means アルゴリズムによるクラスタリングを行った。以下のように $\{1, 2\}$ を並べ替えた順列 $p_\omega(n) (n = 1, 2)$ とクラスター中心 \mathbf{m}_n を更新する。

$$\begin{aligned} p_\omega(n) &\leftarrow \underset{p_\omega(n)}{\text{argmin}} \sum_{\omega} \sum_{n=1}^N \sum_{l=1}^L \gamma_{\omega p_\omega(n)}^l |d_{\omega p_\omega(n)}^l - m_n^l|^2 \quad (19) \\ m_n^l &= \frac{\sum_{\omega} \gamma_{\omega p_\omega(n)}^l d_{\omega p_\omega(n)}^l}{\sum_{\omega} \gamma_{\omega p_\omega(n)}^l} \quad (l = 1, \dots, L), \quad (20) \end{aligned}$$

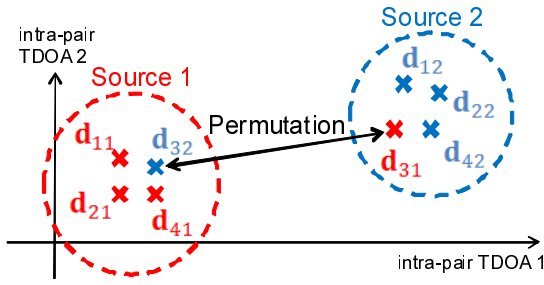


図2 ペア内到来時間差のクラスタリング

ここで $\gamma_{\omega p\omega}^l, d_{\omega p\omega}^l$ の添え字 l は l 列目のベクトルの要素、 m_n^l は n 番目の音源についての l 番目のマイクロホンペアのセントロイドである。これはペア内到来時間差の推定を示す。

3.3 スケールの任意性の解決

分散型マイクロホンアレーを観測に用いた場合、各マイクロホンでのそれぞれ音源の間のパワーの比は配置に依存するので、スケールを各ペア内でそれぞれ独立に合わせると、ICA の音源分離の性能にもばらつきが生じ得る。一般に、復元したい音源の入力 SN 比が大きければ音源分離を行った後の出力信号も SN 比も大きくなる。そこで本研究では、各マイクロホンペアごとに分離行列を更新し、パーミュテーション解決した後に、復元したい音源についての入力信号の SN 比を推定し、SN 比の最も大きなマイクロホンペアの片チャンネルの観測信号に全てのマイクロホンペアの出力信号を合わせる手法を行う。各音源ごとに、最も入力 SN 比の大きなペアにスケールを合わせることで推定信号の SN 比も最大にすることが期待される。また、この手法により後に用いる元信号のパワースペクトル推定値とステアリングベクトルの同時反復推定アルゴリズムの初期値として妥当な値の設計も可能となる。

まず、それぞれの音源について、入力信号の SN 比が最大となっているマイクロホンペアを推定する手法を考える。前節の式 (16) のようにペア内で Minimal Distortion Principle を用いたスケールの任意性解決を行うことで、復元したい音源のみの各マイクロホンでの観測信号の推定値 $\mathbf{Y}_l(\tau, \omega) = [Y_{l1}(\tau, \omega), Y_{l2}(\tau, \omega)]^H$ が得られる。これを用いて入力 SN 比が最大であるペアは音源 1、音源 2 についてそれぞれ

$$\begin{cases} \tilde{l}_1 = \operatorname{argmax}_l \sum_{\omega} \left(\frac{E[|Y_{l1}(\tau, \omega)|^2]}{E[|Y_{l2}(\tau, \omega)|^2]} \right) \\ \tilde{l}_2 = \operatorname{argmax}_l \sum_{\omega} \left(\frac{E[|Y_{l2}(\tau, \omega)|^2]}{E[|Y_{l1}(\tau, \omega)|^2]} \right) \end{cases} \quad (21)$$

