Unaliasing of Recorded Signals Based on Blind Source Separation

Yuki Nakamura[†,‡], Taishi Nakashima[‡], Nobutaka Ono[‡], Ryoichi Miyazaki[†]
[†]*National Institute of Technology, Tokuyama College, Yamaguchi, Japan*
i17nakamura1y@tokuyama.kosen-ac.jp
[‡]*Tokyo Metropolitan University, Tokyo, Japan*

***Abstract*—In this paper, we propose a method of restoring the original signal from aliased signals. Generally, it is challenging to resolve aliasing in post-processing for already aliased signals. However, the proposed method solves aliasing only with a post-processing approach by increasing the number of channels. First, we model the aliased signal as a mixture of the original and aliased components. Next, these components are separated by applying blind source separation to the mixture. Finally, the separated components are converted in the frequency domain to recover the original signal accurately. We confirm the effectiveness of our method through experimental evaluations.**

***Index Terms*—unaliasing, blind source separation, signal restoration**

I. Introduction

Aliasing is a well-known phenomenon that arises when sampling a continuous-time domain signal that includes frequency components exceeding the Nyquist frequency, which is half the sampling frequency. The Nyquist-Shannon sampling theorem dictates that in order to reconstruct the original signal, the sampling frequency must be at least twice the maximum frequency component in the signal. Failure to meet this criterion results in aliasing, where the frequency components above and below the Nyquist frequency become indistinguishable in the discrete-time domain (Fig. 1(a)). A conventional approach to mitigating aliasing is to apply a low-pass filter before sampling, thereby sufficiently attenuating frequency components above the Nyquist frequency (Fig. 1(b)) [1]–[3]. However, even with the use of a low-pass filter, aliasing can still arise in practice, particularly when the filter's attenuation is insufficient owing to device configuration or resource constraints. In such cases, it may be necessary to employ more sophisticated signal processing techniques, such as oversampling, undersampling, and adaptive filtering, to reduce the generated aliasing artifacts.

In this paper, we propose a new approach to resolving aliasing issues in a scenario where multiple microphones observe a single sound source (Fig. 1(c)). Specifically, we note that the observed aliasing signal is a composite of frequency components located below and above the Nyquist frequency band. By utilizing blind source separation, we can separate and restore the original components of the signal, including frequency components exceeding the Nyquist frequency. The effectiveness of our proposed method is established through objective evaluation experiments.

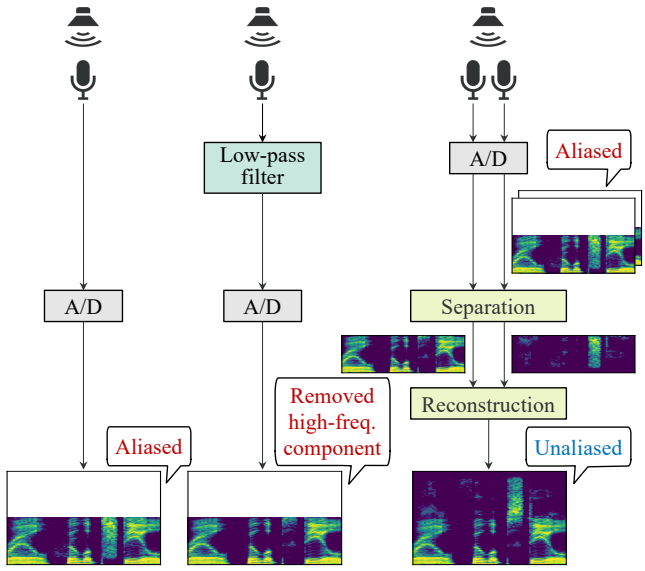

Fig. 1. Overview of procedures of conventional and proposed methods.

