A Planar Microphone Array for Spatial Coherence-based Source Separation

Abdullah Fahim*, Prasanga N. Samarasinghe†, Thushara D. Abhayapala‡, and Hanchi Chen§

Research School of Engineering, The Australian National University, Canberra, Australia

Email: {*abdullah.fahim, †prasanga.samarasinghe, ‡thushara.abhayapala, §hanchi.chen}@anu.edu.au

Abstract—We proposed a spatial coherence-based PSD estimation and source separation technique in [1] using a 32-channel spherical microphone array. While the proposed spherical microphone-based method exhibited a satisfactory performance in separating multiple sound sources in a reverberant environment, the use of a large number of microphones remains an issue for some practical considerations. In this paper, we investigate an alternative array structure to achieve spatial coherence-based source separation using a planar microphone array. This method is particularly useful in separating a limited number of sound sources in a mixed acoustic scene. The simplified array structure we used here can easily be integrated with many commercial acoustical instruments such as smart home devices to achieve better speech enhancements.

Index Terms—Planar array, PSD estimation, source separation, spherical harmonics

I. INTRODUCTION

Source separation is an important technique used in many practical devices such as intelligent home assistants, video telephony, and other automatic speech recognition systems. Furthermore, these devices often operate in a reverberant room where the room reflections affect speech quality and intelligibility [2], [3]. Due to the compact nature and cost model of most contemporary commercial products, it is important to attain such speech enhancements with a simple and small microphone array. In this paper, we propose a simple planar microphone array to separate multiple sound sources in both reverberant and non-reverberant environments.

Many multi-channel source separation techniques are found in the literature which employ a microphone array to enhance the speech signal in a desired manner. Beamforming is one of the most fundamental techniques that use a microphone array to boost signals from a desired direction [4], [5]. A beamformer is often complemented with a Wiener filter for a better signal enhancement [6]. Conventionally, a multi-channel signal enhancement is performed with a multi-channel Wiener filter (MWF) [7] which requires the explicit knowledge of the undesired signal power spectral density (PSD) at each microphone position, especially in a reverberant room where the desired and undesired signals are correlated. The authors of [8], [9] used a combination of beamformers and single-channel Wiener filter to separate multiple sources in a non-reverberant room. However, the authors considered the sources to be on the same plane as the microphone array. In the scenario when

the sources lie on a 3D plane, beamforming with a planar array is a challenging task and often results in performance degradation.

We proposed a solution to separate sound sources on a 3D plane using a 32-channel spherical microphone array [1]. While the technique described in [1] is capable of separating a large number of sources in a reverberant environment, it is not always commercially viable to use a large number of microphones or a spherical array, especially when dealing with a small number of sources or a non-reverberant environment. In this paper, we solve this issue by proposing a hybrid planar array with a circular microphone array and an additional microphone at the origin. The experimental validations were carried out with 6 microphones which offers an attractive solution for a commercial product. We estimate the PSD components at the origin using a multichannel PSD estimation technique [1] and employ a single-channel Wiener filter to the received signal at the origin. We measure the performance of the proposed method in practical and simulated environments and compare it with other contemporary techniques.

II. PROBLEM FORMULATION

Considering an omni-directional microphone located at $\boldsymbol{x}_q = (r_q, \theta_q, \phi_q)$, the expression for the recorded sound field is given by

$$p(\boldsymbol{x}_q, t) = \sum_{\ell=1}^{L} s_\ell(t) * \left(h_\ell^{(d)}(\boldsymbol{x}_q, t) + h_\ell^{(r)}(\boldsymbol{x}_q, t) \right) \quad (1)$$

where t is the discrete time index, $\ell \in [1, L]$ with L being the total number of sound sources, $s_\ell(t)$ is the ℓth sound source excitation, $*$ denotes the convolution operation, and $h_\ell^{(d)}(\cdot)$ and $h_\ell^{(r)}(\cdot)$ are respectively the direct and reverberation path components of the room impulse response (RIR) between the ℓth source and the microphone positions. Converting (1) to frequency domain using short-time Fourier transform (STFT), we obtain

$$P(\boldsymbol{x}_q, \tau, k) = \sum_{\ell=1}^{L} S_\ell(\tau, k) \left(H_\ell^{(d)}(\boldsymbol{x}_q, \tau, k) + H_\ell^{(r)}(\boldsymbol{x}_q, \tau, k) \right) \quad (2)$$

where $\{P, S, H\}$ represent the corresponding signals of $\{p, s, h\}$ in the STFT domain, τ is the time frame index, $k = 2\pi f/c$, f is the center frequency of the corresponding frequency bin, and c is the speed of sound propagation. Given

the measured sound pressure $p(\boldsymbol{x}_q, t) \ \forall q$, where q denotes the microphone index in an array, we want to estimate the source signals $s_\ell(t) \ \forall \ell$ with a single-channel Wiener filter driven by a multi-channel PSD estimator.

