

スペクトル伸縮モデルと複素正規分布音源モデルに基づく 複数マイクロホンの同期

Synchronization of multiple microphones based on spectral warping and complex Gaussian source models

糸山 克寿^{1*} 中臺 一博^{1,2}
Katsutoshi Itoyama¹ Kazuhiro Nakadai^{1,2}

¹ 東京工業大学

¹ Tokyo Institute of Technology

² (株) ホンダ・リサーチ・インスティテュート・ジャパン

² Honda Research Institute Japan Co., Ltd.

Abstract: This paper describes a method of synchronizing microphones with independent A/D converters using audio signals which come from multiple sound sources observed by the microphones. The proposed model consisting a sound source model represented by a complex Gaussian distribution, a spatial transfer model represented by a steering vector, and an observation model represented by a spectrum warping matrix is constructed. Synchronization is realized as maximum a posteriori estimation of model parameters for the observed audio signals.

1 はじめに

マイクロホンアレイは、複数のマイクロホンと A/D コンバータ、およびコンバータを同期するためのクロックを生成する発振器からなる音響信号録音装置である。マイクロホンアレイを用いた音響信号処理では、録音された信号が全チャンネルの全サンプルで同期されていることを前提に、音の伝わり方を表すステアリングベクトルや空間相関行列を用いて音源方向に基づく音源定位や音源分離などを実現する。独立した A/D コンバータをもつ複数のマイクロホンを用いた場合はマイクロホンアレイ音響信号処理を行うことはできない。同じメーカーの同じ製品であったとしても、録音された信号のサンプルレベルでの同期は仮定できないためである。

本稿では、独立した A/D コンバータをもつ複数のマイクロホンで録音された非同期音響信号の確率的生成モデルによるチャンネル間の同期および音源定位・分離手法について述べる。確率的生成モデルは、音源スペクトルの生成過程を表す音源モデル、音源からマイクロホンへの伝達過程を表す空間伝達モデル、各マイクロホンのサンプリング周波数に基づいてスペクトルを変調させるスペクトル伸縮モデルの 3 つのモデルからなる。同一の音響信号が異なるサンプリング周波数で

動作する複数のマイクロホンで録音された場合、その録音された音響信号は、サンプリング定理に基づいて各々のサンプリング周波数でサンプリングされたものとみなすことができる。リサンプリングは伸縮に相当する線形変換で近似できるため、音響信号のフーリエ変換で得られるスペクトルもまた、元のスペクトルの伸縮によって近似できる。この生成過程の逆問題を解くことにより、音響信号からサンプリング周波数と音源スペクトルが推定され、音源定位と分離が実現される。

2 関連研究

非同期分散マイクロホンアレイ [1] もしくはアドホックマイクロホンアレイ [2] に関する研究における主要な課題は (1) マイクロホン位置推定, (2) 音源位置推定および分離, (3) チャンネル間同期, である。課題 (1) はマイクロホンアレイのキャリブレーションとも呼ばれ、観測音エネルギーの差に基づく手法 [3,4], 位相差に基づく手法 [5], ビームフォーミングのクラスタリングに基づく手法 [6], 双線型写像の解に基づく手法 [7], 到達時間差 (time difference of arrival; TDOA) に基づく手法 [8], 距離行列の低ランク性を利用する手法 [9] などが報告されている。課題 (2) に関しては、TDOA に基づく手法 [10] やチャンネル間同期と同時に解く手法 [11] が報告されている。(1) と (2) は類似の定式化が行える

*連絡先: 東京工業大学 工学院 システム制御系
〒152-8552 東京都目黒区大岡山 2-12-1
E-mail: itoyama@ra.sc.e.titech.ac.jp

