ESTIMATION OF SAMPLING FREQUENCY MISMATCH BETWEEN DISTRIBUTED ASYNCHRONOUS MICROPHONES UNDER EXISTENCE OF SOURCE MOVEMENTS WITH STATIONARY TIME PERIODS DETECTION

Shoko Araki[1]*, Nobutaka Ono*[2]*, Keisuke Kinoshita*[1]*, Marc Delcroix*[1]

[1] NTT Communication Science Laboratories, NTT Corporation,
2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
[2] Faculty of System Design, Tokyo Metropolitan University
6-6, Asahigaoka, Hino-shi, Tokyo 191-0065, Japan

ABSTRACT

In this paper, we propose a method of estimating the sampling frequency mismatch among asynchronous recording devices, even when the sources sometimes move. For a spatially stationary source, there is a method of estimating the sampling frequency mismatch, which appears in the drift of the time difference among the observed digitized signals. When the source moves, however, the change of its location also affects the drift, and the method fails to estimate the mismatch. In the meantime, looking at the practical recording situations, sources sometimes move but sometimes do not move. That is, there should be a set of time frames in which we can assume the spatial stationarity of sources, and to which we are still able to apply the sampling frequency mismatch estimation method. Based on this idea, our proposed method first detects a set of time frames where we can assume the spatial stationary by clustering the time frames using the covariance matrix of each recording device, and then estimates the mismatch by using the detected stationary time frames. Using real recordings with several IC recorders, we show that the proposed method can estimate the sampling frequency mismatch accurately even when the sources sometimes move.

Index Terms— Asynchronous microphones, distributed microphones, synchronization, moving sources, spatial stationarity

1. INTRODUCTION

In the multi-channel speech enhancement research field, many successful speech enhancement techniques with a synchronized microphone array have been proposed, e.g., [1–5]. However, it is sometimes difficult to obtain synchronous multi-channel recordings because all the microphones must be connected to the same analog-to-digital converter, which may be expensive and/or impractical in many applications. In contrast, it is easy to obtain asynchronous multi-channel recordings owing to the widespread availability of voice recording devices including smartphones. However, it then becomes difficult to apply beamforming approaches directly to such asynchronous recordings, because those techniques are severely affected by the synchronization misalignment including the sampling frequency mismatch and the difference in the recording start time among recording devices.

Signal processing for asynchronous distributed microphones has been studied in recent years [6–17], and it has been reported that well-established microphone array techniques can be employed if recordings with distributed microphones are synchronized successfully [6–9, 13–19]. However, most of these papers addressed the sampling frequency mismatch compensation only for fixed sound sources, which do not move during recording, and the sampling frequency mismatch estimation and compensation for moving-source recordings still remains as an open problem. If the sampling frequency mismatch is precisely estimated, we can compensate it with, e.g., recently proposed mismatch compensation methods [20, 21].

In this paper, we focus on a blind synchronization method proposed by Miyabe et al. [8, 13] which has been proposed for a fixed source, and we propose a way to apply it to a case where sources sometimes move. When a source does not move, the sampling mismatch between two recording devices causes a drift of the time difference among their observed digitized signals. This method assumes this drift to be constant within each time frame but to vary proportionally to the time frame index, and estimates the sampling frequency mismatch in the short time Fourier transform (STFT) domain. This method also assumes the spatial stationarity of sound sources and works reasonably well when the source locations are fixed during the recording [8, 13]. If the source moves, however, this movement also causes a drift of the time difference among the recordings, making it difficult to distinguish between the sampling frequency mismatch and the change in the source location. As a result, this method cannot estimate the sampling frequency mismatch for moving sources.

Because the conventional method [8,13] assumes the spatial stationarity of the sources, even if sources sometimes move, we may still be able to employ it if we can obtain the time frames where stationarity assumption is hold. Looking at the practical recordings, sources sometimes move but sometimes do not move (e.g., speakers at a conversation at a meeting), and there should be the time frames where the spatial stationarity assumption is hold. Considering such a situation in real recordings, we first attempt to detect the time frames where the source can be assumed to be stationary. Here, we assume that each recording device has at least two channels, and we propose a method for clustering the time frames by assuming that the covariance matrix of each recording device has stationary periods. It will be shown that, by applying the conventional synchronization method to the detected stationary time frames, we can estimate the sampling frequency mismatch among devices even under the existence of source movements during recording.

