

教師信号を用いた非同期分散型マイクロホンアレーによる音源分離*

坂梨龍太郎*¹ 小野順貴*^{2,*3} 宮部滋樹*¹
山田武志*¹ 牧野昭二*¹

【要旨】 近年、独立に動作する複数の録音機器を用いた非同期マイクロホンアレーが検討されている。非同期マイクロホンアレーは多チャンネルの A/D 変換器を必要としないので、従来より安価であり、マイクロホンの配置を柔軟に行えるという利点がある。一方、各チャンネルの信号が時間的に同期して録音されておらず、また異なる A/D 変換器を用いているために、録音開始時刻オフセットやサンプリング周波数ミスマッチ（以下ではミスマッチパラメータと総称する）が生じてしまう。従来、ミスマッチパラメータを推定して時間同期補償を行うために幾つかの手法が提案されている。しかし、処理が複雑であるために長時間の録音に対しては多大な処理時間を要するという問題がある。そこで本論文では、高速な時間同期補償と高性能な音源分離を実現するために、ミスマッチパラメータ推定と音源分離の両方を共通の教師信号を用いて行う枠組みを提案する。教師信号には、ある話者、あるいはある音源のみが音を発している区間である単一音源区間の信号を用いる。提案法では、時間的に離れた位置にある二つの単一音源区間の信号を手がかりに、ミスマッチパラメータを推定して時間同期補償を行う。更に、単一音源区間の信号を音源分離の教師信号として用い、またマイクロホンの分散型配置が可能という特徴を活用するように、音源分離手法である SN 比最大化ビームフォーマと Duong 法を拡張する。実験の結果、提案法により十分な精度での時間同期補償が可能であり、また高い音源分離性能が得られることを確認した。

キーワード 音源分離, 非同期マイクロホンアレー, 時間同期補償, 録音開始時刻オフセット, サンプリング周波数ミスマッチ

Speech enhancement, Ad hoc microphone array, Sampling frequency mismatch

1. はじめに

複数のマイクロホンを用いたマイクロホンアレーは、各マイクロホンに到達する音波の位相差から音源の空間的情報を得ることにより様々な音響信号処理を行うことができる。実環境におけるハンズフリー音声認識やコンピュータによる音環境理解などへの応用を目的として、これまでに様々な研究が行われてきた [1, 2]。これらのマイクロホンアレーによる音響信号処理技術は、マイクロホンは近接して配置され、各マイクロホンの信号は多チャンネル A/D 変換器で各チャンネルが時間的に同期された状態で録音されていることが前提となっている。しかし、多チャンネル A/D 変換器は高価

であることや、マイクロホンと A/D 変換器を接続するケーブルの長さによる制約などから、大規模なマイクロホンアレーを構成することが困難であるなど、拡張性の面で欠点がある。その解決策として、近年、独立に動作する複数の録音機器を分散配置して録音し、その後それぞれの録音信号を集めて処理を行う、非同期マイクロホンアレーを用いた音響信号処理が検討されている [3, 4]。

非同期マイクロホンアレーを用いた音響信号処理は、多チャンネル A/D 変換器を必要としない、多チャンネル化が容易である、マイクロホン配置が柔軟であるなど、これまでの同期マイクロホンアレーにはない魅力的な利点を持つ。しかし、音源やマイクロホンの位置が未知である、各録音機器における録音開始時刻が異なる（録音開始時刻オフセットが発生する）、公称サンプリング周波数が同じであっても録音機器間においてサンプリング周波数にわずかな個体差（サンプリング周波数ミスマッチ）がある、などの問題点も存在する。特に録音開始時刻オフセットとサンプリング周波数ミスマッチは、デジタル領域において波形のシフトと伸縮を引き起こしてしまう。これは、録音機器間で位相差が変化し、音源位置が疑似的に変化することに相当

* Audio source separation with asynchronous ad hoc microphone array using teaching signals, by Ryutaro Sakanashi, Nobutaka Ono, Shigeki Miyabe, Takeshi Yamada and Shoji Makino.

*¹ 筑波大学

*² 国立情報学研究所

*³ 総合研究大学院大学

(問合せ：牧野昭二 〒305-8577 つくば市天王台 1-1-1 筑波大学生命領域学際研究センター e-mail: maki@tara.tsukuba.ac.jp)

(2016年4月27日受付, 2016年11月10日採録決定)

するため、各音源は移動せず固有の位相差を持つということを前提とする、大多数のマイクロホンアレーによる音源分離手法を破綻させてしまう。

非同期マイクロホンアレーの信号処理方式のアプローチとして、振幅領域でウィーナフィルタを生成する方法 [5] や振幅領域で伝達関数ゲイン基底 NMF を用いる方法 [6] があるが、位相情報を用いないため目的音がひずみ易いという欠点がある。

一方、センサ間の同期を合わせる方法がある。非同期マイクロホンアレーを扱うためには、録音開始時刻オフセットとサンプリング周波数ミスマッチの値（以降、特に区別する必要がなく録音開始時刻オフセットとサンプリング周波数ミスマッチの両方を指す場合にはミスマッチパラメータという表記を用いる）の推定と、各録音信号の時間同期補償が必要不可欠であり、これまでに幾つかの手法が提案されている [7-11]。また、近年では、最尤法を用いてミスマッチパラメータをブラインドで高精度に推定する手法も提案されている [12, 13]。しかし、この手法には、音源は定常であり移動しないという仮定が必要である。

そこで本論文では、非同期マイクロホンアレーによる音源分離を想定し、高速な時間同期補償と高性能な音源分離を実現するために、ミスマッチパラメータ推定と音源分離の両方を共通の教師信号を用いて行う枠組みを提案する。教師信号には、ある話者、あるいはある音源のみが音を発している区間である単一音源区間の信号を用いる。提案法では、時間的に離れた位置にある二つの単一音源区間の信号を手がかりに、ミスマッチパラメータを推定して時間同期補償を行う。実用上は、例えば会議の開会、閉会の挨拶や、初めの自己紹介など、簡潔な特定話者の音声や、又はチャープ信号のような目印となる音を録音の最初と最後に鳴らすなどして、これを単一音源区間情報として扱う。単一音源区間情報を利用することにより、ブラインドな従来の手法と比べ録音時の操作（単一音源の録音）や音源分離の前処理（単一音源区間情報の特定）が増える形となるが、従来手法と比べて録音開始時刻オフセットやサンプリング周波数ミスマッチの算出はより容易になると考えられる。更に本論文では、単一音源区間の信号を音源分離の教師信号として用い、またマイクロホンの分散型配置が可能という特徴を活用するように、音源分離手法である SN 比最大化ビームフォーマ [14, 15] と Duong 法 [16] を拡張する。実験の結果、提案法により十分な精度での時間同期補償が可能であり、また高い音源分離性能が得られることを示す [17]。

本論文の構成を以下に示す。第 2 章では単一音源区間情報を用いたミスマッチパラメータ推定と時間同期

補償について述べる。第 3 章では音源分離手法である SN 比最大化ビームフォーマと Duong 法の概要を説明し、第 4 章ではこれらの手法の拡張を行う。第 5 章では提案手法の有効性を検証し、第 6 章で本論文の結論をまとめる。