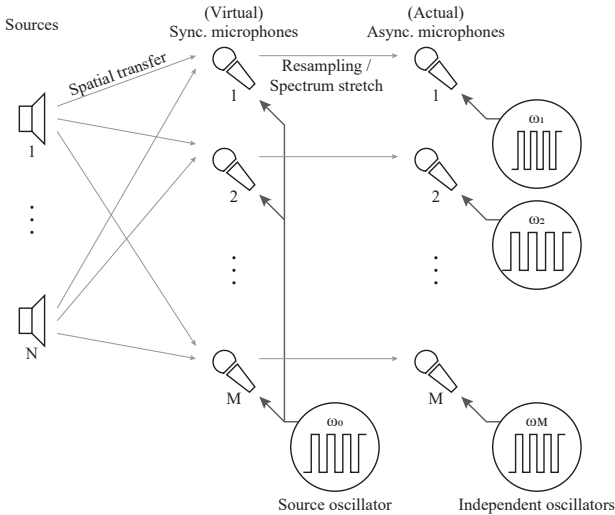


図 1: 非同期マイクロホンアレイで録音された音響信号の生成過程。複数音源からの音響信号もしくはスペクトルは空間特性にしたがった変調を受けながら各マイクロホンに到達し、サンプリング周波数と時間オフセットによってリサンプリングを受ける。

ことを利用してこれらを同時に解く手法 [3,4] や SLAM (simultaneous localization and mapping) の枠組みを用いた手法 [12,13] も報告されている。

課題 (3) に対しては、観測 TDOA に基づいて同期を行う手法 [14]、サンプリング周波数のずれが線型な位相変化を引き起こすことに基づく手法 [15] などが報告されている。また、SLAM の枠組みに基づいてサンプリング周波数とマイクロホン位置を推定する手法 [16]、サンプリング時刻のオフセットとマイクロホン位置を推定する手法 [17] が報告されている。一方で、センサネットワークの分野では、マイクロホン (センサ) 間の無線通信による同期を実現する手法が報告されている [18–22]。

従来研究の多くは、環境からの雑音が十分に小さく無視できるとの仮定の下でアレイの周囲からの拍手音などを入力とすること、すなわち単独音が常に観測されることを前提として、同期が実現できることを報告している。ただし、環境からの雑音や外来音を制御できない実際の環境では、このような同期プロセスは実用的ではない。本研究では、複数同時音源の混合音を入力としたうえでマイクロホン間の同期を実現し、音源定位および分離を行うことを目的とする。

3 手法

本節では、本稿で提案する非同期分散マイクロホンの同期手法について述べる。提案する手法の入力と出力は以下の通りである。

入力 非同期分散マイクロホンで録音された音響信号のフーリエ変換で得られるマルチチャンネルのスペクトログラム $X \in \mathbb{C}^{MTF}$

出力 各マイクロホンのサンプリング周波数 $\omega \in \mathbb{R}^M$ とオフセット $\tau \in \mathbb{R}^M$

仮定 マイクロホンアレイから各方向に対するステアリングベクトル $\mathbf{g}_{lf} \in \mathbb{C}^M$ は既知

さらに、これに基づく音源定位・分離手法についても述べる。

3.1 音源モデル

音源スペクトルを s_{ntf} で表現する。 n は音源のインデックス、 t と f は時間フレームと周波数ビンを表す。 s_{ntf} は平均がゼロ、分散が σ_s^2 の複素正規分布から生成される [23] とする。

$$s_{ntf} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_s^2). \quad (1)$$

3.2 空間伝達モデル

M をマイクロホン数、 y_{mtf} を m 番目の仮想的な同期マイクロホンで観測されたスペクトルとする。マイクロホンの位置は既知であり、周波数領域での瞬時混合過程にしたがって複数音源が観測される。 y_{mtf} は音源スペクトル y_{ntf} と n 番目の音源のステアリングベクトル $\mathbf{a}_{nf} \in \mathbb{C}^M$ を用いて以下で表現される。

$$y_{mtf} = \sum_{n=1}^N a_{nfm} s_{ntf} + \epsilon_{mtf} \quad (2)$$

ϵ_{mtf} は観測ノイズであり、複素正規分布 $\mathcal{N}_{\mathbb{C}}(0, \sigma_y^2)$ から生成されるとすると、 y_{mtf} もまた以下の複素正規分布に従って生成される。

$$y_{mtf} \sim \mathcal{N}_{\mathbb{C}}\left(\sum_{n=1}^N a_{nfm} s_{ntf}, \sigma_y^2\right). \quad (3)$$