2. PROBLEM DESCRIPTION

Suppose we have two stereo recording devices with slightly different sampling frequencies. Let $\mathbf{x}_1[t] = [x_{1L}[t], x_{1R}[t]]^T$ and $\mathbf{x}_2[t] = [x_{2L}[t], x_{2R}[t]]^T$ be their *continuous* time domain observations at the left and right channels. For simple notation, let x_{1B}

denote x_{1L} or x_{1R}, and define x_{2B} similarly. Assuming that the sampling frequency of each device is time-invariant, their *discretized* signals $x_{1B}(t)$ and $x_{2B}(t)$ (B=L or R) are modeled as follows.

$$x_{1B}(t) = x_{1B}\left[\frac{t}{f_s}\right], \quad (1)$$

$$x_{2B}(t) = x_{2B}\left[\frac{t}{(1+\epsilon)f_s} + T\right], \quad (2)$$

where T and ϵ are the parameters that represent the offset time (= difference between the recording start times) and the sampling frequency mismatch between the two devices, respectively.

The objective of this paper is to estimate the sampling frequency mismatch parameter ϵ from the two stereo recordings $x_{1B}(t)$ and $x_{2B}(t)$, even in a dynamic recording condition where sound sources may move. Here, we assume that the offset time T has already been compensated by, e.g., finding the peak of the cross-correlation between $x_{1B}(t)$ and $x_{2B}(t)$ [13].

3. BLIND SYNCHRONIZATION: REVIEW

To estimate ϵ, in this paper, we employ the blind synchronization technique proposed in [8, 13]. Here, we review this technique in considering to apply it to two stereo devices.

First, we obtain the STFT domain representation of $x_{2B}(t)$ as

$$X_{2B}(f,n) = \sum_{l=0}^{L-1} w(l) x_{2B}(l+n-\frac{L}{2}) \exp\left(-\frac{\jmath 2\pi f l}{L}\right), \quad (3)$$

where $w(l)$ is a window function of the length L, f is the discrete frequency index, n is the central sample of the analysis time frame, and $\jmath = \sqrt{-1}$. Then, approximating the time difference between channels caused by the sampling frequency mismatch ϵ is constant within a time frame, the sampling frequency mismatch is compensated by a linear phase shift in the STFT domain as follows:

$$\hat{X}_{2B}(f,n;\epsilon) = X_{2B}(f,n) \exp\left(\frac{\jmath 2\pi f n \epsilon}{L}\right). \quad (4)$$

If all the sources do not move and stationary, the synchronized multi-channel observation

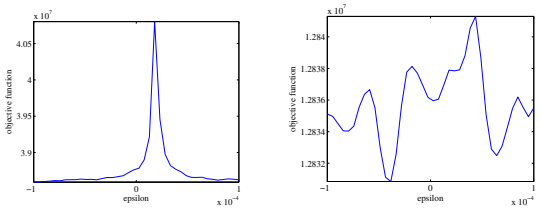
$$\mathbf{Y}(f,n;\epsilon) = [X_{1L}(f,n), X_{1R}(f,n), \hat{X}_{2L}(f,n;\epsilon), \hat{X}_{2R}(f,n;\epsilon)]^T \quad (5)$$

is also regarded as being stationary. Therefore, we assume that the compensated observation vector $\mathbf{Y}(f,n;\epsilon)$ with an accurate ϵ follows the zero-mean multivariate normal distribution. The log-likelihood function is given by

$$J(\epsilon) = \sum_{f,n} [-\log\det(\pi \mathbf{V}_{\mathbf{Y}}(f)) - \mathbf{Y}^H(f,n;\epsilon)\mathbf{V}_{\mathbf{Y}}^{-1}(f)\mathbf{Y}(f,n;\epsilon)] \quad (6)$$

$$= -|\forall n| \sum_f [D(1+\log\pi) + \log\det \mathbf{V}_{\mathbf{Y}}(f)], \quad (7)$$

