Audio Engineering Society

Conference Paper

Presented at the 2022 International Conference on
Audio for Virtual and Augmented Reality
2022 August 15–17, Redmond, WA, USA

Towards Blind Localization of Room Reflections with Arbitrary Microphone Arrays

Yogev Hadadi[1], Vladimir Tourbabin[2], Paul Calamia[2], and Boaz Rafaely[1]

[1]*School of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel*
[2]*Reality Labs Research at Meta, Redmond, WA USA*

Correspondence should be addressed to Yogev Hadadi (yogevhad@post.bgu.ac.il)

ABSTRACT

Blind estimation of the direction of arrival (DOA) of early room reflections, without a priori knowledge of the room impulse response or of the source signal, may be beneficial in many applications. Recently, a method denoted PHALCOR (PHase ALigned CORrelation) was developed for DOA estimation of early reflections, which displayed superior performance compared to previous methods. However, PHALCOR was developed and evaluated only for spherical microphone arrays with a frequency-independent steering matrix, with input signals in the spherical harmonics domain. This paper extends the formulation of PHALCOR by introducing a focusing process that removes the frequency dependence of the steering matrix, and provides performance analysis of the estimation of the recorded signal from a spherical array, operating in the microphone signals domain, compared to the performance of PHALCOR, operating on signals in the spherical harmonics domain.

1 Introduction

Many signal processing tasks, such as speech enhancement and dereverberation [1], [2], source separation [3], optimal beamforming [4] and room geometry inference [5] may benefit from information about the directions of arrival (DOAs) and delays of room reflections. In particular, early reflections play a key role in sound perception by improving speech intelligibility and the sense of listener envelopment, as well as the ability to assess source width, loudness and distance [6], [7]. Thus, methods that exploit early reflections may advance spatial audio signal processing [8], [9].

DOA estimation methods are categorized as blind and non-blind methods. Non-blind methods assume a priori knowledge regarding the signal, such as the room impulse response or a clean anechoic recording of the sound source. Blind estimation methods operate directly on microphone signals, which is the more realistic case in practical audio signal processing applications. The approach which is adopted in this work is blind estimation.

One technique employed to blindly estimate the DOAs of the early reflections is spatial filtering (beamforming). This method can also separate reflection signals from the direct sound, which enables delay estima-

tion using cross-correlation analysis [5]. However, beamformers may have inadequate spatial resolution in practical scenarios where the spatial density of early reflections may be high [10]. Higher resolution can be achieved using subspace methods, such as MUSIC or ESPRIT [11]-[14], but these assume uncorrelated sources, which is clearly inappropriate for early reflections that are delayed copies of the direct sound, and are thus highly correlated. The source signals could be decorrelated using frequency smoothing, but this is not possible when the reflections have similar delays. Furthermore, these methods require that the number of microphones is larger than the number of sources and reflections, which is not the case for many practical arrays. Another method is implemented by formulating the problem as an under-determined linear system and using sparse recovery [15]; however, this method can only detect the first few reflections. Other methods are based on modeling the sources signals as deterministic unknowns, but these demand difficult non-linear optimization, and can have poor spatial resolution [11, 16]-[20].

PHALCOR [21] is a recently proposed method that overcomes the limitations of the previous methods, by exploiting the property that the reflections are delayed copies of the direct sound. It formulates a transform that can separate reflections over time and space, which enables the detection of more reflections than could previously have been detected. However, in its current form, PHALCOR is only compatible with spherical microphone arrays, with a formulation that assumes a frequency-independent steering matrix. This limitation means that it is not applicable for arbitrary arrays such as wearable arrays.

This work make a first step toward the generalization of PHALCOR, therefore enabling it to work with arbitrary microphones arrays, by using frequency focusing [22] to obtain a frequency-independent steering matrix. The input to the proposed method is a steering matrix that may depend on frequency; a focusing transformation is applied, which converts the matrix into a frequency-independent form within the focusing frequency range.

The paper is organized as follows. Section 2 present the system model and the PHALCOR algorithm. Section 3 presents the proposed algorithm. Section 4 presents a simulation study and Section 5 concludes.