II. Modeling of Aliased Observed Signal

Let $x(\tau)$ be an observed signal in the continuous-time domain and $f_{\max}$ be the maximum frequency of $x(\tau)$ where τ indicates the continuous-time variable. Let $x_{\mathrm{ali}}(n)$ be the observed signal sampled at $f_{\mathrm{s,ali}}$ where n indicates the discrete-time index. If $f_{\mathrm{s,ali}} < 2f_{\max}$, *aliasing* happens. We need a higher sampling frequency than $2f_{\max}$ to avoid the aliasing. Let $\tilde{x}_{\mathrm{org}}(n)$ be the observed signal sampled at the sampling frequency $f_{\mathrm{s,org}} = \alpha f_{\mathrm{s,ali}}$ where

$$\alpha = \left\lceil \frac{f_{\max}}{f_{\mathrm{s,ali}}/2} \right\rceil \quad (\alpha \in \mathbb{N}), \tag{1}$$

and $\lceil \cdot \rceil$ indicates the ceiling function. The aliasing does not happen in this case. We put $\tilde{}$ or $\hat{}$ to the signal sampled by the frequency $f_{\mathrm{s,org}}$ to avoid confusion of signals sampled by the different frequencies $f_{\mathrm{s,ali}}$ and $f_{\mathrm{s,org}}$. Since α is an integer, we have

$$x_{\mathrm{ali}}(n) = \tilde{x}_{\mathrm{org}}(\alpha n). \tag{2}$$

Hereafter, we refer to $x_{\mathrm{ali}}(n)$ and $\tilde{x}_{\mathrm{org}}(n)$ as the aliased and original observed signals, respectively. The problem we consider in this paper is how to restore $\tilde{x}_{\mathrm{org}}(n)$ from $x_{\mathrm{ali}}(n)$ when we have multi-channel observations.

Then, let's consider the relationship between $x_{\mathrm{ali}}(n)$ and $\tilde{x}_{\mathrm{org}}(n)$ in the frequency domain. $\tilde{X}_{\mathrm{org}}(f_{\mathrm{org}})$ $(f_{\mathrm{org}} = 0, ..., N-1)$, which is the discrete Fourier transform (DFT) of $\tilde{x}_{\mathrm{org}}(n)$ $(n = 0, ..., N-1)$, can be expressed as

$$\tilde{X}_{\mathrm{org}}(f_{\mathrm{org}}) = \sum_{n=0}^{N-1} \tilde{x}_{\mathrm{org}}(n) W_N^{-nf_{\mathrm{org}}}, \tag{3}$$

where f_{org} indicates the discrete frequency index, $W_N = e^{j\frac{2\pi}{N}}$ indicates the rotational operator of DFT, and j indicates the imaginary unit. For simplicity, we assume that N/α is an integer. Moreover, $X_{\mathrm{ali}}(f_{\mathrm{ali}})$ $(f_{\mathrm{ali}} = 0, ..., \frac{N}{\alpha} - 1)$, which is the DFT of $x_{\mathrm{ali}}(n)$ $(n = 0, ..., \frac{N}{\alpha} - 1)$, can be expressed as

$$X_{\mathrm{ali}}(f_{\mathrm{ali}}) = \frac{1}{\alpha} \sum_{l=0}^{\alpha-1} \tilde{X}_{\mathrm{org}}\left(f_{\mathrm{ali}} + \frac{l}{\alpha}N\right). \tag{4}$$

In the following, we consider the scenario where $\alpha = 2$, indicating a situation where the signal is folded only once owing to aliasing. For simplicity, we use f to denote f_{ali} $(f_{\mathrm{ali}} = 0, ..., \frac{N}{2} - 1)$, then the aliased observed signal is expressed as

$$X_{\mathrm{ali}}(f) = \frac{1}{2}\left(\tilde{X}_{\mathrm{org}}(f) + \tilde{X}_{\mathrm{org}}\left(f + \frac{N}{2}\right)\right). \tag{5}$$

We define the first and second terms in Eq. (5) as *original* and *aliased components*, respectively.

