SOUND SOURCE LOCALIZATION IN A REVERBERANT ROOM USING HARMONIC BASED MUSIC

Lachlan Birnie, Thushara D. Abhayapala, Hanchi Chen, Prasanga N. Samarasinghe

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

ABSTRACT

The localization of acoustic sound sources is beneficial to signal processing applications of speech enhancement, dereverberation, separation and tracking. Difficulties in position estimation arise in real world environments due to coherent reflections degrading performance of subspace localization techniques. This paper proposes a method of multiple signal classification (MUSIC) subspace localization, which is suitable for reverberant rooms. The method is based on the modal decomposition of a room's region-to-region transfer function, which is assumed to be known. We perform a numerical simulation of four sound sources in a reverberant room, and show that the localization method exhibits increased spatial resolution and distance focusing abilities when the region-to-region transfer function is incorporated. The proposed method is successful in estimating three-dimensional sound source positions without distortion due to reverberation.

Index Terms— Coherent MUSIC, source localization, reverberant environment, modal decomposition.

1. INTRODUCTION

The ability to estimate locations of sound source in reverberant environments is an active problem in acoustic signal processing. Source localization is required in areas of research involving speech enhancement [1], [2], diarization [3] and source separation [4]. Additionally, practical applications of robot audition [5], [6], surveillance systems, speaker tracking, and telephone conferencing, can benefit greatly from positional knowledge of sound sources.

A diverse range of methods have been developed for source localization, such as beam forming [7], binaural [8], time delay estimation [9], as well as subspace techniques including CSSM [10], ESPRIT [11] and WAVES [12]. MUSIC [13] could be considered as the best known form of subspace localization, and recently it has been adapted to use spherical harmonic based approaches [14], [15], [16]. Despite this, reverberant environments still cause issues of performance degradation in subspace localization due to coherent multipath reflections.

This research is supported by an Australian Government Research Training Program (RTP) Scholarship and the Australian Research Council (ARC) Discovery Project DP180102375.

In this paper, we propose a novel method for sound source localization within reverberant rooms. The derivation of this method is facilitated by the spherical harmonic based region-to-region room transfer function presented in [17]. By expressing the room's transfer function as a set of modal coupling coefficients, we show how a MUSIC subspace algorithm can be modified to avoid spectra distortion from a reflected sound field. We verify the proposed method via a simulated shoebox room, and demonstrate increased spatial resolution and distance focusing improvements to coherent sound source localization.

2. PROBLEM FORMULATION

Let there be L point sources within a reverberant room, each with a position $\boldsymbol{y}_\ell \equiv (r_\ell, \theta_\ell, \phi_\ell)$, $\ell = 1, \cdots, L$ for range, elevation and azimuth with respect to the origin O, producing a signal $S_\ell(k)$, where $k = 2\pi f/c$ is the wave number, f is the frequency and c is the speed of sound propagation. Consider a higher order microphone (such as a spherical microphone array) comprised of Q omnidirectional sensors placed about the same origin O, where each sensor has a position of $\boldsymbol{x}_q \equiv (r_q, \theta_q, \phi_q)$, $q = 1, \cdots, Q$. The microphone has a maximum aperture radius of R_q, which is denoted as the receiver region. Similarly, a source region is denoted by R_ℓ, shown in Fig. 1.

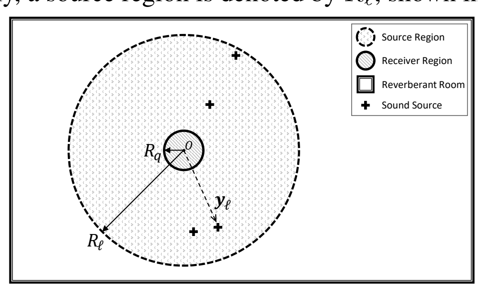

Fig. 1. Concentric source and receiver regions in a room

The signal received by the q^{th} sensor due to L sources is

$$P(k, \boldsymbol{x}_q) = \sum_{\ell=1}^{L} H(k, \boldsymbol{x}_q, \boldsymbol{y}_\ell) S_\ell(k) + \mathcal{N}(k), \qquad (1)$$

where $P(k, \boldsymbol{x}_q)$ is the pressure, $H(k, \boldsymbol{x}_q, \boldsymbol{y}_\ell)$ is the room transfer function between the q^{th} sensor and the ℓ^{th} source, and $\mathcal{N}(k)$ is the noise. The problem discussed in this paper is to estimate $\boldsymbol{y}_\ell = (r_\ell, \theta_\ell, \phi_\ell)$ of the L point sources from the sampled pressure and a model of the room's transfer function, which we explain next.

