AN EFFECTIVE DOA ESTIMATION BY EXPLORING THE SPATIAL SPARSE REPRESENTATION OF THE INTER-SENSOR DATA RATIO MODEL

Yuexian Zou[1], *Yifan Guo*[1], *Weiqiao Zheng*[1], *C. H. Ritz*[2] and *Jiangtao Xi*[2]

[1]ADSPLAB, School of Electronic Computer Engineering, Peking University, Shenzhen 518055, China
[2]School of Electrical, Computer, and Telecommunications Engineering, University of Wollongong, Australia
{ zouyx@pkusz.edu.cn, guoyifan@sz.pku.edu.cn, critz@uow.edu.au and jiangtao@uow.edu.au }

ABSTRACT

[1]This paper investigates speaker direction of arrival (DOA) estimation using a single acoustic vector sensor (AVS). With the definition of the inter-sensor data ratio (ISDR) in the time-frequency (TF) domain and the use of the high local signal-to-noise ratio (HLSNR) TF points, an effective ISDR data model is derived, which determines the relationship between the ISDR and the AVS manifold vector. With the spatial sparse representation of the ISDR data, the DOA estimation is formulated by recovering the sparse matrix and locating the peak of the power spectrum of the reconstructed sparse matrix. Preliminary experimental results using simulations and real AVS recordings show that the proposed DOA estimation method is able to achieve high elevation and azimuth estimation accuracy for all angles when the SNR is above 10dB, avoiding the spatial aliasing problem and suppressing the adverse impact of the room reverberation. It is expected that the proposed DOA estimation method may find wide applications in portable devices due to its small compact physical size and superior performance.

Index Terms—Direction of arrival estimation, acoustic vector sensor, spatial sparse representation, inter-sensor data ratio, time-frequency sparsity

1. INTRODUCTION

Direction of arrival (DOA) estimation of the spatial speech source is a key technique in hands-free communication applications such as the audition system of the service robot, which is of significant application value. Compared to common microphone array based techniques for DOA estimation [1] , the acoustic vector sensor (AVS) has a smaller size and a spatial compact structure making it attractive for mobile speech applications [2][3].

In our previous work [4], the high resolution DOA estimation using AVS array under a spatial sparsity representation (SSR) framework was developed by making use of the relationship between the received data model of the AVS array and its subarray manifold. The proposed DOA estimation algorithm provided better DOA estimation accuracy. In [4], 8 AVS units with spacing of half of the source wavelength are used for data capture, which limits its applications when size is the main concern. To reduce the size of the data capture system, a single AVS based DOA estimation system and algorithm have been developed [5], which is able to estimate the DOAs of the multiple spatial speech sources by exploiting the time-frequency (TF) sparsity of the speech and the DOA information provided by the inter-sensor data ratios (ISDRs) of AVS sensors. As a result, the trigonometric relations between the ISDRs at the TF points with high local SNR (HLSNR TF points) and the DOAs have been established and the DOAs can be obtained by estimating the mean values of ISDR data using a clustering method [5]. In practice, the ISDRs extracted at the HLSNR TF points have increased variance compared to ideal conditions due to corruption by strong noise or competing speech and room reverberation. Hence, DOA estimation using the clustering method may be biased away from the true DOA. Highly reverberant environment is a serious problem for many existing DOA estimation methods, where the DOA estimation accuracy will be degraded [6].

In this paper, we propose a new approach using the ISDR data for DOA estimation. The motivation lies on four aspects: 1) The DOA estimation methods developed under the SSR framework usually are able to estimate the elevation and azimuth angle at the same time and achieve higher DOA estimation accuracy than traditional DOA estimation methods [5][7]. 2) Using the HLSNR TF point extraction strategy, the ISDR data model can be formulated to ignore the impact of the room reverberation, which essentially leads to the superior performance under heavy room reverberation. 3) ISDRs of an AVS are independent of the source frequencies, so there will be no need to consider the spatial aliasing problem; 4) A spatial sparse representation can be formulated with the ISDR data model and the solution can be obtained by the well-known l_1-SVD method [8]. The rest of this paper is organized as follows. The problem formulation is introduced in Section 2. The proposed DOA estimation algorithm is presented in Section 3. The experiments and results are given in Section 4 and the conclusions are drawn in Section 5.