音源ステアリングベクトル \mathbf{a}_{nf} について掘り下げる。空間は L 個の部分空間に分割されており、各部分空間を代表する方向のステアリングベクトル $\mathbf{g}_{lf} = (g_{lf1}, \dots, g_{lfM})^T$ が計測もしくは幾何計算により与えられているとする。音源はこれらの部分空間のいくつかにまたがって存在し、その存在比を $\mathbf{r}_n = (r_{n1}, \dots, r_{nL})^T$ ($0 \leq r_{nl} \leq 1, \sum_l r_{nl} = 1$) とする。この存在比 \mathbf{r}_n を用いて、音源ステアリングベクトル \mathbf{a}_{nf} を以下で定義する。

$$\mathbf{a}_{nf} = \sum_{l=1}^L r_{nl} \mathbf{g}_{lf} \quad (4)$$

存在比 \mathbf{r}_n の事前分布はディリクレ分布 $\mathcal{D}(\alpha)$ とする。

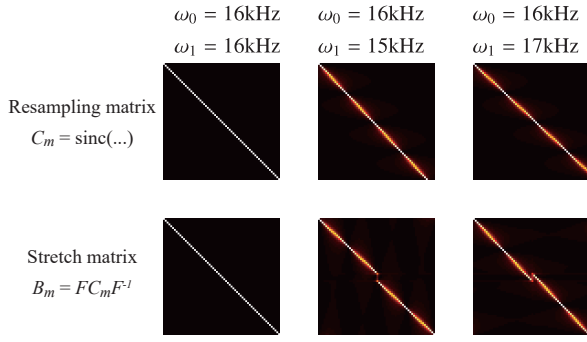


図 2: リサンプリング行列とスペクトル伸縮行列.

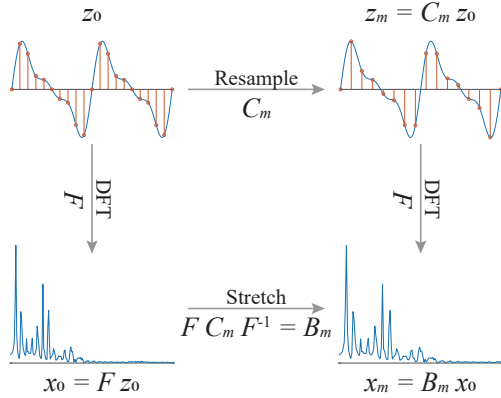


図 3: 時間領域での信号のリサンプリングと周波数領域でのスペクトルの伸縮の関係.

3.3 スペクトル伸縮モデル

サンプリングレート ω_0 で録音された長さ L の音響信号を $\mathbf{y}_0 = (y_{0,0}, \dots, y_{0,L-1})^T$, 同じ音響信号をサンプリングレート ω_1 で録音した音響信号を $\mathbf{y}_1 = (y_{1,0}, \dots, y_{1,L-1})^T$ とする. $y_{0,0}$ のサンプル時刻を 0, $y_{1,0}$ のサンプル時刻を τ_1 とする. この τ_1 はフレームのオフセットと解釈できる. 同じ音響信号を録音しているため一方はもう一方のリサンプリングで表されると考えると, \mathbf{y}_0 と \mathbf{y}_1 の間には以下の関係が成り立つ.

$$z_{mt} = \sum_{t'=1}^T \text{sinc}\left(\frac{\pi}{\omega_0}((t-1)\omega_m + \tau_m - (t'-1)\omega_0)\right) z_{0t'} \quad (5)$$

$$\text{sinc}(t) = \frac{\sin(t)}{t} \quad (6)$$

これを行列形式で記述すると以下となる.

$$\mathbf{z}_m = C_m \mathbf{z}_0 \quad (7)$$

B は以下で定義されるリサンプリング行列である.

$$C_m = \left(\text{sinc}\left(\frac{\pi}{\omega_0}(t\omega_m + \tau_m - t'\omega_0)\right) \right)_{\substack{t=0, \dots, T-1 \\ t'=0, \dots, T-1}} \quad (8)$$

\mathbf{y}_0 と \mathbf{y}_1 をそれぞれ離散フーリエ変換して得られるスペクトルを \mathbf{x}_0 と \mathbf{x}_1 とする.