III. Application of Blind Source Separation to Aliased Observed Signals

A. Problem formulation

We consider an observation model in which a two-channel microphone array observes a single source. Let $\tilde{x}_{\mathrm{org},i}(n)$ and $\tilde{s}(n)$ be the original observed signal of channel i and the source signal, respectively. Furthermore, let $\tilde{X}_{\mathrm{org},i}(f_{\mathrm{org}}, t)$ and $\tilde{S}(f_{\mathrm{org}}, t)$ be the short-time Fourier transforms (STFTs) of these signals, respectively. Then, the observation model can be expressed as

$$\tilde{\mathbf{X}}_{\mathrm{org}}(f_{\mathrm{org}}, t) = \tilde{\mathbf{A}}(f_{\mathrm{org}})\tilde{S}(f_{\mathrm{org}}, t), \tag{6}$$

where $\tilde{\mathbf{X}}_{\mathrm{org}}(f_{\mathrm{org}}, t) = [\tilde{X}_{\mathrm{org},1}(f_{\mathrm{org}}, t), \tilde{X}_{\mathrm{org},2}(f_{\mathrm{org}}, t)]^\top$, $\tilde{\mathbf{A}}(f_{\mathrm{org}}) = [\tilde{A}_1(f_{\mathrm{org}}), \tilde{A}_2(f_{\mathrm{org}})]^\top$, $\cdot^\top$ indicates the transpose of the vector, $\tilde{A}_i$ indicates the transfer function from the source to channel i, and t indicates the discrete-time frame index. By substituting Eq. (6) into Eq. (5), we can express the observation model for the aliased observed signal as

$$\begin{aligned}\mathbf{X}_{\mathrm{ali}}(f, t) &= \frac{1}{2}\left\{\tilde{\mathbf{A}}(f)\tilde{S}(f, t) + \tilde{\mathbf{A}}\left(f + \frac{N}{2}\right)\tilde{S}\left(f + \frac{N}{2}, t\right)\right\} \\ &= \mathbf{A}(f)\mathbf{S}(f, t),\end{aligned} \tag{7}$$

where $\mathbf{X}_{\mathrm{ali}}(f, t) = [X_{\mathrm{ali},1}(f, t), X_{\mathrm{ali},2}(f, t)]^\top$, $X_{\mathrm{ali},i}(f, t)$ indicates the STFT of $x_{\mathrm{ali},i}(n)$, and $x_{\mathrm{ali},i}(n)$ indicates the aliased observed signal of channel i in the discrete-time domain. In addition, $\mathbf{A}(f) = \frac{1}{2}\left[\tilde{\mathbf{A}}(f), \tilde{\mathbf{A}}\left(f + \frac{N}{2}\right)\right]$ and $\mathbf{S}(f, t) = [\tilde{S}(f, t), \tilde{S}\left(f + \frac{N}{2}, t\right)]^\top$. The aliased observed signals are expressed by adding the two components, each multiplied by the value of a different frequency bin of the same transfer function. In the following subsection, we will explain that $\tilde{A}(f)$ and $\tilde{A}(f + \frac{N}{2})$ will have different values if the source is not in front of the two microphones, at least, in the plane wave case.

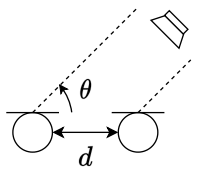

Fig. 2. Arrival of a sound source at two microphones.