3. SPHERICAL HARMONIC DECOMPOSITION OF A REVERBERANT SOUND FIELD

In this section, we present a modal decomposition of $H(k, \boldsymbol{x}_q, \boldsymbol{y}_\ell)$ of (1), such that it is applicable to any two arbitrary points from a predefined source region and receiver region (see Fig.1).

3.1. Region-to-region transfer function

The region-to-region transfer function between any arbitrary source of $|\boldsymbol{y}_\ell| \leq R_\ell$ and any arbitrary receiver of $|\boldsymbol{x}_q| \leq R_q$ as shown in Fig. 1, is given by the modal decomposition [17]

$$H(k, \boldsymbol{x}_q, \boldsymbol{y}_\ell) = \frac{e^{ik||\boldsymbol{x}_q - \boldsymbol{y}_\ell||}}{4\pi||\boldsymbol{x}_q - \boldsymbol{y}_\ell||} + \sum_{\nu=0}^{V} \sum_{\mu=-\nu}^{\nu} j_\nu(k|\boldsymbol{x}_q|) \times Y_{\nu\mu}(\hat{\boldsymbol{x}}_q) \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{\nu\mu}^{nm}(k) \Big(ikj_n(k|\boldsymbol{y}_\ell|) Y_{nm}^*(\hat{\boldsymbol{y}}_\ell) \Big), \quad (2)$$

where $j_\nu(\cdot)$ denotes the spherical Bessel function of order ν, $Y_{\nu\mu}(\cdot)$ denotes the spherical harmonic [18] of order ν and mode μ, $|\cdot| \equiv r$, $(\hat{\cdot}) \equiv (\theta, \phi)$, $V = \lceil kR_q \rceil$ is the truncation limit [19] of the receive region whose harmonics are indexed by order ν and mode μ, $N = \lceil kR_\ell \rceil$ is the truncation limit of the source region whose harmonics are indexed by order n and mode m, and $\alpha_{\nu\mu}^{nm}(k)$ are the room coupling coefficients. The coupling coefficients are considered to be a-prior known (pre-recorded [17] or modeled [20]) constant properties of the reverberant environment. They describe the room response incident at the receiver region due to the sound propagating away from the source region.

In (2) the first term describes the incoming direct path field, while the second term describes the outgoing path field reflections. For the case of $|\boldsymbol{x}_q| < |\boldsymbol{y}_\ell|$, the direct path field can be defined as [18]

$$\frac{e^{ik||\boldsymbol{x}_q - \boldsymbol{y}_\ell||}}{4\pi||\boldsymbol{x}_q - \boldsymbol{y}_\ell||} = \sum_{\nu=0}^{V} \sum_{\mu=-\nu}^{\nu} j_\nu(k|\boldsymbol{x}_q|) Y_{\nu\mu}(\hat{\boldsymbol{x}}_q) \Big(ikh_\nu(k|\boldsymbol{y}_\ell|) Y_{\nu\mu}^*(\hat{\boldsymbol{y}}_\ell) \Big), \quad (3)$$

where $h_\nu(\cdot)$ denotes the spherical Hankel function of the first kind and ν^{th} order. Therefore, (3) along with the coupling term of (2), provides a complete modal decomposition of the room transfer function in terms of a source position $\boldsymbol{y}_\ell$.

3.2. Reverberant sound field within a room

The spherical harmonic decomposition of the signal received by the q^{th} sensor due to L sources within a reverberant room is given by substituting (3), (2) and (1) as

$$P(k, \boldsymbol{x}_q) = \sum_{\nu=0}^{V} \sum_{\mu=-\nu}^{\nu} \gamma_{\nu\mu}(k) j_\nu(k|\boldsymbol{x}_q|) Y_{\nu\mu}(\hat{\boldsymbol{x}}_q) + \mathcal{N}(k), \quad (4)$$

where

$$\gamma_{\nu\mu}(k) = \sum_{\ell=1}^{L} S_\ell(k) \times \left(\Psi_{\nu\mu}(k, \boldsymbol{y}_\ell) + \sum_{n=0}^{N} \sum_{m=-n}^{n} \alpha_{\nu\mu}^{nm}(k) \beta_{nm}(k, \boldsymbol{y}_\ell) \right), \quad (5)$$