として推定できる。マイクロホンペア \tilde{l}_n にスケールを合わせるには、入力 SN 比が最大であるペアでの片チャンネルでの観測信号 $X_{\tilde{l}_n 1}(\tau, \omega)$ とスケールを合わせる為の係数を並べたベクトル $\mathbf{b}_{l_n}(\omega)$ を用いて

$$\min E \left[\left| X_{\tilde{l}_n 1}(\tau, \omega) - \mathbf{b}_{l_n}^H(\omega) \mathbf{Y}_l(\tau, \omega) \right|^2 \right] \quad (22)$$

を満たす $\mathbf{b}_{l_n}(\omega)$ を用いて

$$\mathbf{Y}_l(\tau, \omega) \leftarrow \operatorname{diag}(\mathbf{b}_{l_n}^H(\omega)) \mathbf{Y}_l(\tau, \omega) \quad (23)$$

とすればよい。ここで式 (22) を解けば

$$\mathbf{b}_{l_n}(\omega) = \left(E \left[\mathbf{Y}_l(\tau, \omega) \mathbf{Y}_l^H(\tau, \omega) \right] \right)^{-1} E \left[X_{\tilde{l}_n 1}^*(\tau, \omega) \mathbf{Y}_l(\tau, \omega) \right] \quad (24)$$

となるので、このスケールを合わせる為の係数 $\mathbf{b}_{l_n}(\omega)$ を式 (23) に用いればよい。

4. 指向性の制御とマルチチャネル Wiener フィルタ

ペア毎での ICA により得られた複数の出力信号を用い、マルチチャネル Wiener フィルタを設計することで、拡散性雑音の抑制を試みる。以下ではペア毎での ICA により得られた出力信号を統合するフィルタを音源ごとに設計するので、ICA の出力信号を並び替えた形で改めて定式化を行う。音源ごとに ICA の出力信号を並べたベクトル $\tilde{\mathbf{Y}}_n(\tau, \omega) = [Y_{1n}(\tau, \omega), \dots, Y_{Ln}(\tau, \omega)]^T$ は ICA により抑制不十分な残響成分から成る拡散性雑音を含んだ形で

$$\tilde{\mathbf{Y}}_n(\tau, \omega) = S_n(\tau, \omega) \mathbf{v}_n(\omega) + \mathbf{N}_n(\tau, \omega) \quad (n = 1, 2) \quad (25)$$

のように定式化できる。ここで、 n は音源のインデックス、 $\mathbf{v}_n(\omega)$ はステアリングベクトル、 $\mathbf{N}_n(\tau, \omega)$ は n 番目の音源以外の拡散性雑音成分である。元信号と雑音は無相関が仮定できるので、短時間でのエルゴード性を仮定すれば式 (25) は

$$\Phi_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}^n(\tau, \omega) = \phi_{SS}^n(\tau, \omega) \mathbf{v}_n(\omega) \mathbf{v}_n^H(\omega) + \Phi_{NN}^n(\tau, \omega) \quad (26)$$

のように表せる。ただし

$$\Phi_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}^n(\tau, \omega) = E \left[\tilde{\mathbf{Y}}_n(\tau, \omega) \tilde{\mathbf{Y}}_n^H(\tau, \omega) \right] \quad (27)$$

$$\phi_{SS}^n(\tau, \omega) = E \left[|S_n(\tau, \omega)|^2 \right] \quad (28)$$

$$\Phi_{NN}^n(\tau, \omega) = E \left[\mathbf{N}_n(\tau, \omega) \mathbf{N}_n^H(\tau, \omega) \right] \quad (29)$$

である。各マイクロホンペア間の距離は十分に大きくとることができるので、ペア間の拡散性雑音成分は無相関であるとみなすことができる。この性質を用い、ポストフィルタによる拡散性雑音成分の抑制を目指した。