2. PROBLEM FORMULATION

This section presents the AVS data model and proposes a robust DOA estimation approach using a single AVS based on both TF domain sparsity and spatial sparsity.

2.1. AVS Data Mode

[1] This work is partially supported by National Natural Science Foundation of China (No: 61271309) and the Shenzhen Science & Technology Fundamental Research Program (No: JC201105170727A)

Generally each AVS unit consists of an omnidirectional sensor (*o*-sensor) and three orthogonally oriented directional sensors (named as *u*-, *v*-, *w*-sensor, respectively), where particle velocity sensors or differential microphones are often used as the directional sensors according to [9]. In this study, to address the problem formulation, the DOA estimation of one spatial speaker is investigated. When the speech signal $s(t)$ impinges upon an AVS with DOA (θ_s,ϕ_s), the manifold vector can be denoted as [9]

$$\boldsymbol{a}(\theta_s,\phi_s)=[u_s,v_s,w_s,1]^T,\boldsymbol{a}\in R^{4\times 1} \quad (1)$$

where $[.]^T$ denotes the vector/matrix transposition, $\theta_s\in[0,180°]$, $\phi_s\in[0,360°)$ are the elevation and the azimuth angle, respectively. Elements u_s, v_s and w_s are respectively the *x*-, *y*-, *z*-axis direction cosines given by:

$$u_s=\sin\theta_s\cos\phi_s, v_s=\sin\theta_s\sin\phi_s, w_s=\cos\theta_s \quad (2)$$

The data captured by the AVS at time t is expressed as:

$$x_u(t)=u_s s(t)*h(t)+n_u(t) \quad (3)$$

$$x_v(t)=v_s s(t)*h(t)+n_v(t) \quad (4)$$

$$x_w(t)=w_s s(t)*h(t)+n_w(t) \quad (5)$$

$$x_o(t)=s(t)*h(t)+n_o(t) \quad (6)$$

where $x_u(t)$, $x_v(t)$, $x_w(t)$ and $x_o(t)$ represent the output of the *u*-, *v*-, *w*- and *o*-sensor, respectively. $h(t)$ represents the room impulse response. $n_u(t)$, $n_v(t)$, $n_w(t)$ and $n_o(t)$ are the additive zero-mean Gaussian noise at the *u*-, *v*-, *w*- and *o*-sensor, respectively, which are assumed uncorrelated to each other, and uncorrelated to the speech signal.

2.2 Data Model of Inter-Sensor Data Ratio

It is widely accepted that speech signals have sparsity in the TF domain [10]. This indicates that only one speech source with the highest energy dominates at a specific TF point (τ, ω) while the contributions from other speech sources can be negligible. With this assumption, the ISDRs of the AVS defined in the frequency domain can be expressed as follows [5]:

$$I_{uo}(\tau,\omega)=X_u(\tau,\omega)/X_o(\tau,\omega) \quad (7)$$

$$I_{vo}(\tau,\omega)=X_v(\tau,\omega)/X_o(\tau,\omega) \quad (8)$$

$$I_{wo}(\tau,\omega)=X_w(\tau,\omega)/X_o(\tau,\omega) \quad (9)$$

where $I_{uo}(\tau,\omega)$, $I_{vo}(\tau,\omega)$ and $I_{wo}(\tau,\omega)$ are the ISDRs between *u*- and *o*-sensor, *v*- and *o*-sensor, *w*- and *o*-sensor, respectively. $X_u(\tau,\omega)$, $X_v(\tau,\omega)$, $X_w(\tau,\omega)$, and $X_o(\tau,\omega)$ are the short-time Fourier transform (STFT) of (3)-(6):

$$X_u(\tau,\omega)=u_s S(\tau,\omega)H(\omega)+N_u(\tau,\omega) \quad (10)$$

$$X_v(\tau,\omega)=v_s S(\tau,\omega)H(\omega)+N_v(\tau,\omega) \quad (11)$$

$$X_w(\tau,\omega)=w_s S(\tau,\omega)H(\omega)+N_w(\tau,\omega) \quad (12)$$

$$X_o(\tau,\omega)=S(\tau,\omega)H(\omega)+N_o(\tau,\omega) \quad (13)$$

where $H(\omega)$ is the Fourier transform of $h(t)$.

