MULTIPLE SOUND SOURCE LOCALIZATION WITH RIGID SPHERICAL MICROPHONE ARRAYS VIA RESIDUAL ENERGY TEST

Mert Burkay Çöteli and Hüseyin Hacıhabiboğlu

Graduate School of Informatics
Middle East Technical University (METU)
Ankara, Turkey, TR-06800
{mert.coteli, hhuseyin}@metu.edu.tr

ABSTRACT

The estimation of the directions-of-arrival (DOAs) of multiple sound sources is a fundamental stage in acoustic scene analysis. Many application areas such as robot audition and object-based audio (OBA) broadcast require that DOA estimation is computationally efficient to allow real-time operation. We propose a new DOA estimation approach based on a sparse representation of recorded sound fields as a linear combination of spatially bandlimited impulses in this paper. The proposed algorithm operates on a time-frequency representation of the spherical harmonic components of the sound field. We describe a residual energy test that can identify time-frequency bins with a single active source. DOA estimation is carried out at each time-frequency bin by seeking a single-source dictionary atom which provides the best match to the steered response function calculated at the selected bins. We demonstrate the accuracy of the proposed method via a set of emulations using acoustic impulse responses measured in a highly reverberant room.

Index Terms— direction-of-arrival estimation, rigid spherical microphone arrays, orthogonal matching pursuit

1. INTRODUCTION

Sound source localization is essential in a variety of contexts including but not limited to object-based audio (OBA), robot audition, and acoustic surveillance [1]. Most if not all of these problems require direction-of-arrival (DOA) estimation algorithms that can effectively and accurately estimate source directions in or close to real-time.

DOA estimation typically involves the processing of microphone array recordings to extract the spatial information. Rigid spherical microphone arrays (RSMAs) allow DOA estimation using the spherical harmonic decomposition of the sound field which they trivially afford. There exist several different DOA estimation methods proposed for RSMAs with different computational demands and different levels of accuracy [2]. These methods range from those that are computationally efficient but have low DOA estimation accuracy such as pseudointensity vectors (PIV) [3] to those that are computationally costly while providing high DOA estimation accuracy such as eigenbeam multiple signal classification (EB-MUSIC) [4] and eigenbeam estimation of signal parameters via rotational invariance (EB-ESPRIT) [5]. Direct-path dominance (DPD) test addresses two of the inherent problems with EB-MUSIC: the necessity to find the dimensions of the noise subspace and to reduce the overall computational cost [6]. Recently proposed methods such as subspace PIV (SSPIV) [7] and hierarchical grid refinement (HiGRID) [8] provide a high estimation accuracy at a substantially lower computational cost. Combination of EB-MUSIC with the latter as a preprocessing stage for source counting was also shown to work well [9]. Similar bin selection approaches have also been proposed [10, 11, 12].

We propose a computationally efficient DOA estimation method which complements a source separation method that we recently proposed [13]. The method is based on a sparse representation of the steered response functional of a recorded sound field obtained using an RSMA. The method we propose uses the residual energy test (RENT) as a preprocessing stage to select time-frequency bins containing only a single sound source. RENT involves finding the best fitting atom from an overcomplete dictionary to the steered response functional (SRP) vector and using the ratio of the energies of the residual vector and the SRP vector as a direct path test. For bins that pass RENT, the DOA is identified and registered as the index of the best fitting atom. We then employ the histogram of DOA estimates to estimate the source DOAs.

This paper is organised as follows: Sec. 2 presents background information. We propose the new, dictionary-based DOA estimation method in Sec. 3. Sec. 4 presents an assessment of the proposed method's DOA estimation accuracy. Finally, Sec. 5 concludes the paper.

2. BACKGROUND

2.1. Spherical Harmonic Decomposition

A pressure distribution on a spherical surface can be represented as a linear combination of spherical harmonic functions, such that:

$$p(k, r, \Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} p_{nm}(kr) Y_n^m(\Omega) \tag{1}$$

using the coefficients:

$$p_{nm}(kr) = \iint_{\Omega \in S} f(\Omega)[Y_n^m(\Omega)]^* d\Omega \tag{2}$$

where $Y_n^m(\Omega)$ are the spherical harmonic functions of degree $n \in \mathbb{N}$ and order $m \in \mathbb{Z}$, $\Omega = (\theta, \phi)$ is a direction on the unit sphere, S, with θ and ϕ, the elevation and azimuth angles, respectively and $p_{nm}(kr)$ are the spherical harmonic decomposition (SHD) coefficients with $k = 2\pi f/c$ being the wave number and r being the radius. The compact representation afforded by SHD facilitates its

The work reported in this paper is supported by the Turkish Scientific and Technological Research Council (TÜBİTAK) Research Grant 113E513 "Spatial Audio Reproduction Using Analysis-based Synthesis Methods" (2014-2018).

use in many different fields of physics and signal processing, and in particular in the analysis of acoustic scenes [14].

