ACOUSTIC DIRECTION FINDING IN HIGHLY REVERBERANT ENVIRONMENT WITH SINGLE ACOUSTIC VECTOR SENSOR

Metin Aktas, Toygar Akgun, Huseyin Ozkan

ASELSAN, Ankara, Turkey
{maktas,takgun,huozkan}@aselsan.com.tr

ABSTRACT

We propose a novel wideband acoustic direction finding method for highly reverberant environments using measurements from a single Acoustic Vector Sensor (AVS). Since an AVS is small in size and can be effectively used within the full acoustic frequency bands, the proposed solution is suitable for wideband acoustic source localization. In particular, we introduce a novel approach to extract the signal portions that are not distorted with multipath signals and noise. We do not make any stochastic and sparseness assumptions regarding the underlying signal source. Hence, our approach can be applied to a wide range of wideband acoustic signals. We present experiments with acoustic signals that are specially exposed to long reverberations, where the Signal-to-Noise Ratio is as low as 0 dB. In these experiments, the proposed method reliably estimates the source direction with less than 5 degrees of error even under the introduced significantly high reverberation conditions.

Index Terms**—** Acoustic Vector Sensor, Under-determined Direction Finding, Reverberation, Time-Frequency Analysis.

1. INTRODUCTION

In real communication environments, the transmitted signal is often received via multiple paths due to reflection, diffraction, and scattering by objects in the transmission medium [1]. In such cases, each path is considered to generate a virtual source that is highly correlated with the actual source [2], which degrades the detectability of the actual source, i.e., it generates -in a sense- a concealing effect. The signal power and total number of these virtual sources are directly related to the reverberation in the medium. For instance, as the reverberation increases, the total number of incident signals including the real and virtual sources is more likely to exceed the number of sensors. Under these adverse conditions, namely, under high reverberation, we study the acoustic source localization problem and propose a novel and robust direction of arrival (DOA) estimation method without any stochastic assumptions regarding the signal source.

When the total number of sources (including the virtual ones) exceeds the number of acoustic sensors, the actual source localization can be performed by solving an under-determined DOA estimation problem. Since the acoustic signals are typically wideband, this problem is addressed by analyzing the signals in time-frequency domain. Several methods have been proposed to determine the time-frequency bins with single active source and estimate the DOAs only at those bins [3–5]. In this case, the under-determined DOA estimation problem becomes an over-determined DOA estimation problem at specific time-frequency bins. The most critical part in this respect is to correctly determine the time-frequency bins, where only a single source is active. In [5], subspace based time-frequency selection procedure is proposed with the assumption that signals and noise are statistically independent. In [3], only the most powerful source is assumed to be active in the time-frequency bins with high signal power. The methods in [3] and [5] are jointly used for single source selection in [4]. All of these selection methods strictly depend on the time-frequency sparseness of source signal and are mostly applied to speech sources. However, the assumptions used in these methods are not valid under high reverberation, where the DOA estimation performances are known to seriously degrade [6, 7].

On the contrary, we propose a novel 2-D robust wideband acoustic source localization method for highly reverberant environment, which utilizes a single Acoustic Vector Sensor (AVS) that is composed of only 4 sensors [8]. In the proposed method, the time-frequency bin selection for a single active source is performed without any assumption about the signal power distribution and signal statistics. Our approach performs the DOA estimation and the time-frequency bin selection jointly in one framework. We first assume that there is only a single active source in the medium and estimate DOA at each time-frequency bin with a computationally efficient direct solution. Then, we select the time-frequency bins, where the estimated DOA is consistent with the received sig-

M. Aktas and T. Akgun are in Transportation, Security, Energy & Automation Systems Business Sector (UGES) in ASELSAN.

H. Ozkan is in Microelectronics, Guidance & Electro-Optics Business Sector (MGEO) in ASELSAN. Huseyin Ozkan is also with the Department of Electrical and Electronics Engineering at Bilkent University, Ankara Turkey. Email: huseyin@ee.bilkent.edu.tr

This study is partially supported by TUBITAK TEYDEB 1509 Program with project number 9130020 and by ARTEMIS JU Program with GA number 332913.

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

nals by exploiting the special structure of the AVS array response. The proposed method is more reliable compared to the previous methods (as also shown by our experiments) that perform the time-frequency bin selection and DOA estimation separately.