Taking $I_{uo}(\tau,\omega)$ as an example, the relation between the ISDR and the DOA of the speaker can be derived as follows. Substituting (10) and (13) into (7), we obtain

$$I_{uo}(\tau,\omega)=\frac{u_s S(\tau,\omega)H(\omega)+N_u(\tau,\omega)}{S(\tau,\omega)H(\omega)+N_o(\tau,\omega)}=\frac{u_s+N_u(\tau,\omega)/[S(\tau,\omega)H(\omega)]}{1+N_o(\tau,\omega)/[S(\tau,\omega)H(\omega)]} \quad (14)$$

Simply, Eqn. (14) can be rewritten as follows

$$I_{uo}(\tau,\omega)=\alpha(\tau,\omega)u_s+\eta_u(\tau,\omega) \quad (15)$$

with

$$\alpha(\tau,\omega)=\frac{1}{1+N_o(\tau,\omega)/[S(\tau,\omega)H(\omega)]} \quad (16)$$

$$\eta_u(\tau,\omega)=\frac{N_u(\tau,\omega)/[S(\tau,\omega)H(\omega)]}{1+N_o(\tau,\omega)/[S(\tau,\omega)H(\omega)]} \quad (17)$$

Eqn. (15) can be viewed as the data model of the ISDR between *u*- and *o*-sensor. Similarly, the data models of ISDRs between *v*- and *o*-sensor, *w*- and *o*-sensor can be modeled as follows, respectively.

$$I_{vo}(\tau,\omega)=\alpha(\tau,\omega)v_s+\eta_v(\tau,\omega) \quad (18)$$

$$I_{wo}(\tau,\omega)=\alpha(\tau,\omega)w_s+\eta_w(\tau,\omega) \quad (19)$$

where

$$\eta_v(\tau,\omega)=\frac{N_v(\tau,\omega)/[S(\tau,\omega)H(\omega)]}{1+N_o(\tau,\omega)/[S(\tau,\omega)H(\omega)]} \quad (20)$$

$$\eta_w(\tau,\omega)=\frac{N_w(\tau,\omega)/[S(\tau,\omega)H(\omega)]}{1+N_o(\tau,\omega)/[S(\tau,\omega)H(\omega)]} \quad (21)$$

To simplify the notation, the data model of ISDRs can be expressed in a compact form:

$$\boldsymbol{I}(\tau,\omega)=\alpha(\tau,\omega)\boldsymbol{b}(\theta_s,\phi_s)+\boldsymbol{\varepsilon}(\tau,\omega) \quad (22)$$

where

$$\boldsymbol{I}(\tau,\omega)=[I_{uo}(\tau,\omega),I_{vo}(\tau,\omega),I_{wo}(\tau,\omega)]^T \quad (23)$$

$$\boldsymbol{b}(\theta_s,\phi_s)=[u_s,v_s,w_s]^T \quad (24)$$

$$\boldsymbol{\varepsilon}(\tau,\omega)=[\eta_u(\tau,\omega),\eta_v(\tau,\omega),\eta_w(\tau,\omega)]^T \quad (25)$$

From (24), it can be seen that $\boldsymbol{b}(\theta_s, \phi_s)$ is the manifold vector of the *u*-, *v*-and *w*-sensor. The DOA estimation can be achieved by estimating u_s, v_s and w_s from (22). In [5], a DOA estimation method based on clustering has been developed with the assumption of an anechoic environment.

In this paper, we propose a novel method to estimate DOA with (22). Firstly, it is noted that Eqn. (22) is valid for each TF point. For the DOA estimation task, it is beneficial to use the HLSNR TF points, where the effects of the additive noise component $\boldsymbol{\varepsilon}(\tau,\omega)$ will be smaller. Secondly, the HLSNR TF points for speech signals can be effectively extracted by the Sinusoidal tracks extraction (SinTrE) method [11]. In our study with an AVS, the HLSNR TF points are estimated by using $X_o(\tau,\omega)$. From (13), assuming (τ,ω) is a HLSNR TF point extracted by SinTrE, then we have $S(\tau,\omega)H(\omega)>>N_o(\tau,\omega)$. From (16), we can get $\alpha(\tau,\omega)\approx 1$. Accordingly, from (22), the data model of the ISDRs can be reformulated as

$$\boldsymbol{I}(\tau,\omega)=\boldsymbol{b}(\theta_s,\phi_s)+\boldsymbol{\varepsilon}_1(\tau,\omega) \quad (26)$$

where $\boldsymbol{\varepsilon}_1(\tau,\omega)$ can be viewed as the residual error caused by additive Gaussian noise, room reverberation, and SSR model mismatch.