is denoted as the received reverberant coefficients,

$$\Psi_{\nu\mu}(k, \boldsymbol{y}_\ell) = ikh_\nu(k|\boldsymbol{y}_\ell|) Y_{\nu\mu}^*(\hat{\boldsymbol{y}}_\ell), \quad (6)$$

are the incoming coefficients due to the direct path field, and

$$\beta_{nm}(k, \boldsymbol{y}_\ell) = ikj_n(k|\boldsymbol{y}_\ell|) Y_{nm}^*(\hat{\boldsymbol{y}}_\ell), \quad (7)$$

are the outgoing coefficients of the source region which are related to the reflected path field through the known $\alpha_{\nu\mu}^{nm}(k)$ coupling coefficients.

It is observed from (4) that the receiver is independent of the reverberant coefficients and sound source positions in (5). For an ideal microphone array, pressure measurements can be decomposed into the reverberant coefficients by [21]

$$\gamma_{\nu\mu}(k) = \frac{1}{j_\nu(k|\boldsymbol{x}|)} \int P(k, \boldsymbol{x}) Y_{\nu\mu}^*(\hat{\boldsymbol{x}}) d\hat{\boldsymbol{x}}. \quad (8)$$

Consider a microphone capable of solving (8) up to the V^{th} order, expanding the harmonic decomposition of (5) for this microphone gives

$$\boldsymbol{\gamma}(k) = \Big(\boldsymbol{\Psi}(k) + \boldsymbol{\alpha}(k)\boldsymbol{\beta}(k) \Big) \mathbf{S}(k) + \bar{\boldsymbol{\mathcal{N}}}(k), \quad (9)$$

where $\boldsymbol{\gamma}(k) = [\gamma_{00}(k), \cdots, \gamma_{VV}(k)]^{\text{T}}$ is the set of measured reverberant coefficients, $\mathbf{S}(k) = [S_1(k), \cdots, S_L(k)]^{\text{T}}$ are the source signals, $\bar{\boldsymbol{\mathcal{N}}}(k) = [\bar{\mathcal{N}}_{00}(k), \cdots, \bar{\mathcal{N}}_{VV}(k)]^{\text{T}}$ is the noise carried through (8),

$$\boldsymbol{\Psi}(k) = \begin{bmatrix} \Psi_{00}(k, \boldsymbol{y}_1) & \cdots & \Psi_{00}(k, \boldsymbol{y}_L) \\ \vdots & \ddots & \vdots \\ \Psi_{VV}(k, \boldsymbol{y}_1) & \cdots & \Psi_{VV}(k, \boldsymbol{y}_L) \end{bmatrix}, \quad (10)$$

is a matrix with L column vectors describing the direct path field of each source,

$$\boldsymbol{\alpha}(k) = \begin{bmatrix} \alpha_{00}^{00}(k) & \cdots & \alpha_{00}^{NN}(k) \\ \vdots & \ddots & \vdots \\ \alpha_{VV}^{00}(k) & \cdots & \alpha_{VV}^{NN}(k) \end{bmatrix}, \quad (11)$$

describes the coupling between the receiver and the outgoing field due to each source, where the outgoing field is expressed with L column vectors as

$$\boldsymbol{\beta}(k) = \begin{bmatrix} \beta_{00}(k, \boldsymbol{y}_1) & \cdots & \beta_{00}(k, \boldsymbol{y}_L) \\ \vdots & \ddots & \vdots \\ \beta_{NN}(k, \boldsymbol{y}_1) & \cdots & \beta_{NN}(k, \boldsymbol{y}_L) \end{bmatrix}. \quad (12)$$

In the next section we show how to retrieve the source positions from a set of measured reverberant coefficients (9) with a MUSIC subspace localization method.