where $\mathbf{V}_{\mathbf{Y}}(f) = \sum_n \mathbf{Y}(f,n;\epsilon)\mathbf{Y}^H(f,n;\epsilon)/|\forall n|$ is the maximum likelihood estimation of the spatial covariance matrix, D is the dimension of $\mathbf{Y}$ (therefore $D=4$ here), and $|\forall n|$ is the number of frames. The sampling frequency mismatch ϵ is estimated by maximizing this log-likelihood function. Since the values of estimation of ϵ that maximizes the likelihood cannot be obtained analytically, we use a golden-section search approach [8, 13].



(a) $\hat{\epsilon}$=18.5 (ppm) (b) $\hat{\epsilon}$=41.0 (ppm)

Fig. 1. Example plots of J for (a) a fixed source and (b) a moving source (Dynamic 1 scenario in Sec. 5.1). The oracle is ϵ_o=18.4 (ppm), and the estimated ($\hat{\epsilon}$) values of ϵ are shown under the plots.

This approach assumes the spatial stationarity of the sources, and it works very well for a fixed source. Figure 1 (a) shows an example plot of J for a fixed source. We can see that the log-likelihood J is monomodal and the method can accurately estimate ϵ values. On the other hand, when the source moves, the assumption of spatial stationarity no longer holds. Therefore, the shape of the log-likelihood function collapses (Fig. 1 (b)), and we cannot accurately estimate ϵ value. This is the motivation of this paper.

4. PROPOSED APPROACH TO APPLY BLIND SYNCHRONIZATION TO DYNAMIC CONDITIONS

In this section, we explain how to handle dynamic cases. We first find a set of time frames for which we can assume spatial stationarity, and then estimate ϵ by using only these found time frames.

Let $\mathbf{X}_d(f,n) = [X_{dL}(f,n)\ X_{dR}(f,n)]^T$ be the stereo observation of the dth microphone ($d=1$ or 2) at (f,n) time frequency bin. We classify each time frame into several classes such that $\mathbf{X}_d(f,n)$ can be more spatially stationary in each class. It is expected that each class would represent for example, a fixed source, a stationary mixture of fixed sources, a source activity when a moving source is located at a certain position, background noise, and so on. To obtain such classes, we assume that $X_d(f,n)$ follows a complex multivariate distribution with zero mean and a spatial covariance matrix $\mathbf{V}_{dk}(f)$ depending on the class. Then, the log-likelihood when a time frame n belongs to a class k can be written as

$$\mathcal{L}_c(n;k) = \sum_d \sum_f [-\log\det(\pi \mathbf{V}_{dk}(f)) - \mathbf{X}_d^H(f,n)\mathbf{V}_{dk}^{-1}(f)\mathbf{X}_d(f,n)]. \quad (8)$$

Equation (8) is similar to eq.(6), however, eq.(8) consists of only the covariance matrices of the stereo observation at each device. Therefore, the sampling frequency mismatch between the devices does not affect eq.(8).

Thus, we have the objective function

$$\mathcal{L} = \sum_k \sum_{n \in C_k} \mathcal{L}_c(n;k), \quad (9)$$

where C_k represents the set of indices of time frames that belong to the class k. The objective function in terms of $\mathbf{V}_{dk}$ and C_k can be increased and ideally maximized by applying the following update rules iteratively.

Update of spatial covariance matrix of each class

$$\mathbf{V}_{dk}(f) = \frac{1}{|\forall C_k|} \sum_{n \in C_k} \mathbf{X}_d(f,n)\mathbf{X}_d^H(f,n) \quad (10)$$

Update of each class

$$C_k = \{n \mid \mathcal{L}(n;k) \geq \mathcal{L}(n;k')\ ^\forall k' \neq k\} \quad (11)$$

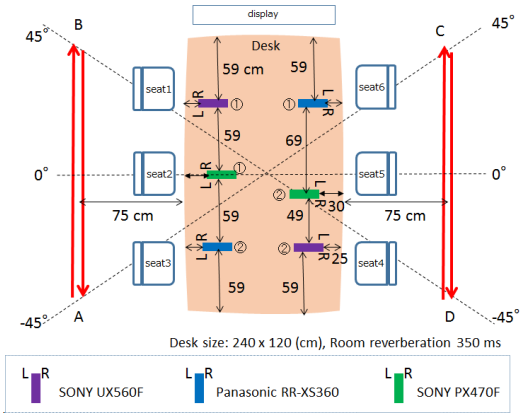


Fig. 2. Recording setup.

