Reactive Intensity Vector Based Direct Path Detection for DoA Estimation on a Single Acoustic Vector Sensor

Bing Zhu and Wen Zhang
Center of Intelligent Acoustics and Immersive Communications School of Marine Science and Technology
Northwestern Polytechnical University, Xi'an, China

***Abstract*—As a compact device, an acoustic vector sensor (AVS) has been widely used for source direction of arrival (DOA) estimation in 3D space. However, when deployed in complex environments, such as with strong noise, reverberation, and multiple sources, the DOA estimation accuracy degrades significantly. In this paper, we propose an efficient method for multi-source DOA estimation based on an AVS in a reverberant room environment. The method is based on exploiting a feature from the AVS signals, i.e., the reactive intensity vector, to detect the direct-path dominated parts of the received signals in time-frequency bins. Then, intensity-vector based DOA estimation is performed, which has low computational complexity. As demonstrated through experiments, the proposed method outperforms the state of the art methods even in highly reverberant environments.**

***Index Terms*—Acoustic vector sensor, DOA, multi-source, reactive intensity vector**

I. Introduction

Due to its small size and light weight, the acoustic vector sensor (AVS) has been recently researched and developed for acoustic wave capturing and acoustic scene analysis. It consists of one omni-directional pressure sensor and three co-located directional velocity sensors. With these four sensors, an AVS can estimate the sound source direction of arrival (DOA) in three dimensional space [1].

The most commonly used DOA methods based on an AVS can be roughly divided into the intensity vector-based method, the subspace-based method and the beamformer-based method. The intensity vector-based method estimates the source DOA by calculating the intensity vector from the pressure signal and velocity signals [2], [3], which performs well in a free-field environment but has significantly degraded performance in a multi-source, reverberant and noisy environment. The subspace-based methods that exploit the second-order signal statistics for source DOA estimation are applied on the four-channel AVS signals straightforwardly, resulting in algorithms such as MUSIC ESPRIT-based method, and velocity-covariance-based method [4]–[7]. These methods are effective to distinguish noise and non-coherent interference but cannot deal with early room reflections that generate coherent interference. One of the representative beamformer-based methods is the maximum steered response power (MSRP) method [8]. It uses the received four-channel AVS signals to form a first-order beamformer to search DOA with the maximum power. And its enhanced method, the maximum likelihood (ML) method [9], finds the optimal parameter under noisy environment.

To deal with room reverberation and interference, several methods have been recently proposed [10]–[15], with a pre-processing step of the AVS signals to identify the direct path dominant (DPD) components for the follow-up DOA estimation. The earliest work is the coherent test method [10], in which the spatial correlation matrix is calculated and the rank-1 points are selected as the DPD points. Its improvement, the DPD test [11], [12] method, applies the time-frequency smoothing on the spatial covariance matrix and uses the ratio of the first two eigenvalues as the effective rank. As frequency smoothing can decorrelate the signal, the DPD test method improves the detection accuracy. These methods are based on analyzing the spatial covariance matrix of the received AVS signal within a time-frequency (TF) zone to detect the direct path dominated components, thus are characterized as TF zone level (or block-wise) detection methods. However, most of the time both the true DPD point and its neighborhood points are detected because the DPD points dominate the detection results, leading to reduced DOA estimation accuracy.

Another kind of representative approach is the TF point level detection method, which detects the direct path dominant TF bins for the DOA estimation. For example, in [14], the low reverberant single source (LRSS) TF bins are detected by comparing the absolute directions from the real and imaginary parts of the received AVS signals in the STFT domain. When these two directions are close enough, the corresponding TF bins are assumed to be the LRSS points. However, it cannot distinguish the TF point that is due to a signal plus small-level interference or two-closely spaced propagation directions.

In this paper, we propose a point-level detection method for AVS-based DOA estimation, which is based on the characteristics of the intensity vector. The sound intensity vector describes the direction and rate of the acoustic energy flow per unit area, which can be divided into the active part and the reactive part [16]. We show that for single-source DPD points the reactive intensity approximates zero, while for multi-source and reverberation-dominated points, this value becomes significantly larger. Through analysis and experimental results, we show that the proposed method achieves more accurate DPD detection and DOA estimation, especially for some challenging situations such as small-level interference and

reflections with its propagation direction close to the source DOA that is hard to extinguish.