4. SOURCE LOCALIZATION

Let us define the harmonic based covariance matrix of reverberant coefficients as

$$\begin{aligned}\boldsymbol{R}_{\boldsymbol{\gamma}}(k) &\triangleq \mathrm{E}\left\{\boldsymbol{\gamma}(k)\boldsymbol{\gamma}(k)^H\right\} \\ &= \mathbf{A}(k)\boldsymbol{R}_{\boldsymbol{s}}(k)\mathbf{A}(k)^H + \boldsymbol{R}_{\boldsymbol{n}}(k),\end{aligned} \tag{13}$$

where $\mathrm{E}\{\cdot\}$ denotes the statistical expectation, $(\cdot)^H$ is a conjugate transpose, and

$$\mathbf{A}(k) = \boldsymbol{\Psi}(k) + \boldsymbol{\alpha}(k)\boldsymbol{\beta}(k) \tag{14}$$

is the frequency dependent steering matrix of L column vectors. The signal $\boldsymbol{R}_{\boldsymbol{s}}(k)$ and noise $\boldsymbol{R}_{\boldsymbol{n}}(k)$ covariance matrices are given by

$$\boldsymbol{R}_{\boldsymbol{s}}(k) \triangleq \mathrm{E}\{\mathbf{S}(k)\mathbf{S}(k)^H\}, \tag{15}$$

$$\boldsymbol{R}_{\boldsymbol{n}}(k) \triangleq \mathrm{E}\{\boldsymbol{\mathcal{N}}(k)\boldsymbol{\mathcal{N}}(k)^H\}. \tag{16}$$

In practice (13) is estimated for a narrowband frequency by averaging (9) over T time frames, given as

$$\hat{\boldsymbol{R}}_{\boldsymbol{\gamma}}(k) = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\gamma}(t,k)\boldsymbol{\gamma}(t,k)^H. \tag{17}$$

Due to the coupling of frequency and angular components in (14), frequency smoothing techniques are difficult to utilize, and so for simplicity we solve the covariance matrix for a narrowband frequency bin of k [12], [10]. The covariance matrices are decomposed into signal and noise subspaces by singular value decomposition:

$$\hat{\boldsymbol{R}}_{\boldsymbol{\gamma}}(k) = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^H = \begin{bmatrix}\boldsymbol{U}_s & \boldsymbol{U}_n\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_s & 0\\ 0 & \boldsymbol{\Sigma}_n\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_s^H\\ \boldsymbol{U}_n^H\end{bmatrix}, \tag{18}$$

where the right hand side frequency dependency is omitted here for notation convenience. The signal subspace $\boldsymbol{U}_s(k)$ is of the size $[(V+1)^2$ by $L]$, while the noise subspace $\boldsymbol{U}_n(k)$ is of the size $[(V+1)^2$ by $\big((V+1)^2 - L\big)]$. For the purposes of this paper we assume that L is known, in practice the number of sources can be determined from the eigenvalues of (18). The MUSIC spectrum for a given frequency is then computed by

$$M(k,\boldsymbol{y}_s) = \frac{1}{||\boldsymbol{U}_n^H(k)\boldsymbol{a}(k,\boldsymbol{y}_s)||^2}, \tag{19}$$

where $\boldsymbol{a}(k,\boldsymbol{y}_s)$ is a $[(V+1)^2$ by $1]$ column vector of $\mathbf{A}(k)$ for a steering position $\boldsymbol{y}_s \equiv (r_s,\theta_s,\phi_s)$. The inclusion of coupling coefficients in $\boldsymbol{a}(k,\boldsymbol{y}_s)$ is the difference to traditional MUSIC algorithms which allows for the suppression of multipath spectra distortions.

Typically the coupling coefficients of a room's transfer function are measured over a broadband frequency, and therefore it is of interest to utilize them in source localization. The easiest method to enhance localization is to simply average spatial spectra over multiple frequency bins, expressed as

$$\begin{aligned}\hat{M}(\boldsymbol{y}_s) &= \frac{1}{J}\sum_{j=1}^{J} M(k_j,\boldsymbol{y}_s) \\ &= \frac{1}{J}\sum_{j=1}^{J}\frac{1}{||\boldsymbol{U}_n^H(k_j)\boldsymbol{a}(k_j,\boldsymbol{y}_s)||^2}\end{aligned} \tag{20}$$

where k_j denotes the wave number of the j^{th} frequency bin and J is the total number of bins. It may be possible to adapt more advanced frequency smoothing techniques for further enhancement [15], [12].