2 Mathematical Background

2.1 Signal Model

The considered acoustic scene comprises a single source and a microphone array with Q microphones in a shoe box room. The direct sound signal in the frequency domain is denoted as $s(f)$ (where f represents the frequency index), with a DOA Ω_0 relative to the array. Assuming K early reflections, the k'th reflection of the direct signal is modeled as a separate source $s_k(f)$ with a DOA Ω_k. This reflection is assumed to be a delayed and attenuated copy of the direct signal [23]:

$$s_k(f) = \alpha_k e^{-i2\pi f \tau_k} s(f) \tag{1}$$

where τ_k is the delay relative to the direct signal, and α_k is an attenuating factor. The delay and attenuating factor of the direct sound signal are normalized such that $\tau_0 = 0$ and $\alpha_0 = 1$. The delays are sorted such that $\tau_{k-1} \leq \tau_k$.

Let $\mathbf{p}(f) := [p_1(f), p_2(f), \ldots, p_Q(f)]^T$ denote the captured pressure signal at the array, leading to the array equation:

$$\mathbf{p}(f) = \mathbf{H}(f,\Omega)\mathbf{s}(f) + \mathbf{n}(f) \tag{2}$$

where:

$$\mathbf{s}(f) := [s_0(f), \ldots, s_K(f)]^T \tag{3}$$

$$\mathbf{H}(f,\Omega) := [\mathbf{h}(f,\Omega_0), \ldots, \mathbf{h}(f,\Omega_K)] \tag{4}$$

$\mathbf{n}(f)$ is the noise captured by the array, and $\mathbf{h}(f,\Omega_k)$ is the steering vector in a free field of the k'th reflection and the array.

2.2 PHALCOR

In a spherical array $\mathbf{H}(f,\Omega)$ can be formulated as [24, 25]:

$$\mathbf{H}(f,\Omega) = \mathbf{Y}(\Omega_{mic})\mathbf{B}(f,r)\mathbf{Y}(\Omega) \tag{5}$$

where $\mathbf{Y}(\Omega_{mic})$ is a $Q \times (N+1)^2$ spherical harmonics matrix in the microphone angles, N is the spherical harmonics order, and $\mathbf{B}(f,r)$ is an $(N+1)^2 \times (N+1)^2$ diagonal matrix holding array radial functions [25]. $\mathbf{Y}(\Omega)$ is an $(N+1)^2 \times K$ spherical harmonics matrix in the DOA angles. Substituting Equation (5) into Equation (2) and using plane wave decomposition (PWD) [24] leads to a frequency-independent steering matrix:

$$\mathbf{a}_{nm}(f) = \mathbf{Y}(\Omega)\mathbf{s}(f) + \mathbf{n}'(f) \tag{6}$$

where $\mathbf{a_{nm}}(f) = [a_{00}(f), a_{1(-1)}(f), a_{10}(f), \ldots, a_{NN}(f)]$ is the $(N+1)^2 \times 1$ PWD coefficients [24] and $\mathbf{Y}(\Omega) = [\mathbf{y}(\Omega_0), \ldots, \mathbf{y}(\Omega_K)]$ is the steering matrix, which now is frequency-independent, with $\mathbf{y}(\Omega_k)$ representing a steering vector in free field with a DOA of Ω_k. $\mathbf{n}'(f)$ is now the noise term in $\mathbf{a}_{nm}(f)$, and it is modified relative to the original $\mathbf{n}(f)$ in Equation (2). The spatial correlation matrix (SCM) $\mathbf{R}(f)$ is calculated next:

$$\mathbf{R}(f) := E[\mathbf{a}_{nm}(f)\mathbf{a}_{nm}(f)^H] \tag{7}$$

Substituting Equation (6) in Equation (7) leads to:

$$\mathbf{R}(f) = \mathbf{Y}(\Omega)\mathbf{M}(f)\mathbf{Y}(\Omega)^H + \mathbf{N}'(f) \tag{8}$$

where $\mathbf{M}(f) = E[\mathbf{s}(f)\mathbf{s}(f)^H]$ and $\mathbf{N}'(f) = E[\mathbf{n}'(f)\mathbf{n}'(f)^H]$. The phase-aligned transform is defined in PHALCOR in order to enhance an entry in the SCM, formulated as:

$$\overline{\mathbf{R}}(\tau, f) := \sum_{j=0}^{J_f - 1} \omega_j \mathbf{R}(f_j) e^{i2\pi\tau j\Delta f} \tag{9}$$