$$\mathbf{x}_0 = \mathcal{F} \mathbf{z}_0, \quad \mathbf{x}_m = \mathcal{F} \mathbf{z}_m \quad (9)$$

$$\mathcal{F} = \frac{1}{\sqrt{T}} \left(e^{-\frac{2\pi i t t'}{L}} \right)_{t=0, \dots, T-1, t'=0, \dots, T-1} \quad (10)$$

したがって, 上記を合わせると以下が成り立つ.

$$\mathbf{x}_m = B_m \mathbf{x}_0, \quad B_m = \mathcal{F} C_m \mathcal{F}^{-1} \quad (11)$$

この A をスペクトル伸縮行列と定義する. 図 2 にリサンプリング行列とスペクトル伸縮行列の例を示す.

伸縮されたスペクトル \mathbf{x}_{mt} は \mathbf{y}_{mt} の線型変換であり, \mathbf{y}_{mt} は式 (3) で示すように複素正規分布に従って生成されるため, \mathbf{x}_{mt} もまた以下の複素正規分布に従って生成される.

$$\mathbf{x}_{mt} \sim \mathcal{N}_{\mathbb{C}}(\bar{\mathbf{x}}_{mt}, \sigma_y^2 B_m B_m^H) \quad (12)$$

$\bar{\mathbf{x}}_{mt}$ は音源スペクトル, ステアリングベクトル, スペクトル伸縮行列によって定まる \mathbf{x}_{mt} の予測値である.

$$\bar{\mathbf{x}}_{mt} = B_m \left(\sum_{n=1}^N a_{n1m} s_{nt1}, \dots, \sum_{n=1}^N a_{nFm} s_{ntF} \right)^T \quad (13)$$

3.4 パラメータ推定と音源定位・分離

サンプリング周波数 ω_m とオフセット τ_m の頑健な推定を実現するため, これらの事前分布を導入する.

$$\omega_m \sim \mathcal{N}(\hat{\omega}_m, \sigma_\omega^2), \quad \tau_m \sim \mathcal{N}(0, \sigma_\tau^2) \quad (14)$$

これにより, モデル全体は以下で表現される.

$$\begin{aligned} p(X, S, R, \omega, \tau) & \quad (15) \\ &= \prod_{t=1}^T \prod_{m=1}^M p(\mathbf{x}_{mt} | \mathbf{s}_{1t}, \dots, \mathbf{s}_{Nt}, R, \omega_m, \tau_m) \\ &= \prod_{n=1}^N \prod_{t=1}^T \prod_{f=1}^F p(s_{ntf}) \prod_{n=1}^N p(\mathbf{r}_n) \prod_{m=1}^M p(\omega_m) \prod_{m=1}^M p(\tau_m) \end{aligned}$$

パラメータと観測スペクトルの対数同時確率は以下で表わされる.

$$\log p(X, S, R, \omega, \tau) \quad (16)$$

$$\begin{aligned}
&= -\sum_{m=1}^M T \log |\sigma_y^2 B_m B_m^H| \\
&\quad - \sum_{t=1}^T \sum_{m=1}^M (\mathbf{x}_{mt} - \bar{\mathbf{x}}_{mt})^H (\sigma_y^2 B_m B_m^H)^{-1} (\mathbf{x}_{mt} - \bar{\mathbf{x}}_{mt}) \\
&\quad - \sum_{n=1}^N \sum_{t=1}^T \sum_{f=1}^F \frac{\|s_{ntf}\|_2^2}{\sigma_s^2} + \sum_{n=1}^N \sum_{l=1}^L \alpha \log r_{nl} \\
&\quad - \sum_{m=1}^M \frac{(\omega_m - \hat{\omega}_m)^2}{2\sigma_\omega^2} - \sum_{m=1}^M \frac{\tau_m^2}{2\sigma_\tau^2} + \text{const.}
\end{aligned}$$

サンプリング周波数 ω_m とオフセット τ_m の事後分布は正規分布を提案分布としたメトロポリス法によって推定する。音源方向 \mathbf{r}_n の事後分布はディリクレ分布を提案分布としたメトロポリス・ヘイスティングス法によって推定する。