2. 単一音源区間情報を用いた非同期録音信号の時間同期

2.1 非同期信号の時間領域モデル

非同期録音の時間領域モデルについて議論する。以下では 2 チャネル間の非同期信号に限定するが、すべてのチャネルを特定の 1 チャネルに時間同期することで多チャネルにおいても容易に時間同期を実現できる。

まず、二つのマイクロホンが観測した連続信号 $x_1(t)$, $x_2(t)$ (t は連続時間) が異なる A/D 変換器により離散信号 $x_1(n_1)$, $x_2(n_2)$ (n_1, n_2 はサンプル番号) にサンプリングされたとする。ここで $x_1(n_1)$ のサンプリング周波数は f_s , $x_2(n_2)$ のサンプリング周波数は未知のミスマッチ ϵ により $(1 + \epsilon)f_s$ であるとする。このとき離散信号 $x_i(n)$ は連続信号 $x_i(t)$ ($i = 1, 2$) により以下のように表される。

$$x_1(n_1) = x_1\left(\frac{n_1}{f_s}\right) \quad (1)$$

$$x_2(n_2) = x_2\left(\frac{n_2}{(1 + \epsilon)f_s} + D_{21}\right) \quad (2)$$

ここで、 D_{21} は $x_1(n_1)$ に対する $x_2(n_2)$ の連続時間における録音開始時刻の差を表す。このとき、時刻 t を示す両チャネルのサンプル番号 n_i ($i = 1, 2$) は

$$n_1 = tf_s \quad (3)$$

$$n_2 = (1 + \epsilon)(t - D_{21})f_s \quad (4)$$

と与えられる。よって n_2 と n_1 の関係は以下のように表される。

$$n_2 = (1 + \epsilon)n_1 - (1 + \epsilon)D_{21} \quad (5)$$

ここで、 $D_{21} = D_{21}f_s$ はチャネル 2 の録音開始時刻を 0 としたときのチャネル 1 の離散時刻を表し、この値を録音開始時刻オフセットと呼ぶ。

また、 n_1 と n_2 の差は、

$$g(n_{12}) = n_2 - n_1 = \epsilon n_1 - (1 + \epsilon)D_{21} \quad (6)$$

のように表すことができ、チャネル間のサンプル差 $g(n_{12})$ は時刻が進むにつれ拡大していくことが分かる。このことからサンプリング周波数ミスマッチによる時間差のドリフトは、音源位置が疑似的に変化することに相当するため、サンプリング周波数ミスマッチの推定と補償は到来時間差 (TDOA) を利用する従来

のマイクロホンアレー技術に不可欠である。また、SN比最大化ビームフォーマをはじめとする音源方向を陽に与えないマイクロホンアレー手法においては、数サンプル程度の録音開始時刻オフセットは大きな問題とならないが、各チャンネルの同一番号のフレーム内ではほぼ同じ時刻の信号を参照する程度の補償は必要である。なお、今後、非同期マイクロホンアレーの議論であっても、単にサンプル番号を表す場合は n と表し、チャンネル間の同一連続時刻におけるサンプル番号を比較する場合のみ n_1, n_2 という表記を用いることとする。

2.2 単一音源区間情報を用いた mismatch パラメータの推定

ここでは、単一音源区間情報を用いた mismatch パラメータの推定法を提案する。単一音源区間とは、ある特定の話者、又は音源のみが音を発している区間であり、提案手法では、この単一音源区間を信号の最初と最後に録音することでこれを手がかりとし mismatch パラメータの推定を行う。mismatch パラメータの推定の議論の前に、録音開始時刻オフセット D_{21} とサンプリング周波数 mismatch ϵ が、同じアナログ時間に対応する二つのチャンネル間ペア $\{n_{A1}, n_{B1}\}$ と $\{n_{A2}, n_{B2}\}$ (図-1) が与えられることで推定可能となることを示す。これらの離散時刻の間には式 (5) より以下の関係が成り立つ。

$$n_{A2} = (1 + \epsilon)(n_{A1} - D_{21}) \quad (7)$$

$$n_{B2} = (1 + \epsilon)(n_{B1} - D_{21}) \quad (8)$$

この連立方程式を解くことにより、mismatch パラメータ D_{21} と ϵ は以下のように得られる。

$$D_{21} = \frac{n_{A1}n_{B2} - n_{A2}n_{B1}}{n_{B2} - n_{A2}} \quad (9)$$

$$\epsilon = \frac{n_{B2} - n_{A2}}{n_{B1} - n_{A1}} - 1 \quad (10)$$

従って、異なる 2 時刻における各チャンネルのサンプル番号 $\{n_{Ai}, n_{Bi}\}$ ($i = 1, 2$) を推定することにより、録音開始時刻オフセット D_{21} とサンプリング周波数 mismatch ϵ を同定することができる。これらのチャンネル間ペアの正確なサンプル番号を推定することは困難だが、ある程度の精度で推定することができれば、後段のマイクロホンアレー処理に十分な精度の録音開始時刻オフセット D_{21} とサンプリング周波数 mismatch ϵ の推定が可能となる。

次に、 n_{Ai} と n_{Bi} ($i = 1, 2$) の推定について説明する。図-2 で示す、単一音源区間情報の中心サンプル番号をそれぞれ $\{n_{Ai}, n_{Bi}\}$ ($i = 1, 2$) とし、このサンプル番号の推定を考える。 n_{Ai} が推定できれば n_{Bi} ($i = 1, 2$) は同様に推定可能であるため、ここではま

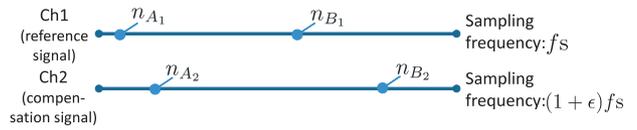


図-1 mismatch モデル

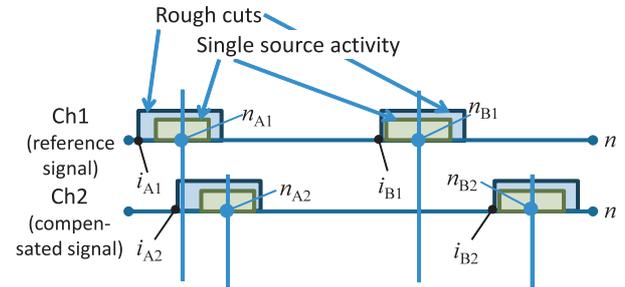


図-2 異なる 2 時刻における単一音源区間情報

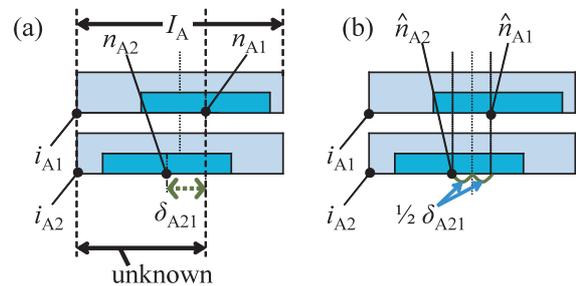


図-3 切り出し範囲における相互相関

ず n_{Ai} の推定方法を説明する。各単一音源区間情報の中心サンプル番号 n_{Ai} ($i = 1, 2$) は未知だが、単一音源がアクティブである大まかな位置は既知であるため、各チャンネルで単一音源区間情報を含む長さ I_A の大まかな切り出しを行う。チャンネル 1 の切り出し範囲は $i_{A1}, \dots, i_{A1} + I_A - 1$ 、チャンネル 2 の切り出し範囲は $i_{A2}, \dots, i_{A2} + I_A - 1$ である。図-2 より、切り出し範囲における話者がアクティブである時刻のずれは以下のように推定される。

$$\delta_{A21} = \arg \max_{\tau} \sum_{q=0}^{I_A-1} x_1(i_{A1} + q)x_2(i_{A2} + q - \tau) \quad (11)$$