5. SIMULATION AND DISCUSSION

In this section, we illustrate the performance of our proposed method for a simulated $4 \times 6 \times 3$ m shoebox room with wall reflection coefficients of [0.9, 0.75, 0.95, 0.7, 0.6, 0.8]. We considered a $R_\ell = 1$ m sized source region about the origin $O = [2, 3, 1.5]$ m with respect to the front-left-bottom corner of the room. A third order open spherical microphone array with a $Q = 32$ sensor Eigenmike sampling scheme [22] and size $R_q = 0.042$ m was also placed at O. For such a receiver, measurements allow for the reverberant coefficient decomposition (8) to be approximated with [21]

$$\gamma_{\nu\mu}(k) \approx \sum_{q=1}^{Q} w_q \frac{P(k,\boldsymbol{x}_q)Y_{\nu\mu}^*(\hat{\boldsymbol{x}}_q)}{j_\nu(k|\boldsymbol{x}_q|)}, \tag{21}$$

where w_q, $q = 1, \cdots, Q$ are a set of suitable sampling weights, in this case $w_q = 1$. The pressures $P(k,\boldsymbol{x}_q)$ were simulated with the point-to-point image source method [23] of tenth order depth. Measurements were performed over the frequency band of 1000 Hz to 2000 Hz at $J = 10$ equally spaced bins, for a time period of $T = 100$ windows and a microphone SNR of 20 dB. The coupling coefficients $\alpha_{\nu\mu}^{nm}(k)$ were simulated by a modal image source method [20] for each frequency bin, with a source order of $N = \lceil kR_\ell \rceil$, a receiver order of $V = 3$ and a third order image depth.

We first consider four ($L = 4$) sound sources positioned within the room with respect to O at $\boldsymbol{y}_1 = (0.4\text{ m}, 60°, 50°)$, $\boldsymbol{y}_2 = (0.8\text{ m}, 120°, 300°)$, $\boldsymbol{y}_3 = (0.8\text{ m}, 140°, 320°)$ and $\boldsymbol{y}_4 = (1\text{ m}, 60°, 50°)$, where each source produced a random Gaussian signal $S_\ell(k)$ for the $1-2$ kHz band. We show the MUSIC spectrum of the proposed method after frequency averaging at four focused distances in Fig. 2.

Results of Fig. 2 show that the proposed method is able to successfully identify the position of the four sound sources, especially when the distance focusing parameter D matches the true distance to the source. This is clear with Fig. 2(a) and the first source with $D = 0.4$ m, and Fig. 2(c) with the second and third sources with $D = 0.8$ m, and finally with Fig. 2(d) and the fourth source with $D = 1$ m. We note that the MUSIC spectrum is seen to decrease in clarity as the focusing approaches the source region boundary of R_ℓ, resulting in the peak of the furthest source to be less defined. However, distance focusing manages to successfully distinguish between $\boldsymbol{y}_1$ and $\boldsymbol{y}_4$ which are in the same direction but at different radii (see Fig. 2(a) and (d)). Additionally, the proposed method is seen to be resistant to sound reflections as no significant stray peaks are observed throughout the MUSIC spectra.

In Fig. 3 we present the same scenario as shown in Fig. 2 with the assumption that the room's coupling coefficients are unknown, such that $\alpha_{\nu\mu}^{nm}(k) = 0$ and the spatial steering

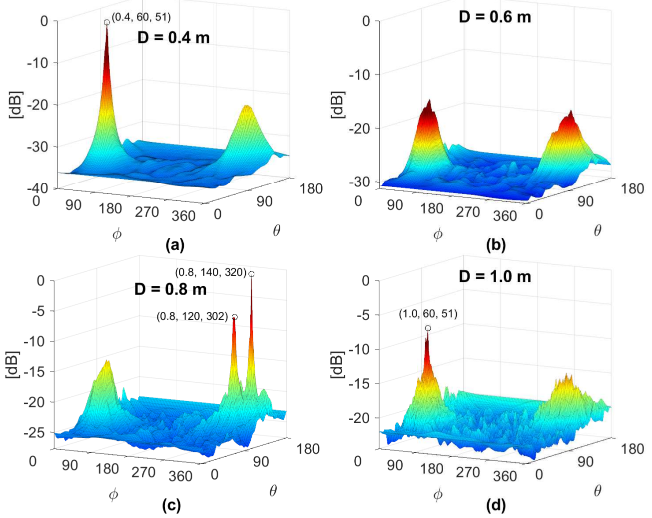

Fig. 2. Proposed method MUSIC spectrum as in (20) with focus distance D and sources at $(0.4\,\text{m},60°,50°)$, $(0.8\,\text{m},120°,300°)$, $(0.8\,\text{m},140°,320°)$, and $(1\,\text{m},60°,50°)$.