After the convergence of this iteration, we choose the most stationary class k by the maximum likelihood criterion as

$$k = \arg\max_{k'} \sum_{n \in C_{k'}} \mathcal{L}_c(n; k'). \tag{12}$$

Then, by maximizing eq.(6) over not all the time frames but only the time frames $n \in C_k$, we can estimate the sampling frequency mismatch ϵ even under a dynamic source condition.

5. EXPERIMENTS

5.1. Recording setup

Our recording was conducted in a small office with reverberation time of 350 ms. There was low level of room noise from, e.g., the personal computers and air conditioners. As the recording devices, we utilized three types of IC recorders, Panasonic RR-X360 ("pana" for short), Sony ICD-PX470F ("sony47"), and Sony ICD-UX560F ("sony56"). We used six IC recorders in total, and their positions and orientations are shown in Fig. 2. The sampling rate was 44,100 Hz.

We recorded speech signals for following three scenarios:

1. Fixed scenario: One speaker was seated and spoke about 1 minute at seat $s (s = 1, \cdots, 6)$ (see Fig. 2).
2. Dynamic 1 scenario: One speaker spoke while walking. The speaker followed one of the following one way routes: A→B, B→A, C→D, or D→C (see Fig. 2).
3. Dynamic 2 scenario: One speaker spoke while walking. The speaker made two round-trips of one of the following routes: A→B→A→B→A, B→A→B→A→B, C→D→C→D→C or D→C→D→C→D (see Fig. 2).

We employed one male speaker and one female speaker for all the recording scenarios.

5.2. Experimental conditions

We estimated ϵ for each scenario and each IC recorder, and compared the estimated values with the oracle ϵ values. To estimate the oracle values of ϵ_0 for each IC recorder, we used a time marking at the start and end timings of the recordings. For the time marking, we played a chirp signal (a time-stretched pulse of 0–7 kHz with the duration of 1 sec.). From the number of samples between the start and end points, we can obtain the oracle ϵ. The reference device $\mathbf{x}_1$ in this paper was sony56①.

We compared the estimation error for the conventional method with that for the proposed methods. As the conventional method, we used the entire data for each recording, and applied the original blind synchronization approach [8,13]. As the proposed method, we used the clustering method described in Sec. 4, where the number of classes was 5[1]. For both methods, the frame length and frame shift were 4096 and 2048, respectively. We used the following evaluation measures:

- Average value of the estimated ϵ for each IC recorder,
- Absolute error a defined by $a = |\epsilon\_***-\epsilon\_\text{oracle}|$, where *** denotes conv (conventional) or prop (proposed). We counted the cases where the error $a > 1$ (ppm) for each scenario.