しかし図-3(a) に示すとおり、依然として中心サンプル番号 n_{Ai} ($i = 1, 2$) は不明であるため、おおよその推定として、

$$\hat{n}_{A1} = i_{A1} + I_A/2 + \frac{1}{2}\delta_{A21} \quad (12)$$

$$\hat{n}_{A2} = i_{A2} + I_A/2 - \frac{1}{2}\delta_{A21} \quad (13)$$

のように近似を行い、これを n_{Ai} ($i = 1, 2$) の推定として扱う (図-3(b))。同様に n_{Bi} ($i = 1, 2$) の推定も行うことで、異なる 2 時刻における各チャンネルのサンプル番号 $\{n_{Ai}, n_{Bi}\}$ ($i = 1, 2$) を得ることができる。

これらの推定値を式 (9)(10) に代入することにより、ミスマッチパラメータ D_{21} と ϵ を求める。

2.3 短時間フレーム内におけるサンプリング周波数ミスマッチモデル

多くのアレー信号処理は STFT 領域で行われるため、本論文では STFT 領域でのサンプリング周波数ミスマッチ補償方法について述べる。STFT 分析に進む前に、短時間フレーム内におけるサンプリング周波数ミスマッチの影響について議論し、短時間フレーム内ではサンプリング周波数ミスマッチが無視できチャンネル間の時間ドリフトが起こらないと見なせることを示す。

チャンネル 1 の第 $(n_1 + m)$ サンプルと同期する第 2 チャンネルの離散時刻は、式 (5) より $n_2 = \phi_{21}(n_1; D_{21}, \epsilon)$ という表記を用いて、

$$\begin{aligned} \phi_{21}(n_1 + m; D_{21}, \epsilon) \\ &= (1 + \epsilon)(n_1 - D_{21}) + (1 + \epsilon)m \\ &= \phi_{21}(n_1; D_{21}, \epsilon) + (1 + \epsilon)m \end{aligned} \quad (14)$$

のように表される。 $|\epsilon| \ll 1$ より $|m\epsilon| \ll 1$ が成り立ち、この条件のもとで式 (14) は

$$\phi_{21}(n_1 + m; D_{21}, \epsilon) \approx \phi_{21}(n_1; D_{21}, \epsilon) + m \quad (15)$$

と近似することができ、ある時刻ペア n_1 と n_2 が同期されている場合、これらから少しだけ離れた時刻 $n_1 + m$ と $n_2 + m$ はほぼ同期していると思なすことができる。

従って、以下で与えられる時刻 n_i を中心とする $x_i(n)$ のフレーム長 $L \ll 1/\epsilon$ (以下の定式化はすべて L が偶数であるとして扱う) のフレーム信号 $x_i^{\text{fr}}(l, n_i)$, $l = 0, \dots, L-1$ は以下のように表される。

$$x_i^{\text{fr}}(l, n_i) = w(l)x_i\left(l + n_i - \frac{L}{2}\right) \quad (16)$$

ここで $w(l)$ は窓関数を表し、チャンネル番号 $i = 1, 2$ の信号はこのフレーム内ではほぼ同期していると思なすことができる。典型的な機器間のサンプリング周波数ミスマッチの値 ϵ は 10^{-5} オーダであり [3, 4], またマイクロホンアレー信号処理における典型的な窓幅は 0.1 s オーダであることから、この近似によるフレーム内の最大の誤差 $|\epsilon L|/2$ は多くの場合に $1 \mu\text{s}$ オーダに収まる。そして、このように誤差が大きくなるのはフレームの両端においてであり、窓関数 $w(l)$ に Hanning 窓のようなフレームの両端の振幅が小さくなる典型的なものを用いることで、誤差の大きいフレームの両端のサンプルの影響は軽減される。

以上、 n_1 と n_2 が同期しているという前提においては、STFT フレーム内ではチャンネル間における時刻ドリフトが無視できること、また式 (16) のように各フ

レームごとにチャンネル間の時間同期が行えることを示したが、サンプリング周波数ミスマッチ ϵ は微小な値であるため、チャンネル 2 における STFT フレームの中心サンプル番号 n_2 は多くの場合において非整数値を示す。デジタル信号処理において非整数サンプル番号を扱うことは困難であり、実際の処理において式 (16) で表される非整数点シフトによるフレーム分析をそのまま行うことはできない。そのため本論文では整数点フレームシフトと小数点位相シフトを組み合わせることにより非整数点シフトを実現する。

2.4 非整数点シフトのフレーム分析によるミスマッチ補償

ここでは、録音開始時刻オフセット D_{21} とサンプリング周波数ミスマッチ ϵ が与えられたという条件の元で、 $x_2^{\text{fr}}(l, n_2)$ の近似の STFT 表現を整数点フレームシフトと小数点位相シフトを用いて求める方法について議論する。

第 i チャンネルの中心サンプルを n とする STFT 分析は以下で与えられる。

$$X_i(k, n) = \sum_{l=0}^{L-1} x_i^{\text{fr}}(l, n) \exp\left(-\frac{2\pi jkl}{L}\right) \quad (17)$$

ここで $k = -L/2, \dots, L/2 - 1$ は離散周波数インデックスを表す。ただし実際の処理では高速フーリエ変換によりこの信号変換をより効率的に行う。式 (5) より、 $X_1(k, n_1)$ の中心サンプル n_1 と同期する第 2 チャンネルの離散時刻は $n_2 = \phi_{21}(n_1; D_{21}, \epsilon)$ で与えられる。この離散時刻を中心とした第 2 チャンネルのフレーム分析により、チャンネル間のフレーム分析の同期が近似的に達成できるが、一般には n_2 は非整数となり、正確なフレーム分析は困難になる。そこで、このようなフレーム分析を、まず最近傍整数サンプルのフレーム分析で近似し、更にフレーム中心の丸めによる誤差を線形位相フィルタを用いた円状時間シフトにより近似的に補償する。これらの操作により、整数点シフト以上に正確なフレームシフトを実現することで、音源位置の擬似的な変化を回避する。

まずフレーム中心 $\phi_{21}(n_1; D_{21}, \epsilon)$ と最も近い整数値 $\bar{\phi}_{21}(n_1; D_{21}, \epsilon)$ を求める。

$$\bar{\phi}_{21}(n_1; D_{21}, \epsilon) = \arg \min_{n \in \mathbb{Z}} |\phi_{21}(n_1; D_{21}, \epsilon) - n| \quad (18)$$