音源スペクトル s_{ntf} の事後分布の推定、すなわち音源分離は線型フィルタリングによって実現される。スペクトル伸縮を補償する行列 \bar{B}_m はサンプリング周波数とオフセットによって以下で定まる。

$$\bar{C}_m = \left(\text{sinc} \left(\frac{\pi}{\omega_m} (t\omega_0 - \tau_m - t'\omega_m) \right) \right)_{\substack{t=0, \dots, T-1 \\ t'=0, \dots, T-1}} \quad (17)$$

$$\bar{B}_m = \mathcal{F} \bar{C}_m \mathcal{F}^{-1} \quad (18)$$

この行列を観測スペクトル \mathbf{x}_{mt} に適用することで、伸縮を補償したスペクトルが推定される。

$$\mathbf{y}_{mt} = \bar{B}_m \mathbf{x}_{mt} \quad (19)$$

さらにウィナーフィルタを適用することで、音源スペクトルの推定値 \hat{s}_{ntf} が得られる

$$\hat{s}_{ntf} = \frac{\mathbf{a}_{nf}^H}{\mathbf{a}_{nf}^H \mathbf{a}_{nf}} \bar{B}_m \mathbf{x}_{mt} \quad (20)$$

4 評価実験

提案手法を評価するためシミュレーション実験を行った。図4に示すように、半径10cmの円形4チャンネルマイクロホンアレイを用いて、マイクロホンアレイの中心から300cmの距離に、5°刻みの位置にランダムに音源を配置した。各音源の配置に対して、ステアリングベクトルを幾何的に計算した。信号は同期された状態で録音され、その後各マイクロホンのサンプリング周波数に応じてリサンプリングすることで非同期的に録音された観測音響信号を生成した。音源数 N は2、同期状態のサンプリング周波数は16kHz、フレーム長は512とした。音源スペクトルは複素正規分布から生成されたホワイトノイズを用いた。

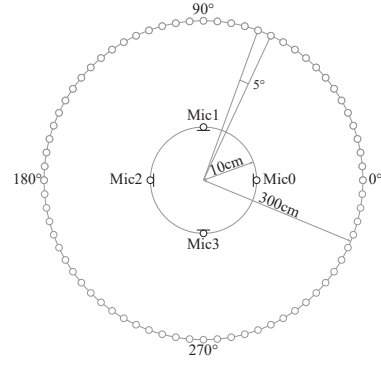


図4: 実験環境。

表1: サンプリング周波数の組み合わせ

ID	サンプリング周波数
S1	16010, 16000, 16000, 16000
S2	16010, 15990, 16000, 16000
S3	16010, 15990, 16005, 16000
S4	16010, 15990, 16005, 15980

表2: サンプリング周波数推定の実験結果。値はそれぞれ推定されたサンプリング周波数の平均値（上段）とその標準誤差（下段、括弧内）を表す。

組み合わせ	ω_0 [Hz]	ω_1 [Hz]	ω_2 [Hz]	ω_3 [Hz]
S1	16008.1	15997.4	15998.1	15997.2
	(1.81)	(1.78)	(1.83)	(2.05)
S2	16009.6	15990.5	16001.1	15997.8
	(2.35)	(3.19)	(2.05)	(2.48)
S3	16007.0	15991.0	16003.9	15996.9
	(3.53)	(3.89)	(3.01)	(3.33)
S4	16007.2	15993.4	16004.0	15980.0
	(2.74)	(2.81)	(2.00)	(4.90)

表3: 音源方向の組み合わせ

ID	音源方向
D1	0°, 60°
D2	0°, 90°
D3	0°, 120°
D4	0°, 150°

表4: 音源方向推定の実験結果。値はそれぞれ推定された音源方向誤差の平均値とその標準誤差（括弧内）を表す。

組み合わせ	誤差の平均値（標準誤差）
D1	8.42° (3.40°)
D2	11.0° (4.79°)
D3	8.48° (3.69°)
D4	8.50° (3.32°)