2.2. Rigid Spherical Microphone Arrays

Rigid spherical microphone arrays (RSMAs) are particularly useful for obtaining the spherical harmonic decompositions (SHD) of sound fields subject to an order limitation dictated by the number of microphone elements in the array. For an RSMA with $Q \geq (N+1)^2$ elements positioned at $\{\Omega_q\}_{q=1\cdots Q}$, the SHD coefficients up to a degree N can be calculated using a numerical quadrature. Defining $\mathbf{p}_{nm} = [p_{00}(kr)\ p_{1-1}(kr)\ p_{10}(kr)\ p_{11}(kr) \cdots p_{NN}(kr)]^T$ as the vector containing the SHD coefficients, $\mathbf{W} = \mathrm{diag}\{w_q\}_{q=1\cdots Q}$ as the diagonal matrix of quadrature weights, $\mathbf{p} = [p_1\ p_2 \cdots p_Q]^T$ as the $Q \times 1$ vector containing the and $\mathbf{Y}$ as the $L \times (N+1)^2$ matrix with the l-th row given by:

$$\mathbf{y}(\Omega_q) = [Y_0^0(\Omega_q), Y_1^{-1}(\Omega_q), Y_1^0(\Omega_q), ..., Y_N^N(\Omega_q)], \quad (3)$$

the SHD coefficients can be calculated as:

$$\mathbf{p}_{nm} = \mathbf{Y}^H \mathbf{W} \mathbf{p}. \quad (4)$$

Note that discrete sampling limits the maximum order of spherical harmonic coefficients to N and the SHD coefficients obtained this way will converge to the true SHD coefficients only if the sampling scheme satisfies the *discrete orthonormality condition* [15].

While the expressions above are given in wave number domain, the assumption is made here that they also hold for a time-frequency representation of the signals.

2.3. Steered Response Functional

An important advantage of SHD is that it allows decoupling the frequency and direction dependent components of monochromatic plane waves. The SHD coefficients due to monochromatic plane wave incident from the direction Ω_s on a rigid sphere of radius r_a is given as:

$$p_{nm}(k, \Omega_s) = 4\pi i^n b_n(kr_a)[Y_n^m(\Omega_s)]^* \quad (5)$$

with the frequency-dependent component due to the acoustic scattering from the sphere given as:

$$b_n(kr) = j_n(kr) - \frac{j_n^{'}(kr_a)}{h_n^{(2)'}(kr_a)} h_n^{(2)}(kr). \quad (6)$$

where $j_n(\cdot)$, $h_n^{(2)}(\cdot)$, and $h_n^{(2)'}(\cdot)$ are, the spherical Bessel function of the first kind, spherical Hankel function of the second kind and its derivative with respect to its argument, respectively. Sound fields can be expressed as a combination of multiple plane waves. Let us assume that the sound field consists of S plane waves. The SHD coefficients for such a sound field are:

$$p_{nm}(k) = \sum_{s=1}^{S} \alpha_s p_{nm}(k, \Omega_s) \quad (7)$$

where $\alpha_s \in \mathbb{C}$ is the complex valued amplitude of each plane wave that constitutes the sound field. Simplification by normalizing the SHD coefficients allows eliminating the frequency dependence as well as the order-dependent quadrature phase shift, resulting in:

$$\widetilde{\mathbf{y}}^H = \mathbf{B}^{-1}\mathbf{P}\mathbf{a} \quad (8)$$

where $\mathbf{a} = [\alpha_1\ \alpha_2 \cdots \alpha_S]^T$ is the column vector of complex-valued amplitudes, $\mathbf{B}$ is a $(N+1)^2 \times (N+1)^2$ diagonal matrix containing the terms $4\pi i^n b_n(kr_a)$ in its diagonal and $\widetilde{\mathbf{y}} = [\widetilde{y}_{00}\ \widetilde{y}_{1-1} \cdots \widetilde{y}_{NN}]$ with:

$$\widetilde{y}_{nm} = \sum_{s=1}^{S} \alpha_s^* Y_n^m(\Omega_s). \quad (9)$$

and $\mathbf{P}$ is a $(N+1)^2 \times S$ matrix consisting of columns of the form $\mathbf{p}_s = [p_{00}(k, \Omega_1)\ p_{1-1}(k, \Omega_1) \cdots p_{NN}(k, \Omega_1)]^T$.