この整数サンプル $\bar{\phi}_{21}(n_1; D_{21}, \epsilon)$ を中心とするフレーム $x_2^{\text{fr}}(l, \bar{\phi}_{21}(n_1; D_{21}, \epsilon))$ は、非整数離散時刻 $\phi_{21}(n_1; D_{21}, \epsilon)$ よりも中心時刻が

$$\tilde{\phi}_{21}(n_1; D_{21}, \epsilon) = \phi_{21}(n_1; D_{21}, \epsilon) - \bar{\phi}_{21}(n_1; D_{21}, \epsilon) \quad (19)$$

サンプルだけ遅延しているため、これを線形位相フィルタで補償することにより、近似的な STFT 表現

$$\begin{aligned} \hat{X}_2(k, \phi_{21}(n_1; D_{21}, \epsilon)) = \\ X_2(k, \bar{\phi}_{21}(n_1; D_{21}, \epsilon)) \exp\left(\frac{2\pi j k \bar{\phi}_{21}(n_1; D_{21}, \epsilon)}{L}\right) \end{aligned} \quad (20)$$

が得られる。また、このようにして近似的に同期した STFT 分析 $\hat{X}_2(k, \phi_{21}(n_1; D_{21}, \epsilon))$ を $X_1(k, n_1)$ と同じフレームシフトの逆 STFT に施すことにより、ミスマッチ補償信号 $\hat{x}_2(n_1; D_{21}, \epsilon)$ が得られる。

3. 教師信号を用いたマイクロホンアレー音源分離の従来法

3.1 問題設定と観測モデル

ここでは、本論文で扱う音源分離における観測モデルについての説明を行う。まず、実環境において J 人の音声信号 $s_j (j = 1, \dots, J)$ が I 個のマイクロホンで観測されたとする。このときの観測信号 $x_i(t)$ は以下のようにモデル化できる。

$$x_i(t) = \sum_{j=1}^J \sum_{\tau} h_{ij}(\tau) s_j(t - \tau), (j = 1, \dots, J) \quad (21)$$

ここで h_{ij} は音源 j からマイクロホン i へのインパルス応答、 $s_j(t)$ は j 番目の原音源の連続時間領域表現である。本論文における非同期信号の同期補償と音源分離には、時間領域での観測信号に短時間フーリエ変換 (STFT) を適用し時間周波数領域で表現する。時間周波数領域では観測信号 $\mathbf{x}(r, f)$ は以下のように表現できる。

$$\begin{aligned} \mathbf{x}(r, f) = [x_1(r, f), \dots, x_I(r, f)]^T \\ \approx \sum_{j=1}^J \mathbf{h}_j(f) s_j(r, f) \end{aligned} \quad (22)$$

$$\mathbf{h}_j(f) = [h_{1j}(f), \dots, h_{Ij}(f)]^T \quad (23)$$

ここで r は時間フレーム番号、 f は周波数ビン番号を表す。また、マイクロホン位置における、残響などの空間特性を含んだ音源信号を音像 (source image) と呼び [18]、 j 番目の音源の各マイクロホン位置における音像を以下のように定義する。

$$c_j(r, f) = [c_{1j}(r, f), \dots, c_{Ij}(r, f)]^T \quad (24)$$

$$= \mathbf{h}_j(f) s_j(r, f) \quad (25)$$

3.2 SN 比最大化ビームフォーマ

SN 比最大化ビームフォーマは、出力信号中の目的信号と非目的信号のパワー比を最大化するように指向

特性を形成する。ここでパワー比 $\lambda(f)$ は

$$\lambda(f) = \frac{\mathbf{w}_j^H(f) \mathbf{R}_T^j(f) \mathbf{w}_j(f)}{\mathbf{w}_j^H(f) \mathbf{R}_I^j(f) \mathbf{w}_j(f)} \quad (26)$$

のように表される。なお、 \mathbf{R}_T^j は目的信号区間、 \mathbf{R}_I^j は非目的信号区間それぞれの共分散行列であり、

$$\mathbf{R}_T^j(f) = \sum_{r \in \Theta_T} \mathbf{x}_T(r, f) \mathbf{x}_T^H(r, f) \quad (27)$$

$$\mathbf{R}_I^j(f) = \sum_{r \in \Theta_I} \mathbf{x}_I(r, f) \mathbf{x}_I^H(r, f) \quad (28)$$

と表される。ここで、 Θ_T は目的信号区間、 Θ_I は非目的信号区間のそれぞれ時間フレームの集合である。このパワー比 $\lambda(f)$ を最大化する $\mathbf{w}_j(f)$ は、以下の一般化固有値問題の最大固有値に対応する固有ベクトルに相当する。

$$\mathbf{R}_T^j(f) \mathbf{w}_j(f) = \lambda(f) \mathbf{R}_I^j(f) \mathbf{w}_j(f) \quad (29)$$

なお、SN 比最大化ビームフォーマ $\mathbf{w}_j(f)$ はスケールを決定するため次の補正を行う。

$$\mathbf{w}_j(f) \leftarrow b_j(f) \mathbf{w}_j(f) \quad (30)$$

ただし、 $b_j(f)$ は $\mathbf{b}(f)$ の i 番目のある任意の要素であり以下のように表される。

$$\mathbf{b}(f) = \frac{\mathbf{R}_x(f) \mathbf{w}_j(f)}{\mathbf{w}_j^H(f) \mathbf{R}_x(f) \mathbf{w}_j(f)} \quad (31)$$

$$\mathbf{R}_x(f) = \sum_{r=1}^T \mathbf{x}(r, f) \mathbf{x}^H(r, f) \quad (32)$$

SN 比最大化ビームフォーマの出力信号 $y_j(r, f)$ は、各マイクロホンの観測信号をフィルタリングして足し合わせた $y_j(r, f)$ であり、ビームフォーマ $\mathbf{w}_j(f)$ を用いて以下の式で得られる。

$$y_j(r, f) = \mathbf{w}_j^H(f) \mathbf{x}(r, f) \quad (33)$$

3.3 Duong 法

Duong 法は音源と伝達関数の積である音像が多変量正規分布に従うという仮定により、時不変の空間相関行列と時変の分散を EM アルゴリズムにより推定し、それらのパラメータを用いた多チャネルウィーナフィルタにより音源分離を行う手法である [16]。文献 [16] にはその最尤推定が明示的に定式化されていないが、Togami が指摘しているように、その EM アルゴリズムの Q 関数では隠れ変数として扱われる音像が多チャネルの複素振幅のまま重ね合わせられて観測信号を生成するというモデル化となっていて、近似的にしか成り立たないスパース性の仮定を排している [19]。

まず適当な初期値 $v_j(r, f)$ と $\mathbf{R}_j(f)$ ($j = 1, \dots, J$)