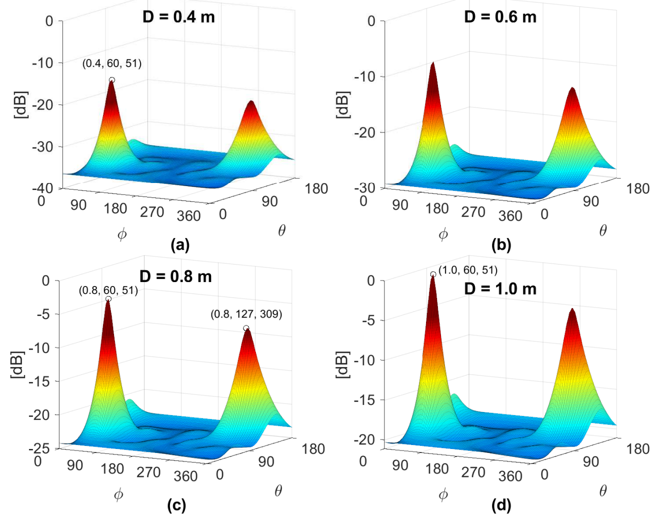

Fig. 3. Traditional method MUSIC spectrum with focus distance D and sources at $(0.4\,\text{m},60°,50°)$, $(0.8\,\text{m},120°,300°)$, $(0.8\,\text{m},140°,320°)$, and $(1\,\text{m},60°,50°)$.

$\boldsymbol{a}(k_j, \boldsymbol{y}_s)$ is given by a column vector of (10). This scenario is somewhat the same as traditional MUSIC algorithms, which do not account for room response. Comparing Fig. 2 to Fig. 3 provides insight on how subspace localization benefits from the use of region-to-region coupling coefficients. Two key observations are made from the comparison. First is on distance focusing capability, where we observe how Fig. 3 fails to distinguish between the two sources with the same direction of $(60°, 50°)$ at $\boldsymbol{y}_1$ and $\boldsymbol{y}_4$. It is also unclear whether a source exists at 0.4 m due to peak magnitudes being below -10 dB. This is in contrast to Fig. 2(a), where the closest source is sharply distinguished by the proposed method.

Our second observation is on spatial resolution, where we draw attention to the sources at $\boldsymbol{y}_2$ and $\boldsymbol{y}_3$ in Fig. 3(c). Here the MUSIC spectrum fails to localize the two nearby sources as separate peaks, and instead shows a single peak between them. Fig. 2(c) shows that with the incorporation of room coupling coefficients to the MUSIC subspace localization, the two nearby sources can be individualized as separate peaks.

Finally, we consider the case of a sound source moving beyond the predefined source region. Fig. 4 shows the proposed method spectra for the same scenario as in Fig. 2, with the furthest source moved to a position of $\boldsymbol{y}_4 = (1.8\text{ m}, 60°, 50°)$ and the maximum focus distance increased to 1.8 m. The $\alpha_{\nu\mu}^{nm}(k)$ coupling coefficients are still defined for $R_\ell = 1$ m, and as a result we expect the proposed method to suffer from truncation error while focusing beyond this distance.

It can be seen in Fig. 4(d) that the proposed method has difficulties distinguishing the fourth sound source, as evident by the blunt peak in the $(60°, 50°)$ direction, relative to elevations in the spectrum caused by room reflections and coupling truncation. However, sources within the predefined region are still able to be localized, as shown by the peaks in Fig. 4(a) and (c) being comparable to those in Fig. 2(a) and (c).

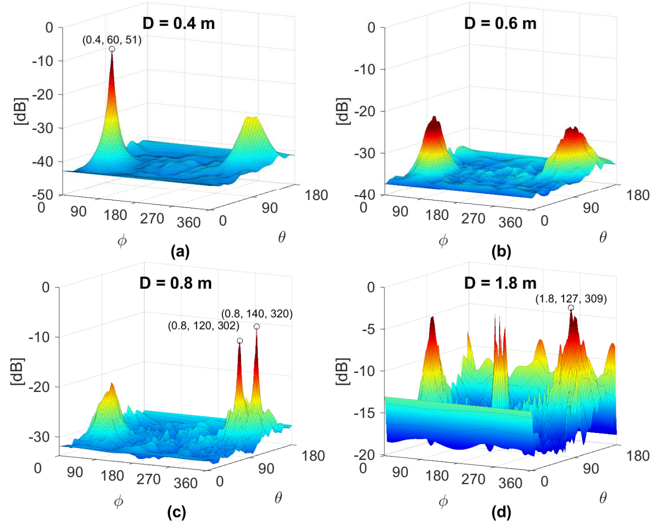

Fig. 4. Proposed method MUSIC spectrum with focus distance D and sources at $(0.4\,\text{m},60°,50°)$, $(0.8\,\text{m},120°,300°)$, $(0.8\,\text{m},140°,320°)$, and $(1.8\,\text{m},60°,50°)$.

