Synchronization of Signals with Sampling Rate Offset and Missing Data Using Dynamic Programming Matching

Hayato Takeuchi*, Takao Kawamura*, Nobutaka Ono* and Shoko Araki†
* Tokyo Metropolitan University, Tokyo, Japan
† NTT Corporation, Japan

***Abstract*—In this paper, we propose a blind synchronization method for signals with sampling rate offset (SRO) and missing data, which occasionally occurs in distributed recording for acoustic scene classification. In our method, the correspondence between short-time frames is first estimated using cross-correlation and dynamic programming (DP) matching. Then, two methods for producing synchronized signals are compared. The first method is based on the overlap-add along the DP path, while the second method uses the DP path only to identify missing data positions and compensates for the SRO with a linear phase model. The performance of these methods is evaluated through experiments. The results are promising, and further applications to acoustic scene classification are expected.**

I. Introduction

Distributed microphone array consists of multiple independent recording devices such as smartphones, voice recorders, and notebook PCs, which do not require wired connections and allow for flexible placement of devices [1]. Furthermore, the use of devices with communication capabilities increases the convenience of the system [2]. In the distributed microphone array, we can use spatial information obtained from these devices for blind source separation [3], [4], source localization [5], speech activity detection [6], and acoustic scene classification [7]–[10].

However, when recording with different devices, recorded signals are not synchronized due to a sampling time offset (STO) and sampling rate offset (SRO). This may occur because the acoustic signal is converted to a digital signal by the AD converters in each device. In acoustic scene classification (ASC), this lack of synchronization degrades processing performance because time difference information plays an important role [7]–[9]. In addition, when there is also missing data in multi-channel signals due to microphone malfunctions, packet loss in network errors, and faulty connections of the microphone cable, ASC performance is significantly degraded because the multi-channel signals with missing data differ from those without missing data [11], [12].

Many blind synchronization methods have been proposed using only recorded signals [13]–[16]. On the other hand, a synchronization method that can cope with various types of missing data is desirable, although a study [17] that performs online SRO estimation based on a probability model of packet loss and a study [18] that detects the presence of missing data using CNNs have been proposed.

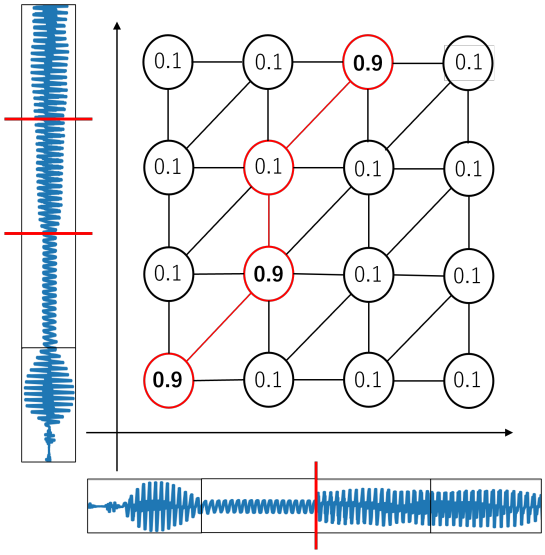

Fig. 1. Conceptual diagram of correspondence estimation based on DP matching. Each node indicates a cross-correlation of short-time frame signals, and the path represents the estimated correspondence. Note that the actual matching is performed on an overlapped-frame sequence.

Synchronization of asynchronous signals is generally performed using frame-by-frame correlations between signals, but under conditions where signals include missing data, it is not possible to simply add and accumulate the correlations between frames. For periodic signals, dynamic addition using their periodicity has been proposed [19], but this method cannot be used directly when the target is arbitrary.

In this study, we propose a new method to synthesis synchronous signals by blindly estimating the frame-by-frame correspondence of asynchronous signals including SROs and missing data by combining frame-by-frame correlation between signals and dynamic programming (DP) matching [20] (see Fig. 1). We confirm the effectiveness of the proposed method through evaluation experiments.