4.1 元信号のパワースペクトルとステアリングベクトルの同時推定

ポストフィルタとして最適なマルチチャネル Wiener フィルタの設計には、元信号のパワースペクトルとステアリングベクトル推定が必要である。拡散性雑音を含むペア毎での ICA の出力信号から、これらを推定する手掛かりは ICA の出力信号のクロススペクトルの非対角成分である。拡散性雑音の無相関性が仮定できるので、 $\Phi_{NN}^n(\tau, \omega)$ の非対角成分はゼロに近い値となるので、式 (26) の非対角成分を考えれば

$$\phi_{\tilde{\mathbf{Y}}_i \tilde{\mathbf{Y}}_j}^n(\tau, \omega) \simeq \phi_{SS}^n(\tau, \omega) v_{ni}(\omega) v_{nj}^*(\omega) \quad (i \neq j) \quad (30)$$

が成り立つ。ただし $\phi_{\tilde{\mathbf{Y}}_i \tilde{\mathbf{Y}}_j}^n(\tau, \omega)$ は $\Phi_{\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}}^n(\tau, \omega)$ の $\{ij\}$ 成分、 $v_{ni}(\omega)$ は $\mathbf{v}_n(\omega)$ の i 行目の成分である。これを用いて、式 (30) の両辺の差の二乗和を目的関数とし、

$$J_n(\omega) = \sum_{i,j,i \neq j} E \left[\left| \phi_{\tilde{\mathbf{Y}}_i \tilde{\mathbf{Y}}_j}^n(\tau, \omega) - \phi_{SS}^n(\tau, \omega) v_{ni}(\omega) v_{nj}^*(\omega) \right|^2 \right] \quad (31)$$

を最小にすることで最適な元信号のパワースペクトルが

推定できる。式 (31) の目的関数は非線形なので解析解は得られないが、文献[9]の手法を用いれば、反復的に解を推定できる。 $\phi_{SS}^n(\tau, \omega) = \phi_{SS}(\tau)$, $\phi_{\hat{Y}_i \hat{Y}_j}^n(\tau, \omega) = \phi_{\hat{Y}_i \hat{Y}_j}(\tau)$, $\mathbf{v}_n(\omega) = \mathbf{v} = [v_1, \dots, v_L]^T$ と改めて書けば、元信号のパワースペクトル推定値 $\hat{\phi}_{SS}^n(\tau)$ とステアリングベクトルの推定値 $\hat{\mathbf{v}}$ は

$$\hat{\phi}_{SS}^n(\tau) \leftarrow \frac{\sum_{i,j,i \neq j} \Re \left[\phi_{\hat{Y}_i \hat{Y}_j}(\tau) \hat{v}_i^* \hat{v}_j \right]}{\sum_{i,j,i \neq j} |\hat{v}_i|^2 |\hat{v}_j|^2} \quad (32)$$

$$\hat{v}_l \leftarrow \frac{\mathbb{E} \left[\hat{\phi}_{SS}^n(\tau) \sum_{l,l \neq i} \phi_{\hat{Y}_i \hat{Y}_l}(\tau) \hat{v}_i \right]}{\mathbb{E} \left[\hat{\phi}_{SS}^n(\tau) \sum_{l,l \neq i} |\hat{v}_i|^2 \right]} \quad (33)$$

$(l = 1, \dots, L)$

$$\hat{\mathbf{v}} \leftarrow \hat{\mathbf{v}} / \hat{v}_1 \quad (34)$$

と更新することで同時推定が可能となる。ここで $\Re[\cdot]$ は実部をとる操作である。ここでは観測信号のみからステアリングベクトルを推定するので、スケールを定める為に、式 (34) で常に基準の周波数伝達特性を $\hat{v}_1 = 1$ としている。また、ステアリングベクトルの初期値として式 (23) で一つのマイクロホンにスケールを揃えたため、近似的に各マイクロホンペアまでの周波数伝達特性は揃っているものとして $\mathbf{v}_0 = [1, \dots, 1]^T$ を与えている。