を与え、以下のように初期化する。

$$\mathbf{R}_{c_j}(r, f) = v_j(r, f)\mathbf{R}_j(f) \quad (34)$$

$$\mathbf{R}_x(r, f) = \sum_{j=1}^J \mathbf{R}_{c_j}(r, f) \quad (35)$$

この初期化の後、E-step では以下の更新

$$\mathbf{W}_j(r, f) = \mathbf{R}_{c_j}(r, f)\mathbf{R}_x^{-1}(r, f) \quad (36)$$

$$\hat{c}_j(r, f) = \mathbf{W}_j(r, f)\mathbf{x}(r, f) \quad (37)$$

$$\hat{\mathbf{R}}_{c_j}(r, f) = \hat{c}_j(r, f)\hat{c}_j^H(r, f) + (\mathbf{I} - \mathbf{W}_j(r, f))\mathbf{R}_{c_j}(r, f) \quad (38)$$

を行い、M-step では以下の更新

$$v_j(r, f) = \frac{1}{I} \text{tr}(\mathbf{R}_j^{-1}(f)\hat{\mathbf{R}}_{c_j}(r, f)) \quad (39)$$

$$\mathbf{R}_j(f) = \frac{1}{N} \sum_{n=1}^N \frac{1}{v_j(r, f)} \hat{\mathbf{R}}_{c_j}(r, f) \quad (40)$$

$$\mathbf{R}_{c_j}(r, f) = v_j(r, f)\mathbf{R}_j(f) \quad (41)$$

$$\mathbf{R}_x(r, f) = \sum_{j=1}^J v_j(r, f)\mathbf{R}_j(f) \quad (42)$$

を行う。これを繰り返すことによりパラメータを推定する。ここで \mathbf{I} は単位行列である。

Duong 法による出力は、各マイクロホンの観測信号をフィルタリングして音像 $\mathbf{c}(r, f)$ を求めたものであり、マルチチャンネルウィーナフィルタ $\mathbf{W}_j(r, f)$ を用いて以下の式で得られる。

$$\mathbf{c}_j(r, f) = \mathbf{W}_j(r, f)\mathbf{x}(r, f) \quad (43)$$

4. 教師信号を用いたマイクロホンアレー音源分離手法の分散型配置への拡張

4.1 拡張の狙い

第2章で提案したミスマッチパラメータ推定と時間同期補償により、従来の音源分離手法を非同期マイクロホンアレーにおいてそのまま適用することが可能となる。ここで会議録音に着目すると、各話者に録音機器をそれぞれ近接して配置することが可能であり、各マイクロホンではある一人の話者音声のパワーが大きく観測されることになる。また、単一音源区間の信号を音源分離の教師信号として用いることが可能である。そこで、これらの特徴を活用して音源分離性能を更に向上させるために、第3章で述べた音源分離手法の拡張を行う。

4.2 分散型配置における音源分離手法の目的信号

本論文で扱う非同期分散型マイクロホンアレーでは各マイクロホンを各話者に近接配置し録音を行う。このた

め、各マイクロホンはある一人の話者音声のパワーが大きく観測されるという性質を持ち、本論文ではこの性質を利用して音源分離を行う。本論文では、近接する音源番号とマイクロホン番号は一致 ($j = i$) しているとする。SN 比最大化ビームフォーマにおける目的信号は第3章の説明と同じくビームフォーマ \mathbf{w}_j により $y_j(r, f) = \mathbf{w}_j^H(f)\mathbf{x}(r, f)$ の形で得られ、Duong 法では、上記の性質から音像 $\mathbf{c}_j(r, f) = [c_{1j}(r, f), \dots, c_{Ij}(r, f)]^T$ のうち $c_{jj}(r, f)$ を目的信号として扱う。

4.3 SN 比最大化ビームフォーマの拡張

本論文で想定している分散型アレーでは、チャンネルによりゲインが大きく異なるため、各マイクロホンを接近して設置させた従来のマイクロホンアレー以上に式 (30) の補正フィルタを目的話者ごとに以下のように対応させる。

$$\mathbf{w}_j(f) \leftarrow b_i(f)\mathbf{w}_j(f) \quad (44)$$

4.4 Duong 法の拡張

4.4.1 初期値設定

Duong 法はスパース性を仮定しない音像重ね合わせモデルを用いることで高品質な分離を実現するが、初期値依存性が高いことが知られており、バイナリマスクによる出力から Duong 法の初期値を生成する手法がこれまでに提案されている [20]。本論文では、教師信号を用いて以下のように EM アルゴリズムの初期値を設定する。

$$\mathbf{R}_j^{\text{init}}(f) = \frac{1}{N} \sum_n \mathbf{x}(r, f)\mathbf{x}(r, f)^H \quad (45)$$

$$v_j^{\text{init}}(r, f) = 1 \quad (46)$$

このように教師信号を用いて空間相関行列を設定することで、高い性能が期待できる。

4.4.2 ウィーナフィルタの作成

以上の初期値を用いて式 (36) で表されるウィーナフィルタ $\mathbf{W}_j(r, f) = \mathbf{R}_{c_j}(r, f)\mathbf{R}_x^{-1}(r, f)$ を作成する。本論文では初期値と Duong 法によるパラメータ更新の程度により以下3パターンのウィーナフィルタ [20] を作成する。

4.4.3 固定係数型ウィーナフィルタ

固定係数型ウィーナフィルタは、初期値を EM アルゴリズムで更新せず作成した以下の $\mathbf{R}_{c_j}(f)$ と $\mathbf{R}_x(r, f)$ を用いて作成する。

$$\mathbf{R}_{c_j}(f) = v_j^{\text{init}}(r, f)\mathbf{R}_j^{\text{init}}(f) \quad (47)$$

$$\mathbf{R}_x(r, f) = \sum_{j=1}^J v_j^{\text{init}}(r, f)\mathbf{R}_j^{\text{init}}(f) \quad (48)$$

これは分散 $v_j(r, f)$ を更新しないため、推定された音

像はパワーが重み付けされていない形となる。

4.4.4 空間相関固定の Duong 型ウィーナフィルタ

空間相関固定の Duong 型ウィーナフィルタは、初期値を EM アルゴリズムで更新し、更新された重み $v_j(r, f)$ と初期値の空間相関行列 $\mathbf{R}_j^{\text{init}}(f)$ により作成した $\mathbf{R}_{c_j}(f)$ と $\mathbf{R}_x(r, f)$ を用いて作成する。

$$\mathbf{R}_{c_j}(f) = v_j(r, f)\mathbf{R}_j^{\text{init}}(f) \quad (49)$$

$$\mathbf{R}_x(r, f) = \sum_{j=1}^J v_j(r, f)\mathbf{R}_j^{\text{init}}(f) \quad (50)$$

これは教師信号を用いて EM アルゴリズムによる推定を行うため、推定された音像はパワーが理想的に重み付けされた値であると考えられる。

4.4.5 全パラメータ更新の Duong 型ウィーナフィルタ

全パラメータ更新の Duong 型ウィーナフィルタは、初期値を EM アルゴリズムで更新し、推定された分散 $v_j(r, f)$ と空間相関行列 $\mathbf{R}_j(f)$ により作成した $\mathbf{R}_{c_j}(f)$ と $\mathbf{R}_x(r, f)$ を用いて作成する。

$$\mathbf{R}_{c_j}(f) = v_j(r, f)\mathbf{R}_j(f) \quad (51)$$

$$\mathbf{R}_x(r, f) = \sum_{j=1}^J v_j(r, f)\mathbf{R}_j(f) \quad (52)$$

これは空間相関固定の Duong 型ウィーナフィルタと比べ、更に更新された空間相関行列を用いている。教師信号を与えた場合、空間相関行列は初期値の段階で正解値を持っているが、本論文では EM アルゴリズムのパラメータ更新を調査するためこの形のウィーナフィルタを作成した。