2.3. Spatial Sparse Representation Model of ISDRs

In (26), we have established the relation between the ISDRs of the AVS and its direction manifold vector $\boldsymbol{b}(\theta_s, \phi_s)$. In the following subsection, a novel DOA estimation method

under the SSR framework by using the ISDR data model in (26) will be presented in details.

Firstly, the azimuth angle range and the elevation angle range are uniformly divided into N_1 and N_2 grids (N_1, N_2 >>1), respectively. As a result, the whole spatial space is divided by M grids (M=N_1×N_2). Moreover, N_1 and N_2 can be selected corresponding to different application requirements. Accordingly, a predefined angle set Θ= {(θ_i,ϕ_j), i=1,…,N_1, j=1,…,N_2 } is formed. Correspondingly, an overcomplete manifold matrix of the u-, v- and w-sensor can be constructed according to Θ:

$$\boldsymbol{\Psi} = \left[\boldsymbol{b}(\theta_1,\phi_1),\ldots,\boldsymbol{b}(\theta_i,\phi_j),\ldots,\boldsymbol{b}(\theta_{N_1},\phi_{N_2})\right], \boldsymbol{\Psi} \in R^{3\times M} \quad (27)$$

where $\boldsymbol{b}(\theta_i, \phi_j)$ is given in (24). Therefore, with the assumption of a sufficiently small grid spacing, using $\boldsymbol{\Psi}$ instead of $\boldsymbol{b}(\theta_s,\phi_s)$, the ISDR data model in (26) can be re-formulated as follows:

$$\boldsymbol{I}(\tau,\omega) = \boldsymbol{\Psi z} + \boldsymbol{\varepsilon}_1(\tau,\omega), \boldsymbol{\Psi} \in R^{3\times M}, \boldsymbol{z} \in R^{M\times 1} \quad (28)$$

where $\boldsymbol{z}$ is called sparse vector and there is only one nonzero element corresponding to the DOA (θ_s,ϕ_s). Essentially, (28) is the spatial sparse representation model derived from (22) and it is termed the AVS-ISDR-SSR model. Therefore, the estimation of DOA (θ_s,ϕ_s) can be achieved by locating the nonzero element in the reconstructed $\boldsymbol{z}$. Moreover, the DOA estimation accuracy is affected by the choice of N_1 and N_2. Obviously, a larger value of N_1 or N_2 leads to a smaller grid spacing. This gives higher probability of matching the true DOA of the speaker with the predefined angle in set Θ.

To make the presentation clear, we denote the number of extracted HLSNR TF points as L. Besides, it is noted that, for each HLSNR TF point, the sparse vector $\boldsymbol{z}$ maintains the same sparse structure given in (28). Utilizing this property, we form a joint SSR-ISDR model as follows

$$\boldsymbol{A} = \boldsymbol{\Psi Z} + \boldsymbol{E} \quad (29)$$

$$\boldsymbol{A} = [\boldsymbol{I}(\tau_1,\omega_1),\ldots,\boldsymbol{I}(\tau_L,\omega_L)], \boldsymbol{A} \in R^{3\times L} \quad (30)$$

$$\boldsymbol{E} = [\boldsymbol{\varepsilon}_1(\tau_1,\omega_1),\ldots,\boldsymbol{\varepsilon}_1(\tau_L,\omega_L)], \boldsymbol{E} \in R^{3\times L} \quad (31)$$

$$\boldsymbol{Z} = [z_1,\ldots,z_L], \boldsymbol{Z} \in R^{M\times L} \quad (32)$$

where (τ_i,ω_i) (i=1,…,L) denotes the i^{th} HLSNR TF point extracted by the SinTrE method [11]. The vector $\boldsymbol{z}_i$ represents the sparse vector associated with the i^{th} HLSNR TF point (τ_i,ω_i), which satisfies the relation in (28). In (29), for estimating the DOA of a speaker, we can see that the matrix $\boldsymbol{Z}$ should have only one nonzero row corresponding to the DOA (θ_s,ϕ_s). As a result, the DOA (θ_s,ϕ_s) estimation problem now turns to locating the index of the nonzero row of the reconstructed matrix $\hat{\boldsymbol{Z}}$ via (29).