4.2 マルチチャンネル Wiener フィルタ

二乗誤差を最小にする意味で最適なフィルタ処理として、式 (33) で推定したステアリングベクトルを用いた最小分散ビームフォーマ (MVDR: Minimum Variance Distortionless Response) の出力

$$Z_n(\tau, \omega) = \frac{\hat{\mathbf{v}}_n^H(\omega) \Phi_{\hat{\mathbf{Y}} \hat{\mathbf{Y}}}^{-1}(\tau, \omega) \tilde{\mathbf{Y}}_n(\tau, \omega)}{\hat{\mathbf{v}}_n^H(\omega) \Phi_{\hat{\mathbf{Y}} \hat{\mathbf{Y}}}^{-1}(\tau, \omega) \hat{\mathbf{v}}_n(\omega)} \quad (35)$$

に対して、更に後段に Wiener フィルタを用いる。

式 (32) で推定した元信号のパワースペクトルの推定値 $\hat{\phi}_{SS}^n(\tau)$ を用い、Wiener フィルタ

$$H_n(\tau, \omega) = \hat{\phi}_{SS}^n(\tau, \omega) / \phi_{ZZ}^n(\tau, \omega) \quad (36)$$

を設計した。ここで、 $\phi_{ZZ}^n(\tau, \omega)$ は MVDR の出力パワースペクトル $\mathbb{E}[|Z_n(\tau, \omega)|^2]$ である。更に事前情報 $0 \leq H_n(\tau, \omega) \leq 1$ を用いた処理

$$\begin{cases} H_n(\tau, \omega) \leftarrow 1 & (H_n(\tau, \omega) > 1) \\ H_n(\tau, \omega) \leftarrow 0 & (H_n(\tau, \omega) < 0) \end{cases} \quad (37)$$

を行い、最終的な出力信号を

$$\tilde{Z}_n(\tau, \omega) = \frac{\hat{\phi}_{SS}^n(\tau, \omega)}{\phi_{ZZ}^n(\tau, \omega)} \frac{\hat{\mathbf{v}}_n^H(\omega) \Phi_{\hat{\mathbf{Y}} \hat{\mathbf{Y}}}^{-1}(\tau, \omega) \tilde{\mathbf{Y}}_n(\tau, \omega)}{\hat{\mathbf{v}}_n^H(\omega) \Phi_{\hat{\mathbf{Y}} \hat{\mathbf{Y}}}^{-1}(\tau, \omega) \hat{\mathbf{v}}_n(\omega)} \quad (38)$$

として求める。

5. 実験

5.1 実験目的および条件

提案手法の有用性を示す目的で、マイクロホンの数を増やせる利点を生かした例としてワイアレスマイクロホンを用いた実

表 1 実環境実験の条件

Sources	TIMIT database (female)
Microphones	SANYO HM-W300
Room size	6.3 m × 7.7 m × 2.7 m
Reverberation time	300 ms
Sampling rate	16 kHz
Data length	5.0 s
Window function	Hamming
Frame length (ICA)	4096 points(256 ms)
Frame shift (ICA)	1024 points
Frame length (Post filter)	1024 points(64 ms)
Frame shift (Post filter)	64 points

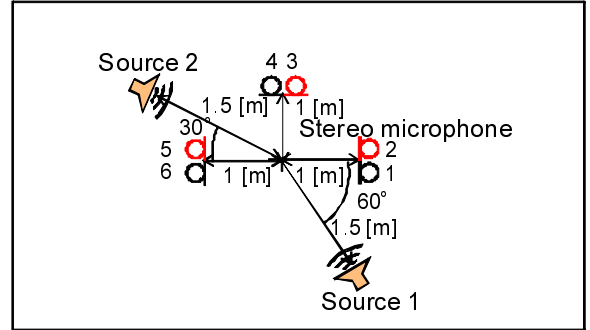


図 3 音源とマイクロホン配置

環境実験を行った。表 1 のような実験条件に対し、図 3・図 4 のようにワイアレスシステムにより全てのマイクロホン間の時間同期はとれている元で、中心から 1.5 [m] の距離にある 2 つの音源に対して、マイクロホンペア 3 組 (単一マイクロホン 6 個) を中心から 1.0 [m] の位置に 3 か所配置した。