4.4.6 伝達関数ゲインベースのパーミュテーション解法

Duong 法では、各周波数が個別に取り扱われるために生じるパーミュテーションという問題が存在する。これは推定信号の各周波数成分は周波数ごとそれぞれ異なる音源番号の順に出力されるという問題である。そのため分離信号を周波数ごとに並び替える必要があり、ここではそのパーミュテーション解法について説明する。

本論文では各マイクロホンはある一人の話者音声のパワーが大きく観測されるという分散型マイクロホンアレーの性質を利用した、伝達関数ゲインベースのパーミュテーション解法を採用する。具体的には、本論文ではマイクロホン番号と、目的音である音源番号は一致していると考えられるため ($i = j$)、周波数ごとに「素子数 × 音源数」の正方行列の形で表される時間定常の伝達関数において、対角成分和が最大となる音源番号

表-1 設定したミスマッチパラメータ

	Recording start time offset	Sampling frequency
Mic1	0	16,000 Hz
Mic2	-24,000 samples	16,001 Hz
Mic3	+24,000 samples	16,002 Hz
Mic4	+48,000 samples	16,003 Hz

の組み合わせを探すことでパーミュテーションを解決する。しかし Duong 法から伝達関数そのものは推定されないため、本論文では Duong 法による推定音像 $\mathbf{c}_j(r, f)$ から以下のように時間方向にパワー 2 乗合計を取った正方行列 $\mathbf{A}_j(f)$ を求める。また、対角成分和を求めるため音源ごとにパワーを規格化する。

$$\mathbf{A}_j(f) = \frac{1}{I} \sum_n \|\mathbf{c}_j(r, f)\|^2 \quad (53)$$

ここで $\mathbf{A}_j(f)$ は各周波数ごと時間方向に推定音像のパワー 2 乗合計を取った行列であり、伝達関数として見なすものである。

5. 実験

本章では、提案手法の有効性と教師信号が音源分離の学習に利用できる時、どこまで非同期分散型マイクロホンアレーにおける音源分離の性能を向上させられるか調査するための実験を行う。5.2 節では提案手法である単一音源区間情報を用いたミスマッチ推定精度の調査を行い、5.3 節では提案手法を用いて時間同期した観測信号に第 4 章で説明した音源分離手法を適用することで、教師信号を用いた非同期分散型マイクロホンアレーにおける音源分離手法の評価、検討を行う。

5.1 実験 1：単一音源区間情報を用いたミスマッチ推定精度の調査

5.1.1 実験条件

実験 1 では、単一音源区間情報を用いたミスマッチ推定精度の調査を行う。実験目的は本論文で提案した、単一音源区間情報を用いたミスマッチ推定精度と、基準チャンネルにおける二つの単一音源区間の間の長さの関係を調べることである。本提案手法は観測信号に単一音源区間情報が存在すれば推定が行えるので、無音区間に単一音源区間のみを付加した信号を 4 チャンネル作成し、2 チャンネル以降に録音開始時刻オフセットを付加し、わずかなりサンプリングを行った。基準信号のサンプリング周波数は 16,000 Hz である。そして信号長による推定精度を調査するため単一音源区間の長さを適宜変更しミスマッチ推定を行った。各チャンネルにおける録音開始時刻差の値、リサンプリング周波数を表-1 に示す。チャンネル 3 以降のサンプリング周波数ミスマッチの大きさは、実際の機器間のミスマッチ

表-2 実験条件

マイクロホン素子数	4
音源数	4
サンプリング周波数	16,000 Hz
フレーム長	4,096 samples
シフト長	2,048 samples
単一区間間長	330 s
教師信号区間長	20 s
目的音分離区間長	20 s
Duong 法 EM アルゴリズム反復回数	30 回

表-3 録音条件

AD/DA	Steinberg UR824
マイクロホン	SHURE SM57
Power amp.	YAMAHA XM4080
話者	日本人男性 4 人
残響時間	600 ms
暗騒音	44.4 dB (A)
気温	18.4°C
湿度	37%

ンアレーが非同期録音でどのように動作をするか検討を行うため、以下の 4 条件

- 1) 非同期録音（録音開始時刻オフセット+サンプリング周波数 mismatch ともに未補償, 図-6)
- 2) 録音開始時刻のみ同期（サンプリング周波数 mismatch が未補償, 図-7)
- 3) 時間同期（提案手法, 図-8)
- 4) 同期録音（録音開始時刻オフセット, サンプリング周波数 mismatch なし)

に対し、以下の音源分離手法

- 未処理 (Baseline)
 - SN 比最大化ビームフォーマ
 - 固定係数型ウィーナフィルタ
 - 空間相関固定の Duong 型ウィーナフィルタ
 - 全パラメータ更新の Duong 型ウィーナフィルタ
- を適用し、各観測音ごとに評価を行う。評価尺度として SDR (Signal to Distortion Ratio), SIR (Source to Interference Ratio) の評価尺度を使用する [21]。録音データは図-9 のようにマイクロホンと話者を配置し、各マイクロホンが同期された状態で話者ごとに録音を行い、録音後に表-1 に示した条件でチャンネルごとに時間シフトとサンプリングを行い、人工的に非同期データを作成した。また、非同期信号の mismatch 推定は、実験 1 における単一区間の時間長が 5 分 30 秒での条件で推定を行い、それを用いて非同期信号の同期補償を行った。その他の実験条件を表-2、録音条件を表-3 に示す。

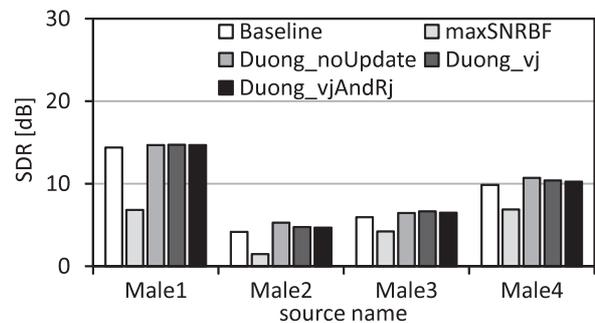


図-10 実験結果 (1. 非同期録音, SDR)

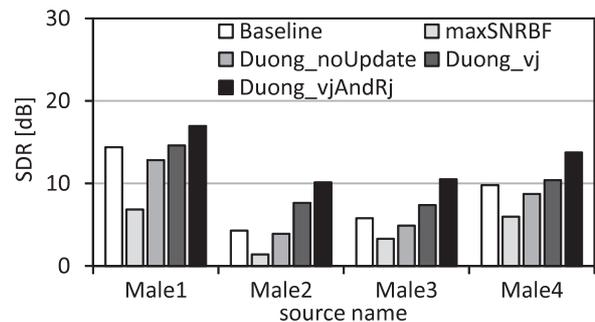


図-11 実験結果 (2. 録音開始時刻のみ同期, SDR)