B. Phase difference model of original and aliased components

Let $\phi_{\mathrm{low}}(f)$ and $\phi_{\mathrm{high}}(f)$ be the phase differences of the original and aliased components, respectively. Here, the phase difference indicates the phase difference between the two channels of the aliased observed signal. We derive models for $\phi_{\mathrm{low}}(f)$ and $\phi_{\mathrm{high}}(f)$. In our study, we assume that the sound arrives as a plane wave at the speed of sound, c, from a direction represented by θ. We consider two microphones placed at an interval d, as shown in Fig. 2, to record the sound. When the arrival time difference between the two channels is $\delta = \frac{f_s d \cos\theta}{c}$, $A_2(f)$ can be expressed as

$$A_2(f) = A_1(f)\ W_N^{-f\delta} \tag{8}$$

using $A_1(f)$, and the phase difference of the original components, $\phi_{\mathrm{low}}(f)$, can be expressed as

$$\phi_{\mathrm{low}}(f) = \frac{2\pi f}{N}\frac{f_s d \cos\theta}{c}. \tag{9}$$

On the other hand, from Eq. (8), $A_2(f + \frac{N}{2}) = A_1(f + \frac{N}{2})W_N^{-(f+\frac{N}{2})\delta}$. Therefore, the phase difference of the aliased components $\phi_{\mathrm{high}}(f)$ can be expressed as

$$\phi_{\mathrm{high}}(f) = \frac{2\pi f}{N}\frac{f_s d \cos\theta}{c} + \pi\frac{f_s d \cos\theta}{c}. \tag{10}$$

C. Separation of original and aliased components

Equations (9) and (10) show that the original and aliased components display different phase differences and arrive with different transfer functions except for the case when θ is 90 deg. We separate these two components using an established source separation method.

Common source separation techniques, such as beamforming [4], [5], involve blind source separation (BSS) [6]–[11]. Beamforming is an approach to separating sound sources by amplifying sources in a specific direction. However, beamforming requires prior knowledge of the sound source's direction, which is impractical in our scenario. BSS methods

such as frequency domain independent component analysis (FDICA) [7], [8] and independent vector analysis (IVA) [9]–[11] are more appropriate because they separate mixtures without prior knowledge of source direction based on their independence. IVA separates sources on the basis of differences in activity or the time-varying pattern of signal strength. However, since the sound sources considered in this study are low-bandwidth and high-bandwidth folded components of the same source, their activities are expected to be similar, making IVA less effective for this scenario. Therefore, we opt for FDICA. In this study, we assume that the original and aliased components are independent of each other, and verify this assumption by experiments.

FDICA separates sources by frequency bins, but reliably determining the order of separated signals across different frequency bins is an issue known as the permutation problem. In this paper, we assume that the impact of reverberation is insignificant, and the plane wave propagation model is the appropriate approach. To address this issue, we verify whether the phase difference between the separated signals in each frequency bin satisfies either Eq. (9) or (10).

IV. Proposed Method

In this study, we recover the original observed signal from the aliased observed signals. The procedure is as follows.

1) By applying FDICA [7] to the aliased observed signal (Eq. (7)), we obtain the separated signal $\mathbf{Y}(f,t) = [Y_1(f,t), Y_2(f,t)]^\top$.
2) Divide the source direction θ into discrete intervals between 0 and 180 deg, and calculate the corresponding objective function values $\sum_f J(\theta, f)$ for each θ,

$$J(\theta,f) = \begin{cases} J_1(\theta,f) \ (\text{if} J_1(\theta,f) \leq J_2(\theta,f)) \\ J_2(\theta,f) \ (\text{otherwise}) \end{cases} \tag{11}$$

$$J_1(\theta,f) = (|\phi_1(f) - \phi_{\text{org}}(f)|^2 + |\phi_2(f) - \phi_{\text{ali}}(f)|^2)$$
$$J_2(\theta,f) = (|\phi_1(f) - \phi_{\text{ali}}(f)|^2 + |\phi_2(f) - \phi_{\text{org}}(f)|^2)$$

   We select the permutation for each frequency bin f and each θ, so that the objective function is minimized. Among the separated signals in which permutations have been resolved, let $Y_{\text{low}}(f,t)$ and $Y_{\text{high}}(f,t)$ be the signals composed of the original and aliased components, respectively.
3) We concatenate the frequency bands, ensuring that the original components are under the Nyquist frequency and that the inverted and aliased components exist above the Nyquist frequency, as

$$\hat{X}_{\text{prop}}(f,t) = Y_{\text{low}}(f,t), \tag{12}$$

$$\hat{X}_{\text{prop}}\left(f + \frac{N}{2}, t\right) = Y_{\text{high}}\left(\frac{N}{2} - 1 - f, t\right). \tag{13}$$

By following these steps, we obtain the restored signal $\hat{X}_{\text{prop}}(f_{\text{org}}, t)$.