II. Related Works

Several blind SRO estimation methods using only recorded signals without prior information have been proposed [13]–

[16], [21], [22]. In these methods, SRO is estimated using the methods based on Linear phase drift (LPD) model. The methods based on coherence drift (CD) calculate the complex coherence between two consecutive time frames in frequency [21]. In the correlation maximization (CM)-based methods [14], the optimal SRO is searched to maximize the correlation between a pair of signals after resampling. There are also methods based on maximum likelihood (ML) [13]. Meanwhile, online SRO estimation methods assuming a statistical model for partial missing samples in recorded signals have been proposed [17].

III. Problem Settings

We consider the problem of synchronizing two discrete signals, $x[n]$ and $y[n]$. We assume that these signals capture the same audio signals but are not synchronized, specifically due to the time drift caused by sample rate offset (SRO) and partial data loss caused by packet loss or other reasons. The problem here is to produce $\hat{y}[n]$, which is synchronized to $x[n]$, from $y[n]$ by stretching and compressing the time axis.

Even if SRO or missing samples occur, there should be pairs of short-time frames in the two signals that have correlation. Then, we consider synchronizing two signals $x[n], y[n]$ by estimating the correspondence between short-time frame signals $x_i[m](i = 0, \ldots, I-1)$ and $y_j[m](j = 1, \ldots, J-1)$, which are split from $x[n], y[n]$ with frame length L and frame shift S. The variables i and j are short time frame indices, and $m = 0, \ldots, L-1$ is the local discrete-time index in the frame. The relation between the short-time frame signal and the original signal is expressed using an analysis window $w_a[m]$ of length L as follow:

$$x_i[m] = w_a[m]x[iS + m]. \tag{1}$$

The frame signal $y_j[m]$ is calculated in the same way as $x_i[m]$.

IV. Proposed Method

A. Estimation of correspondence between short-time frames

1) DP matching: DP matching [20] is a well-known method for finding the correspondence between two sequences that include insertions, deviations, stretches, and contractions. In this study, we consider the use of DP matching for synchronizing $x_i[m]$ and $y_j[m]$ in the presence of missing or SROs. Specifically, we seek the path (i_k, j_k) that maximizes the cumulative score D calculated using the node score $d_{i,j}$ and the edge score $e_{(i_{k-1},j_{k-1}),(i_k,j_k)}$. When the optimal path contains (i_k, j_k), $x_{i_k}[m]$ and $y_{j_k}[m]$ are in correspondence. The score D is given as

$$D = \sum_{k=1}^{K} d_{i_k,j_k} + e_{(i_{k-1},j_{k-1}),(i_k,j_k)}, \tag{2}$$

where K is the optimal path length, $(i_0, j_0) = (0, 0)$ and $(i_K, j_K) = (I-1, J-1)$, respectively. The k-th correspondence (i_k, j_k) satisfies the following constraints (see Fig. 2).

$$\begin{aligned}(i_k, j_k) \in \{&(i_{k-1}+1, j_{k-1}),\\ &(i_{k-1}, j_{k-1}+1),\\ &(i_{k-1}+1, j_{k-1}+1)\}.\end{aligned} \tag{3}$$

Algorithm 1 Algorithm of Dynamic Programming Matching

Input: d, e, I, J
Output: $(i_k, j_k)_{k=0}^{K}$

$D_{0,0} \leftarrow d_{0,0}$
for $i = 0, \ldots, I-1$ **do**
  for $j = 0, \ldots, J-1$ **do**
    if $(i, j) \neq (0, 0)$ **then**
      $\Lambda \leftarrow \{(i-1, j), (i, j-1), (i-1, j-1)\}$
      $(\hat{i}, \hat{j}) \leftarrow \underset{(i',j') \in \Lambda, i' \geq 0, j' \geq 0}{\arg\max} D_{i',j'} + e_{(i',j'),(i,j)}$
      $D_{i,j} \leftarrow D_{\hat{i},\hat{j}} + d_{i,j} + e_{(\hat{i},\hat{j}),(i,j)}$
      $\mathrm{prev}_{i,j} \leftarrow (\hat{i}, \hat{j})$
    end if
  end for
end for
$(i, j) \leftarrow (I-1, J-1)$
$K \leftarrow 0$
while $(i, j) \neq (0, 0)$ **do**
  $(i, j) \leftarrow \mathrm{prev}_{i,j}$
  $K \leftarrow K + 1$
end while
$(i_K, j_K) \leftarrow (I-1, J-1)$
for $k = K-1, \ldots, 0$ **do**
  $(i_k, j_k) \leftarrow \mathrm{prev}_{i_{k+1},j_{k+1}}$
end for