II. Signal model and Problem Formulation

Considering a reverberant room environment with M sources, the signals received by a single AVS can be modeled as

$$\mathbf{x}[t] = \sum_{m=1}^{M} s_m[t] * \mathbf{h}_m[t] + \mathbf{n}[t], \tag{1}$$

where $\mathbf{x}(t) = [x_o(t), x_{d_x}(t), x_{d_y}(t), x_{d_z}(t)]^T$ is a vector of the received signals, including the pressure signal $x_o(t)$ picked up by the omnidirectional sensor, and $[x_{d_x}(t), x_{d_y}(t), x_{d_z}(t)]$, the particle velocity signals received by the three directional velocity sensors along the x, y and z axis. $s_m(t)$ is the mth source signal. $\mathbf{h}_m(t)$ denotes the room impulse response (RIR) between the mth source and each individual sensor of the AVS. The additive term $\mathbf{n}(t)$ represents the additive noise.

The signal is normally processed in the STFT domain and can be represented as follows [14], [17], [18]

$$\underline{\mathbf{x}}(k,n) = \sum_{m=1}^{M}\sum_{l=0}^{L} \underline{\mathbf{h}}_m(k,l)\, \underline{s}_m(k,n-l) + \mathbf{e}(k,n) + \underline{\mathbf{n}}(k,n), \tag{2}$$

where $\underline{\mathbf{x}}(k,n)$ and $\underline{s}_m(k,n)$ is the STFT domain representation of the received signal and the mth source signal, with n and k denoting the time frame index and frequency bin index, respectively. $\underline{\mathbf{h}}_m(k,l)$ models the acoustic transfer function (ATF) between the mth source and each individual sensor of the AVS in the STFT domain. The additive term $\mathbf{e}(k,n)$ and $\underline{\mathbf{n}}(k,n)$ represents the modeling error and additive noise, respectively.

III. Review of Existing detection methods

In this section, we review two state-of-the-art DPD points detection methods, i.e., the DPD test method [11] that is a TF zone level detection method and the LRSS method [14] that is a TF point level detection method.

A. TF zone level detection

As a TF zone level detection method, the DPD test method applies the time-frequency smoothing to get the decorrelated spatial correlation matrix as shown in the following,

$$\begin{aligned}\tilde{\mathbf{R}}(k,n) = \frac{1}{J_k J_n}\sum_{j_k=0}^{J_k-1}\sum_{j_n=0}^{J_n-1} &\underline{\mathbf{x}}(k-j_k, n-j_n) \\ &\times \underline{\mathbf{x}}^H(k-j_k, n-j_n),\end{aligned} \tag{3}$$

where J_k and J_n denote the size of the frequency-domain and time-domain smoothing window, respectively. The effective rank ζ_{DPD} is calculated as the ratio of the first eigenvalue δ_1 to the second eigenvalue δ_2 of $\tilde{\mathbf{R}}$,

$$\zeta_{DPD}(k,n) = \delta_1(k,n)/\delta_2(k,n). \tag{4}$$

When the effective rank is larger than the threshold ϵ_{DPD}, the central point of the zone is detected as a single-source DPD point and used for the DOA estimation, where Γ_{DPD} denotes the set of the detected points,

$$\Gamma_{DPD} = \{(k,n) : \zeta_{DPD}(k,n) \geq \epsilon_{DPD}\}. \tag{5}$$

The above TF zone level (or a block-wise) detection methods usually detect both the DPD points and its neighborhood points for the follow-up DOA estimation, resulting in degraded estimation accuracy [14].