2. PROBLEM STATEMENT

We assume that a wideband acoustic source is impinging on a four element Acoustic Vector Sensor (AVS) [9] in a highly reverberant environment. The level of the reverberation is measured with the reverberation time RT_{60}, which describes the time it takes for the sound to decay by 60 dB after the sound is generated.[1] If the source signal is received from L different paths, then the signal at the output of an AVS can be expressed as

$$\mathbf{x}(t) = \sum_{i=0}^{L} \mathbf{a}(\theta_i, \phi_i)\gamma_i e^{-\alpha\tau_i} s(t-\tau_i) + \mathbf{w}(t), \tag{1}$$

where $s(t)$ and $\mathbf{w}(t)$ are the wideband acoustic source signal and 4x1 vector contains the noise signals at sensor outputs, respectively. The attenuation coefficient in sound propagation is denoted by α and γ_i is the reflection coefficient for the i'th path. Also, θ_i and ϕ_i are the azimuth and elevation angles of the i'th path, respectively. The time delay corresponding to the i'th path relative to the direct path $(i = 0)$ is represented by τ_i. Since the direct path is the shortest path between the source and an AVS, $\tau_i > \tau_0 = 0$ is valid for any non-zero i. The array response vector $\mathbf{a}(\theta, \phi)$ is defined as [9]

$$\mathbf{a}(\theta, \phi) = \left[\begin{array}{cccc} 1 & \cos(\theta)\cos(\phi) & \sin(\theta)\cos(\phi) & \sin(\phi) \end{array}\right]^T. \tag{2}$$

Our goal is to estimate the 2D DOA angles of an acoustic source signal for a given microphone outputs in highly reverberant environment, i.e., large number of multipaths, small attenuation coefficient and large reflection coefficient without knowing environmental and signal characteristics.

3. DIRECTION FINDING IN REVERBERANT ENVIRONMENT

The proposed algorithm -to detect the direction of a wideband acoustic source- consists of four main steps. In the first step, the array output in time domain is transformed into time-frequency domain via an overlapped short-time Fourier transform (STFT). In the second step, we first estimate the DOA angles at each time-frequency bin by assuming that there is only one active source. Then, we check whether the estimated DOA is consistent with the received signal components and select the time-frequency bins satisfying the consistency condition. In the third step, the DOA estimates at the selected time-frequency bins are grouped into clusters, where the estimated DOAs are close to each other. Each cluster is associated with one of the paths and we select a cluster corresponding to the direct path based on the fact that the direct path is the shortest path. In the final step, we estimate the source direction as the weighted average of the DOA estimates at the selected cluster.

The details of each step are explained in the following sections.

3.1. Transform Into Time-Frequency Domain

The initial step of the proposed algorithm is to transform the signal in time domain into time-frequency domain with overlapped STFT by

$$\begin{aligned}\mathbf{X}(m, \omega_k) &= \sum_{i=0}^{L} \mathbf{a}(\theta_i, \phi_i)\gamma_i e^{-\alpha_i\tau_i} S(m-\kappa_i, \omega_k) e^{j\frac{2\pi k}{K}\xi_i} \\ &\quad + \mathbf{W}(m, \omega_k),\end{aligned} \tag{3}$$

where $m = 0, \ldots, M-1$ is the time block index, $\omega_k = 2\pi k/K, k = 0, \ldots, K-1$ is the frequency bin index, $S(m, \omega_k)$ and $\mathbf{W}(m, \omega_k)$ are the STFT of $s(n)$ and $\mathbf{w}(n)$, respectively. The time shift for the i'th path is expressed as $\tau_i = \kappa_i FT + \xi_i T$, where $\kappa_i \in \mathbb{Z}^+$ and $\xi_i \in \mathbb{Z}$ are the time bin shift and the sample shift inside the time bin, respectively. The sampling period and the offset between successive time blocks in samples are denoted by T and F, respectively. We assume that without loss of generality, $\tau_0 = \kappa_0 = \xi_0 = 0$ since time shifts are defined with respect to the direct path.