比較として、6 チャンネル全てのマイクロホンを用いた BSS と各ペアごとの BSS を実装した。また提案法において、ワイアレスシステムにより正確な時間同期がとれているので、相互相関関数に基づく録音開始時刻を合わせるキャリブレーションの操作は行わなかった。補助関数を用いた ICA の反復回数は 30 回とし、提案法での元信号のパワースペクトルとステアリングベクトルの同時推定アルゴリズムも 30 回反復とした。提案法の最小分散ビームフォーマの為に $\Phi_{\hat{\mathbf{Y}} \hat{\mathbf{Y}}}^n(\tau, \omega)$ には $\tilde{\mathbf{Y}}(\tau, \omega) \tilde{\mathbf{Y}}^H(\tau, \omega)$ の全時間フレームの平均を用い、Wiener フィルタの為に $\phi_{\hat{Y}_i \hat{Y}_j}^n(\tau, \omega)$ には $\tilde{Y}_i(\tau, \omega) \tilde{Y}_j^*(\tau, \omega)$ の前後 5 フレームの平均を用いた。

5.2 実験結果

表 2 に SIR(Single-to-Interference Ratio) による評価を示す。各ペアごとの BSS の結果は最も入力 SN 比がよいペアに対する音源分離の結果である。各ペアごとの BSS の結果と比べて、提案法は 3 [dB] 程度の性能向上を確認することができた。また図 5 に出力信号スペクトログラムを示す。提案法がストフィルタにより妨害音の残響成分の消し残り成分の一部が抑制できていることが確認できる。

6. まとめと今後の展望

本研究ではマイクロホンペアを分散させる枠組みの中でブラインド音源分離を行う手法に加え、ペア間の拡散性雑音成分の



図 4 実環境実験の写真

表 2 実験結果: SIR による評価 [dB]

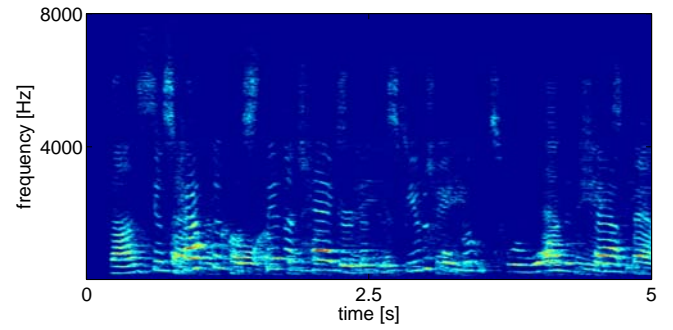
SIR [dB]	Only ICA with All Mics	Only ICA with Each Pair	Post-filter Proposed
Source 1	10.4	14.7	17.6
Source 2	8.0	14.3	17.3

無相関性に着目したポストフィルタを設計し、分離性能の向上を検討した。ICA によるペア毎での分離信号を統一するパーミュテーションとスケールの任意性の解決を行った。また、ペア毎での出力信号のモデル化を行い、ステアリングベクトルと元信号のパワースペクトルを同時推定し、最小分散ビームフォーマと Wiener ポストフィルタを設計した。更に実環境実験により、3 [dB] 程度の性能向上を確認することができた。

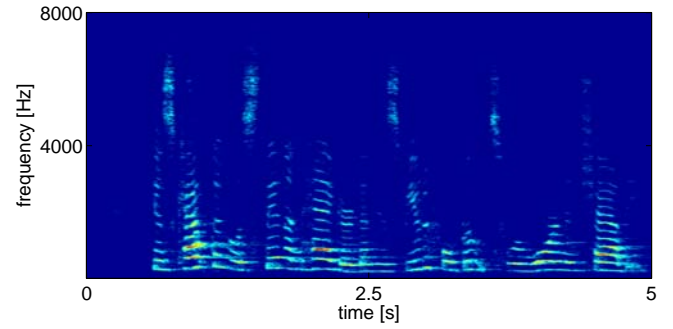
今後は、より多くのマイクロホンを用いた条件での実験を行いマイクロホン数と性能向上の度合いを考察したいと考えている。