5.3. Results

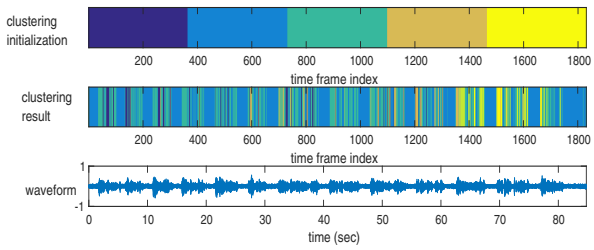
First, we show the clustering result with our proposed method for the Fixed and Dynamic 1 scenarios. Figures 3 and 4 show (0) the clustering result of the time frames in each recording, and (i)–(v) the log-likelihood function $J(\epsilon)$ in the blind synchronization in each class. In the Fixed scenario, Fig. 3 (0), the second (light blue) and third (light green) clusters correspond to the noise and voice segments, respectively. The clustering log-likelihood for these five classes was $\sum_{n \in C_k} \mathcal{L}_c(n; k) = -7, 4e5, 3.7e6, -1.3e6, -8.2e5, -6.1e5$ $(k = 1, \cdots, 5)$, and the second class, which had the maximum likelihood, was selected in this case. From the shape of the likelihood functions shown in Figs. 3 (i)–(v), it appears that we can obtain reasonable ϵ values from all classes for the Fixed scenario. Actually, for the oracle $\epsilon_0 = -3.92$ ppm, the estimated ϵ values for classes 1–5 were -3.62, -3.73, -3.61, -3.64 and -2.97 (ppm), respectively. Next, we move to the dynamic scenario, which is shown in Fig. 4. From Fig. 4 (0), we can see that the fifth cluster (yellow) corresponded to the noise period, and no class corresponded to the voice segment due to the movement of the source. The clustering log-likelihood for these five classes was $\sum_{n \in C_k} \mathcal{L}_c(n; k) = -1.5e5, 1.9e4, -2.1e5, -1.4e5, 5.7e5$ $(k = 1, \cdots, 5)$, and the fifth class, which has the maximum likelihood, was selected in this case. The estimated ϵ from the fifth class was -3.90 ppm, where the oracle was $\epsilon_0 = -3.92$ ppm. It might be worth mentioning that our proposed approach selected quite often the noise class, that provided accurate estimation of ϵ as discussed in the next paragraph. Using the noise regions might be reasonable practically when the source is always moving, because we can expect spatially stationary noise sources in real recordings.

Tables 1–3 show the summaries of the evaluation results for each IC recorder for the Fixed, Dynamic 1, and Dynamic 2 scenarios, respectively. In the tables, *_ora, *_conv, and *_prop denote the values for the oracle, conventional method, and proposed method, respectively. With the conventional method, the ϵ values were estimated accurately for the Fixed scenario (Table 1). However, for the dynamic scenarios, the conventional method could no longer estimate the correct ϵ values, especially for the Dynamic 1 scenario. On the other hand, the proposed method almost always worked reasonably well for both Fixed and Dynamic conditions. The estimation error of the Dynamic 2 scenario is almost the same as that of the Fixed scenario, and even for the Dynamic 1 scenario, the proposed method estimated ϵ values with small errors.

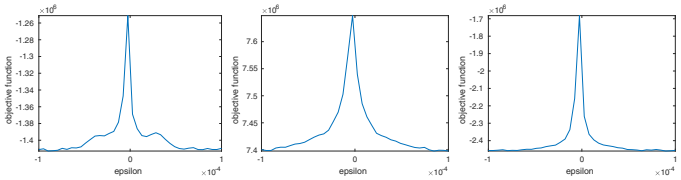
6. CONCLUSION

In this paper, we proposed a method of estimating the sampling frequency mismatch among asynchronous recording devices, even

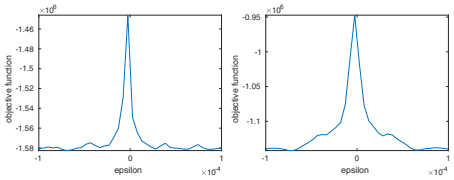
[1]We also tried 3 and 4 classes, and confirmed that they also worked.



(0) Clustering initialization, clustering result and waveform at channel L.

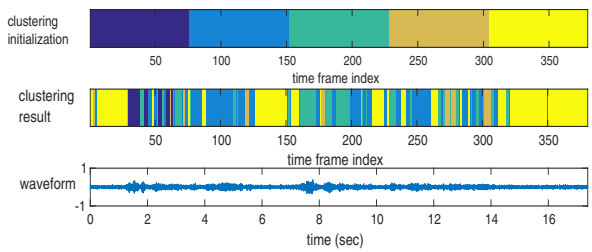


(i) J for class 1. (ii) J for class 2. (iii) J for class 3.

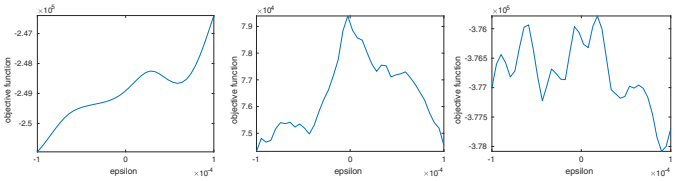


(iv) J for class 4. (v) J for class 5.