3. THE PROPOSED DOA ESTIMATION ALGORITHM

Research shows that the sparse matrix $\boldsymbol{Z}$ in (29) can be recovered by solving the following optimization problem

$$\hat{\boldsymbol{Z}} = \arg\min_{\boldsymbol{Z}} \left\|\boldsymbol{A} - \boldsymbol{\Psi Z}\right\|_2^2 + \lambda \left\|\boldsymbol{Z}\right\|_1 \quad (33)$$

where the l_2-term forces the reconstruction error to be small, whereas the l_1-term enforces sparsity of the representation. The regularization parameter λ controls the tradeoff between the sparsity and the reconstruction error. It is a commonly used approach to employ the l_1-SVD technique [8] to solve (33). In our study (DOA estimation), λ is selected as 30 according to several experiments, which emphasizes the spatial sparsity and suppress the adverse impact of the noise on the reconstruction error. The merits of l_1-SVD technique lie on its computational efficiency and robustness to noise. In our study, the sparse matrix $\boldsymbol{Z}$ in (33) is estimated by using the optimization software CVX [12]. For DOA estimation, we compute the following:

$$P_{\boldsymbol{Z}}(i) = 10\log\sum_{j=1}^{L}\hat{\boldsymbol{Z}}^2(i,j), i = 1,\ldots,M \quad (34)$$

$$i_p = \arg\max_i P_{\boldsymbol{Z}}(i) \quad (35)$$

where i_p represents the index of the dominant row of the sparse matrix $\hat{\boldsymbol{Z}}$, which is determined by locating the peak of the $\boldsymbol{Pz}$. Hence, the index (i,j) of the grid corresponds to the estimated DOA can be computed from i_p and we have

$$\hat{\theta}_s = \theta_i, \hat{\phi}_s = \phi_j, where\ \theta_i, \phi_j \in \Theta \quad (36)$$

To simplify the notation in the following context, the proposed DOA estimation algorithm is termed as the **AVS-ISDR-SSR** algorithm, which addresses the DOA algorithm developed under the SSR framework with the ISDR data model using a single AVS. The AVS-ISDR-SSR algorithm is summarized as follows:

1) Segment $x_u(t)$, $x_v(t)$, $x_w(t)$ and $x_o(t)$; Calculate the STFT;
2) Extract L HLSNR TF points [11];
3) Get the ISDRs by (7)-(9) for L HLSNR TF points ;
4) Construct the data matrix $\boldsymbol{A}$ in (30);
5) Construct the overcomplete manifold matrix $\boldsymbol{\Psi}$ in (27);
6) Utilize l_1-SVD technique to determine $\hat{\boldsymbol{Z}}$ in (33);
7) Compute $\boldsymbol{Pz}$ and i_p by (34) and (35);
8) Compute the grid index from i_p and obtain the estimated DOA by (36).

4. EXPERIMENTAL RESULTS

In this section, the performance of our proposed AVS-ISDR-SSR algorithm is evaluated and compared with that of the GMDA-Laplace method [10]. The simulation parameters are set as follows: 1) 3 seconds of male speech sampled at 32kHz; 2) A 1024-point DFT using a Hamming window of 30ms duration and 20ms overlapping. For the AVS-ISDR-SSR algorithm, θ_s∈[0,180°], ϕ_s∈[0,180°], N_1=N_2=180. For the GMDA-Laplace algorithm, following the setup in [10], two microphones are placed along the z-axis with 8cm spacing. The absolute error (AER) and the root mean squared error (RMSE) are taken as the performance metrics, which are defined respectively as

$$AER = (|\hat{\theta} - \theta| + |\hat{\phi} - \phi|)/2 \quad (37)$$

$$RMSE = 0.5\sqrt{\sum_{i=1}^{N_T}\left((\hat{\theta}_i - \theta)^2 + (\hat{\phi}_i - \phi)^2\right)/N_T} \quad (38)$$

where N_T is the number of independent trails.