The signal in time-frequency domain (3) can be rewritten in terms of direct path, i.e., $i = 0$, and the multipaths as

$$\mathbf{X}(m, \omega_k) = \mathbf{D}(m, \omega_k) + \mathbf{M}(m, \omega_k), \tag{4}$$

where $\mathbf{D}(m, \omega_k)$ is the direct path signal at time-frequency bin (m, ω_k) and $\mathbf{M}(m, \omega_k)$ is the interference signal that contains both the noise signal and the multipath components of the source signal. Namely,

$$\mathbf{D}(m, \omega_k) = \mathbf{a}(\theta_0, \phi_0) S(m, \omega_k), \tag{5}$$

$$\begin{aligned}\mathbf{M}(m, \omega_k) &= \sum_{i=1}^{L} \mathbf{a}(\theta_i, \phi_i)\gamma_i e^{j\frac{2\pi k}{K}\xi_i} e^{-\alpha_i\tau_i} S(m-\kappa_i, \omega_k) \\ &\quad + \mathbf{W}(m, \omega_k).\end{aligned} \tag{6}$$

It is important to note that in the first $m < \min(\kappa_i)$ time blocks, the interference signal contains only the noise signal since $S(m, \omega_k) = 0, \forall m < 0$. Hence, there is no multipath component during these time blocks. To guarantee that there is at least one time block in that condition, the window size of STFT should be selected by considering all possible time delays as

$$N < \left(\min_{1 \le i \le L} \tau_i\right) \frac{1}{T}. \tag{7}$$

[1] We consider that in this paper, the reverberation is "high" when the reverberation time is more than a second.

3.2. Single Active Source Detection

To eliminate the multipath distortion, we estimate the source DOA at the time-frequency bins, where only the direct path signal is active, i.e., $\mathbf{X}(m, \omega_k) = \mathbf{D}(m, \omega_k)$. Since the source signal and the environmental characteristics are unknown, this case can not be evaluated directly. On the other hand, the array output normalized with its first entry is equal to the AVS array response when only the direct path signal is active, cf. (2) and (5). Therefore, we first estimate DOA at each time-frequency bin assuming that there is only one active source with

$$\hat{\theta}(m, \omega_k) = tan^{-1}\left(\frac{X_3(m, \omega_k)}{X_2(m, \omega_k)}\right), \quad (8)$$

$$\hat{\phi}(m, \omega_k) = tan^{-1}\left(\frac{X_4(m, \omega_k)}{\sqrt{X_2^2(m, \omega_k) + X_3^2(m, \omega_k)}}\right). \quad (9)$$

Then, we check the consistency of DOA estimates with the array outputs and identify the time-frequency bins with single active source as

$$P(m, \omega_k) = \begin{cases} 1, & Q(m, \omega_k) \leq \epsilon, |X_1(m, \omega_k)|^2 > \psi \\ 0, & \text{otherwise} \end{cases}, \quad (10)$$

where ϵ and ψ are user specified thresholds for error and signal power levels, respectively. The consistency error $Q(m, \omega_k)$ is defined as

$$Q(m, \omega_k) = \left\| \bar{\mathbf{X}}(m, \omega_k) - \mathbf{a}(\hat{\theta}(m, \omega_k), \hat{\phi}(m, \omega_k)) \right\|^2, \quad (11)$$

where $\bar{\mathbf{X}}(m, \omega_k) = \frac{1}{X_1(m, \omega_k)}\mathbf{X}(m, \omega_k)$ is the array output signal in time-frequency bin (m, ω_k) normalized with omni-directional sensor signal $X_1(m, \omega_k)$. Note that, $P(m, \omega_k) = 1$ corresponds to the case, where there is only one active source at the corresponding time-frequency bin. Since the consistency error measures the similarity between array output and AVS array response structure, it may also take a small value for noise only signal. Therefore, we also check the signal power constraint at each time-frequency bin to conclude that there is only one active source.

3.3. Clustering

After determining the time-frequency bins with single active source, we group them into clusters by considering the closeness (in the euclidean sense) of direction estimates at those bins as

$$C_i = \left\{(m, \omega_k) \middle| \sqrt{E(m, \omega_k)} < \Gamma, P(m, \omega_k) = 1 \right\}, \quad (12)$$

where Γ is a threshold and $E(m, \omega_k)$ is the deviation of the estimated directions in each cluster which is defined as

$$E(m, \omega_k) = \left|\hat{\theta}(m, \omega_k) - \hat{\theta}(m_i, \omega_{ki})\right|^2 + \left|\hat{\phi}(m, \omega_k) - \hat{\phi}(m_i, \omega_{ki})\right|^2 \quad (13)$$

with $(m_i, \omega_{ki}) \in C_i$. Each cluster corresponds to a path with different direction in multipath propagation. In order to localize the acoustic source correctly, the cluster corresponding to the direct path is required to be identified. Since the direct path is the smallest path among all the other paths, in the first $m \leq \min(\kappa_i)$ time bins, only the source signal is active. On the other hand, in the later time bins, there may be more than one signal from different paths at different frequency bins. Therefore, we define the cluster corresponding to the direct path as a cluster with minimum time bin such that there are no other clusters on that time bin.