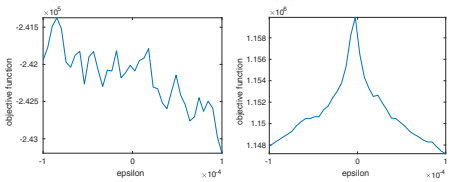
Fig. 3. Examples of (0) Clustering result, and (i-v) their log-likelihood $J(\epsilon)$ for a Fixed scenario. The log-likelihood for clustering of time frames $\sum_{n \in C_k} \mathcal{L}_c(n;k) = -7.4e5, +3.7e6, -1.3e6, -8.2e5, -6.1e5$ (class 2 was selected).



(0) Clustering initialization, clustering result and waveform at channel L.



(i) J for class 1. (ii) J for class 2. (iii) J for class 3.



(iv) J for class 4. (v) J for class 5.

Fig. 4. Examples of (0) Clustering result, and (i-v) their log-likelihood $J(\epsilon)$ for a Dynamic 1 scenario. The log-likelihood for clustering of time frames $\sum_{n \in C_k} \mathcal{L}_c(n;k) = -1.5e5, +1.9e4, -2.1e5, -1.4e5, +5.7e5$ (class 5 was selected).

when the sources sometimes move. The proposed method first finds the stationary time frames by clustering the time frames using the covariance matrix of each recording device, and then applies the blind synchronization to the found time frames to estimate the sampling frequency mismatch. We confirmed that the proposed clustering method can find the stationary time frames required for blind synchronization, and that the proposed method can estimate the sampling frequency mismatch accurately. Our future work includes the application of our proposed method to asynchronous distributed microphone array processing, such as, speech enhancement or speech separation with multiple asynchronous recording devices.

Acknowledgements: This work was partially supported by a Grant-in-Aid for Scientific Research (A) (KAKENHI Grant Number 16H01735) from Japan Society for the Promotion of Science (JSPS). We thank Dr. Emmanuel Vincent for the preliminary discussion of this work.

Table 1. Estimated ϵ and absolute errors for Fixed source scenario,

ICrecorder	pana①	pana②	sony47①	sony47②	sony56②
ϵ_ora	-3.925	18.424	11.400	-0.540	-1.109
ϵ_conv	-4.006	18.500	11.449	-0.518	-1.025
ϵ_prop	-3.907	18.463	11.364	-0.523	-1.154
a_conv	0/12	0/12	0/12	1/12	1/12
a_prop	0/12	0/12	0/12	0/12	0/12

Table 2. Estimated ϵ and absolute errors for Dynamic 1 scenario,

ICrecorder	pana①	pana②	sony47①	sony47②	sony56②
ϵ_ora	-3.925	18.424	11.400	-0.540	-1.109
ϵ_conv	13.443	-10.111	9.548	3.689	-5.539
ϵ_prop	-3.985	18.221	10.736	-0.356	-1.231
a_conv	6/8	8/8	8/8	8/8	8/8
a_prop	0/8	0/8	1/8	0/8	0/8

Table 3. Estimated ϵ and absolute errors for Dynamic 2 scenario,

ICrecorder	pana①	pana②	sony47①	sony47②	sony56②
ϵ_ora	-4.586	18.159	11.224	-0.661	-1.334
ϵ_conv	2.822	19.334	9.643	-0.905	-3.703
ϵ_prop	-4.598	18.024	11.201	-0.702	-1.281
a_conv	2/4	3/4	2/4	1/4	1/4
a_prop	0/4	0/4	0/4	0/4	0/4

7. REFERENCES

[1] M. Brandstein and D. Ward, Eds., *Microphone Arrays*, Springer, 2001.

[2] S. Makino, Ed., *Audio Source Separation*, Springer, 2018.

[3] E. Vincent, T. Virtanen, and S. Gannot, Eds., *Audio Source Separation and Speech Enhancement*, Wiley, 2018.