サンプリング周波数推定精度は、表 1 に示すサンプリング周波数の組み合わせおよびランダムに与えられたオフセットと音源方向を用いて評価した。オフセットと音源方向を既知として、メトロポリス法を用いてサンプリング周波数の事後分布に従うサンプルを生成し、事後確率が最大となるサンプルを推定値とした。

音源定位精度は、表 3 に示す音源方向の組み合わせおよびランダムに与えられたサンプリング周波数とオフセットを用いて評価した。観測スペクトルのサンプリング周波数とオフセットを既知として、音源方向の重みパラメータをメトロポリス・ヘイスティングス法で生成し、事後確率が最大となるサンプルを音源方向の推定値とした。

実験結果を表 2 および表 4 に示す。サンプリング周波数推定値の平均誤差は最大で 3.4Hz であり、提案手法の有効性を一定の範囲で示している。16kHz からずれたサンプリング周波数のマイクロホンが増えると、推定値の標準誤差が増大している。音源定位誤差は D1, D3, D4 で類似した傾向を示し、平均誤差はおよそ 8.5°であった。一方で、D2 では他の組み合わせよりも大きな誤差を示した。これは、音源からマイクロホンへの位相差がゼロとなる組み合わせが多くなったためだと考えられる。

5 おわりに

本稿では、非同期分散マイクロホンアレイでのスペクトル伸縮モデルに基づいてチャンネル間の同期と音源定位を行う手法について述べた。複数の音源スペクトルと観測スペクトルとの関係性を表すモデルを確率的生成過程に基づいて表現し、その逆問題を解くことでサンプリング周波数、オフセット、音源方向、音源スペクトルを推定する。数値実験により、提案手法がサンプリング周波数と音源方向を推定狩野であることが示された。今後の課題は、低ランクもしくは深層学習に基づく音源モデル [24, 25] を導入し音源分離を実現することである。

謝辞

本研究の一部は JSPS 科研費 16H02884, 17K00365 および 19K12017 の助成を受けた。

参考文献

[1] 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イに基づく音響信号処理. 日本音響学会誌, Vol. 70, No. 7, pp. 391–396, 2014.

[2] Alexander Bertrand, Simon Doclo, Sharon Gannot, Nobutaka Ono, and Toonvan Waterschoot. Special issue on wireless acoustic sensor networks and ad hoc microphone arrays. *Signal Processing*, Vol. 107, pp. 1–3, 2015.

[3] Zicheng Liu, Zhengyou Zhang, Li-Wei He, and Phil Chou. Energy-based sound source localization and gain normalization for ad hoc microphone arrays. In *ICASSP*, Vol. 2, pp. 761–764, 2007.

[4] Minghua Chen, Zicheng Liu, Li-Wei He, Phil Chou, and Zhengyou Zhang. Energy-based position estimation of microphones and speakers for ad hoc microphone arrays. In *WASPAA*, pp. 22–25, 2007.

[5] Marius Hennecke, Thomas Plötz, Gernot A. Fink, Jörg Schmalenströer, and Reinhold Häb-Umbach. A hierarchical approach to unsupervised shape calibration of microphone array networks. In *SSP*, pp. 257–260, 2009.

[6] Ivan Himawan, Iain McCowan, and Sridha Sridharan. Clustered blind beamforming from ad-hoc microphone arrays. *IEEE Trans. Audio, Speech, and Language Process.*, Vol. 19, No. 4, pp. 661–676, 2011.

[7] Marco Crocco, Alessio Del Bue, and Vittorio Murino. A bilinear approach to the position self-calibration of multiple sensors. *IEEE Trans. Signal Process.*, Vol. 60, No. 2, pp. 660–673, 2012.

[8] Yubin Kuang and Kalle Åström. Stratified sensor network self-calibration from TDOA measurements. In *EUSIPCO*, pp. 1–5, 2013.

[9] Mohammad J. Taghizadeh, Reza Parhizkar, Philip N. Garner, HervéBourlard, Afsaneh Asaei. Ad hoc microphone array calibration: Euclidean distance matrix completion algorithm and theoretical guarantees. *Signal Processing*, Vol. 107, pp. 123–140, 2015.