B. TF point level detection

As a TF point level detection method, the LRSS method compares the absolute direction of the real part and the imaginary part of the received signal in the STFT domain. The detection rule is shown as follows,

$$\zeta_{LRSS}(k,n) = \frac{\left|\Re\{\underline{\mathbf{x}}^T(k,n)\}\Im\{\underline{\mathbf{x}}(k,n)\}\right|}{\|\Re\{\underline{\mathbf{x}}^T(k,n)\}\|_2 \|\Im\{\underline{\mathbf{x}}(k,n)\}\|_2}, \tag{6}$$

which is actually the absolute cosine distance of the real part and imaginary part of the received vector signal, i.e., between $\Re\{\underline{\mathbf{x}}(k,n)\}$ and $\Im\{\underline{\mathbf{x}}(k,n)\}$. Here, $|\cdot|$ represents the absolute value, and $\|\cdot\|_2$ represents the $l2$ norm. If the absolute cosine distance ζ_{LRSS} is larger than a preset threshold ϵ_{LRSS}, the two directions are assumed to be close enough so that the corresponding TF point is determined as the LRSS point. Γ_{LRSS} denotes the set of detection points,

$$\Gamma_{LRSS} = \{(k,n) : \zeta_{LRSS}(k,n) \geq \epsilon_{LRSS}\}. \tag{7}$$

The LRSS method determines the DPD points at the time-frequency bin level, which suits the W-disjoint orthogonality assumption widely adopted in speech processing. Thus, the LRSS method can achieve more accurate detection of DPD points. However, it cannot distinguish the TF point that is due to a signal with small-level interference or two-closely spaced propagation directions, as shown in the following analysis.

IV. The proposed method

A. Reactive intensity vector

When the direct path is dominant, (2) can be approximated as

$$\underline{\mathbf{x}}(k,n) \approx \underline{s}_m(k,n) e^{-\jmath\omega_k\tau_m}\mathbf{d}_m, \tag{8}$$

in which the pressure signal and the velocity vector can be expressed as

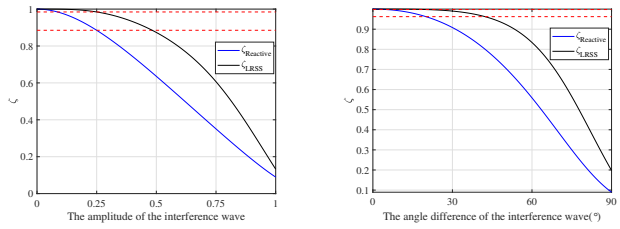
$$\underline{x}_o(k,n) = \underline{s}_m(k,n) e^{-\jmath\omega_k\tau_m} \tag{9}$$

$$[\underline{x}_{d_x}(k,n), \underline{x}_{d_y}(k,n), \underline{x}_{d_z}(k,n)] = \underline{s}_m(k,n) e^{-\jmath\omega_k\tau_m}\mathbf{u}_m. \tag{10}$$

Here, $\mathbf{u}_m = [\sin\theta_m\cos\phi_m, \sin\theta_m\sin\phi_m, \cos\theta_m]$ is a unit vector pointing to the source position, with θ_m and ϕ_m denoting the elevation and azimuth of source[1]. Then, the intensity vector can be expressed as follows

$$\begin{aligned}\mathbf{I}(k,n) &= \underline{x}_o^*(k,n)[\underline{x}_{d_x}(k,n), \underline{x}_{d_y}(k,n), \underline{x}_{d_z}(k,n)] \\ &= \underline{s}_m^*(k,n) e^{\jmath\omega_k\tau_m}\underline{s}_m(k,n) e^{-\jmath\omega_k\tau_m}\mathbf{u}_m \\ &= |\underline{s}_m(k,n)|^2\,\mathbf{u}_m\end{aligned} \tag{11}$$

[1]The azimuth angle is measured counter clockwise from the x-axis within the range of $[0^\circ, 360^\circ]$ and the elevation is measured counter clockwise from the z-axis within the range of $[0^\circ, 180^\circ]$



(a) (b)

Fig. 1. (a) Comparison of ζ_{LRSS} and $\zeta_{Reactive}$ with increasing amplitude of the interference wave. (b) Comparison of ζ_{LRSS} and $\zeta_{Reactive}$ with increasing angle difference of the interference wave from the source direction. A unit-amplitude plane wave with a direction of $\theta_d = 45°$ and $\phi_d = 45°$ is simulated.