This definition of direct path clustering is prone to noise in the signal and its direct application may generate spurious clusters. However, since the noise -in general- is arbitrary and unstructured, those spurious clusters are often small in size and spread on time-frequency bins. Therefore, we can eliminate such issues by updating single active source mapping $P(m, \omega_k)$ with regards to these spurious cluster characteristics in two stages. We first discard the time-frequency bins, which are the members of a cluster with smaller number of direction estimates by

$$P(m, \omega_k) = \left\{0 \middle| (m, \omega_k) \in C_i, N_{C_i} < \rho \max_{1<i<V} (N_{C_i}) \right\}, \quad (14)$$

where V is the total number of determined clusters and N_{C_i} is the number of direction estimates in cluster C_i and $0 \leq \rho \leq 1$ is the scale factor for cluster elimination. Then, we discard all frequency bins in the time bin, where there are more than one cluster, i.e.,

$$P(m, :) = \{0 \,| (m, \omega_k) \in C_i, (m, \omega_l) \in C_j, \exists i \neq j, k \neq l\}, \quad (15)$$

where $P(m, :) \in \{P(m, \omega_0), P(m, \omega_1), \ldots, P(m, \omega_{K-1})\}$ is a single source map corresponding to all frequency bins in m'th time bin.

3.4. Direct Path Direction Estimation

Once the possible direct path clusters are robustly identified, the clustering is performed for the remaining time-frequency bins with single source as in (12) and a cluster with smallest time bin is selected. Then, the source direction is estimated as a weighted sum of the direction estimations in time-frequency bins corresponding to the selected cluster by

$$\hat{\theta}_d = \sum_{(m,\omega_k) \in C^d} \frac{\beta(m, \omega_k)}{\sum\limits_{(m,\omega_k) \in C^d} \beta(m, \omega_k)} \hat{\theta}(m, \omega_k), \quad (16)$$

$$\hat{\phi}_d = \sum_{(m,\omega_k) \in C^d} \frac{\beta(m, \omega_k)}{\sum\limits_{(m,\omega_k) \in C^d} \beta(m, \omega_k)} \hat{\phi}(m, \omega_k), \quad (17)$$

where C^d is the selected cluster for direct path and $\beta(m, \omega_k)$ is the weighting coefficient for time-frequency bin (m, ω_k)

and defined as

$$\beta(m, \omega_k) = \frac{|X_1(m, \omega_k)|^2}{Q(m, \omega_k)}. \tag{18}$$

The weighting coefficients are selected as in (18), since the direction estimation in the time-frequency bin with higher signal power and lower consistency error $Q(m, \omega_k)$ is more reliable.

4. PERFORMANCE RESULTS

The performance of the proposed algorithm is evaluated for the direction estimation errors and failure conditions in various scenarios. In our simulations, the array output is generated from the recorded gunshot sound signal with 360 ms duration and 22.050 Hz sampling rate as in (1). The reverberant environment is modeled by considering that the signals are received from $L = 100$ multipaths with $30 \leq \tau_i \leq 300$ ms time delays. The attenuation coefficient of a sound signal is determined as $\alpha = -\ln(10^{-3})/RT_{60}$ and a perfect reflection is assumed, i.e., $\gamma_i = 1$ for all paths. For transforming signal in time domain into time-frequency domain, $K = 512$ frequency bands and $N = 128$ Hanning window with 75% overlap ratio is used. The functional parameters for the proposed method are selected as $\psi = 20$ dB, $\rho = 0.6$, $\epsilon = 0.05$ and $\Gamma = 5^o$. The performance results are averaged over 500 trials. At each trial, arbitrarily selected source direction, multipath directions, time delays between signal paths and additive white Gaussian noise signal are used for signal generation. The DOA estimation errors are measured as the averaged root mean squared error (RMSE) as

$$RMSE = \sqrt{\frac{1}{N_s}\sum_{i=1}^{N_s} |\theta_e^i - \theta_c^i|^2 + |\phi_e^i - \phi_c^i|^2}, \tag{19}$$

where (θ_e^i, ϕ_e^i) and (θ_c^i, ϕ_c^i) are the estimated and correct value of the direction estimations in i'th trial, respectively. N_s is the number of successive trials after removing the failure conditions. The algorithm is considered to fail when the DOA estimation error is more than 20 degrees.