The algorithm for DP matching is shown in Algorithm 1. The inputs are all node scores d, all edge scores e, and the number of frames I, J for each signal. The output is a path $(i_k, j_k)_{k=0}^{K}$ that maximizes Eq. (2).

2) Node scores: The node score of $x_i[m]$ and $y_j[m]$ is the similarity between their frames. The variables $d_{i,j}$ and $u_{i,j}$ are the maximum value (node score) and the position of maximum value (time difference) of the normalized cross-correlation function $c_{i,j}[\tau]$, respectively.

$$d_{i,j} = \max_{0 \leq |\tau| \leq \frac{S}{2}} \{c_{i,j}[\tau]\}, \tag{4}$$

$$u_{i,j} = \operatorname*{argmax}_{0 \leq |\tau| \leq \frac{S}{2}} \{c_{i,j}[\tau]\}, \tag{5}$$

$$c_{i,j}[\tau] = \frac{\sum_{m=0}^{L-1} x_i[\tau + m] y_j[m]}{\sqrt{\sum_{m=0}^{L-1} x_i^2[m]} \sqrt{\sum_{m=0}^{L-1} y_j^2[m]}}. \tag{6}$$

The reason why the range of $|\tau|$ is restricted to $S/2$ is that if $|\tau|$ exceeds $S/2$, the target frame is considered to correspond to another frame. The frame signal $x_i[m]$ is defined for any integer a as

$$x_i[m] = x_i[m + aL]. \tag{7}$$

3) Edge scores: Edge scores are given as

$$e_{(i_{k-1},j_{k-1}),(i_k,j_k)} = \begin{cases} 0 & \text{if } (i_k, j_k) = (i_{k-1}+1, j_{k-1}), \\ 0 & \text{if } (i_k, j_k) = (i_{k-1}, j_{k-1}+1), \\ p & \text{if } (i_k, j_k) = (i_{k-1}+1, j_{k-1}+1). \end{cases} \tag{8}$$

The score of $(i_k, j_k) = (i_{k-1}+1, j_{k-1}+1)$ as p is based on the prior information that the time axis of the signals $x[n]$ and $y[n]$ proceed in the same way due to almost same sampling

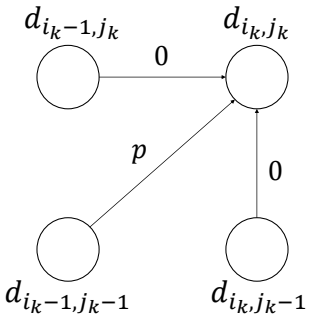

Fig. 2. Edge score (Eq. (8)). d and p indicate edge and node score, respectively.

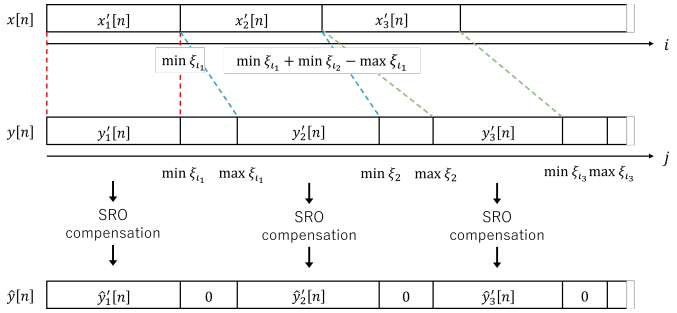

Fig. 3. Framework of missing intervals estimation and linear phase compensation. $x'[n]$ and $y'[n]$ are non-missing intervals of asynchronous signal and reference signal. $\hat{y}'[n]$ is a synchronized signal using $x'[n]$ after SRO compensation.

frequency. Here, a diagram of the edge scores is shown in Fig. 2.