It can be seen that when the direct path dominates, the acoustic intensity vector is a real number, in other words its imaginary part or the reactive intensity vector approximates to zero. This is due to the fact that the reactive intensity vector is perpendicular to wave propagation direction (or acoustic energy flow direction), describing the non-propagating local energy exchange that depends on the sound waves [19]. Therefore, the existence of strong direction paths leads to a large active intensity vector while the reactive intensity vector becomes very small.

Thus, in this paper, we make use of the acoustic intensity vector, i.e., the product of the conjugation pressure and particle velocity vector, as a feature for DPD point detection.

B. Reactive intensity vector based DPD points detection

The reactive-intensity based DPD detection rule is proposed as follows

$$\zeta_{Reactive}(k,n) = \frac{1}{e^{\|\Im\{\mathbf{I}(k,n)\}\|_2 / \|\Re\{\mathbf{I}(k,n)\}\|_2}}, \tag{12}$$

where $\Im\{\mathbf{I}(k,n)\}$ and $\Re\{\mathbf{I}(k,n)\}$ represent the reactive intensity vector and active intensity vector, respectively. Here, $\|\cdot\|_2$ denotes the $l2$ norm to calculate the energy of the reactive and active intensity vector. e is a natural constant, which is used to normalize the ratio between the reactive and active intensity vector within the range of 0 and 1. If the amplitude of the reactive intensity vector approaches zero, $\zeta_{Reactive}(k,n) \approx 1$, and the corresponding TF bin is determined as DPD point. With the increase of the amplitude of the reactive intensity vector, $\zeta_{Reactive}(k,n)$ becomes smaller than 1 but still larger than 0.

In this work, if the value $\zeta_{Reactive}(k,n)$ is larger than a preset threshold $\epsilon_{Reactive}$, the corresponding TF bin is detected as a DPD point. Here, $\Gamma_{Reactive}$ denotes the sets of the detected DPD points,

$$\Gamma_{Reactive} = \{(k,n) : \zeta_{Reactive}(k,n) \geq \epsilon_{Reactive}\}. \tag{13}$$

To make the proposed method robust to noise, a time-frequency domain mask based on the IMCRA method [20] is used to eliminate the low SNR points.

Both the proposed method and the LRSS method are TF point-level detection methods. To demonstrate the superiority of the proposed method, a simulation is conducted. We generate a direct path wave of unit amplitude with a direction of $\theta_d = 45°$ and $\phi_d = 45°$. In Fig. 1 (a), an interference wave with a direction of $\theta_i = 135°$ and $\phi_i = 135°$ is generated and the amplitude of the interference wave increases from 0 to 1. As shown in Fig. 1 (a), even for small amplitudes interference waves (i.e., amplitudes less than 0.25), the proposed method has a significant declining curve above the value $\zeta_{Reactive}$, compared with the value of ζ_{LRSS} used in the LRSS method. In Fig. 1 (b), an interference wave of unit amplitude is generated while its 2D angle difference from the source increases from $0°$ to $90°$. For small angle differences (i.e., angle differences less than $20°$), the same trend as in Fig. 1 (a) can be found. In conclusion, the LRSS method, which is based on (6) to detect DPD points, cannot distinguish small-level interference or closely-spaced interference. On the other hand, the proposed method has a much clearer variation to set up the threshold value.

C. DOA estimation

The intensity vector-based DOA estimation is performed from the detected DPD TF points as follows

$$\phi = \arctan\left(\frac{\Re\left\{\underline{x}_o^* \underline{x}_{d_y}\right\}}{\Re\left\{\underline{x}_o^* \underline{x}_{d_x}\right\}}\right)$$
$$\theta = \arccos\left(\frac{\Re\{\underline{x}_o^* \underline{x}_{d_z}\}}{\sqrt{\Re\{\underline{x}_o^* \underline{x}_{d_x}\}^2 + \Re\{\underline{x}_o^* \underline{x}_{d_y}\}^2 + \Re\{\underline{x}_o^* \underline{x}_{d_z}\}^2}}\right). \tag{14}$$

Given instantaneous DOA estimates from multiple detected DPD points, the fusion step is adopted to obtain the final DOA estimation result by utilizing a histogram with two-dimensional Gaussian smoothing [14], [21].