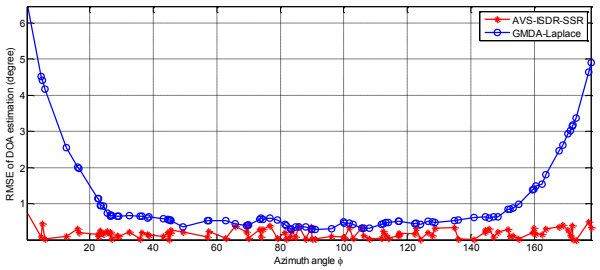

Fig. 1 RMSE versus different source DOA

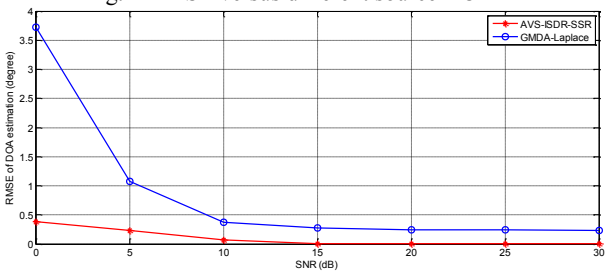

Fig. 2 RMSE versus different SNR

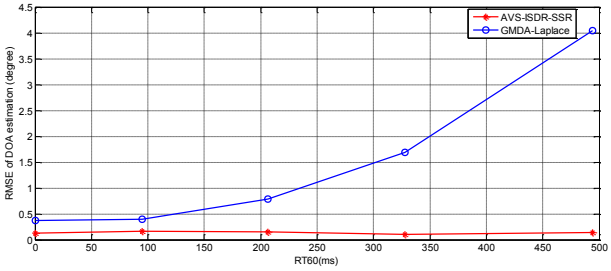

Fig. 3 RMSE versus RT_{60}

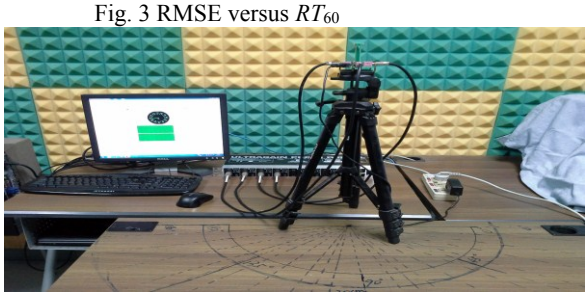

Fig 4. Experimental setup in real scenario

Experiment* 1: DOA estimation accuracy at different angles*

This simulation has been performed to test the DOA estimation accuracy of our proposed AVS-ISDR-SSR algorithm at different angles, where SNR=10dB without reverberation, θ=60° and ϕ is randomly generated between 0° to 180° for each trial. With this setup, the DOA of the speaker has high probability of mismatch with the grids in Θ and may cover the full range from 0° to 180°. The AER is obtained for 100 different trials and the simulation results are plotted in Fig. 1. It is clear that the AER of AVS-ISDR-SSR is superior to that of the GMDA-Laplace for all angles, especially when the DOA at the range of 0°-20° and 160°-180°. This result shows that our proposed AVS-ISDR-SSR algorithm is able to achieve an RMSE of about 0.5° under this simulation condition.

Experiment* 2: RMSE versus different noise levels*

The simulation aimed at evaluating the robustness of the AVS-ISDR-SSR to additive noise without reverberation. θ_s=60°, ϕ_s=45°, SNR varies from 0dB to 30dB. The RMSE results shown in Fig. 2 are obtained by 100 independent trials (N_T=100) for each SNR. It can be seen that the RMSE of AVS-ISDR-SSR is much smaller than that of GMDA-Laplace method for all SNR values. It is encouraging to see that when the SNR<5dB, the RMSE of AVS-ISDR-SSR is about 1° and when SNR>15dB, the RMSE of AVS-ISDR-SSR goes to 0°. It shows that our proposed AVS-ISDR-SSR is not sensitive to additive noise and is able to obtain good DOA estimation accuracy when SNR> 5dB.