A maximally directive steered beam in a given direction $\overline{\Omega}_d$ can be formed multiplying $\widetilde{\mathbf{y}}^H$ with the steering vector $\mathbf{y}(\overline{\Omega_d})$ such that:

$$\begin{aligned} s(\Omega_d) &= \mathbf{y}(\overline{\Omega_d})\widetilde{\mathbf{y}}^H = \mathbf{y}(\overline{\Omega_d})\mathbf{B}^{-1}\mathbf{P}\mathbf{a} \quad (10) \\ &= \sum_{s=1}^{S}\sum_{n=0}^{N}\sum_{m=-n}^{n} \alpha_s Y_n^m(\overline{\Omega_d}) Y_n^m(\Omega_s)^*. \end{aligned}$$

More generally, a *steered response functional* (SRF) vector can be obtained by calculating the steered response in D discrete directions such that:

$$\mathbf{s} = \mathbf{Y}_s \widetilde{\mathbf{y}}^H \quad (11)$$

where $\mathbf{Y}_s$ is a $D \times (N+1)^2$ matrix whose rows consist of the steering vectors $\{\mathbf{y}(\Omega_d)\}_{d=1\cdots D}$. Note that the steered response power (SRP) map at the same steering directions can be obtained from the SRF vector as $\mathrm{diag}\{\mathbf{s}\mathbf{s}^H\}$ and would have positive, real-valued terms as opposed to the SRF vector which will, in general, have complex-valued terms.

3. TIME-FREQUENCY BIN SELECTION AND DOA ESTIMATION VIA RENT

The method we propose consists of two stages: (1) joint time-frequency bin selection and DOA estimation, and (2) clustering for multiple DOA estimation. These stages are described below.

3.1. Residual Energy Test (RENT)

It may be shown via the spherical harmonic addition theorem [16] that:

$$\begin{aligned} \lambda_N(\overline{\Omega}, \Omega_s) &= \sum_{n=0}^{N}\sum_{m=-n}^{n} Y_n^m(\overline{\Omega}) Y_n^m(\Omega_s)^* \quad (12) \\ &= \frac{P_{N+1}(\cos\Theta) - P_N(\cos\Theta)}{P_1(\cos\Theta) - P_0(\cos\Theta)} \end{aligned}$$

where Θ is the angle between the unit vectors in directions $\overline{\Omega}$ and Ω_s. Note that $\lim_{N\to\infty} \lambda_N(\overline{\Omega}, \Omega_s) = \delta(\overline{\Omega} - \Omega_s)$, an impulse at $\overline{\Omega} = \Omega_s$, and also that $\lambda_N(\overline{\Omega}, \Omega_s)$ is a spatially bandlimited pulse localized at $\Omega = \Omega_s$. Henceforth we will call $\lambda_N(\overline{\Omega}, \Omega_s)$ as a *Legendre pulse* of order N. We can express steered response given in (10) as a linear combination of Legendre pulses such that:

$$s(\overline{\Omega}_d) = \sum_{s=1}^{S} \alpha_s \lambda_N(\overline{\Omega}_d, \Omega_s). \quad (13)$$

Using the expression in (13) it is possible to express the SRF sampled at D discrete directions n matrix form as:

$$\mathbf{s} = \mathbf{L}_{DS}\mathbf{a}_S \quad (14)$$

where $\mathbf{L}_{DS} \in \mathbb{R}^{D\times S}$ is a matrix with the columns given as $\mathbf{\Lambda}_s = [\lambda(\overline{\Omega}_1, \Omega_s)\ \lambda(\overline{\Omega}_2, \Omega_s) \cdots \lambda(\overline{\Omega}_D, \Omega_s)]^T$, and $\mathbf{a} \in \mathbb{C}^{S\times 1}$ is the column vector containing complex amplitudes. Note that while $\mathbf{s}$ can be trivially calculated, usually neither Ω_s nor S is known. Note also that while S can potentially be very large in a reverberant environment with a strong diffuse component, only a few complex coefficients, α_s, will have a large magnitude.