B. Method 1 for synthesis of a synchronized signal: Overlap-add on DP path

We propose two methods to produce the synchronized signal in this paper. The first method is to apply the overlap-add technique along the DP path. In this method, using the correspondence (i_k, j_k) estimated by DP matching, time difference u_{i_k,j_k}, and $x[n]$, The synchronization signal $\hat{y}[n]$ is estimated by

$$\hat{y}[n] = \sum_{j=0}^{J} \tilde{y_j}[n - jS]. \tag{9}$$

Here, $\tilde{y_j}[m]$ is the frame obtained by using the correspondence and time difference, and it is defined as

$$\tilde{y_j}[m] = \begin{cases} 0 & \text{if } \mathcal{I}_j = \mathcal{I}_{j-1}, \\ w_s[m] x_{\mathcal{I}_j}[m + u_{\mathcal{I}_j,j}] & \text{else,} \end{cases} \tag{10}$$

where $w_s[m]$ is a synthesis window for $w_a[m]$. The variable $\mathcal{I}_j$ denotes the indice of the frame that has the largest node score among the frames that correspond to $y_j[m]$, and it is denoted as

$$\mathcal{I}_j = \underset{i_k, k \in \{k | j_k = j\}}{\operatorname{argmax}} \{d_{i_k,j}\}. \tag{11}$$

C. Method 2 for synthesis of a synchronized signal: Missing interval identification and linear phase compensation

The other method to synthesize a synchronization signal is to identify a missing interval using the DP correspondence and performing SRO compensation with linear phase compensation for the non-missing interval (see Fig. 3).

First, the set of indices of frames estimated as missing intervals is given as

$$\psi = \{j \mid \mathcal{I}_j = \mathcal{I}_{j-1}\}. \tag{12}$$

For any i, the set of j that corresponds to the missing intervals and is contained in ψ is given as

$$\xi_i = \{j \mid (\mathcal{I}_j = i) \wedge (j \in \psi)\}. \tag{13}$$

The set of i represented as the missing location in $x[n]$ is denoted as

$$\iota = \{i \mid \xi_i \neq \phi\}, \tag{14}$$

where ϕ indicates the empty set. The set ι is sorted in ascending order, with elements denoted by $\iota_1, \iota_2, \cdots, \iota_q, \cdots, \iota_Q$, where Q is the number of estimated missing intervals and $Q+1$ non-missing intervals. The partial signals $x'_q[n]$ and $y'_q[n]$ are the q-th $(1 \leq q \leq Q+1)$ non-missing interval in $x[n]$ and $y[n]$, respectively. The frame indices range of $x'_q[n]$ and $y'_q[n]$ are given as

$$\sum_{s=1}^{q-1} \min \xi_{\iota_s} - \sum_{s=1}^{q-2} \max \xi_{\iota_s} \leq i < \sum_{s=1}^{q} \min \xi_{\iota_s} - \sum_{s=1}^{q-1} \max \xi_{\iota_s}, \tag{15}$$

$$\max \xi_{\iota_{q-1}} \leq j < \min \xi_{\iota_q}. \tag{16}$$

We apply SRO compensation [13] for $x'_q[n]$ and $y'_q[n]$.

V. Experiments

A. Experimental condition

In this study, we chose speech signals with few silent intervals as a relatively straightforward example for our experiments. This choice allows us to evaluate the fundamental performance of the proposed method. Extending the experiments to more challenging scenarios, such as using environmental sounds for real-world acoustic scene classification (ASC), will be addressed in future work. Specifically, we used five speech signals from Japanese Newspaper Article Sentences (JNAS) [23] of Acoustical Society of Japan (ASJ). The sampling frequency was 16000 Hz. We conducted two cases: (a) using the same utterance, and (b) using convolutive mixtures simulated by [24].