V. Experiment Results

A. Simulation setup

Simulations are conducted to evaluate the proposed method. The image source model [22] is used to generate the room impulse response. Device noise and diffuse noise are simulated to evaluate the algorithms. The T_{60} is set as 0.3s, 0.6s, 0.7s, and 0.8s, while the SNR is set as 5dB, 10dB, and 15dB. Three sizes of rooms (small, middle, and large) are simulated, with two active sources put in 20 positions with an angular spacing of $60°$. Clean speech of 15 men and 15 women from the TIMIT [23] database are used in the simulation, with a sampling rate of 16kHz and a duration of 3s. Root mean square angular error (RMSAE) [14] defined as $\sqrt{\mathrm{E}[e^2]}$ to qualify the performance across signal blocks is used as the metric to evaluate the performance of different methods. The average angular error e can be expressed as follows

$$e = \frac{1}{M}\sum_{m}^{M} 2\sin^{-1}(\|\hat{\mathbf{u}}_m - \mathbf{u}_m\|/2), \tag{15}$$

where $\hat{\mathbf{u}}_m$ and $\mathbf{u}_m$ denotes the estimated and the actual direction of the mth source, respectively. We compare our

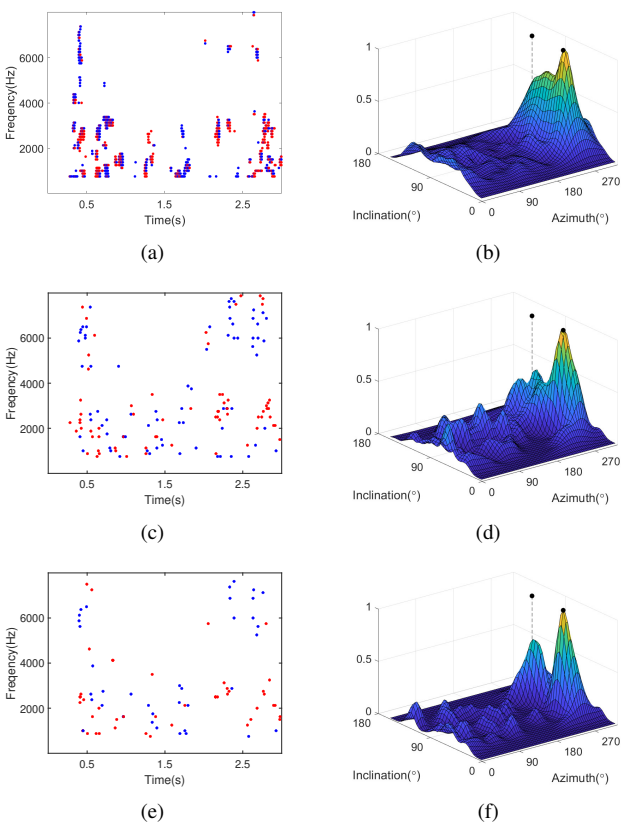


Fig. 2. Detection results of different methods at SNR=5dB and $T_{60} = 0.7s$. Two sources are located 1.5m away from the AVS and the DOA of these two sources are $\phi_1 = 300°, \theta_1 = 60°$ and $\phi_2 = 300°, \theta_2 = 120°$, respectively. The DPD method (a) and (b), the LRSS method (c) and (d), and the proposed intensity vector-based method (e) and (f).

method with the state-of-the-art methods, i.e. the DPD test method and LRSS method. The threshold for the proposed method, LRSS method, and the DPD test method is set as 0.87, 0.97, and 6, respectively. The RMSAE are shown in Table I and Table II.