[4] J. Barker, R. Marxer, E. Vincent, and S. Watanabe, "The third 'CHiME' speech separation and recognition challenge: Analysis and outcomes," *Computer Speech and Language*, vol. 46, no. 2017, pp. 605–626, 2016.

[5] E. Vincent, S. Watanabe, J. Barker, and R. Marxer, "The 4th CHiME speech separation and recognition challenge," 2016, http://spandh.dcs.shef.ac.uk/chime_workshop/presentations/CHiME_2016_Vincent_overview.pdf.

[6] S. Wehr, W. Kellermann, R. Lienhart, and I. Kozintsev, "Synchronization of acoustic sensors for distributed ad-hoc audio networks and its use for blind source separation," in *IEEE Sixth International Symposium on Multimedia Software Engineering (IEEE-MSE2004)*, Dec 2004, pp. 18–25.

[7] S. Markovich-Golan, S. Gannot, and I. Cohen, "Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming," in *Proc. of IWAENC 2012*, 2012.

[8] S. Miyabe, N. Ono, and S. Makino, "Blind compensation of inter-channel sampling frequency mismatch with maximum likelihood estimation in STFT domain," in *ICASSP2013*, 2013, pp. 674–678.

[9] R. Sakanashi, N. Ono, S. Miyabe, T. Yamada, and S. Makino, "Speech enhancement with ad-hoc microphone array using single source activity," in *Proc. IPSIPA 2013*, 2013.

[10] Y. Uezu, K. Kinoshita, M. Souden, and T. Nakatani, "On the robustness of distributed EM based BSS in asynchronous distributed microphone array scenarios," in *Proc. of Interspeech 2013*, 2013, pp. 3298–3301.

[11] H. Chiba, N. Ono, S. Miyabe, Y. Takahashi, T. Yamada, and S. Makino, "Amplitude-based speech enhancement with nonnegative matrix factorization for asynchronous distributed recording," in *Proc. of IWAENC 2014*, 2014, pp. 204–208.

[12] M. Souden, K. Kinoshita, M. Delcroix, and T. Nakatani, "Location feature integration for clustering-based speech separation in distributed microphone arrays," *IEEE Trans. Audio, Speech and Language Processing*, vol. 22, no. 2, pp. 354–367, 2014.

[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 Processing*, vol. 107, pp. 185–196, 2015.

[14] L. Wang and S. Doclo, "Correlation maximization-based sampling rate offset estimation for distributed microphone arrays," *IEEE Trans. Audio, Speech and Language Processing*, vol. 24, no. 3, pp. 571–582, March 2016.

[15] D. Cherkassky and S. Gannot, "Blind synchronization in wireless acoustic sensor networks," *IEEE Trans. Audio, Speech and Language Processing*, vol. 25, no. 3, pp. 651–661, 2017.

[16] M. H. Bahari, A. Bertrand, and M. Moonen, "Blind sampling rate offset estimation for wireless acoustic sensor networks through weighted least-squares coherence drift estimation," *IEEE Trans. Audio, Speech and Language Processing*, vol. 25, no. 3, pp. 674–686, 2017.

[17] R. M. Corey and A. C. Singer, "Speech separation using partially asynchronous microphone arrays without resampling," in *Proc. IWAENC2018*, Sept. 2018, pp. 111–115.

[18] K. Ochi, N. Ono, S. Miyabe, and S. Makino, "Multi-talker speech recognition based on blind source separation with ad hoc microphone array using smartphones and cloud storage," in *Proc. of Interspeech2016*, 2016, pp. 3369–3373.

[19] S. Araki, N. Ono, K. Kinoshita, and M. Delcroix, "Meeting recognition with asynchronous distributed microphone array," in *Proc. of ASRU2017*, 2017, pp. 32–39.

[20] J. Schmalenstroeer and R. Haeb-Umbach, "Efficient sampling rate offset compensation - an overlap-save based approach," in *Proc. EUSIPCO2018*, Sept. 2018, pp. 504–508.

[21] A. Chinaev, P. Thune, and G. Enzner, "Low-rate farrow structure with discrete-lowpass and polynomial support for audio resampling," in *Proc. EUSIPCO2018*, Sept. 2018, pp. 480–484.