PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Data-driven threshold selection for direct path dominance test

Orhun Olgun and Hüseyin Hacıhabiboğlu

[1]Graduate School of Informatics, Middle East Technical University (METU), Ankara, TR-06800, Turkey

Abstract

Direction-of-arrival estimation methods, when used with recordings made in enclosures are negatively affected by the reflections and reverberation in that enclosure. Direct path dominance (DPD) test was proposed as a pre-processing stage which can provide better DOA estimates by selecting only the time-frequency bins with a single dominant sound source component prior to DOA estimation, thereby reducing the total computational cost. DPD test involves selecting bins for which the ratio of the two largest singular values of the local spatial correlation matrix is above a threshold. The selection of this threshold is typically carried out in an ad hoc manner, which hinders the generalisation of this approach. This selection method also potentially increases the total computational cost or reduces the accuracy of DOA estimation. We propose a DPD test threshold selection method based on a data-driven statistical model. The model is based on the approximation of the singular value ratio distribution of the spatial correlation matrices as a generalised Pareto distribution and allows selecting time-frequency bins based on their probability of occurrence. We demonstrate the application of this threshold selection method via emulations using acoustic impulse responses measured in a highly reverberant room with a rigid spherical microphone array.

Keywords: direction-of-arrival estimation, spherical microphone arrays, direct path dominance test (DPD)

1 INTRODUCTION

Direction of arrival (DOA) estimation is one of the essential stages of many different audio and acoustic signal processing applications. Rigid spherical microphone arrays (RSMA) can be used to obtain the spherical harmonic decomposition of sound fields which enables the development of 3D DOA estimation methods.

Several DOA estimation methods have been proposed for RSMAs. For example steered response power (SRP) maps [8] are obtained by steering a maximally directive beam and identifying directions that maximize the output power as source DOAs. Other methods such as pseudo-intensity vector (PIV) [7] and its improved version, augmented intensity vectors (AIV) [5] are based on the energetic analysis of sound fields. A subspace-based extension of PIV method was also proposed [11]. EB-MUSIC [18] and EB-ESPRIT [19] are methods that are based on the adaptation of high resolution spectrum estimation methods to the spherical harmonic domain. More recently, hierarchical grid refinement (HiGRID) was proposed as a spatial entropy-based DOA estimation method [1]. It was shown that HiGRID has a lower computational cost for time-frequency bins that contain a single sound source.

One of the main problems with EB-MUSIC is the necessity for prior information on the number of sources in a time-frequency bin. Direct path dominance (DPD) test which can identify time-frequency bins with a single dominant source was proposed to address this problem [12]. DPD test has since been used together with PIV and SSPIV to improve the DOA estimation accuracy of these methods [10].

DPD test is based on the calculation of the ratio of the effective rank of the spatial correlation matrix and selecting the time-frequency bins that have unit effective rank. This is done by calculating the ratio of the largest two singular values of the spatial correlation matrix. Time-frequency bins for which this ratio is above a given threshold are selected as bins containing contributions only from a single source and used for DOA estimation. DPD test is very effective in selecting time-frequency bins with a single sound source when the employed threshold is selected appropriately. When it is selected to be higher than necessary, it eliminates bins that contain useful DOA information, potentially resulting in the detection of only the most prominent sources in multiple source localisation scenarios. In contrast, carrying out the DPD test with a lower than necessary threshold results in selecting time-frequency bins that are possibly contaminated by other sources or coherent reflections, degrading the DOA estimation accuracy, and increasing the computational cost of DOA estimation.

We present a data-driven threshold selection method based on the approximation of the long-term statistical distribution of singular value ratios as a generalised Pareto distribution. The method allows selecting bins which have a prescribed probability of occurrence. It also allows an approximate specification of the number of bins that will be selected, thereby making it possible to adapt the DOA estimation process to the available computational power. We evaluate the capabilities of the method using emulations obtained with acoustic impulse responses measured using an RSMA and with HiGRID as the DOA estimation method.

This paper is organized as follows. Sec. 2 presents a brief overview of spherical harmonic decomposition. Sec. 3 explains DPD test and DPD test threshold selection using the generalized Pareto distribution (GPD). Sec. 4 presents an evaluation of DPD threshold with the proposed approach for bin selection and DOA estimation using the hierarhical grid refinement (HIGRID) algorithm for DOA estimation. Sec. 5 concludes the paper.