where $f_j = f + j\Delta f$,J_f is an integer parameter, Δf is the frequency resolution, such that $J_f \Delta f = B_w$ where B_w is the bandwidth. $\omega_0, \ldots, \omega_{J_f - 1}$ are non-negative weights proportional to $tr(\mathbf{R}(f_j))$. The transform is a weighted version of the inverse DFT. Therefore, as $M(\tau, f)$ is composed of delayed copies of the direct sound, its inverse DFT has the form of a Dirichlet kernel with peaks in τ which corresponded to reflections [21]. Substituting Equation (8) in Equation (9), and using the steering matrix, which is frequency-independent, allows the phase aligned transform to be applied on $\mathbf{M}(f)$ and $\mathbf{N}'(f)$:

$$\overline{\mathbf{R}}(\tau, f) = \mathbf{Y}(\Omega)\overline{\mathbf{M}}(\tau, f)\mathbf{Y}(\Omega)^H + \overline{\mathbf{N}}'(\tau, f) \tag{10}$$

where $\overline{\mathbf{M}}(f)$ and $\overline{\mathbf{N}}'(f)$ are obtained by applying the phase alignment transform from (9) on $\mathbf{M}(f)$ and $\mathbf{N}'(f)$, respectively. Rewriting (10) by separating reflections leads to:

$$\overline{\mathbf{R}}(\tau, f) = \sum_{k=0}^{K} \sum_{k'=0}^{K} [\overline{\mathbf{M}}(\tau, f)]_{k,k'} \mathbf{y}(\Omega_k)\mathbf{y}(\Omega_{k'})^H + \overline{\mathbf{N}}'(\tau, f) \tag{11}$$

In PHALCOR it was shown that $\overline{\mathbf{M}}(\tau, f)$ has a maximum when $\tau = \tau_k - \tau_{k'}$. With k' representing the direct sound, $\tau_{k'}$ will be the time of arrival (normalized to be 0), and $\tau_k - \tau_{k'}$ represents the delay between the k'th reflection and the direct sound signal. It is assumed that for the delay $\tau = \tau_k - \tau_{k'}$ the transformation enhances a single entry. It is clear from Equation (11) that in this case $\overline{\mathbf{R}}(\tau)$ is a rank 1 matrix. The 1-rank approximation of $\overline{\mathbf{R}}(\tau)$ is denoted by $\overline{\mathbf{R}}_1(\tau)$ and is computed by truncating its singular-value decomposition (SVD):

$$\overline{\mathbf{R}}_1(\tau) = \sigma_\tau \mathbf{u}_\tau \mathbf{v}_\tau^H \tag{12}$$

where σ_τ is the first singular value, and $\mathbf{u}_\tau$ and $\mathbf{v}_\tau$ denote the left and right singular vectors, respectively. The 1-rank approximation also performs denoising. If there are a number of reflections with the same delay, $\mathbf{v}_\tau$ will be approximately equal to $\mathbf{y}(\Omega'_k)$ (the direct signal), and $\mathbf{u}_\tau$ will be a combination of some $\mathbf{y}(\Omega_k)$ (the reflections).

A way to detect τ values with reflections is developed by looking at $\rho(\tau)$ and $\hat{\Omega}'(\tau)$:

$$\rho(\tau) = \max_{\Omega' \in S^2} \left| \mathbf{y}(\Omega')^H \mathbf{v}_\tau \right| \tag{13}$$

$$\hat{\Omega}'(\tau) = arg \max_{\Omega' \in S^2} \left| \mathbf{y}(\Omega')^H \mathbf{v}_\tau \right| \tag{14}$$

where S^2 represents the unit sphere, $\mathbf{y}(\Omega')$ and $\mathbf{v}_\tau$ are unit vectors, and, using the Cauchy-Schwartz inequality, $\rho(\tau) \leq 1$. $\hat{\Omega}'(\tau)$ is an estimation of Ω_0 (the direct signal DOA). For each τ with $\rho(\tau)$ higher than a threshold ρ_{min} (set empirically), and $\hat{\Omega}'(\tau)$ close to Ω_0 up to a threshold Ω_{th} (set empirically), the DOA is computed using the orthogonal matching pursuit (OMP) algorithm [26], which finds the smallest set of directions $\hat{\Omega}_1, \ldots, \hat{\Omega}_S$ and coefficients $x_1, \ldots, x_S$ such that:

$$\| \sum_{s=1}^{S} x_s \mathbf{y}(\hat{\Omega}_s) - \mathbf{u}_\tau \|^2 \leq \varepsilon_u \tag{15}$$

where $\varepsilon_u \in (0, 1)$ is a threshold, and the norm is an L^2 norm. S is a parameter limited to be less than S_{max}, which is a threshold that determines the maximum number of reflections to detect by the OMP method.