B. Experiment results

First, we compare the DPD detection results of different methods at the condition of $T_{60} = 700ms$, SNR=5dB. There are two sources located 1.5m away from the AVS and the DOA of these two sources are $\phi_1 = 60°, \theta_1 = 300°$ and $\phi_2 = 120°, \theta_2 = 300°$, respectively. Fig. 2(a), (c), (e) show the detection results of the DPD method, the LRSS method, and the proposed test method, respectively. The red dots belong to the first source and the blue dots belong to the second source. Fig. 2(b), (d), (f) show the DOA results using the detected points of (a), (c), and (e), respectively. Obviously, the proposed method has two distinct peaks corresponding to the source directions. And the DOA estimation error for each method is $15.0°, 8.8°, 3.3°$, respectively. The proposed method has the lowest error. Comparing the TF zone-level method to the LRSS method and the proposed method, it is evident that the former detects results in blocks, including the DPD point and its neighboring points. On the other hand, the latter two methods perform TF point-level detections, resulting in sparser detection outcomes. The proposed method demonstrates the most precise detection results.

Table I shows RMSAE for two sources placed at 20 positions with an angular spacing of $60°$ in different environments. For each case, the result is calculated from 100 trials of signal blocks. It can be seen from the result that all the method degrades as the T_{60} increases and the SNR decreases, i.e., in a more reverberant and noisy environment. The proposed method always has the lowest RMSAE. As the room size becomes smaller, all the methods degrade due to the fact that a small size room usually has stronger reverberation under the same T_{60}. Table II shows RMSAE for three sources with an angular spacing of $40°, 60°, 80°$. Similar results as in the two-source case can be observed.

In summary, the original intensity vector (IV) method, which does not incorporate any DPD detection, has the largest error. The TF point level detection methods, i.e., the LRSS method and the proposed method, have improved performance compared with the TF zone level detection method, i.e., the DPD test method. Especially, the proposed method has the smallest error among all conditions.

VI. Conclusion

This paper proposed a multi-source DOA estimation method based on an AVS in a reverberant room environment. The reactive intensity vector is used as the feature to detect single-source direct path dominated TF points. The intensity vector based DOA estimation is then adopted, which is computationally more efficient and suits real-time applications. Experimental results confirmed the effectiveness and the superiority of the proposed method compared to the state of the art methods. Note that the detection threshold setting plays a critical role in the proposed method and requires careful consideration. Future research will aim to develop a more robust selection technique for determining the threshold.

Acknowledgment

This work was supported in part by the National Key Research and Development Program of China under Grant No. 2018AAA0102200 and in part by the Key Program of National Science Foundation of China (NSFC) under Grants 61831019, 62192713, and 62271401.

References

[1] J. Cao, J. Liu, J. Wang, and X. Lai, "Acoustic vector sensor: reviews and future perspectives," *IET Signal Process.*, vol. 11, no. 1, pp. 1-9, 2017.

[2] D. P. Jarrett, E. A. P. Habets and P. A. Naylor, "3D source localization in the spherical harmonic domain using a pseudointensity vector," in *Proc. Eur. Signal Process. Conf.*, Aalborg, Denmark, Aug. 2010, pp. 442–446.

[3] A. Nehorai and E. Paldi,"Acoustic vector-sensor array processing," *IEEE Trans. on Signal Process.*, vol. 42, no. 9, pp. 2481-2491, Sept. 1994.

[4] K. T. Wong and M. D. Zoltowski,"Self-initiating MUSIC-based direction finding in underwater acoustic particle velocity-field beamspace," *IEEE J. Ocean. Eng.*, vol. 25, pp. 262-273, 2000.

[5] A. H. Moore, C. Evers and P. A. Naylor, "Direction of arrival estimation in the spherical harmonic domain using subspace pseudointensity vectors," *IEEE/ACM Trans. on Audio, Speech, Lang. Process.*, vol. 25, no. 1, pp. 178-192, 2017.

TABLE I
RMSAE FOR THE CASE OF 2 SOURCES WITH AN ANGULAR SPACING OF 60° IN DIFFERENT CONDITIONS.