Figure 3 shows examples of spectrograms. The target signal is a chirp signal that transitions from 0 Hz to 8 kHz over 10 s. From (b), aliased components are mixed in the band below the Nyquist frequency owing to aliasing. Then, from (c) and (d), the aliased components are separated from the original components. Furthermore, (e) shows that (a) is restored by reconstructing these components.

V. Experiments

A. Experimental summary

To confirm the effectiveness of our proposed method, we conducted two experiments: (i) the separation of the original components from the aliased components and (ii) the restoration of the original source signal. We generated the aliased observed signals for the experiments by decimating the original observed signals by a factor of two. We first confirmed the effectiveness of our permutation-solving method, which is based on the phase difference model. To verify the accuracy of the permutations, we compared those generated by our proposed method with those generated by assuming the correct value of the permutation solving to be the original signal, which cannot be directly obtained.

Nest, we evaluated the performance of our proposed method in the discrete-time and STFT domains. We compared our proposed method with a classical approach that uses a low-pass filter with a cutoff frequency of 4 kHz to prevent aliasing. The low-pass filter used in this experiment was a 30th-order FIR filter with a Hamming window. The signal obtained by this approach is referred to as the *low-pass signal*.

B. Experimental conditions

We conducted simulation experiments using Pyroomacoustics [13]. The distance between the sound source and the microphones was 2.12 m, and the interval between the microphones was 0.03 m. In the first experiment, the direction of arrival was changed from 0 deg to 90 deg at 5 deg intervals. In the second experiment, the direction of arrival was set to 45 deg. The sampling frequency was 16 kHz and artificially generated aliased observed signals were folded back at 4 kHz. We set the reverberation time (RT60) to two different types: 0 ms and 100 ms. We set the number of frames for the STFT to 1024 points when RT60 is 0 ms and to 4096 points when RT60 is 100 ms. The frame shift length was half of the frame length in both cases. We utilized a total of 100 speech signals from the Carnegie Mellon University (CMU) Arctic speech databases, comprising 50 male and 50 female voices [12]. For FDICA, we used the auxiliary function method to estimate the separation matrix [8].

We used the scale-invariant signal-to-distortion ratio (SI-SDR) [14] and perceptual evaluation of speech quality (PESQ) score [15] to evaluate the sound quality objectively. We objectively evaluated the reproducibility of the source signal in the STFT domain by calculating the log spectral distance (LSD) for each frequency bin. Given the original observed signal $\tilde{x}_{\text{org}}(n)$ and the signal to be evaluated, $\hat{x}(n)$, $\hat{x}(n)$

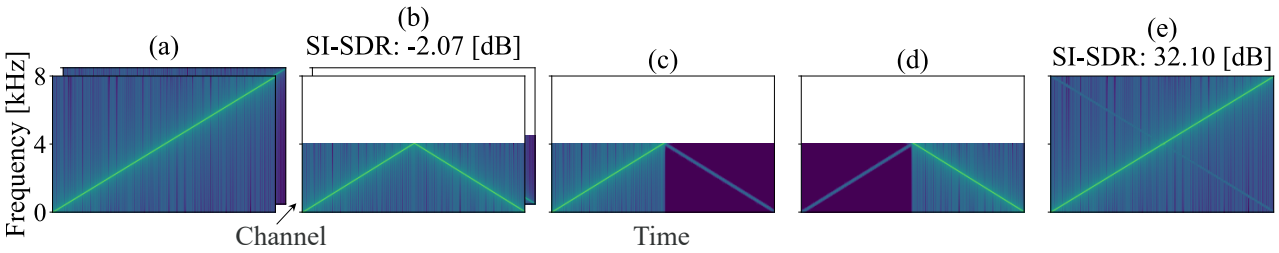

Fig. 3. Examples of spectrograms of proposed method for chirp signal. (a) original observed signals, (b) aliased observed signals, (c) and (d) separated signals (original and aliased components), and (e) restored signal.