III. SOURCE SEPARATION WITH A SPHERICAL MICROPHONE ARRAY

In this section, we review the PSD estimation and source separation technique proposed in [1].

Formulating the room transfer functions in the spatial domain, (2) can be written as

$$
P(\boldsymbol{x}_q, k) = \sum_{\ell=1}^{L} S_\ell(k) \Big(G_\ell^{(d)}(k) \ e^{ik \ \hat{\boldsymbol{y}}_\ell \cdot \boldsymbol{x}_q} +
$$
$$
\int_{\hat{\boldsymbol{y}}} G_\ell^{(r)}(k, \hat{\boldsymbol{y}}) \ e^{ik \ \hat{\boldsymbol{y}} \cdot \boldsymbol{x}_q} \ d\hat{\boldsymbol{y}} \Big) \quad (3)
$$

where $G_\ell^{(d)}(k)$ represents the direct path gain at the origin for the ℓth source, $i = \sqrt{-1}$, $\hat{\boldsymbol{y}}_\ell$ is a unit vector towards the direction of the ℓth source, and $G_\ell^{(r)}(k, \hat{\boldsymbol{y}})$ is the reflection gain at the origin along the direction of $\hat{\boldsymbol{y}}$ for the ℓth source. The time frame index τ is omitted in (3) as well as in the rest of the paper for brevity.

Let us now consider a spherical microphone array of radius r consisting of Q pressure microphones. Hence, the spherical harmonics decomposition of the sound field at q^{th} microphone is given by [10, ch. 6]

$$
P(\boldsymbol{x}_q, k) = \sum_{nm}^{N} \alpha_{nm}(k) \ j_n(kr) \ Y_{nm}(\theta_q, \phi_q) \quad (4)
$$

where $\sum_{nm}^{N} = \sum_{n=0}^{N} \sum_{m=-n}^{n}$, sound field order $N = \lceil kr \rceil$ [11], $\lceil \cdot \rceil$ denotes the ceiling operation, $j_n(\cdot)$ denotes the nth order spherical Bessel function, and $Y_{nm}(\cdot)$ is the spherical harmonics of order n and degree m. The sound field coefficients $\alpha_{nm}(k)$ can be calculated using a spherical microphone array by [12], [13]

$$
\alpha_{nm}(k) = \frac{1}{j_n(kr)} \sum_{q=1}^{Q} w_q \ P(\boldsymbol{x}_q, k) \ Y_{nm}^*(\theta_q, \phi_q) \quad (5)
$$

where w_q is the weight of qth microphone and $Q \geq (N+1)^2$. Furthermore, spherical harmonics decomposition of a plane wave is given by [14, pp. 9–13]

$$
e^{ik \ \hat{\boldsymbol{y}}_\ell \cdot \boldsymbol{x}_q} = \sum_{nm}^{N} 4\pi i^n \ Y_{nm}^*(\hat{\boldsymbol{y}}_\ell) \ j_n(kr) \ Y_{nm}(\theta_q, \phi_q). \quad (6)
$$

Using (4) and (6) in (3), we obtain

$$
\alpha_{nm}(k) = \sum_{\ell=1}^{L} 4\pi i^n \ S_\ell(k) \Big(G_\ell^{(d)}(k) \ Y_{nm}^*(\hat{\boldsymbol{y}}_\ell) +
$$
$$
\int_{\hat{\boldsymbol{y}}} G_\ell^{(r)}(k, \hat{\boldsymbol{y}}) \ Y_{nm}^*(\hat{\boldsymbol{y}}) \ d\hat{\boldsymbol{y}} \Big). \quad (7)
$$