The final part of PHALCOR is clustering, with the objective of eliminating noise effects and clustering all delay and DOA estimates into dominant groups representing reflections. The DBSCAN algorithm [27] is the method which is used.

3 FF-PHALCOR

In PHALCOR, representation of the SCM as in Equation (8) is possible because the steering matrix $\mathbf{H}(f,\Omega)$ is frequency-independent. However, this is not the case for an arbitrary array. As a consequence, the structure in Equation (11), with its spatial - frequency separation, which is the basis of the algorithm, is no longer available.

To overcome this limitation, in this paper, a focusing matrix [22] is used to obtain frequency independence of the steering matrix $\mathbf{H}(f,\Omega)$ in Equation (2). For this, a focusing matrix, $\mathbf{T}(f,f_0)$, is introduced, which satisfies:

$$\mathbf{T}(f,f_0)\mathbf{H}(f,\Omega) = \mathbf{H}(f_0,\Omega) \tag{16}$$

where f_0 is chosen to be the center frequency of a selected focusing frequency band. The center frequency is chosen assuming that the variation in $\mathbf{H}(f,\Omega)$ between this frequency and other frequencies in the band is small. In order to work well, the focusing process requires a finite bandwidth, which limits the bandwidth, B_w, defined in Equation (9). This issue will be discussed later. $\mathbf{T}(f,f_0)$ can be obtained by multiplying both sides of the equation with the pseudo-inverse of $\mathbf{H}(f,\Omega)$:

$$\mathbf{T}(f,f_0) = \mathbf{H}(f_0,\Omega)\mathbf{H}^{\dagger}(f,\Omega) \tag{17}$$

where $(\cdot)^{\dagger}$ is the pseudo-inverse operation. This computation minimizes the mean-square error (MSE) between $\mathbf{H}(f,\Omega)$ and $\mathbf{H}(f_0,\Omega)$. The focusing error is defined as:

$$e = \frac{\|\mathbf{H}(f_0,\Omega) - \mathbf{T}(f,f_0)\mathbf{H}(f,\Omega)\|_F^2}{\|\mathbf{H}(f_0,\Omega)\|_F^2} \tag{18}$$

where the norm is the Frobenius norm. Multiplying Equation (2) by $\mathbf{T}(f,f_0)$ leads to:

$$\tilde{\mathbf{p}}(f) = \mathbf{T}(f,f_0)\mathbf{H}(f,\Omega)\mathbf{s}(f) + \mathbf{T}(f,f_0)\mathbf{n}(f) \tag{19}$$

where $\tilde{\mathbf{p}}(f) = \mathbf{T}(f,f_0)\mathbf{p}(f)$, and substitution of Equation (16) into Equation (19) leads to:

$$\tilde{\mathbf{p}}(f) = \mathbf{H}(f_0,\Omega)\mathbf{s}(f) + \mathbf{T}(f,f_0)\mathbf{n}(f) \tag{20}$$

Now, assuming an ideal focusing process, the steering matrix is frequency-independent, and the SCM can be calculated by:

$$\mathbf{R}(f) = E[\tilde{\mathbf{p}}(f)\tilde{\mathbf{p}}(f)^H] \tag{21}$$

Substituting Equation (20) into Equation (21), and assuming that the signal and the noise are uncorrelated leads to:

$$\mathbf{R}(f) = \mathbf{H}(f_0,\Omega)\mathbf{M}(f)\mathbf{H}(f_0,\Omega)^H + \tilde{\mathbf{N}}(f) \tag{22}$$

where $\mathbf{M}(f)$ is defined similarly to Equation (8), $\tilde{\mathbf{N}}(f) = \mathbf{T}(f,f_0)\mathbf{N}(f)\mathbf{T}(f,f_0)^H$, and $\mathbf{N}(f) = E[\mathbf{n}(f)\mathbf{n}(f)^H]$. Using the phase-aligned transform as in Equation (9) (for a selected bandwidth B_w), and the rank 1 approximation by SVD as in Equation (12) leads to:

$$\rho(\tau) = \max_{\Omega' \in S^2} |\mathbf{h}(\Omega', f_0)^H \mathbf{v}_{\tau)}| \tag{23}$$

$$\hat{\Omega}'(\tau) = arg \max_{\Omega' \in S^2} |\mathbf{h}(\Omega', f_0)^H \mathbf{v}_{\tau)}| \tag{24}$$