A sparse approximation, $\mathbf{s}_G = \mathbf{L}_{DG}\mathbf{a}_G$ using a smaller number of G terms is possible. The optimal selection of G and $\{\Omega_g\}$ corresponds to the minimization of the approximation error:

$$\epsilon = \|\mathbf{s} - \mathbf{s}_G\|^2 + T^2 S \tag{15}$$

where T is a Lagrange multiplier which penalizes the number of terms used in the approximation. While the problem of finding the best G-term representation is NP-hard [17], satisfactory solutions can be obtained using computationally tractable algorithms such as orthogonal matching pursuit (OMP) [18].

We propose the *residual energy test* (RENT) as a time-frequency bin selection mechanism to identify the bins that contain a single source. RENT is based on a single-iteration OMP used to jointly select the time-frequency bins and to estimate source DOAs. The vector to be approximated is the SRF. The employed dictionary consists of Legendre pulses sampled at pixel centroids of a HEALPix grid [19] with the centre directions also corresponding to pixel centroids from a HEALPix grid of the same or higher resolution. Let us express the centroid of a the pixel k from the HEALPix grid at level K as $\overline{\Omega}_{K,p}$ where the number of pixels that completely cover the unit sphere at the same level is $P = 12 \times 2^{2K}$. The centre direction of the Legendre pulses can be sampled from the HEALpix grid with the same or higher resolution $\widetilde{K} \geq K$ resulting in an overcomplete dictionary consisting of the set of $\widetilde{P} = 12 \times 2^{2K}$ atoms $\mathbf{\Phi} = \{\mathbf{\Lambda}_{\widetilde{p}} = [\Lambda(\overline{\Omega}_1, \overline{\Omega}_{\widetilde{p}}) \cdots \Lambda(\overline{\Omega}_P, \overline{\Omega}_{\widetilde{p}})]^T\}_{\widetilde{p}=1\cdots\widetilde{P}}$. RENT identifies time-frequency bins with a single source and estimates its DOA using the following steps:

1. Obtain the windowed Fourier transform $\{P_q(\tau,\kappa)\}_{q=1\cdots Q}$ of microphone array signals, where κ is the frequency index and τ is the time index.
2. Obtain the SHD coefficients for each time-frequency bin, (τ,κ), using (4)
3. For each time-frequency bin, calculate the SRF vector, $\mathbf{s}(\tau,\kappa) \in \mathbb{R}^{P\times 1}$ at a finite number of P directions quasi-uniformly sampled on the HEALPix grid at the resolution level K.
4. Find the dictionary atom which best matches the SRF vector such that:
$$\mathbf{\Lambda}_{\widetilde{p}} = \max_{\mathbf{\Lambda}\in\mathbf{\Phi}} \langle \mathbf{\Lambda}, \mathbf{s}(\tau,\kappa)\rangle \tag{16}$$
5. Calculate the residual error vector orthogonal to the SRF vector:
$$\begin{aligned} \mathbf{r}(\tau,\kappa) &= \mathbf{R}_{\widetilde{p}}\,\mathbf{s}(\tau,\kappa) \qquad (17)\\ &= \left[\mathbf{I} - \mathbf{\Lambda}_{\widetilde{p}}\left(\mathbf{\Lambda}_{\widetilde{p}}^T\mathbf{\Lambda}_{\widetilde{p}}\right)^{-1}\mathbf{\Lambda}_{\widetilde{p}}^T\right]\mathbf{s}(\tau,\kappa). \end{aligned}$$
Note that $\mathbf{R}_{\widetilde{p}}$ can be calculated offline for each atom in the dictionary and stored.
6. Calculate ones' complement of the ratio of the residual energy and the total energy of the SRF vector, such that:
$$\mathcal{R}(\tau,\kappa) = 1 - \frac{\|\mathbf{r}(\tau,\kappa)\|^2}{\|\mathbf{s}(\tau,\kappa)\|^2} \tag{18}$$
where $0 < \mathcal{R}(\tau,\kappa) < 1$. Notice that for a single plane wave in acoustic free field $\mathbf{s}(\tau,\kappa)$ can be represented by a single atom and the energy of the residual vector would be zero resulting in $\mathcal{R}(\tau,\kappa) = 1$ indicating that $\mathcal{R}(\tau,\kappa)$ is close to unity when the time-frequency bin contains a single dominant source.