As the gains of different reflection surfaces are independent in nature and assuming uncorrelated sources, the cross-correlation between the sound field coefficients is derived as[1]

$$
E\{\alpha_{nm}\alpha_{n'm'}^*\}(k) = \sum_{\ell=1}^{L} \Phi_\ell(k) \ C_{nn'} Y_{nm}^*(\hat{\boldsymbol{y}}_\ell) \ Y_{n'm'}(\hat{\boldsymbol{y}}_\ell)
$$
$$
+ \sum_{v=0}^{V} \sum_{u=-v}^{v} \Gamma_{vu}(k) \ C_{nn'} W_{n,n',v}^{m,m',u} \quad (8)
$$

where $\Phi_\ell(k)$ is the ℓth source PSD at the origin, $C_{nn'} = 16\pi^2 i^{n-n'}$, and $\Gamma_{vu}(k)$ is the harmonics power of the reverberation sound field of order v and degree u. The source-independent constant $W_{n,n',v}^{m,m',u}$ is defined as

$$
W_{n,n',v}^{m,m',u} = (-1)^m \sqrt{\frac{(2v+1)(2n+1)(2n'+1)}{4\pi}} \ W_{12} \quad (9)
$$

with W_{12} representing a multiplication between two Wigner-3j symbols [16] as

$$
W_{12} = \begin{pmatrix} v & n & n' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} v & n & n' \\ u & -m & m' \end{pmatrix}. \quad (10)
$$

Considering the cross-correlation between all the available modes $\alpha_{nm}(k)$, we obtain a system of $(N+1)^4$ equations from (8) which can be solved for $\Phi_\ell(k)$, $\Gamma_{vu}(k)$, and hence, the total reverberation power at the origin $\Phi_r(k) = \sqrt{4\pi} \ \Gamma_{00}(k)$, provided that $(N+1)^4 \geq L + (V+1)^2$.

Finally, a Wiener filter is used at the output of a beamformer to estimate source signals by

$$
\hat{S}_\ell(k) = Z_\ell(k) \ \frac{\Phi_\ell(k)}{\sum_{\ell'=1}^{L} \Phi_{\ell'}(k) + \Phi_r(k)} \quad (11)
$$

where $Z_\ell(k)$ is the output of a suitable beamformer steered towards the ℓth sound source.

IV. PROPOSED SOLUTION WITH A PLANAR ARRAY

In this section, we describe our approach to use the aforementioned source separation technique with a planar array.

A. Motivation for a planar array

The motivation for a simple planar array comes from the fact that, though [1] offers a useful technique for PSD estimation and source separation, it requires a minimum $(N + 1)^2$ microphones when used with a spherical microphone array [12], [13] or $\left(2(N + 1)^2 - 2\right)$ omni-directional microphones with a hybrid differential microphone array [17]. Reduction of the number of microphones in an array is desirable from many commercial perspectives, especially when a smaller number of sources are to be considered. Furthermore, a planar array has less design complexity compared to a spherical microphone array. Hence, we design a hybrid planar microphone array which can be used in the source separation technique of [1], but with a significantly smaller number of microphones.

---

[1] For a detailed derivation of (8), please refer to [1], [15].

B. The proposed method

In this section, we describe a simpler array structure that works with [1] with two specific goals - (1) use of a planar microphone array, and (2) eliminate the requirement for a beamformer.

It is evident from (8) that [1] works with any type of microphone arrays that are capable of producing enough number of sound field coefficients $\alpha_{nm}(k)$ such that we have at least $L + (V+1)^2$ spatial correlation coefficients. Hence, we consider a planar circular array on the XY-plane which can extract only the even sound field coefficients as the odd spherical harmonics diminishes on the XY-plane[2]. For an Nth-order sound field, there exists $\big((N+1)(N+2)/2\big)$ active even modes, hence, the necessary condition to solve (8) is

$$\left(\frac{(N+1)(N+2)}{2}\right)^2 \geq L + (V+1)^2. \tag{12}$$

To achieve the second design criteria, i.e. eliminating the beamformer, we propose to use an omni-directional microphone at the center of the microphone array and apply the Wiener filter at the received signal at the origin. The use of the signal at the origin as the input to the Wiener filter also agrees with the definition of $\Phi_\ell(k)$ and $\Phi_r(k)$ which are defined at the origin. Therefore, the estimated source signal under the new model is changed from (11) to

$$\hat{S}_\ell(k) = P(\boldsymbol{x}_0, k) \frac{\Phi_\ell(k)}{\sum\limits_{\ell'=1}^{L} \Phi_{\ell'}(k) + \Phi_r(k)} \tag{13}$$

where $\boldsymbol{x}_0 = (0,0,0)$ indicates the origin.