		Room size $(10 \times 8 \times 6)m^3$				Room size $(8 \times 6 \times 3)m^3$				Room size$(5 \times 4 \times 3)m^3$			
SNR	T_{60}	IV	DPD	LRSS	Proposed	IV	DPD	LRSS	Proposed	IV	DPD	LRSS	Proposed
5	0.3	15.73	10.59	5.52	**3.49**	39.98	23.06	18.58	**10.75**	40.49	25.03	23.1	**12.1**
5	0.6	31.21	22.33	13.06	**7.69**	48.38	26.38	26.72	**18.32**	49.1	27.76	34.77	**21.53**
5	0.7	33.46	20.03	13.38	**7.69**	52.2	28.61	32.48	**21.15**	50.15	32.06	37.68	**24.68**
5	0.8	34.5	20.62	14.85	**8.63**	50.54	27.89	32.25	**21.34**	50.67	32.75	39.44	**27.26**
10	0.3	12.42	9.11	**3.52**	4.39	29.34	16.41	11.3	**6.98**	37.39	16.61	13.16	**7.62**
10	0.6	27.7	17.57	8.5	**5.81**	50.33	24.13	25.28	**14.53**	48.88	22.37	28.15	**14.58**
10	0.7	28.06	17.06	7.94	**5.37**	52.1	26.09	27.94	**16.36**	49.66	26.24	32.38	**18.99**
10	0.8	30.28	16.32	11.01	**6.49**	52.17	26.14	28.26	**17.73**	55.24	27.04	33.27	**21.67**
15	0.3	10.68	7.1	2.69	**2.04**	27.22	12.85	7.95	**4.72**	34.8	14.94	11.25	**5.13**
15	0.6	22.39	12.93	5.78	**2.3**	46.6	25.55	20.74	**11.76**	51.5	17.99	24.49	**13.91**
15	0.7	24.3	13.8	5.41	**3.2**	48.14	23.07	25.43	**14.23**	52.81	21.27	29.33	**16.16**
15	0.8	27.03	12.67	5.45	**2.81**	54.35	28.01	26.18	**15.34**	54.26	24.33	30.46	**18.28**

TABLE II
RMSAE FOR THE CASE 3 SOURCES IN A ROOM SIZE OF $(10 \times 8 \times 6)m^3$ WITH AN ANGULAR SPACING OF $40^\circ, 60^\circ, 80^\circ$.

		Angle spacing 40°				Angle spacing 60°				Angle spacing 80°			
SNR(dB)	$T_{60}(s)$	IV	DPD	LRSS	Proposed	IV	DPD	LRSS	Proposed	IV	DPD	LRSS	Proposed
5	0.3	44.09	39.54	17.88	**16.71**	21.4	15.8	8.19	**7.37**	7.39	7.29	5.62	**5.31**
5	0.6	48.5	42.09	19.33	**17.99**	26.08	24.6	15.96	**12.25**	19.76	17.56	9.75	**7.69**
5	0.7	45.72	40.59	19.95	**16.17**	28.75	24.36	15.53	**12.95**	21.69	18.77	16.24	**11.48**
5	0.8	45.23	40.3	19.85	**16.3**	31.96	23.73	17.13	**13.33**	25.44	21.27	20.46	**10.72**
10	0.3	38.41	34.56	**11.42**	13.07	12.9	8.77	5.46	**4.89**	6.77	5.9	5.11	**4.85**
10	0.6	46.98	42.09	19.16	**15.16**	28.12	17.93	8.7	**8.24**	15.07	8.44	9.77	**6.05**
10	0.7	45.14	38.24	18.62	**15.22**	28.82	18.02	8.45	**8.32**	16.63	11.62	9.7	**6.63**
10	0.8	46.78	40.78	18.64	**15.42**	29.57	18.77	9.81	**8.59**	20.73	14.96	10.14	**7.32**
15	0.3	33.81	35.09	**8.73**	11.63	11.67	6.47	4.61	**4.24**	5.81	5.17	4.69	**4.32**
15	0.6	44.39	39.11	17.1	**15.29**	25.27	12.6	6.77	**5.75**	15.51	7.67	6.23	**5.54**
15	0.7	46.41	40.59	18.69	**14.49**	26.32	12.88	6.95	**6.36**	16.04	7.18	7.06	**5.55**
15	0.8	47.18	44.3	20.33	**16.74**	28.38	14.5	7.61	**7.41**	18.26	7.79	5.49	**5.27**