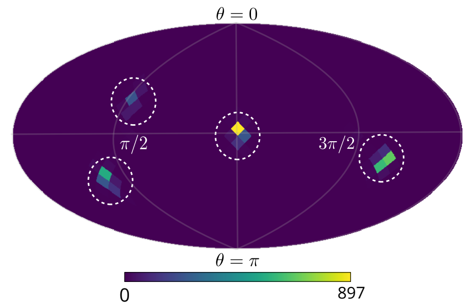

Fig. 1. DOA histogram of four sound sources located at $(118°, 116°)$, $(68°, 90°)$, $(90°, 0°)$, and $(105°, 243°)$. Mollweide projection is used in the figure.

7. Identify the set of time-frequency bins for which (18) is greater than a selected threshold, such that:
$$\mathcal{S}_{\text{RENT}} = \{(\tau,\kappa) : \mathcal{R}(\tau,\kappa) > THR\} \tag{19}$$
8. Register the index of dictionary atom, $\widetilde{p}$, identified for each time-frequency bin found in (16) as the DOA estimate for that time-frequency bin.

After the DOA estimates are obtained for each time-frequency bin that passes RENT, a histogram is formed whose bin centers are aligned with pixel centroids of the employed HEALPix grid.

3.2. Multiple DOA Estimation via Agglomerative Histogram Clustering

RENT is followed by clustering the resulting histogram to obtain the DOA estimates. The main assumption in the development of the histogram clustering approach herein is that sources are spatially separated sufficiently well so that clusters can be identified based on their contiguity. While the proposed clustering approach is similar to neighboring nodes labeling (NNL) approach presented in [8], it is computationally more effective.

The DOA histogram contains the number of occurrences (i.e. counts) that a given pixel on the HEALPix grid was identified as the DOA. (see Fig. 1) The pixels are first sorted into a list according to counts. The pixel with the highest count is removed from the sorted list of selected pixels and is defined as the first cluster. The neighborhood of the cluster is defined as the set of pixels surrounding it. The next pixel in the sorted list is selected and removed from the list. If the selected pixel corresponds to an element already in the cluster, the count of the corresponding element in the cluster is incremented. If the pixel is in the neighborhood of the first cluster, it is added to the cluster and the neighborhood of the cluster is expanded to account for the new pixel. If it is not in the neighborhood, a new cluster is formed. For each pixel from the sorted list, the neighborhoods of all existing clusters are checked and if the new pixel is not in one of the existing clusters or their neighborhoods, a new cluster is formed. The iteration continues until the list of selected pixels is exhausted. The identified clusters provide both source count and DOAs.

4. PERFORMANCE EVALUATION

We simulated scenarios using up to 6 concurrently active sources by convolving real acoustic impulse measurements (AIRs) with anechoic sound samples. The first four sources (S1-S4) were 4s-long

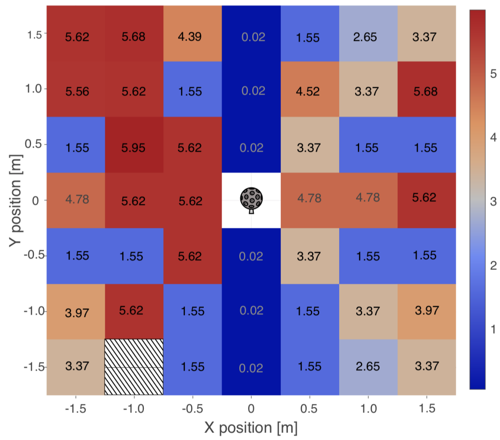

Fig. 2. Heatmap showing the DOA estimation errors for a single source. Eigenmike is positioned at the center. The single source that cannot be localized is denoted by a hatched pattern.

violin tracks from the anechoic recordings of Mahlers Symphony Nr. 1, fourth movement [20]. The remaining two sources (S5-S6) were female and male anechoic speech recordings [21]. The atoms in the dictionary used in RENT are Legendre pulses sampled on a HEALPix grid of resolution level 3 resulting in 768 atoms of size 768×1. This corresponds to $7.33°$ resolution. The frequency range selected for the analysis is between 480 Hz and 3750 Hz. Window size and overlap are 2048 and 50% respectively. Sampling rate is selected as 48 kHZ.