C. Extract the even coefficients using the proposed array structure

The extraction of the even sound field coefficients using multiple circular arrays was first proposed in [18]. With the proposed array structure, we readily calculate $\alpha_{00}(k)$ from the received signal at the origin by setting $q = n = m = 0$ in (4) as

$$\alpha_{00}(k) = \sqrt{4\pi} P(\boldsymbol{x}_0, k). \tag{14}$$

Assuming that the circular array has a radius of R and contains Q omni-directional microphones, we obtain the sound pressure at each microphone using (4) by

$$P(\boldsymbol{x}_{q'}, k) = \sum_{nm}^{N} \alpha_{nm}(k) \, j_n(kR) \, Y_{nm}(\frac{\pi}{2}, \phi_{q'}) \tag{15}$$

where $N = \lceil kR \rceil$ and $q' \in [1, Q]$. From the definition of the spherical harmonics, we know

$$Y_{nm}(\frac{\pi}{2}, \cdot) = \begin{cases} \frac{1}{\sqrt{4\pi}}, & \text{if } n = 0 \\ 0, & \text{if } (n+|m|) \text{ is odd} \\ Y_{nm}(\frac{\pi}{2}, \cdot), & \text{otherwise.} \end{cases} \tag{16}$$

[2]The odd and even coefficients are decided based on the value of the corresponding $(n + |m|)$.
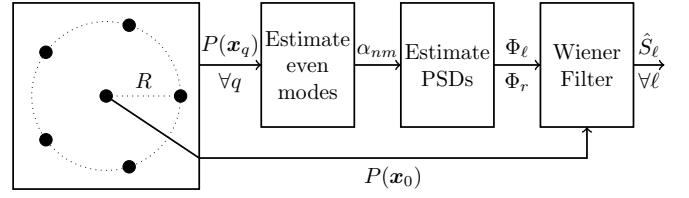

Fig. 1. Block diagram of the proposed method using a planar array with 6 omni-directional microphones. FFT blocks and k-dependency are omitted for brevity.

Hence, using (14) and (16) in (15), we obtain

$$P(\boldsymbol{x}_{q'}, k) - P(\boldsymbol{x}_0, k) \, j_0(kR)$$
$$= \sum_{\substack{nm \\ n \neq 0 \\ n+|m| \text{ even}}}^{N} \bar{\alpha}_{nm}(k) \, j_n(kR) \, Y_{nm}(\frac{\pi}{2}, \phi_{q'}) \tag{17}$$

where $\bar{\alpha}_{nm}(k) = \big\{\alpha_{nm}(k) : n > 0; \text{ and } (n+|m|) \text{ is even}\big\}$. Considering all the microphones on the circular array, we write (17) in a matrix form as

$$\begin{bmatrix} \bar{P}_1 \\ \vdots \\ \vdots \\ \bar{P}_Q \end{bmatrix} = \begin{bmatrix} \Lambda_{1-1}(\phi_1) & \cdots & \Lambda_{NN}(\phi_1) \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ \Lambda_{1-1}(\phi_Q) & \cdots & \Lambda_{NN}(\phi_Q) \end{bmatrix} \begin{bmatrix} \bar{\alpha}_{1-1} \\ \vdots \\ \vdots \\ \bar{\alpha}_{NN} \end{bmatrix} \tag{18}$$

where

$$\bar{P}_{q'} = P(\boldsymbol{x}_{q'}, k) - P(\boldsymbol{x}_0, k) \, j_0(kR) \tag{19}$$

and

$$\Lambda_{nm}(\phi_{q'}) = j_n(kR) \, Y_{nm}(\frac{\pi}{2}, \phi_{q'}). \tag{20}$$

The dependency on k is omitted in (18) for brevity. Note that, the right-most vector of (18) contains $\big((N+1)(N+2)/2-1\big)$ elements, hence, (18) can be solved for all $\bar{\alpha}_{nm}(k)$ as long as

$$Q \geq \frac{(N+1)(N+2)}{2} - 1. \tag{21}$$

D. PSD estimation and source separation

Once we estimate all the even modes $\big[\alpha_{00}(k), \bar{\alpha}_{nm}(k)\big]$ using (14) and (18), we construct and solve (8) considering the even modes only to estimate individual source and reverberant PSDs, subjected to the constraint mentioned in (12). Finally, we employ the single-channel Wiener filter of (13) to reconstruct each source signal separately. Fig. 1 shows the block diagram of the proposed method with the planar array structure.