6. CONCLUSION

In this paper, we have presented a method for acoustic sound source localization in a reverberant room. The method uses region-to-region modal coupling coefficient constants of a room to improve upon a MUSIC subspace localization. Numerical simulation results have shown that the proposed method enhances the spatial resolution and distance focusing ability of source position estimation in multipath environments. The proposed method utilizes coupling coefficients in simple broadband spatial spectra averaging, but in future, modal transfer functions may be exploited in more sophisticated subspace localization techniques.

7. REFERENCES

[1] Y. Peled and B. Rafaely, "Linearly-constrained minimum-variance method for spherical microphone arrays based on plane-wave decomposition of the sound field," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 12, pp. 2532–2540, 2013.

[2] S. Braun, D. P. Jarrett, J. Fischer, and E. AP Habets, "An informed spatial filter for dereverberation in the spherical harmonic domain.," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 669–673.

[3] X. Anguera, S. Bozonnet, N. Evans, C. Fredouille, G. Friedland, and O. Vinyals, "Speaker diarization: A review of recent research," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 2, pp. 356–370, 2012.

[4] F. Asano, M. Goto, K. Itou, and H. Asoh, "Real-time sound source localization and separation system and its application to automatic speech recognition," in *Proc. Eur. Conf. Speech Commun. Technol.*, 2001.

[5] J. M. Valin, F. Michaud, J. Rouat, and D. Létourneau, "Robust sound source localization using a microphone array on a mobile robot," in *Proc. Int. Conf. Intell. Robots Syst.* IEEE, 2003, vol. 2, pp. 1228–1233.

[6] K. Nakadai, D. Matsuura, H. G. Okuno, and H. Kitano, "Applying scattering theory to robot audition system: Robust sound source localization and extraction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.* IEEE, 2003, vol. 2, pp. 1147–1152.

[7] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, 1988.

[8] K. Nakadai, T. Lourens, H. G. Okuno, and H. Kitano, "Active audition for humanoid," in *Proc. Nat. Conf. Artificial Intell.*, 2000, pp. 832–839.

[9] J. Benesty, Y. Huang, and J. Chen, "Time delay estimation via minimum entropy," *IEEE Signal Process. Lett.*, vol. 14, no. 3, pp. 157–160, 2007.

[10] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 4, pp. 823–831, 1985.

[11] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989.

[12] E. D. Di Claudio and R. Parisi, "Waves: Weighted average of signal subspaces for robust wideband direction finding," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2179–2191, 2001.

[13] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[14] T. D. Abhayapala and H. Bhatta, "Coherent broadband source localization by modal space processing," in *Proc. Int. Conf. Telecommun.* IEEE, 2003, vol. 2, pp. 1617–1623.

[15] D. Khaykin and B. Rafaely, "Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.* IEEE, 2009, pp. 221–224.

[16] 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, Language Process.*, vol. 22, no. 10, pp. 1494–1505, 2014.

[17] P. Samarasinghe, T. Abhayapala, M. Poletti, and T. Betlehem, "An efficient parameterization of the room transfer function," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 12, pp. 2217–2227, 2015.

[18] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, (Academic, London, UK), 1999.

[19] R. A. Kennedy, P. Sadeghi, T. D. Abhayapala, and H. M. Jones, "Intrinsic limits of dimensionality and richness in random multipath fields," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2542–2556, 2007.

[20] P. N Samarasinghe, T. D Abhayapala, Y. Lu, H. Chen, and G. Dickins, "Spherical harmonics based generalized image source method for simulating room acoustics," *J. Acoust. Soc. Amer.*, vol. 144, no. 3, pp. 1381–1391, 2018.

[21] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2002, vol. 2, pp. 1949–1952.

[22] MH Acoustics, "Em32 eigenmike microphone array release notes (v17. 0)," *25 Summit Ave, Summit, NJ 07901, USA*, 2013.

[23] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.