[10] A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro. Acoustic source localization with distributed asynchronous microphone networks. *IEEE Trans. Audio, Speech, and Language Process.*, Vol. 21, No. 2, pp. 439–443, 2013.

[11] Nobutaka Ono, Hitoshi Kohno, Nobutaka Ito, and Shigeki Sagayama. Blind alignment of asynchronously recorded signals for distributed microphone array. In *WASPAA*, pp. 161–164, 2009.

[12] Daobilige Su, Teresa Vidal-Calleja, and Jaime Valls Miro. Simultaneous asynchronous microphone array calibration and sound source localisation. In *IROS*, pp. 5561–5567, 2015.

[13] Kouhei Sekiguchi, Yoshiaki Bando, Katsutoshi Itoyama, , and Kazuyoshi Yoshii. Layout optimization of cooperative distributed microphone arrays based on estimation of source separation performance. *J. Robot. Mechatron.*, Vol. 29, No. 1, pp. 83–93, 2017.

[14] Fangyuan Jiang, Yubin Kuang, and Kalle Åström. Time delay estimation for TDOA self-calibration using truncated nuclear norm regularization. In *ICASSP*, pp. 3885–3889, 2013.

[15] Shigeki Miyabe, Nobutaka Ono, and Shoji Makino. Blind compensation of interchannel sampling frequency mismatch for ad hoc microphone array based on maximum likelihood estimation. *Signal Processing*, Vol. 107, pp. 185–196, 2015.

- [16] Hiroaki Miura, Takami Yoshida, Keisuke Nakamura, and Kazuhiro Nakadai. SLAM-based online calibration of asynchronous microphone array for robot audition. In *IROS*, pp. 524–529, 2011.
- [17] Keisuke Hasegawa, Nobutaka Ono, Shigeki Miyabe, and Shigeki Sagayama. Blind estimation of locations and time offsets for distributed recording devices. In *LVA/ICA*, pp. 57–64, 2010.
- [18] Mohamad Hasan Bahari, Alexander Bertrand, and Marc Moonen. Blind sampling rate offset estimation for wireless acoustic sensor networks through weighted least-squares coherence drift estimation. *IEEE/ACM Trans. Audio, Speech, and Language Process.*, Vol. 25, No. 3, pp. 674–686, 2017.
- [19] Alexander Bertrand. Applications and trends in wireless acoustic sensor networks: A signal processing perspective. In *SCVT*, pp. 1–6, 2011.
- [20] Yoshifumi Chisaki, Dan Murakami, and Tsuyoshi Usagaway. Network-based multi-channel signal processing using the precision time protocol. In *APSIPA*, pp. 1–6, 2012.
- [21] Rainer Lienhart, Igor Kozintsev, Stefan Wehr, and Minewa Yeung. On the importance of exact synchronization for distributed audio signal processing. In *ICASSP*, Vol. 4, pp. 840–843, 2003.
- [22] Joerg Schmalenstroeer and Reinhold Haeb-Umbach. Sampling rate synchronisation in acoustic sensor networks with a pre-trained clock skew error model. In *EUSIPCO*, pp. 1–5, 2013.
- [23] Takuma Otsuka, Katsuhiko Ishiguro, Hiroshi Sawada, and Hiroshi G. Okuno. Bayesian non-parametrics for microphone array processing. *IEEE/ACM Trans. Audio, Speech, and Language Process.*, Vol. 22, No. 2, pp. 493–504, 2014.
- [24] Kousuke Itakura, Yoshiaki Bando, Eita Nakamura, Katsutoshi Itoyama, Kazuyoshi Yoshii, and Tatsuya Kawahara. Bayesian multichannel audio source separation based on integrated source and spatial models. *IEEE/ACM Trans. Audio, Speech, and Language Process.*, Vol. 26, No. 4, pp. 831–846, 2018.
- [25] Julio José Carabias-Orti, Joonas Nikunen, Tuomas Virtanen, and Pedro Vera-Candeas. Multichannel blind sound source separation using spatial covariance model with level and time differences and non-negative matrix factorization. *IEEE/ACM Trans. Audio, Speech, and Language Process.*, Vol. 26, No. 9, pp. 1512–1527, 2018.