V. PERFORMANCE EVALUATION

In this section, we evaluate the proposed algorithm with practical experiments as well as through simulations.

Sources	DSB	MBF	SMA	PMA
PESQ				
$L = 8$	1.6	1.85	2.22	1.85
$L = 6$	1.53	1.53	1.95	1.64
$L = 4$	1.6	1.85	2.22	1.85
$L = 2$	1.96	2.42	2.67	2.29
FWSegSNR (dB)				
$L = 8$	4.21	4.95	7.68	5.98
$L = 6$	4.38	5.69	8.75	6.37
$L = 4$	5.64	7.72	10.51	7.77
$L = 2$	8.2	11.14	13.16	10.02

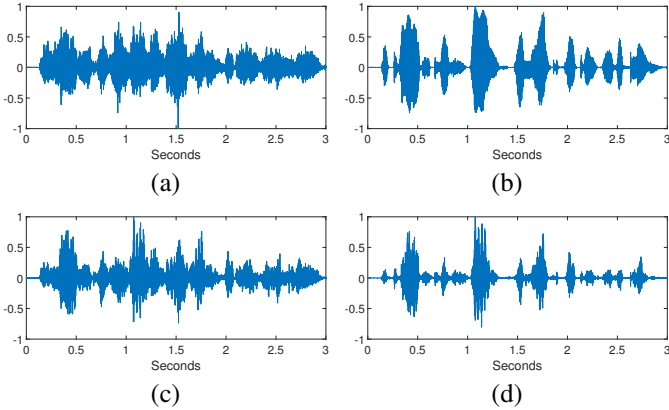Fig. 3. Impact of the spatial aliasing in a non-reverberant case.

Fig. 2. Estimated signal waveform of the first speaker in a 4-speaker non-reverberant environment. (a) Recorded mixed waveform. (b) Original waveform. (c) Estimated waveform at a beamformer output. (d) Estimated waveform with the proposed planar array.

A. Experimental setup

For the experimental validation, we used $N = 2$ for the proposed algorithm. Data processing was performed in the frequency domain with a 8 ms Hanning window, 50% frame overlap, a 128-point fast Fourier transform (FFT), and 8 kHz sampling frequency. The source directions were estimated using a spherical harmonics-based frequency-smoothed MU-SIC algorithm [19]. All the sources are considered to be either above or below the XY-plane, which can be easily ensured with a proper placement of the array. The performance was measured through two objective metrics - frequency-weighted segmental signal to noise ratio (FWSegSNR) [20] and perceptual evaluation of speech quality (PESQ) index [21]. Each of the experiments were performed 20 times with mixed-gender random speech signals and the average values of the objective metrics are presented in the subsequent sections.
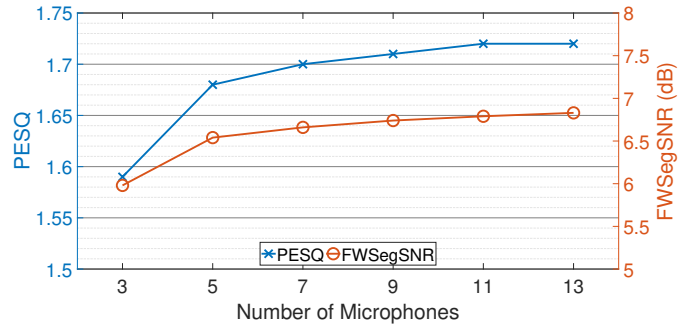
B. Non-reverberant case