includes the aliased observed, low-pass, and restored signals. SI-SDR is defined as

$$\text{SI-SDR} = 10\log_{10}\frac{||\beta\tilde{x}_{\text{org}}(n)||^2}{||\beta\tilde{x}_{\text{org}}(n)-\hat{x}(n)||^2}, \quad \beta = \frac{\hat{x}(n)^\top\tilde{x}_{\text{org}}(n)}{||\tilde{x}_{\text{org}}(n)||^2}.$$

Likewise, the LSD for each frequency bin is defined as

$$\text{LSD}(f_{\text{org}}) = \left(\frac{1}{T}\sum_{t=1}^{T}\left|\log|\tilde{X}_{\text{org}}(f_{\text{org}},t)|^2 - \log|\hat{X}(f_{\text{org}},t)|^2\right|^2\right)^{\frac{1}{2}}, \quad (14)$$

where $\tilde{X}_{\text{org}}(f_{\text{org}},t)$ and $\hat{X}(f_{\text{org}},t)$ indicate the STFTs of $\tilde{x}_{\text{org}}(n)$ and $\hat{x}(n)$, respectively.

Since the number of samples in the aliased observed signal differs from that in the original observed signal, we aligned the sample counts by zero-padding the STFT domain in the frequency range of 4–8 kHz. Additionally, to compute the LSD for both the aliased observed signal and the low-pass signal, we added 10^{-5} to the frequency range of 4–8 kHz in which the signal was zero.

C. Experiment for evaluating permutation correctness rate

Figure 4 shows the accuracy rate of permutations for our proposed permutation-solving method. First, we focus on the case when the reverberation time was 0 ms. The maximum permutation accuracy rate (99.6 %) was achieved at a source arrival direction of the 50 deg case. The results indicate that the proposed method based on phase difference models can accurately estimate the correct permutations when the reverberation effect is negligible. However, the correctness rate decreased when the direction of arrival (DOA) from the sound source was around 90 deg. This is because, when the DOA from the sound source is approximately 90 deg, Eqs. (9) and (10) become identical, making it difficult to determine the permutation accurately. When the reverberation time was 100 ms, the overall permutation accuracy rate decreased, which is believed to be due to the fact that the phase differences of the separated signals no longer adhere to the phase difference models because of the effects of reverberation.

D. Experiment for evaluating restoration performance

We evaluated the quality of sound restoration for the aliased observed, low-pass, and restored signals. First, Table I shows the results of comparing SI-SDR and PESQ for each signal.

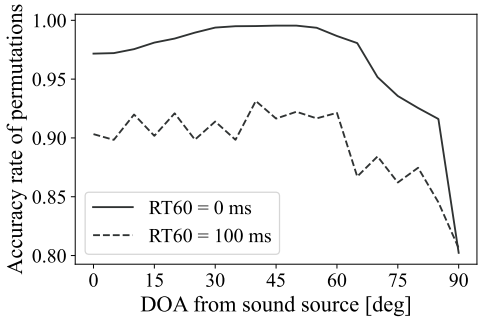

Fig. 4. Percentage of correct permutations of proposed method for each DOA from sound source.