In (a), we made ten mixtures of two speakers using five speech signals of three seconds. For each experiment, we added white noise to the mixture and treated it as a reference signal. For the asynchronous signal, we added white noise, which differed from the reference signal. The signal was resampled to a sampling frequency of 16001.6 Hz by cubic spline interpolation and then missed part of 22500 to 25500

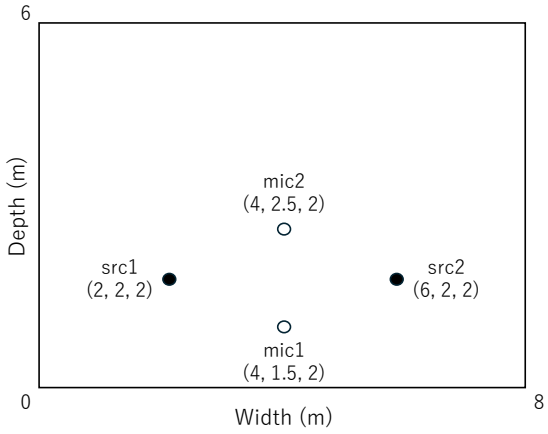

Fig. 4. Arrangement of sound sources and microphones in simulation experiments.The room size is $(8 \times 6 \times 4)$.

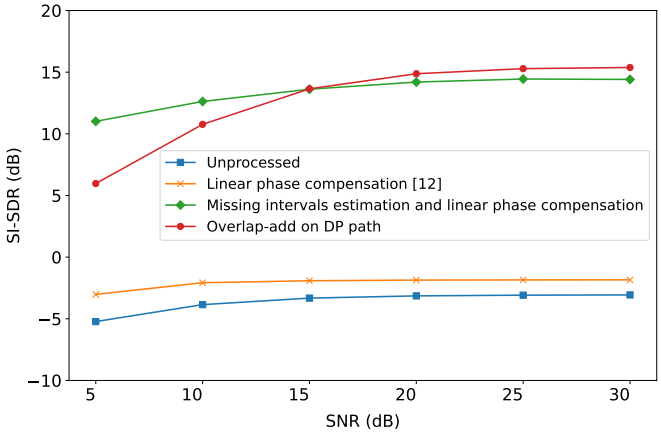

Fig. 5. Average SI-SDR of 10 utterances for each SNR

samples. Each white noise was added so that the signal-to-noise ratio (SNR) to mixtures was $5, 10, 15, 20, 25$, and 30 dB.

In (b), we made ten mixtures of two speakers where we convolved speech signals with an impulse response generated by simulation [24] (see Fig. 4). The room size is $(8 \times 6 \times 4)$, and there are two microphones and two sound sources. The reverberation time is 200 ms, and there is no noise. We regarded the recorded signal of mic1 as a reference signal. We resampled and missed the recorded signal of mic2 as in experiment (a) and regarded it as an asynchronous signal.

To verify the effectiveness of the proposed method, we compared it with linear phase compensation [13]. In the proposed method, frame length $L = 320$ samples, frame shift $S = 80$ samples, window function was Hamming window, and p in (8) was 1.5. In the linear phase compensation method, frame length, frame shift, and window function were the same as the parameters of the proposed method. For the evaluation, we used the scale-invariant signal-to-distortion ratio (SI-SDR) of the asynchronous signal and signal before resampling and missing.

B. Results

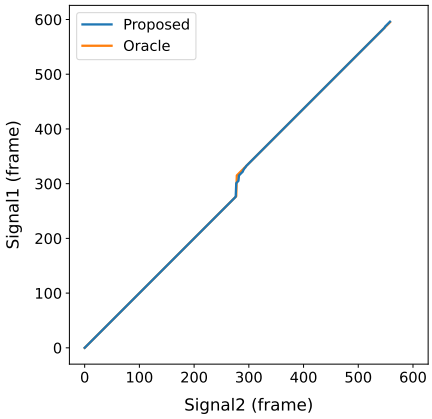

Fig. 6. Correspondence using the proposed method ("proposed") and oracle information ("oracle") for an utterance signal of SNR 5 dB. The signal shown on the horizontal axis include missing interval from 22500 to 25500 samples.