The source direction estimation performance of the proposed algorithm for varying SNR values is illustrated in Figure 1. The performance of the proposed algorithm is compared with the algorithms in [4], labeled as "Tho", and in [5], labeled as "Mohan AVS". For a fair comparison, extra one microphone is added to the co-planar 3 omni-directional microphones in [4] on a perpendicular plane for estimating 2D directions. In "Mohan AVS", the Acoustic Vector Sensor structure is used as in the proposed algorithm. Algorithms are tested in no reverberant ($RT_{60} = 0$ s) and highly reverberant ($RT_{60} = 5$ s) environments. "Tho" is highly sensitive to the reverberation and its performance seriously degrades under high reverberation as demonstrated in Figure 1. On the other hand, both "Mohan AVS" and the proposed algorithm is robust to the reverberation as their performances are comparable in the no reverberant and highly reverberant environments. For low SNR values, "Mohan AVS" performs slightly better than the proposed algorithm. However, its performance saturates after $SNR = 20$ dB, while the proposed algorithm achieves better performance as the SNR increases. Since the direction estimation method in "Mohan AVS" is based on MUSIC spectra search, the resolution of direction estimates is finite for a reasonable computational time. In the simulation, MUSIC spectra is generated with 1 degree resolution in both azimuth and elevation. On the other hand, in the proposed algorithm, the directions are computed directly from data as in (16), (17).

Figure 2 illustrates the failure rates of the algorithms. We observe that the failure rates increase with reverberation for all of the compared methods. On the other hand, the failure rates of the proposed algorithm in highly reverberant environment are as low as the failure rates of the other algorithms in no reverberant environment after $SNR = 5$ dB. This shows that the proposed algorithm is remarkably robust to high reverberation.

The averaged computational times of the algorithms are illustrated in Figure 3. The algorithms are implemented in MATLAB run on Intel Core i7 3.40 GHz CPU with 32 GB of RAM. The computational time of the proposed algorithm is about 1.63 seconds and flat over SNR under high reverberation. On the other hand, in no reverberant environment, its computational time slightly increases with increasing SNR, since the total number of time-frequency bins that satisfy single active source condition increases. While "Tho" has comparable computational time with the proposed algorithm, "Mohan AVS" requires much more time which exponentially increases with increasing SNR in no reverberant environment. These results show the effectiveness of the direct solution for DOA estimation as compared with the computationally heavy MUSIC method as in "Mohan AVS".

5. CONCLUSION

In this paper, a novel method for wideband acoustic source direction finding in highly reverberant environment with single acoustic vector sensor (AVS) is presented. This method exploits the AVS array structure to identify the time-frequency bins where only one signal is active. The proposed time-frequency bin selection procedure does not rely on time-frequency sparseness or any statistical assumptions on signals and noise. Therefore, it can be applied to any wideband acoustic signal, not limited to speech signal. DOA angle at each time-frequency bin is estimated directly from data without an exhaustive search. The estimated DOAs that are close to each other are grouped into clusters to identify the time-frequency bins corresponding to the direct path signal. Highly reverberant environment is modeled with a large

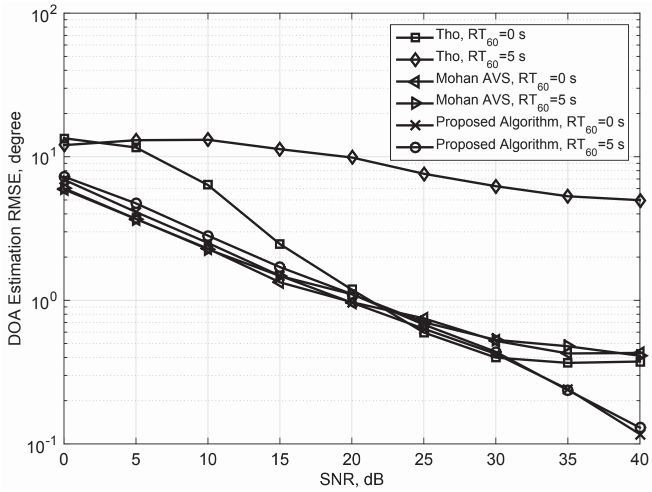

Fig. 1. RMSE of DOA estimation errors of the algorithms for different reverberation times.