Restored signal (prop.) and *Restored signal (true)* refer to the signal obtained by resolving permutations using the phase difference models and the signal obtained by resolving permutations correctly, respectively. The restored signal achieved the highest SI-SDR and PESQ score compared with the aliased observed and low-pass signals when the reverberation time was 0 ms. When the reverberation time was 100 ms, the proposed method yielded a lower SI-SDR and PESQ score than when the reverberation time was 0 ms; specifically, the PESQ score of the proposed method was lower than that of the low-pass signal. This decrease in source separation performance is attributed to the effects of reverberation. Moreover, the difference in SI-SDR and PESQ score between the restored signals (prop.) and (true) when the reverberation time was 100 ms increased as compared with the case when the reverberation time was 0 ms. This can be attributed to the fact that the effects of reverberation no longer solve the permutations. Figure 5 shows examples of spectrograms of (a) the original observed signal, (b) the aliased observed signal, (c) the low-pass signal, and (d) the restored signal when the reverberation time was 0 ms. The target signal is a speech signal from the CMU Arctic speech databases [12]. From Fig. 5, the sound quality of the proposed method is better, as shown in Table I from the following two points: (I) The separation of low-frequency aliasing components (compare Fig. 5 (b) and (d)). (II) Restoration of components above the Nyquist frequency (compare Fig. 5 (c) and (d)).

Next, the results of comparing each signal with the LSD for each frequency bin are shown in Fig. 6. First, we will

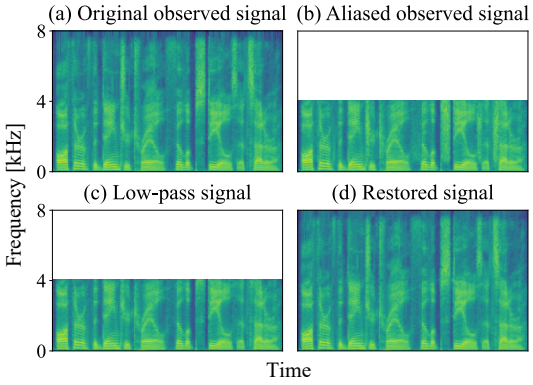

Fig. 5. Examples of spectrograms. (a) original observed signal, (b) aliased observed signal, (c) low-pass signal, and (d) restored signal.

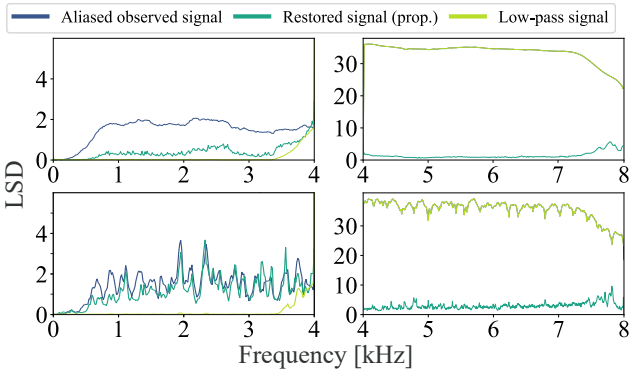

Fig. 6. LSD for each frequency bin at reverberation times of 0 ms (upper panels) and 100 ms (lower panels).

TABLE I
SI-SDR AND PESQ SCORE OF EACH SIGNAL.

	SI-SDR [dB]		PESQ	
	0 ms	100 ms	0 ms	100 ms
Aliased observed signal	17.26	17.12	2.38	2.61
Low-pass signal	20.13	19.97	3.68	3.86
Restored signal (prop.)	**32.86**	**20.86**	**4.27**	**2.63**
Restored signal (true)	32.89	23.02	4.27	2.86

consider the case when the reverberation time was 0 ms. Focusing on the low-frequency band, the proposed method achieved a smaller LSD for the low-pass signal than for the aliased observed signal by effectively separating mixed aliased components, as shown in Fig. 5(d). However, in the high-frequency bandwidth, an overlap between the aliased observed signal and the low-pass signal resulted in a significantly large LSD for these signals. This is because the aliased observed signal lacks high-frequency components, whereas the low-pass signal reduces high-frequency components, as shown in Figs. 5(b) and (c).