Multichannel acoustic impulse responses (AIRs) recorded using an RSMA (Eigenmike em32) in an empty classroom ($T60 \approx 1.12s$) were used in the simulations. The set of AIRs were measured on a set of rectilinear grid points in a 3 m $\times$ 3 m $\times$ 1.2 m volume with 0.5 m separation in the x and y directions and 0.3 m separation in the z direction. Eigenmike was positioned at the center of the grid.

We simulated RSMA recordings using all AIRs in the horizontal plane in order to characterize its performance when only a single source is present. An anechoic violin signal was used while obtaining the simulated recordings. Fig. 2 shows the DOA estimation errors for the 48 directions. The maximum error was $5.95°$. RENT could not localize one of the sources as the error was greater than $10°$. The figure also shows that better accuracy is possible in the vertical axial direction which may be due to the perfect alignment of the source and dictionary atom directions.

For scenarios involving multiple sources, AIRs were randomly selected from the dataset subject only to the constraint that the angular separation between any two sources is greater than $\pi/10$. Random simulations for each source count was repeated 10 times. Fig. 3 shows the results for different number of sources. It may be observed that the average error is always less than $5°$ and that the error does not substantially depend on the number of emulated sources. Table 1 shows the descriptive statistics of the results. The last row shows the average number of identified sources. It may be observed that up to 4 sources, RENT was able to identify all sources correctly. For the five source scenario, RENT missed one source each in 3 out of 10 scenarios. For the six source scenario, RENT missed one source each in 2 out of 10 scenarios. No spurious sources were identified.

We also compared RENT with other state of the art methods: HiGRID [8], SSPIV [7] and DPD-MUSIC [6]. The test involved simulating four sources at diagonal positions at a distance of 1.41 m. The simulated sources were positioned in the horizontal plane at the directions of $(\theta, \phi) = (90°, 45°)$, $(90°, 135°)$, $(90°, 225°)$ and, $(90°, 315°)$. The results shown in Table 2 indicate that RENT provides worse DOA estimations than HiGRID and DPD-MUSIC, but better DOA estimations than SSPIV for the tested condition. However, since the computational cost of RENT is lower than both HiGRID and DPD-MUSIC, it provides an alternative to those methods.

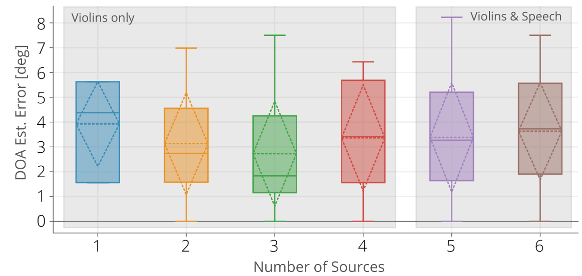

Fig. 3. DOA estimation errors for different number of sources. The first four sources are violins playing in unison and the last two sources are anechoic speech signals. The dotted triangles indicate mean $\pm$ standard deviation points, respectively.

	Number of Sources					
	1	**2**	**3**	**4**	**5**	**6**
Mean	$3.92°$	$3.13°$	$2.72°$	$3.36°$	$3.38°$	$3.63°$
Min.	$1.55°$	$0.02°$	$0.02°$	$0.02°$	$0.02°$	$0.02°$
Max.	$5.62°$	$6.98°$	$7.5°$	$6.42°$	$8.22°$	$7.5°$
Std. Dev.	$1.71°$	$2.05°$	$2.11°$	$2.13°$	$2.20°$	$1.95°$
Avg. Src.	1	2	3	4	4.7	5.8

Table 1. Descriptive statistics of the DOA estimation experiments.

Source	HiGRID	SSPIV	DPD-MUSIC	RENT
1	$1.15°$	$6.6°$	$3.45°$	$3.38°$
2	$1.34°$	$4.11°$	$4.21°$	$3.38°$
3	$0.7°$	$2.31°$	$2.28°$	$5.62°$
4	$1.18°$	$7.2°$	$1.12°$	$3.38°$
Avg.	$1.09°$	$5.06°$	$2.77°$	$3.94°$

Table 2. DOA estimation errors of RENT in comparison with the state of the art DOA estimation methods

5. CONCLUSIONS

Direct path features are used to select time-frequency bins containing substantial contributions from a single source since reliable DOA estimations can be obtained from these bins. The most popular method for selecting bins this way is the direct path dominance (DPD) test which involves calculating the singular value decomposition (SVD) of the time-frequency averaged spatial correlation matrix and is thus computationally costly. We proposed an alternative direct path test and DOA estimation method called residual energy test (RENT) based on orthogonal matching pursuit (OMP). An evaluation of its performance in terms of the DOA estimation accuracy it provides revealed that RENT is both robust to a high level of reverberation and provides DOA estimation accuracy comparable to state of the art. We found that the running time for RENT is more than an order of magnitude less than the state of the art. A thorough evaluation of its computational complexity is left as future work.