The thresholds are defined similarly to in Equation (13) and Equation (14), only now with $\mathbf{h}(\Omega', f_0)$ instead of $\mathbf{y}(\Omega')$, where $\mathbf{h}(\Omega', f_0)$ are the columns of $\mathbf{H}(f_0,\Omega)$ as defined in Equation (4). This change also applies to Equation (15), which is replaced by:

$$\| \sum_{s=1}^{S} x_s \mathbf{h}(\hat{\Omega}_s) - \mathbf{u}_{\tau} \|^2 \leq \varepsilon_u \tag{25}$$

Here ε_u and S are defined as in Equation (15). Now, clustering is applied using the DBSCAN algorithm as in Section.II.

4 Simulation Study

In this section a simulation is presented, comparing the performance of the proposed algorithm to that of the original PHALCOR algorithm. The performance was tested on simulated data using a spherical array. However, it was performed over the signals recorded by the microphones instead of the a_{nm} signals. Analysis and insights will be provided in the next subsections.

4.1 Simulation Setup

The setup of the simulations includes a shoe-box room, a speaker, and a rigid spherical microphone array with 32 microphones, and a radius of $4.2\,cm$ (like the Eigenmike [28]). The room impulse response was simulated using the image method [23], and the speech signal is a 2.5 seconds sample from the TSP Speech Database [29] with 48kHz sampling frequency. The signals were generated with $N = 4$ for the a_{nm} signals and $N = 8$ for the recorded signal p. Two scenarios were tested - a

Table 1: Dimensions, reflection coefficients and reverberation times of the two rooms employed in the simulation study.

Room	*Dimensions* $[m]$	R	$T60\,[s]$
Large	$12\times10\times16$	0.9	1.8
Medium	$9\times7.5\times13.5$	0.9	1.4

Table 2: [x,y,z] coordinates of Source and array positions in the two rooms employed in the simulation study.

Room	*Source*	*Array*
Large	$[8,2.5,1.5]$	$[3,3\frac{1}{3},3.5]$
Medium	$[6,1.875,1.5]$	$[2.25,2.5,3.5]$

large room with the direct sound and 5 early reflections with arrival time less than $20\,ms$ after the direct sound, and a medium-sized room with the direct sound and 10 early reflections with arrival time less than $20\,ms$ after the direct sound. Room details are presented in Table 1. Source and array positions are detailed in Table 2.

4.2 Methodology

An STFT was applied to the signal p recorded by the microphones using a Hanning window of 8192 samples and an overlap of 75%. A frequency range of $[500,5000]$ Hz was chosen for the algorithms, with a bandwidth of $Bw = 2000$ Hz, and $J_f = 8$ creating an overlap of 1748 Hz between bands. J_t, which is defined in PHALCOR in order to create time bands, was set to be $J_t = 8$. It provides time bands of $0.2625s$ and overlap of $0.0375s$, and 66 bands for $2.5s$ long signal. Focusing is performed on the recorded signal, using a steering matrix with 900 directions sampled by the Fliege-Maier method [30] for nearly-uniform sampling. Then the algorithm outlined in Section 3 was applied using the parameters ρ_{min} (as will be discussed later), $\varepsilon_u = 0.63$, $\Omega_{th} = 10^\circ$, $S_{max} = 3$. The clustering parameters γ_Ω,γ_τ which are the weights in the clustering algorithm of PHALCOR, were fine-tuned for each scenario. For PHALCOR, the input signal was chosen to be a_{nm} that was obtained by performing PWD on the recorded microphone signals.