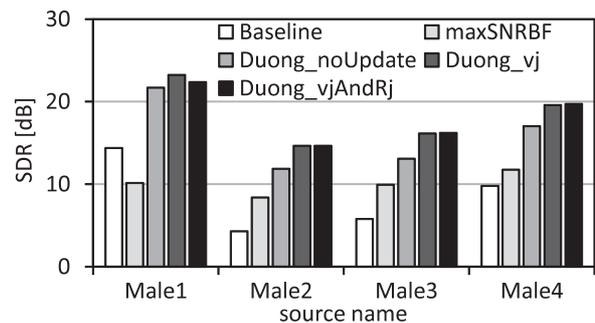


図-12 実験結果 (3. 時間同期, SDR)

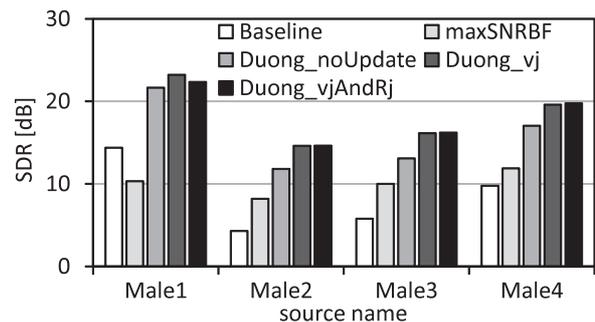


図-13 実験結果 (4. 同期録音, SDR)

5.2.2 結果・考察

非同期録音での SDR を図-10 に、録音開始時刻のみ揃えた状態での SDR を図-11 に、同期補償後の SDR を図-12 に、そして同期録音の SDR を図-13 に示す。また、非同期録音での SIR を図-14 に、録音開始時刻のみ揃えた状態での SIR を図-15 に、同期補償後の SIR を図-16 に、そして同期録音の SIR を図-17 に示す。初めに Baseline に着目すると、SDR は同期の度合

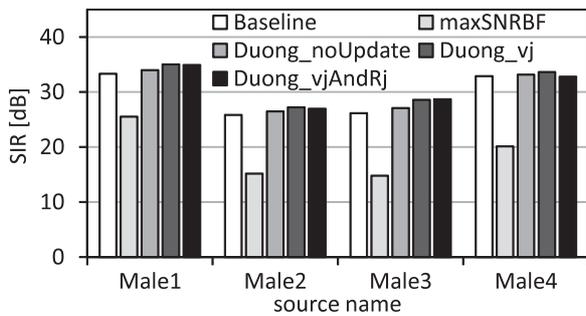


図-14 実験結果 (1. 非同期録音, SIR)

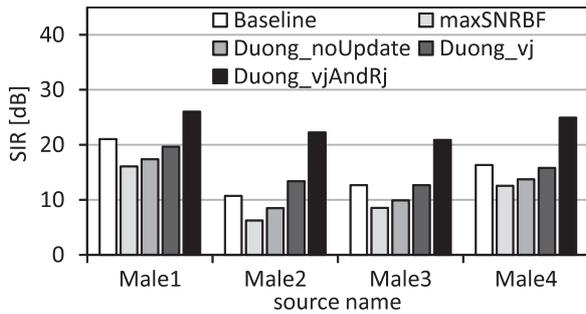


図-15 実験結果 (2. 録音開始時刻のみ同期, SIR)

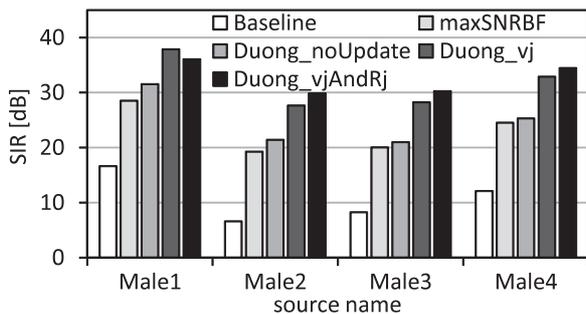


図-16 実験結果 (3. 時間同期, SIR)

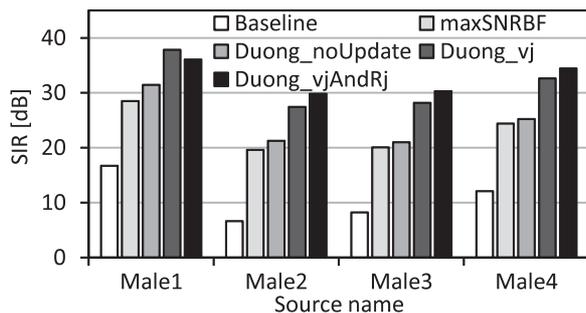


図-17 実験結果 (4. 同期録音, SIR)

いによらずほぼ一定の値を示し, SIR は同期が進むにつれて低い値を示している。これは, SDR は分離音と目的音を比較する評価値であるため, 同期の度合いによらず分離音とリファレンスはそれぞれ時間同期されているためである。しかし SIR は分離音と非目的音を比較する評価値であり, すなわち同期がなされていないほど, 分離音中の非目的音とリファレンス音の相関は下がる。非同期の状態では SIR が高いのはそのため

あり, 非同期時の分離性能として, SIR の値そのものには大して意味はなく, あくまでも各手法の Baseline と比べての改善値を示しているだけと言える。

次に, SN 比最大化ビームフォーマは録音開始時刻オフセットを補償した程度では Baseline よりも低い値を示し, 改善が見られない。サンプリング周波数ミスマッチ補償まで行うことで SDR, SIR とともに改善が見られる。従って SN 比最大化ビームフォーマの指向制御はサンプリング周波数ミスマッチに伴う音源の時間ドリフトに対応できないことが分かる。

次に, Duong 型ウィーナフィルタについての考察を行う。まず, 非同期の状態において, 3 手法に差は見られず, Baseline と比較しても改善はない。これは録音開始時刻オフセットが STFT フレーム長 16,000 点よりも長いため, チャンネル間でフレームの時刻が重ならなくなってしまい, またサンプリング周波数ミスマッチによるドリフトの影響で, フレームごとにチャンネル間時間差が異なってしまうためと考えられる。よって録音開始時刻オフセットが存在する場合, 尤度関数は無意味なものになってしまい, EM アルゴリズムによるパラメータ更新は改善も悪化ももたらすことにならず, 評価値が同程度になると言える。録音開始時刻を揃えた場合, 全パラメータ更新の Duong 型ウィーナフィルタのみわずかな改善が見られる。固定係数型ウィーナフィルタと空間相関固定の Duong 型ウィーナフィルタは非同期時と変わらず, 目立った改善は見られない。これは位相差が変化するドリフトの条件下では固定フィルタであるこれらの手法が非目的音抑圧の効果を与えないためだと考えられる。全パラメータ更新の Duong 型ウィーナフィルタは不完全な初期値ながら正確な空間相関が得られないために精度が高くないものの, 時変の分散重み $v_j(r, f)$ によるマスク効果である程度の分離が得られるため, 改善が見られると考えられる。同期補償を行った場合, 3 手法すべてに改善が見られる。固定係数型ウィーナフィルタと比べ, 他の 2 手法の方が明らかに優位であり, このことから Duong 法の EM アルゴリズムの有効性が確認できる。また, 空間相関固定の Duong 型ウィーナフィルタと全パラメータ更新の Duong 型ウィーナフィルタには目立った差が見られない。これは完全な同期を行った際, 空間相関行列はブラインドでも十分学習が可能でありノンブラインドな場合と同じ形に収束すること, また, EM アルゴリズムにより時変の分散重みが最適な値に収束するためと考えられる。このことから完全な同期を行うことで教師信号による空間相関行列はほぼ正解値を示すことが分かる。