6. REFERENCES

[1] M. Brandstein and D. Ward, "Microphone arrays: signal processing techniques and applications," *Springer Science & Business Media*, 2013.

[2] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, *Theory and Applications of Spherical Microphone Array Processing*, ser. Springer Topics in Signal Processing. Cham: Springer, Aug. 2016, vol. 9.

[3] ——, "3D source localization in the spherical harmonic domain using a pseudointensity vector," in *18th European Signal Process. Conf. (EUSIPCO 2010)*. IEEE, 2010, pp. 442–446.

[4] H. Sun, E. Mabande, K. Kowalczyk, and W. Kellermann, "Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing," *J. Acoust. Soc. Am.*, Apr. 2012.

[5] H. Teutsch and W. Kellermann, "Detection and localization of multiple wideband acoustic sources based on wavefield decomposition using spherical apertures," *IEEE Int. Conf. on Acoust. Speech and Signal Process.*, pp. 5276—5279, Apr. 2008.

[6] O. Nadiri and B. Rafaely, "Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 10, pp. 1494–15 059, 2014.

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

[8] M. B. Çöteli, O. Olgun, and H. Hacıhabiboğlu, "Multiple Sound Source Localization With Steered Response Power Density and Hierarchical Grid Refinement," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 11, pp. 2215–2229, 2018.

[9] O. Olgun and H. Hacıhabiboğlu, "Localization of Multiple Sound Sources In The Spherical Harmonic Domain With Hierarchical Grid Refinement and EB-MUSIC ," *Int. Workshop on Acoust. Sig. Enhancement (IWAENC 2018)*, pp. 101–105, September 2018.

[10] V. Tourbabin and B. Rafaely, "Speaker localization by humanoid robots in reverberant environments," in *2014 IEEE 28th Convention of Electrical Electronics Engineers in Israel (IEEEI)*, Dec 2014, pp. 1–5.

[11] B. Rafaely and K. Alhaiany, "Speaker localization using direct path dominance test based on sound field directivity," *Signal Processing*, vol. 143, no. 2, pp. 42–47, 2018.

[12] V. Tourbabin, D. L. Alon, and R. Mehra, "Space domain-based selection of direct-sound bins in the context of improved robustness to reverberation in direction of arrival estimation," in *Proc. 11th European Congress and Exposition on Noise Control Engineering (EURONOISE'18)*, Crete, Greece, May 27-31 2018, pp. 2589–2596.

[13] M. B. Çöteli and H. Hacihabiboglu, "Acoustic Source Separation Using Rigid Spherical Microphone Arrays Via Spatially Weighted Orthogonal Matching Pursuit," *Int. Workshop on Acoust. Signal Enhancement (IWAENC 2018)*, pp. 81–85, September 2018.

[14] E. G. Williams, *Fourier Acoustics*, ser. Sound Radiation and Nearfield Acoustic Holography. Academic Press, Sept. 1999.

[15] J. Meyer and G. W. Elko, "Spherical microphone arrays for 3d sound recording," in *Audio Signal Processing for Next-generation Multimedia Communication Systems*, B. J. Huang Y, Ed. New York, NY, USA: Springer US, 2004, pp. 67–89.

[16] B. Rafaely, Y. Peled, M. Agmon, D. Khaykin, and E. Fisher, "Spherical Microphone Array Beamforming," in *Speech Processing in Modern Communication*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 281–305.

[17] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Burlington, MA, USA: Academic Press, 2008.

[18] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, USA, November 1993, pp. 40–44.

[19] K. M. Gorski, E. Hivon, A. Banday, B. Wandelt, F. Hansen, M.Reinecke, and M. Bartelmann, "HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere," *Astrophys. J.*, vol. 622, pp. 759–771, 2005.

[20] J. Pätynen, V. Pulkki, and T. Lokki, "Anechoic recording system for symphony orchestra," *Acta Acust. united with Acust.*, vol. 94, no. 6, pp. 856–865, June 2008.

[21] Bang & Olufsen, "Music for Archimedes," Audio CD, 1992.