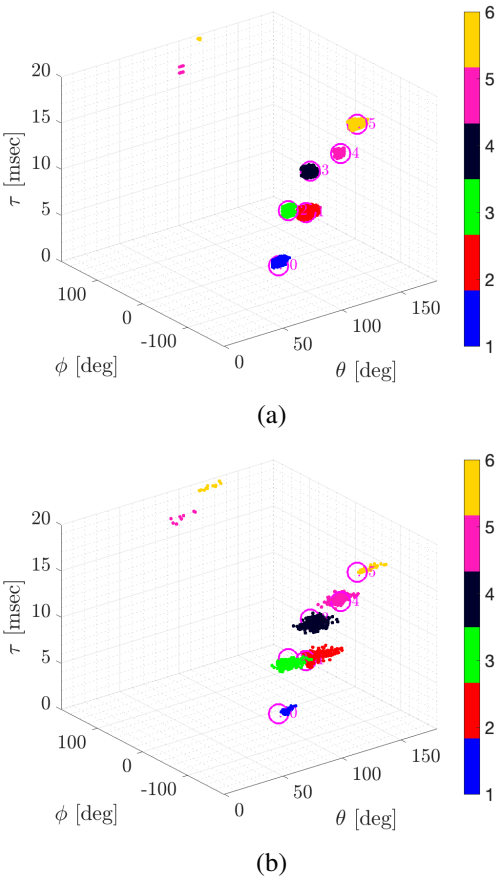


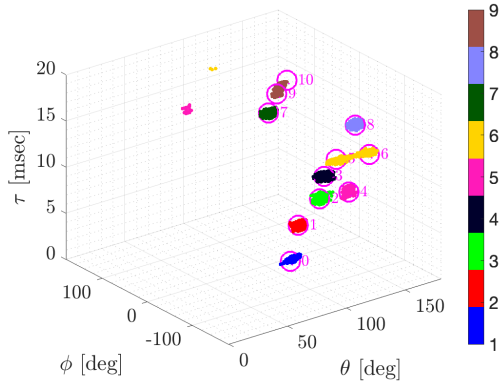
Fig. 1: Clustering of DOA and delay obtained by (a) PHALCOR and (b) FF-PHALCOR in the large room. τ [ms] is the delay, θ [deg] is the elevation and ϕ [deg] is the azimuth. The circles denote the true reflections from the ground-truth.

For both algorithms, a reflection is considered a true positive if its delay and DOA simultaneously match the delay and DOA of a true reflection up to a tolerance. The ground-truth is obtained with the image method, as mentioned above. The tolerances are set to be $500\mu s$ and 15° for the delay and the DOA, respectively. The Probability of Detection (PD) is defined as:

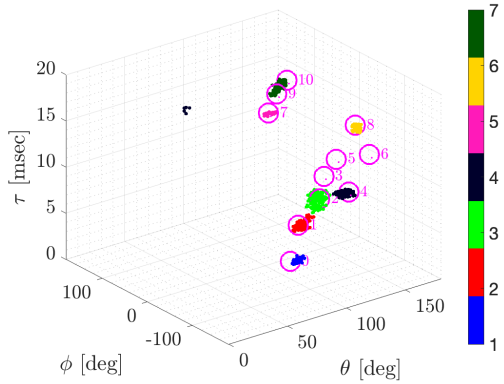
$$PD := \frac{\#\,true\ positive\ detections}{\#reflections\ in\ the\ ground\ truth} \tag{26}$$

The False Alarm is defined as:

$$PFA := \frac{\#\,false\ positive\ detections}{\#reflections\ in\ the\ ground\ truth} \tag{27}$$

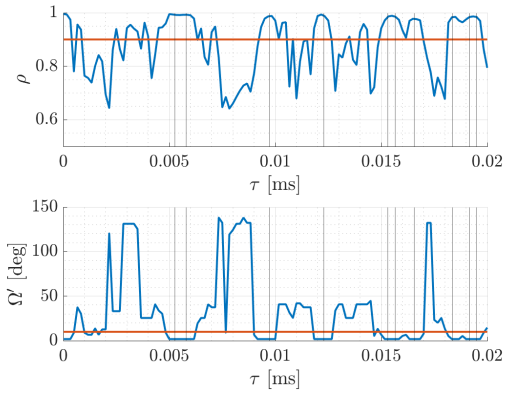


(a)