以上より教師信号を与えられた際の音源分離では,

正確な同期補償を行えばいずれも良好な分離性能が得られることが確認できた。

6. 結 論

本論文では、非同期マイクロホンアレーを用いた音源分離のための、ミスマッチパラメータを推定する手法として、単一音源区間を用いたノンブライントな推定法を提案した。これは非同期録音の開始直後、終了前にある話者、又は音源が音を発している区間である単一音源区間を録音し、チャンネル間で単一音源区間同士の相互相関を取ることでミスマッチパラメータを推定する手法である。推定したパラメータを用いて、非整数点フレームシフトによるSTFT領域での時間同期を行った。また、従来の音源分離手法に対し非同期分散型マイクロホンへの拡張を行い、教師信号を用いた非同期分散型マイクロホンアレーにおける音源分離を行った。その結果、提案法は十分な精度の時間同期が可能であり、高い音源分離性能が得られることを確認した。

謝 辞

本論文は、科学研究費補助金基盤研究(B)(25280069)、基盤研究(A)(16H01735)、及びセコム科学技術振興財団の支援を受けた。

文 献

- [1] O. L. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, 60, 926–935 (1972).
- [2] S. Makino, T. W. Lee and H. Sawada, Eds., *Blind Speech Separation* (Springer, Berlin, 2007).
- [3] 小野順貴, 宮部滋樹, 牧野昭二, "非同期分散マイクロホンアレーに基づく音響信号処理," 音響学会誌, 70, 391–396 (2014).
- [4] 小野順貴, T.-K. Le, 宮部滋樹, 牧野昭二, "アドホックマイクロホンアレー—複数のモバイル録音機器で行う音響信号処理—," 信学会 *Fundam. Rev.*, 7, 336–347 (2014).
- [5] T. Kako, K. Niwa, K. Kobayashi and H. Ohmuro, "Wiener filter design by estimating sensitivities between distributed asynchronous microphones and sound sources," *Proc. WASPAA*, pp. 1–5 (2015).
- [6] 千葉大将, 小野順貴, 宮部滋樹, 高橋 祐, 山田武志, 牧野昭二, "アドホックマイクロホンアレーにおける時間チャンネル領域での非負値行列因子分解を用いた振幅ベースの音声強調," 音響学会誌, 72, 462–470 (2016).
- [7] 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).
- [8] E. Robledo-Arnuncio, T. Wada and B. H. Juang, "On dealing with sampling rate mismatches in blind source separation and acoustic echo cancellation," *Proc. WASPAA*, pp. 21–24 (2007).
- [9] Z. Liu, "Sound source separation with distributed microphone arrays in the presence of clock synchronization errors," *Proc. IWAENC* (2008).
- [10] N. Ono, H. Kohno, N. Ito and S. Sagayama, "Blind alignment of asynchronously recorded signals for distributed microphone array," *Proc. WASPAA*, pp. 161–164 (2009).
- [11] S. Golan, S. Gannot and I. Cohen, "Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming," *Proc. IWAENC* (2012).
- [12] S. Miyabe, N. Ono and S. Makino, "Blind compensation of inter-channel sampling frequency mismatch with maximum likelihood estimation in STFT domain," *Proc. ICASSP*, pp. 674–678 (2013).
- [13] S. Miyabe, N. Ono and S. Makino, "Blind compensation of interchannel sampling frequency mismatch for ad hoc microphone array based on maximum likelihood estimation," *Signal Process.*, 107, 185–196 (2015).
- [14] H. L. Van Trees, Ed., *Optimum Array Processing* (Wiley, New York, 2002).
- [15] S. Araki, H. Sawada and S. Makino, "Blind speech separation in a meeting situation with maximum SNR beamformers," *Proc. ICASSP*, Vol. 1, pp. 41–44 (2007).
- [16] N. Duong, E. Vincent and R. Gribonval, "Underdetermined reverberant audio source separation using a full-rank spatial covariance model," *IEEE Trans. Audio Speech Lang. Process.*, 18, 1830–1840 (2010).
- [17] R. Sakanashi, N. Ono, S. Miyabe, T. Yamada and S. Makino, "Speech enhancement with ad-hoc microphone array using single source activity," *Proc. APSIPA*, pp. 1–6 (2013).
- [18] E. Vincent, R. Gribonval and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio Speech Lang. Process.*, 14, 1462–1469 (2006).
- [19] M. Togami, "Statistical estimation theory considering time-varying nature of systems and source-probability distributions," *Ph. D. thesis*, the University of Tokyo (2011).
- [20] R. Sakanashi, S. Miyabe, T. Yamada and S. Makino, "Comparison of superimposition and sparse models in blind source separation by multichannel Wiener filter," *Proc. APSIPA*, pp. 1–6 (2012).
- [21] E. Vincent, R. Gribonval and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio Speech Lang. Process.*, 14, 1462–1469 (2006).



坂梨龍太郎

2012 筑波大・情報・情報科卒。2014 同大学大学院・シス情・CS 博士前期課程了。修士(工学)。同年株式会社リコーに入社。在学中、アレー信号処理に関する研究に従事。日本音響学会会員。



小野 順貴

2001 東大博士後期課程修了。同年 同大学助手。2005 同大学講師。2011 国立情報学研究所 准教授。アレー信号処理、音源定位、音源分離などの音響信号処理の研究に従事。博士(工学)。IEEE Senior member, 日本音響学会, 電子情報通信学会, 情報処理学会, 計測自動制御学会, 各会員。

**宮部 滋樹**

2007 奈良先端大博士後期課程了。2008 米ジョージア工科大学客員研究員。2009 東大特任研究員。2010 同大助教。2011 筑波大助教。音響信号処理の研究に従事。博士 (工学)。日本音響学会, IEEE, 電子情報通信学会, 各会員。

**山田 武志**

1999 奈良先端大博士後期課程了。同年, 筑波大学講師。現在, 同准教授。音声認識, 音環境理解, 多チャンネル信号処理, メディア品質評価, e ラーニングの研究に従事。博士 (工学)。IEEE, 電子情報通信学会, 情報処理学会, 日本音響学会, 日本語テスト学会, 各会員。

**牧野 昭二**

1981 東北大学院修士課程了。同年日本電信電話公社入社。以来, NTT 研究所において, 電気音響変換器, 音響エコーキャンセラ, ブラインド音源分離などの音響信号処理の研究に従事。工博。現在, 筑波大学生命領域学際研究センター教授。文部科学大臣表彰 (科学技術賞 研究部門), ICA Unsupervised Learning Pioneer Award, IEEE Signal Processing Society Best Paper Award 受賞。IEEE Distinguished Lecturer。IEEE Fellow。電子情報通信学会 Fellow。