On the other hand, the proposed method achieves the smallest LSD owing to the accurate restoration of the high-frequency band of the original observed signal, as shown in Fig. 5(d). In summary, the proposed method outperforms the conventional method in terms of restoration performance in the STFT domain for the problem set up in this paper when the effect of reverberation is sufficiently small. If the high-frequency component of the original observed signal is significant, the difference in LSD will be more notable in the low-frequency band.

When the reverberation time was 100 ms, the overall LSD increased. As with the SI-SDR and PESQ score comparisons, this is likely due to the decrease in source separation performance caused by the effects of reverberation.

VI. Conclusion

In this paper, we proposed a novel method of restoring observed signals that have been degraded by aliasing. The method involved applying a blind source separation technique to the observed signal, followed by signal reconstruction. Our experiments validated the effectiveness of the proposed method. The results demonstrate improvement in restored signal quality, particularly in the high-frequency band.

Acknowledgement

This work was supported by Grant-in-Aid for Scientific Research (A) (Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 20H00613).

References

[1] T. Stilson, "Efficiently-Variable Non-Oversampling Algorithms in Virtual-Analog Music Synthesis — A Root-Locus Perspective," *Ph.D. dissertation, Dept. of Electrical Eng., Stanford Univ., Stanford, CA,* 2006.
[2] V. Valimaki *et al.,* "Antialiasing Oscillators in Subtractive Synthesis," *IEEE Sig. Process. Mag.,* vol. 24, no. 2, pp. 116–125, 2007.
[3] C. Jin *et al.,* "Low Pass Filter for Anti-Aliasing in Temporal Action Localization," *arXiv preprint,* arXiv:2104.11403.
[4] B. D. Van Veen *et al.,* "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.,* vol. 5, no. 2, pp. 4–24, 1988.
[5] H. Cox *et al.,* "Robust Adaptive Beamforming," *IEEE Trans. on ASSP,* vol. 35, no. 10, pp. 1365–1376, 1987.
[6] P. Comon, "Independent Component Analysis, A New Concept?," *Signal Process.,* vol. 36, no. 3, pp. 287–314, 1994.
[7] P. Smaragdis, "Blind Separation of Convolved Mixtures in the Frequency Domain," *Neurocomputing,* vol. 22, pp. 21–34, 1998.
[8] N. Ono *et al.,* "Auxiliary-Function-Based Independent Component Analysis for Super-Gaussian Sources," *Proc. LVA/ICA,* vol. 6365, 2010
[9] T. Kim *et al.,* "Blind Source Separation Exploiting Higher-Order Frequency Dependencies," *IEEE/ACM Trans. on Audio Speech Lang. Process.,* vol. 15, no. 1, pp. 70–79, 2007.
[10] A. Hiroe, "Solution of Permutation Problem in Frequency Domain ICA using Multivariate Probability Density Functions," *Proc. ICA,* pp. 601–608, 2006.
[11] N. Ono, "Stable and Fast Update Rules for Independent Vector Analysis based on Auxiliary Function Technique," *Proc. WASPAA*, pp. 189–192, 2011.
[12] J. Kominek *et al.,* "The CMU Arctic Speech Databases," *Proc. ISCA SSW,* pp. 223–224, 2004.
[13] R. Scheibler *et al,* "Pyroomacoustics: A Python Package for Audio Room Simulation and Array Processing Algorithms," *Proc. ICASSP,* pp. 351–355, 2018.
[14] J. L. Roux *et al.,* "SDR – Half-baked or Well Done?," *Proc. ICASSP,* pp. 626–630, 2019.
[15] A. W. Rix *et al.,* "Perceptual Evaluation of Speech Quality (PESQ) – A New Method for Speech Quality Assessment of Telephone Networks and Codecs," *Proc. of ICASSP,* pp. 749–752, 2001.