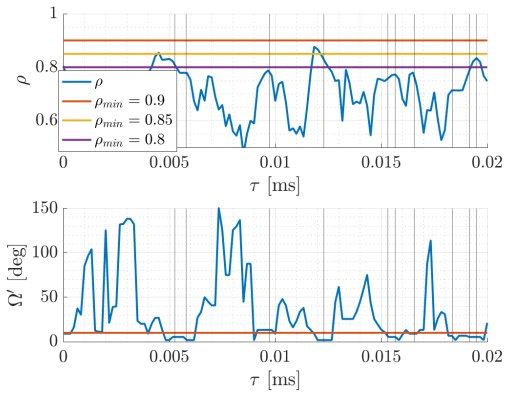


(b)

Fig. 2: Clustering of DOA and delay obtained by (a) PHALCOR and (b) FF-PHALCOR in the medium-sized room. τ [ms] is the delay, θ [deg] is the elevation and ϕ [deg] is the azimuth. The circles are the ground-truths.



(a)



(b)

Fig. 3: $\rho(\tau)$ of (a) a_{nm} and (b) p in the medium-sized room. The vertical lines are the reflections in the ground-truth. The horizontal lines present the threshold ρ_{min}.

4.3 Results and discussion

1. Figure 1 and Figure 2 present the clustering of DOAs and delays for the two algorithms in the large room and medium-sized room respectively. As can be seen, in the large room both algorithms successfully detected the direct sound and all 5 reflections, with $PD = 1$ and $PFA = 0$. However, in the medium-sized room, the algorithm on p found the direct signal and 6 reflections, with $PD = 7/10$ and with $PFA = 0$, while PHALCOR correctly found the direct sound and 8 reflections.

2. Focusing error: the error was computed using Equation (18). It was computed over the operating frequency bands showing very low results. The maximum error at the low frequencies is about $-70dB$, and at the high frequencies about $-39dB$. The minimum error is in the center of the band (the focused frequency). These results show that the focusing worked well for these examples.

3. $\rho(\tau)$ is computed as in Equation (23). A threshold ρ_{min} was defined such that every τ with $\rho(\tau) \geq \rho_{min}$ is consider a detection. In PHALCOR, ρ_{min} was set to be 0.9. For FF-PHALCOR, the threshold ρ_{min} was set to be 0.85, which was found to be more suitable, as can be seen in Figure 3. This graph shows the values of $\rho(\tau)$ in the frequency

band of [1511.7, 3503.9] Hz. It can be seen that $\rho(\tau)$ has fewer values over $\rho_{min} = 0.9$, but more over $\rho_{min} = 0.85$.

5 Conclusions

This paper demonstrated that the proposed algorithm detects all 5 reflections in the large room without false alarms, while in a medium-sized room it managed to detect most reflections but not all. This is a positive step forward in extending PHALCOR to arbitrary arrays. For future work, it is suggested that a more comprehensive investigation will be undertaken, with the aim of providing further insights. Performance analysis for arbitrary arrays is also proposed for future work.

References

[1] Kowalczyk, K., Kacprzak, S., and Ziółko, M., "On the extraction of early reflection signals for automatic speech recognition," in *2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP)*, pp. 351–355, IEEE, 2017.

[2] Peled, Y. and Rafaely, B., "Method for dereverberation and noise reduction using spherical microphone arrays," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 113–116, IEEE, 2010.

[3] Vincent, E., Bertin, N., Gribonval, R., and Bimbot, F., "From blind to guided audio source separation: How models and side information can improve the separation of sound," *IEEE Signal Processing Magazine*, 31(3), pp. 107–115, 2014.

[4] Javed, H. A., Moore, A. H., and Naylor, P. A., "Spherical microphone array acoustic rake receivers," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 111–115, IEEE, 2016.

[5] Mabande, E., Kowalczyk, K., Sun, H., and Kellermann, W., "Room geometry inference based on spherical microphone array eigenbeam processing," *The Journal of the Acoustical Society of America*, 134(4), pp. 2773–2789, 2013.

[6] Catic, J., Santurette, S., and Dau, T., "The role of reverberation-related binaural cues in the externalization of speech," *The Journal of the Acoustical Society of America*, 138(2), pp. 1154–1167, 2015.

[7] Vorländer, M., *Auralization: fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality*, Springer Science & Business Media, 2007.