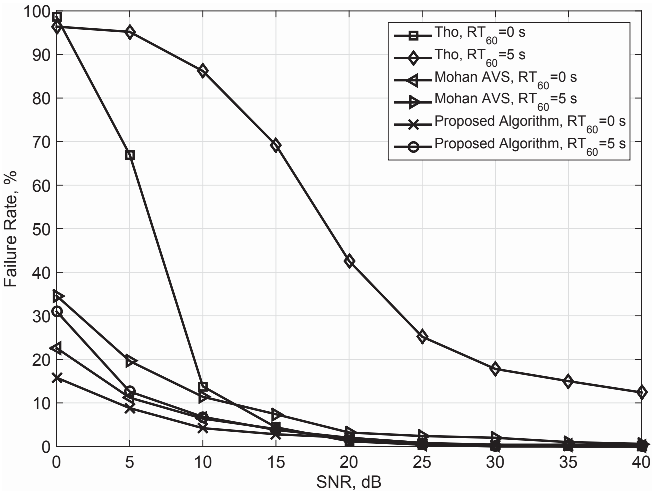

Fig. 2. Failure rate of the algorithms for different reverberation times.

number of multipaths ($L = 100$) and large reverberation time ($RT_{60} = 5$ seconds). The DOA estimation performance and the failure rates are investigated both in no reverberant and highly reverberant environment. The simulation results show that the proposed algorithm can accurately estimate the source direction with a small failure rate even under extremely high reverberation conditions.

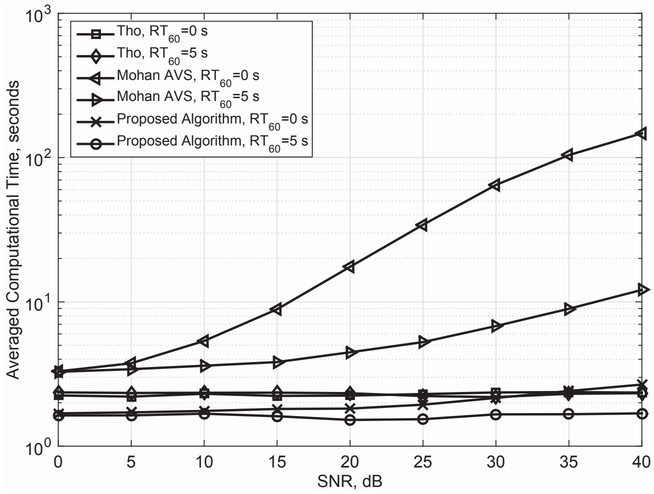

Fig. 3. Averaged computational time of the algorithms for different reverberation times.

REFERENCES

[1] R. A. Kennedy, P. Sadeghi, T. D. Abhayapala, and H. M. Jones, "Intrinsic limits of dimensionality and richness in random multipath fields," in *IEEE Transactions on Signal Processing, Vol. 55, No. 6*, 2007, pp. 2542–2556.

[2] A. Asaei, M. Golbabaee, H. Bourlard, and V. Cevher, "Structured sparsity models for reverberant speech seperation," in *IEEE/ACM Transaction On Audio, Speech, and Language Processing, Vol. 22, No. 3*, 2014, pp. 620–633.

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

[4] N.T.N. Tho, S. Zhao, and D.L. Jones, "Robust doa estimation of multiple speech sources," in *IEEE Internation Conference on Acoustics, Speech and Signal Processing (ICASSP) 2014*, 2014, pp. 2287–2291.

[5] S. Mohan, M.E. Lockwood, M.L. Kramer, and D.L. Jones, "Localization of multiple acoustic sources with small arrays using a coherence test," in *Journal of the Acoustical Society of America Vol. 123 No. 4*, 2008, pp. 2136–2147.

[6] M. Cobos, J.J. Lopez, and S. Spors, "Analysis of room reverberation effects in source localization using small microphone arrays," in *International Symposium on Communications, Control and Signal Processing (ISCCSP) 2010*, 2010, pp. 1–4.

[7] S. Gannot D. Levin, E. A. P. Habets, "On the angular error of intensity vector based direction of arrival estimation in reverberant sound fields," in *The Journal of the Acoustical Society of America*, 2010, pp. 1800–1811.

[8] M. E. Lockwood and D. L. Jones, "Beamformer performance with acoustic vector sensor in air," in *The Journal of the Acoustical Society of America*, 2006, pp. 608–619.

[9] A. Nehorai and E. Paldi, "Acoustic vector-sensor array processing," in *IEEE Transactions on Signal Processing Vol. 42, No. 9*, 1994, pp. 2481–2491.