Experiment* 3: RMSE versus different reverberation levels*

In this experiment, the behavior of the AVS-ISDR-SSR under different reverberation levels is evaluated. The experimental setup is as follows: The room impulse response is simulated by the image method [13] with the virtual room size of 10×5×4 m^3. Five different reverberation times (RT_{60}) were simulated. The distance between the speaker and AVS is 1m. The DOA of the speaker is set as θ_s=60°, ϕ_s=45° and the SNR=10dB. The results averaged over 100 trails for each RT_{60} are shown in Fig. 3. We can see clearly that the curve of the AVS-ISDR-SSR is approximately constant for all RT_{60} conditions. This indicates that our proposed algorithm is not sensitive to room reverberation, which is a very favorable property since many existing DOA estimation algorithms perform badly when heavy room reverberation exists.

***Experiment* 4:** *DOA estimation in a real scenario*

In this experiment, we evaluate the performance of the AVS-ISDR-SSR algorithm in a real scenario using the recorded data by the AVS data capturing system developed in the ADSPLAB (refer to Fig. 4) [5]. The parameters are as follows: the room is about 8.5×3×5 m^3 and an uncontrolled acoustic environment with background noise and reverberation is present. The SNR measured is approximately 20dB. The distance between the speaker and the AVS is 0.5m. The sampling rate is 32kHz, the 1024-point STFT is used, θ_s=90°, we set 5 different azimuth (°): ϕ_s =0, 45, 90, 135, and 180, respectively. The estimated DOAs are (87,4), (90,41), (92,89), (86,135) and (86,180). Obviously, the maximum DOA estimation error is about 4º. These preliminary experimental results further validate the assumptions and derivation of our proposed method.

5. CONCLUSION

In this paper, a novel DOA estimation algorithm (termed as the AVS-ISDR-SSR) has been developed under the spatial sparse representation framework and the speech TF sparsity together with the ISDR data model of a single AVS. Extensive experiments have been carried out with simulated and recorded data. The preliminary results show that the AVS-ISDR-SSR is able to achieve high DOA estimation accuracy compared to existing approaches. It is encouraging to see that the AVS-ISDR-SSR is not sensitive to the room reverberation and additive noise, which are desired properties for possible real applications. Future work will focus on the theoretical analysis of the performance.

6. REFERENCES

[1] M. Hawkes and A. Nehorai, “Acoustic vector-sensor beamforming and Capon direction estimation,” *IEEE Trans. Signal Process*, vol. 46, no. 9, pp. 2291–2304, Sept. 1998.

[2] M. E. Lockwood, D. L. Jones, "Beamformer performance with acoustic vector sensors in air," *The Journal of the Acoustical Society of America*, vol. 119, no. 1, pp. 608-619, Jan. 2006.

[3] M. Shujau, C. H. Ritz, I. S. Burnett, "Designing Acoustic Vector Sensors for localization of sound sources in air," *European Signal Processing Conference,* pp. 849-853, Aug. 2009

[4] B. Li and Y. X. Zou, "Improved DOA estimation with acoustic vector sensor arrays using spatial sparsity and subarray manifold," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2557-2560, Mar. 2012.

[5] Y. X. Zou, W. Shi, B. Li, et al, "Multisource DOA estimation based on time-frequency sparsity and joint inter-sensor data ratio with single acoustic vector sensor," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4011-4015, May 2013.

[6] J. Benesty, J. Chen, and Y. Huang, “*Microphone Array Signal Processing*”, Springer, 2008.

[7] J. Zheng and M. Kaveh, "Direction-of-arrival estimation using a sparse spatial spectrum model with uncertainty," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.2848-2851, May 2011.

[8] D. Malioutov, M. Cetin, and A. Willsky, “A sparse signal reconstruction perspective for source localization with sensor arrays,” *Signal Processing, IEEE Transactions on*， vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[9] K. T. Wong, M. D. Zoltowski, "Closed-form underwater acoustic direction-finding with arbitrarily spaced vector hydrophones at unknown locations," *Oceanic Engineering, IEEE Journal of*, vol. 22, no. 3, pp. 566-575, July 1997.

[10] W. Zhang and B. D. Rao, "A two microphone-based approach for source localization of multiple speech sources," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 1913-1928, Nov. 2010.

[11] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 34, pp. 744-754, 1986.

[12] M. Grant and S. Boyd, "CVX: MATLAB Software for Disciplined Convex Programming," [Online]. Available: http://cvxr.com/ , Apr. 2010.

[13] J. B. Allen, D. A. Berkley, "Image method for efficiently simulating small room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943-950, Apr. 1979.