[6] A. Herzog and E. A. P. Habets, "Eigenbeam-ESPRIT for DOA-vector estimation," *IEEE Signal Process. Let.*, vol. 26, no. 4, pp. 572-576, Apr. 2019.

[7] A. Herzog and E. A. P. Habets, "On the relation between DOA-vector eigenbeam ESPRIT and subspace pseudointensity-vector," in *Proc. Eur. Signal Process. Conf.*, Corua, Spain, Sep. 2019, pp. 1–5.

[8] D. Levin, S. Gannot and E. A. P. Habets, "Direction-of-arrival estimation using acoustic vector sensors in the presence of noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2011, pp. 105–108.

[9] D. Levin, E. A. P. Habets, and S. Gannot, "Maximum likelihood estimation of direction of arrival using an acoustic vector sensor,"*J. Acoust. Soc. Am.*, vol. 131, no. 2, pp. 1240–1248, 2012.

[10] S. Mohan, M. E. Lockwood, M. L. Kramer, and D. L. Jones, "Localization of multiple acoustic sources with small arrays using a coherence test," *J. Acoust. Soc. Am.*, vol. 123, no. 4, pp. 2136-2147, 2008.

[11] O. Nadiri and B. Rafaely, "Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 10, pp. 1494–1505, Oct. 2014.

[12] A. H. Moore, C. Evers, P. A. Naylor, D. L. Alon and B. Rafaely, "Direction of arrival estimation using pseudo-intensity vectors with direct-path dominance test," in *Proc. Eur. Signal Process. Conf.*, 2015, pp. 2296-2300.

[13] L. Madmoni and B. Rafaely, "Direction of arrival estimation for reverberant speech based on enhanced decomposition of the direct sound," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 1, pp. 131–142, Dec. 2018.

[14] K. Wu, G. V. Reju and A. W. H. Khong, "Multisource DOA estimation in a reverberant environment using a single acoustic vector sensor," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 10, pp. 1848–1859, Oct. 2018.

[15] M. Jia, Y. Wu, C. Bao and C. Ritz, "Multi-source DOA estimation in reverberant environments by jointing detection and modeling of time-Frequency Points," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 379-392, 2020.

[16] D. P. Jarrett, E. A. P. Habets and P. A. Naylor, *Theory and applications of spherical microphone array processing.*, Berlin, Germany: Springer, 2016.

[17] A. Jukić, T. van Waterschoot, T. Gerkmann and S. Doclo, "Multi-channel linear prediction-based speech dereverberation with sparse priors," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 9, pp. 1509-1520, 2015.

[18] J. Geng, S. Wang, J. Li, X. Zhang, X. Luo,"Robust multi-source direction of arrival estimation using a single acoustic vector sensor", in *Proc. APCCAS*, 2021, pp. 161-164.

[19] J. Daniel and S. Kitić, "Time domain velocity vector for retracing the multipath propagation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 421–425

[20] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *" IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 466–475, Nov. 2003.

[21] D. Pavlidi, A. Griffin, M. Puigt and A. Mouchtaris, "Real-time multiple sound source localization and counting using a circular microphone array", *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 10, pp. 2193-2206, Oct. 2013.

[22] E. A. P. Habets, "Room impulse response generator," 2010. [Online]. Available:https://www.researchgate.net/publication/259991276_Room_Impulse_Response_Generator.

[23] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "DARPA TIMIT acoustic-phonetic continuous speech corpus CDROM. NIST speech disc 1-1.1," *NASA STI/Recon Techn. Rep. N*, vol. 93., 1993.