To evaluate the performance of the proposed method in a non-reverberant condition, we simulated mixed audio signals with $L = \{2, 4, 6, 8\}$ at random source locations. We used the proposed planar array with $Q = 5$ and $R = 2$ cm. Table I compares the performance of the proposed method (denoted as "PMA") with a conventional delay and sum beamformer (denoted as "DSB"), a multiple beamformer-based method [8] (denoted as "MBF"), and the spatial coherence-based method of [1] using a 32-channel spherical microphone array (denoted as "SMA") of type Eigenmike. For a fair comparison, we used the same number of microphones for the "DSB" and "MBF" methods as we used for the proposed method. From Table I it is obvious that [1] performs better under all the scenarios, which is expected as the method is able to utilize all the available sound field modes while constructing the spatial coherence matrix due to the array structure and a larger number of microphones. The proposed planar array uses the even modes only and always performs better compared to the conventional beamforming-based technique. The performance comparison with [8] reveals that the proposed method exhibits better results in all of the cases except for $L = 2$ where the additional gain achieved with the proposed method compared to "MBF" could not compensate the loss due to spatial aliasing. Furthermore, one of the major drawbacks of "MBF", the rank deficiency issue [8], is less likely yo occur with a less number of beamformers used in $L = 2$ case. It is also observed from Table I that the performance gain of the proposed method over [8] improves as the number of sound sources increases.

The estimated waveform for the first speaker in a 4-speaker system is shown in Fig. 2 which exhibits a good resemblance with the original waveform.

The performance of the proposed method can be affected due to spatial aliasing, especially at the higher frequencies, when (21) is not met. To analyze the impact of spatial aliasing, we increased the array radius to 6 cm and measured the performance with a varying number of microphones. As we observe from Fig. 3, the performance improves with increasing number of microphones which suggests a reduction in spatial aliasing.

Metric	DSB	MBF	DMA	PMA
PESQ	1.94	1.90	2.2	2.22
FWSegSNR	3.57	3.79	4.45	3.97

C. Reverberant case

We evaluated the performance of the proposed technique in a realistic room environment with multiple sound sources. It is worth noting that, as a trade-off for using a small number of microphones in a single plane, we need to restrict the order of the reverberant sound field power to $V \leq 1$ to avoid an underdetermined system of equations of (8). The exclusion of the higher order reverberant sound field power can introduce some artifact noise at the final output, however, the contribution of the higher order modes to the total reverberant power is expected to be less prominent compared to the contribution of the lower order modes.

For the experimental validation, we used a planar array with $M = 5$ and $R = 3$ cm. We compared the performance of the proposed technique with [1] using a 16-channel hybrid differential microphone array [22] (denoted as "DMA") as well as with the "DSB" and "MBF" techniques referred in Section V-B. The results with 2 sound sources are shown in Table II which suggests that the proposed method offers a better performance compared to "DSB" and "MBF", and maintain a comparable performance with "DMA" despite having a fewer number of microphones in the array. Furthermore, Fig. 4 plots the estimated waveforms for the 2-speaker system which exhibits a good resemblance with the original signals. However, with a larger number of sources in a reverberant room, the proposed algorithm suffers from artifact noise at its output. To improve the performance and robustness of the spatial coherence-based source separation technique using a planar array, an array structure with multiple circles, such as [18], can be considered to extract the higher order modes.

VI. CONCLUSION

We proposed a planar array to perform sound source separation utilizing the even harmonics modes of a sound field. The array was found to be capable of separating a significant number of sources in a non-reverberant environment, but the functionality was limited to a small number of sources in a reverberant room. However, due to the simplified design with a smaller number of microphones, the proposed method can be useful in different commercial products. The performance of the proposed algorithm can be enhanced by introducing additional circles of microphones to extract higher order even harmonics modes. A similar concept can be useful in reducing required number of microphones and simplifying the design structure of other sound processing techniques that use distributed higher order microphones [23], [24].
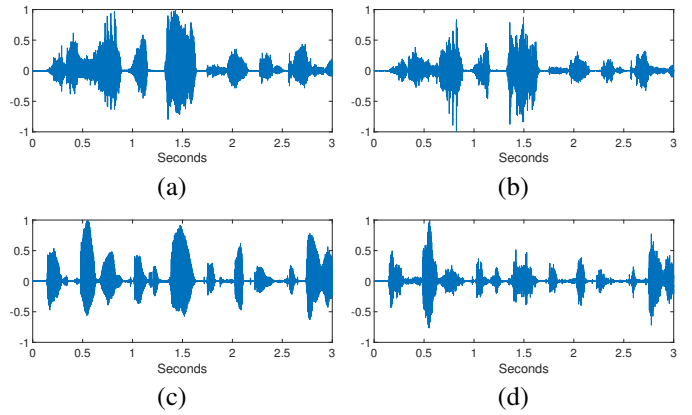

Fig. 4. Estimated signal waveforms using the proposed planar array in a practical reverberant room with 2 speakers. (a) The original waveform of the first speaker. (b) The estimated waveform of the first speaker. (c) The original waveform of the second speaker. (d) The estimated waveform of the second speaker.