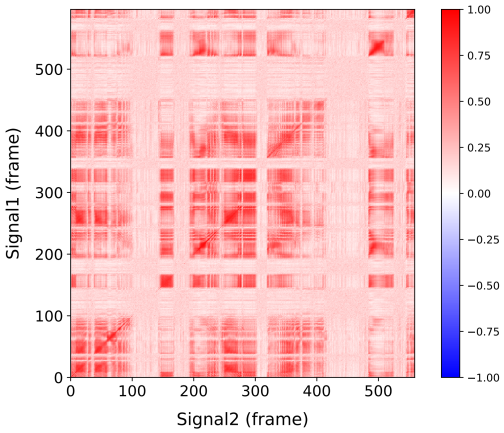

Fig. 7. Node score computed between signals. The signal shown on the horizontal axis include missing interval from 22500 to 25500 samples.

1) Results in (a): Figure 5 shows the average SI-SDR of 10 utterances for each SNR. "Overlap-add on DP path" and "Missing intervals estimation and linear phase compensation" represent the proposed methods described at Section IV-B and Section IV-C, respectively. "Linear phase compensation" is the conventional method [13] and "Unprocessed" is the Unprocessed case. The results show that "Linear phase compensation" performs better than "Unprocessed" due to SRO compensation, but both have a lower. It was considered that these methods could not compensate missing frame correspondence. On the other hand, the proposed methods ("Overlap-add on DP path" and "Missing intervals estimation and linear phase compensation") improved SI-SDR compared to "Unprocessed" and "Linear phase compensation" by more than 10 dB. Thus, we confirmed the effectiveness of the proposed methods. Furthermore, SI-SDR of "Missing intervals estimation and linear phase compensation" was higher than "Overlap-add on DP path" at low SNR. This is thought to be because linear phase compensation robustly worked in the presence of noise. However, in high SNR, "Missing intervals estimation

TABLE I
AVERAGE SI-SDR FOR EACH METHOD.

	Unproc.	Linear Phase comp. [13]	Missing intervals est. and linear phase comp.	Overlap-add on DP path
SI-SDR	-2.66	-2.83	13.41	9.52

and linear phase compensation" was lower than "Overlap-add on DP path". This is thought to be because the missing interval estimation was performed on a frame-by-frame, and the missing part of a sample-by-sample couldn't be considered.

We also showed the example of correspondence and node score for a more detailed analysis of the proposed method. Figures 6 and 7 show the alignment path and node score when the SNR was 5 dB. "Proposed" is the correspondence estimated by DP matching, and "Oracle" is the correspondence based on prior information. As for the estimation of correspondence, Figs. 6 and 7 show that "Proposed" has a path similar to that of "Oracle". It suggests that DP matching considered the local correspondences and allowed proper estimation even when low correlation was caused by noise in some areas.

2) Results in (b): Table I shows the SI-SDR for each method in the simulation experiment. "Unprocessed" corresponds to Unprocessed, "Linear phase comp." corresponds to Linear phase compensation, "Missing intervals est. and linear phase comp." corresponds to estimation and linear phase compensation, and "Overlap-add" corresponds to Overlap-add on DP path.

The table shows that SI-SDR of proposed methods improved more than 10 dB compared with "Unprocessed". We also confirmed that the SI-SDR difference between "Overlap-add on DP path" and "Missing intervals est. and linear phase comp." was about 4 dB. It was considered that the time difference estimation of "Overlap-add on DP path" was affected by the time difference of arrival due to reflected waves of speakers. On the other hand, "Missing intervals est. and linear phase comp." had to estimate a parameter SRO, thus robustly worked.

VI. CONCLUSIONS

In this paper, we proposed a new method for blindly synchronizing signals with SRO and missing data, and describe two methods for synthesizing synchronized signals based on the correspondence between frames estimated by DP matching. The first method involves overlap-adding the corresponding frames along the DP path, while the second method identifies missing intervals and compensates for the SRO using linear phase compensation. Evaluation experiments have demonstrated the effectiveness of these methods. Future work will focus on enhancing the performance of the proposed methods and exploring their application to array signal processing.