[8] Pulkki, V., Delikaris-Manias, S., and Politis, A., *Parametric time-frequency domain spatial audio*, Wiley Online Library, 2018.

[9] Coleman, P., Franck, A., Jackson, P., Hughes, R. J., Remaggi, L., Melchior, F., et al., "Object-based reverberation for spatial audio," *Journal of the Audio Engineering Society*, 65(1/2), pp. 66–77, 2017.

[10] Kuttruff, H., *Room acoustics*, Crc Press, 2016.

[11] Sun, H., Mabande, E., Kowalczyk, K., and Kellermann, W., "Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing," *The Journal of the Acoustical Society of America*, 131(4), pp. 2828–2840, 2012.

[12] Jo, B. and Choi, J.-W., "Robust localization of early reflections in a room using semi real-valued EB-ESPRIT with three recurrence relations and Laplacian constraint," in *International Commission for Acoustics (ICA)*, International Commission for Acoustics (ICA), 2019.

[13] Ciuonzo, D., Romano, G., and Solimene, R., "Performance analysis of time-reversal MUSIC," *IEEE Transactions on Signal Processing*, 63(10), pp. 2650–2662, 2015.

[14] Ciuonzo, D., "On time-reversal imaging by statistical testing," *IEEE Signal Processing Letters*, 24(7), pp. 1024–1028, 2017.

[15] Wu, P. K. T., Epain, N., and Jin, C., "A dereverberation algorithm for spherical microphone arrays using compressed sensing techniques," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4053–4056, IEEE, 2012.

[16] Hu, Y., Lu, J., and Qiu, X., "Direction of arrival estimation of multiple acoustic sources using a maximum likelihood method in the spherical harmonic domain," *Applied Acoustics*, 135, pp. 85–90, 2018.

[17] Van Trees, H. L., *Optimum array processing: Part IV of detection, estimation, and modulation theory*, John Wiley & Sons, 2004.

[18] Dmochowski, J. P., Benesty, J., and Affes, S., "A generalized steered response power method for computationally viable source localization," *IEEE Transactions on Audio, Speech, and Language Processing*, 15(8), pp. 2510–2526, 2007.

[19] DiBiase, J. H., Silverman, H. F., and Brandstein, M. S., "Robust localization in reverberant rooms," in *Microphone Arrays*, pp. 157–180, Springer, 2001.

[20] Do, H. and Silverman, H. F., "SRP-PHAT methods of locating simultaneous multiple talkers using a frame of microphone array data," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 125–128, IEEE, 2010.

[21] Shlomo, T. and Rafaely, B., "Blind Localization of Early Room Reflections Using Phase Aligned Spatial Correlation," *IEEE Transactions on Signal Processing*, 69, pp. 1213–1225, 2021, doi: 10.1109/TSP.2021.3057495.

[22] Beit-On, H. and Rafaely, B., "Focusing and Frequency Smoothing for Arbitrary Arrays With Application to Speaker Localization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, pp. 2184–2193, 2020, doi:10.1109/TASLP.2020.3010098.

[23] Allen, J. B. and Berkley, D. A., "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, 65(4), pp. 943–950, 1979.

[24] Nadiri, O. and Rafaely, B., "Localization of Multiple Speakers under High Reverberation using a Spherical Microphone Array and the Direct-Path Dominance Test," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(10), pp. 1494–1505, 2014, doi:10.1109/TASLP.2014.2337846.

[25] Rafaely, B., *Fundamentals of spherical array processing*, volume 8, Springer, 2015.

[26] Cai, T. T. and Wang, L., "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Transactions on Information theory*, 57(7), pp. 4680–4688, 2011.

[27] Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al., "A density-based algorithm for discovering clusters in large spatial databases with noise." in *Kdd*, volume 96, pp. 226–231, 1996.

[28] Acoustics, M., "EM32 Eigenmike microphone array release notes (v17. 0)," *25 Summit Ave, Summit, NJ 07901, USA*, 2013.

[29] Kabal, P., "TSP speech database," *McGill University, Database Version*, 1(0), pp. 09–02, 2002.

[30] Fliege, J. and Maier, U., "A two-stage approach for computing cubature formulae for the sphere," in *Mathematik 139T, Universitat Dortmund, Fachbereich Mathematik, Universitat Dortmund, 44221*, Citeseer, 1996.