REFERENCES

[1] A. Fahim, P. N. Samarasinghe, and T. D. Abhayapala, "Psd estimation of multiple sound sources in a reverberant room using a spherical microphone array," in *Proc. IEEE WASPAA*. IEEE, 2017, pp. 76–80.

[2] P. Naylor and N. D. Gaubitch, "Introduction," in *Speech dereverberation*. London, UK: Springer Science & Business Media, 2010, pp. 1–15.

[3] J. P. A. Lochner and J. Burger, "The intelligibility of speech under reverberant conditions," *Acta Acustica united with Acustica*, vol. 11, no. 4, pp. 195–200, 1961.

[4] D. H. Johnson and D. E. Dudgeon, *Array signal processing: concepts and techniques*. NJ, USA: Prentice-Hall, Englewood Cliffs, 1993.

[5] J. Bourgeois and W. Minker, *Time-Domain Beamforming and Blind Source Separation*. New York, USA: Springer-Verlag New York Inc., 2010.

[6] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 3, pp. 240–259, 1998.

[7] A. Kuklasiński, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum likelihood psd estimation for speech enhancement in reverberation and noise," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 9, pp. 1595–1608, 2016.

[8] Y. Hioka, K. Furuya, K. Kobayashi, K. Niwa, and Y. Haneda, "Underdetermined sound source separation using power spectrum density estimated by combination of directivity gain," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 6, pp. 1240–1250, 2013.

[9] K. Niwa, T. Kawase, K. Kobayashi, and Y. Hioka, "Psd estimation in beamspace using property of m-matrix," in *Proc. IWAENC*, 2016, pp. 1–5.

[10] E. G. Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*. London, UK: Academic press, 1999.

[11] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 6, pp. 697–707, 2001.

[12] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *Proc. IEEE ICASSP*, vol. 2. IEEE, 2002, pp. II–1949.

[13] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proc. IEEE ICASSP*, vol. 2, 2002, pp. II–1781.

[14] H. Teutsch, *Modal array signal processing: principles and applications of acoustic wavefield decomposition*. Berlin, Germany: Springer-Verlag, 2007.

[15] A. Fahim, P. N. Samarasinghe, and T. D. Abhayapala, "Psd estimation and source separation in a noisy reverberant environment using a spherical microphone array," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 9, pp. 1594–1607, 2018.

[16] F. W. J. Olver, "3j, 6j, 9j symbols," in *NIST Handbook of Mathematical Functions*. Cambridge, UK: Cambridge University Press, 2010, ch. 34.

[17] H. Chen, T. D. Abhayapala, and W. Zhang, "Theory and design of compact hybrid microphone arrays on two-dimensional planes for three-dimensional soundfield analysis," *J. Acoust. Soc. Am.*, vol. 138, no. 5, pp. 3081–3092, 2015.

[18] T. D. Abhayapala and A. Gupta, "Spherical harmonic analysis of wavefields using multiple circular sensor arrays," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 6, pp. 1655–1666, 2010.

[19] D. Khaykin and B. Rafaely, "Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach," in *Proc. IEEE WASPAA*, 2009, pp. 221–224.

[20] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 1, pp. 229–238, 2008.

[21] I.-T. Rec., "Perceptual evaluation of speech quality (pesq): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," *Rec. ITU-T P. 862*, 2001.

[22] P. N. Samarasinghe, H. Chen, A. Fahim, and T. D. Abhayapala, "Performance analysis of a planar microphone array for three dimensional soundfield analysis," in *Proc. IEEE WASPAA*, 2017, pp. 249–253.

[23] A. Fahim, P. N. Samarasinghe, and T. D. Abhayapala, "Sound field separation in a mixed acoustic environment using a sparse array of higher order spherical microphones," in *Proc. HSCMA*. IEEE, 2017, pp. 151–155.

[24] P. Samarasinghe, T. Abhayapala, and M. Poletti, "Wavefield analysis over large areas using distributed higher order microphones," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 3, pp. 647–658, 2014.