ACKNOWLEDGMENT

This work was supported by JST SICORP (JPMJSC2306).

REFERENCES

[1] N. Ono, H. Kohno, N. Ito, and S. Sagayama, "Blind alignment of asynchronously recorded signals for distributed microphone array," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2009, pp. 161–164.

[2] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective," in *Proc. IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT)*, 2011.

[3] S. Araki, N. Ono, K. Kinoshita, and M. Delcroix, "Meeting recognition with asynchronous distributed microphone array using block-wise refinement of mask-based mvdr beamformer," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 5694–5698.

[4] D. Wang, T. Yoshioka, Z. Chen, X. Wang, T. Zhou, and Z. Meng, "Continuous speech separation with ad hoc microphone arrays," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1100–1104.

[5] M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, and B. Lee, "A survey of sound source localization methods in wireless acoustic sensor networks," *Wireless Communications and Mobile Computing*, vol. 2017, pp. 1–24,

[6] P. Giannoulis, A. Brutti, M. Matassoni, *et al.*, "Multi-room speech activity detection using a distributed microphone network in domestic environments," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2015, pp. 1271–1275.

[7] K. Imoto and N. Ono, "Spatial cepstrum as a spatial feature using a distributed microphone array for acoustic scene analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 6, pp. 1335–1343, 2017.

[8] K. Imoto, "Graph cepstrum: Spatial feature extracted from partially connected microphones," *IEICE Transactions on Information and Systems*, vol. E103.D, pp. 631–638, Mar. 2020.

[9] Y. Shiroma, K. Imoto, S. Shiota, N. Ono, and H. Kiya, "Investigation on spatial and frequency-based features for asynchronous acoustic scene analysis," in *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2021, pp. 1161–1166.

[10] T. Kawamura, Y. Kinoshita, N. Ono, and R. Scheibler, "Effectiveness of inter- and intra-subarray spatial features for acoustic scene classification," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[11] K. Imoto and N. Ono, "Acoustic topic model for scene analysis with intermittently missing observations," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 2, pp. 367–382, 2019.

[12] Y. Shiroma, Y. Kinoshita, K. Imoto, S. Shiota, N. Ono, and H. Kiya, "Missing data completion of multi-channel signals using autoencoder for acoustic scene classification," *APSIPA Transactions on Signal and Information Processing*, 2023.

[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/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 3, pp. 571–582, 2016.

[15] A. Chinaev, P. Thüne, and G. Enzner, "Double-cross-correlation processing for blind sampling-rate and time-offset estimation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1881–1896, 2021.

[16] J. Schmalenstroeer, J. Heymann, L. Drude, C. Boeddecker, and R. Haeb-Umbach, "Multi-stage coherence drift based sampling rate synchronization for acoustic beamforming," in *Proc. IEEE International Workshop on Multimedia Signal Processing (MMSP)*, 2017.

[17] A. Chinaev, G. Enzner, T. Gburrek, and J. Schmalenstroeer, "Online estimation of sampling rate offsets in wireless acoustic sensor networks with packet loss," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1110–1114.

[18] T. Raissi, S. Pascual, and M. Omologo, "Sample drop detection for asynchronous devices distributed in space," in *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 815–819.

[19] K. Sumiyoshi, Y. Wakabayashi, and N. Ono, "Dynamic synchronous averaging for enhancement of periodic signal under sampling frequency variation," in *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2020, pp. 863–868.

[20] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.

[21] 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. International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2012, pp. 1–4.

[22] K. Yamaoka, N. Ono, and Y. Wakabayashi, "Sampling frequency mismatch estimation by auxiliary-function-based iterative maximization of double-cross-correlation," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1125–1129.

[23] K. Itou, M. Yamamoto, K. Takeda, *et al.*, "JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research," *Journal of the Acoustical Society of Japan (E)*, vol. 20, no. 3, pp. 199–206, 1999.

[24] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A python package for audio room simulation and array processing algorithms," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 351–355.