文 献

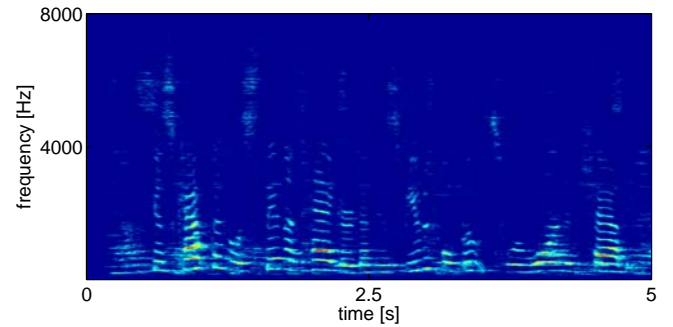
- [1] J. P. Dmochowski, Z. Liu and P. A. Chou, “Blind source separation in a distributed microphone meeting environment for improved teleconferencing,” *Proc. ICASSP*, pp. 89–92, 2009.
- [2] Z. Liu, “Sound Source Separation with Distributed Microphone Arrays in the Presence of Clock Synchronization Errors,” *Proc. IWAENC*, 2008.
- [3] K. Hasegawa, N. Ono, S. Miyabe and S. Sagayama, “Blind Estimation of Locations and Time Offsets for Distributed Recording Devices,” *Proc. LVA/ICA*, pp. 57–64, 2010.
- [4] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.
- [5] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda and F. Itakura, “Evaluation of blind signal separation method using directivity pattern under reverberant conditions,” *Proc. ICASSP*, pp. 3140–3143, 2000.
- [6] H. Sawada, R. Mukai, S. Araki and S. Makino, “A robust approach to the permutation problem of frequency-domain blind source separation,” *Proc. ICASSP*, pp. 381–384, 2003.
- [7] T. Ono, S. Miyabe, N. Ono and S. Sagayama, “Blind Source Separation with Distributed Microphone Pairs Using Permutation Correction by Intra-pair TDOA Clustering,” *Proc. IWAENC*, 2010.
- [8] 小野 拓磨, 宮部 滋樹, 小野 順貴, 嵯峨山 茂樹, “分散配置されたマイクロホンによるブラインド音源分離の検討,” 日本音響学会春季研究発表会講演集, pp. 739–740, 2010.
- [9] N. Ito, N. Ono and S. Sagayama, “Diffuse Noise Suppression Using Crystal-shaped Microphone Arrays,” *IEEE Trans. ASLP* (to appear).
- [10] N. Ono and S. Miyabe, “Auxiliary-function-based Inde-



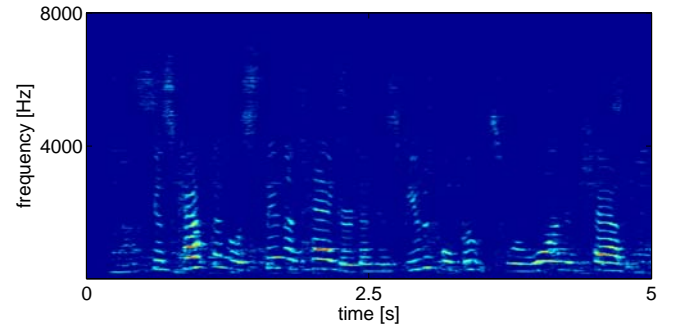
(a) 観測信号



(b) 元信号 (Source 1)



(c) ICA の出力 (Source 1)



(d) 提案法の出力 (Source 1)

図 5 信号のスペクトログラム

- [11] S. Amari, “Natural gradient works efficiently in learning,” *Trans. Neural Computation*, pp. 251–276, 1998.
- [12] S. Ikeda and N. Murata, “A method of ICA in time-frequency domain,” *Proc. ICA*, pp. 365–371, 1999.
- [13] K. Matsuoka and S. Nakashima, “Minimal distortion principle for blind source separation,” *Proc. ICA*, pp. 722–727, 2001.