2 SPHERICAL HARMONIC DECOMPOSITION

It is possible to approximate a spatially band-limited sound field using a linear combination of a finite number of spherical harmonic functions. For a band-limited pressure distribution $p(k,r,\Omega)$ defined on a sphere:

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

where $p_{nm}(k,r)$ are the SHD coefficients that depend on the wave number k and the radius r of sphere, $Y_n^m(\Omega)$ is the spherical harmonic function of order $n \in \mathbb{N}$ and degree $m \in \mathbb{Z}$ evaluated at the direction $\Omega = (\theta, \phi)$, N is the maximum order of the decomposition. SHD coefficients are related to the sound field, such that:

$$p_{nm}(k,r) = \iint_S p(k,r,\Omega) \, [Y_n^m(\Omega)]^* \, d\Omega \tag{2}$$

where S is a spherical surface.

In practice, sound fields can be recorded via spherical microphone arrays. Such arrays sample the sound field at a finite number of Q points, $\{\Omega_q\}_{q=1\cdots Q}$ on a spherical surface. If the sampling scheme satisfies the orthogonality conditions of the spherical harmonic functions [9], the following cubature can be used to obtain the SHD coefficients:

$$p_{nm}(k,r) = \sum_{q=1}^{Q} w_q p(k,r,\Omega_q) \, [Y_n^m(\Omega)]^* \tag{3}$$

where $\{w_q\}_{q=1\cdots Q}$ are the cubature coefficients. The maximum SHD order and the number of microphones on the array are related by $Q \geq (N+1)^2$.

SHD coefficients of the sound field around a rigid sphere due to a single monochromatic plane wave with frequency f, wavenumber $k = 2\pi f$, arrival direction Ω_s, and amplitude $a_l(k) \in \mathbb{C}$ are given as [16]:

$$p_{nm}^{(l)}(k,r) = a_l(k) e^{j\mathbf{k}\cdot\mathbf{r}} = 4\pi i^n a_l(k) b_n(kr) \, [Y_n^m(\Omega_s)]^* \tag{4}$$

where $b_n(kr)$ is a frequency dependent weight that is due to a combination of the incoming and scattered sound. Notice that we omitted the time dependence of the plane wave for simplicity. SHD coefficients can be normalized to obtain the spherical harmonic functions. For a monochromatic plane wave with the complex amplitude $\alpha_l(k) \in \mathbb{C}$, the normalized SHD coefficients are:

$$\tilde{p}_{nm}^{(l)}(k) = \frac{p_{nm}(k,r)}{4\pi i^n b_n(kr)} = \alpha_l(k) \, [Y_n^m(\Omega_s)]^* \tag{5}$$

Sound fields can be represented as linear combination of multiple plane waves. Since SHD is a linear operation, the SHD coefficients of a sound field comprising multiple plane waves will consist of a linear combination of the SHD coefficients of its constituents. For a sound field that consist of L plane waves:

$$\tilde{p}_{nm}(k) = \sum_{l=1}^{L} \alpha_l(k) \left[Y_n^m(\theta, \phi)\right]^*, \tag{6}$$

which can be expressed in matrix notation as $\mathbf{p}_{nm}(k) = \mathbf{Y}_s^H \mathbf{a}(k)$ where $\mathbf{a}(k) = [a_1(k), a_2(k), ..., a_L(k)]^T$ is the $L \times 1$ complex amplitude vector, $\mathbf{Y}_s$ is $L \times (N+1)^2$ beamspace manifold matrix with the l-th row given by:

$$\mathbf{y}(\Omega_l) = [Y_0^0(\Omega_l), Y_1^{-1}(\Omega_l), Y_1^0(\Omega_l), ..., Y_N^N(\Omega_l)] \tag{7}$$

In general terms, DOA estimation with RSMAs corresponds to the identification of Ω_l given $\mathbf{p}_{nm}(k)$ subject to certain practical constraints such as sensor noise, and by using certain assumptions appropriate for the task such as *a priori* knowledge on the number of sound sources.

3 THRESHOLD SELECTION FOR DIRECT PATH DOMINANCE TEST

Direct path dominance (DPD) test is a time-frequency bin selection method that was originally proposed for use with eigenbeam multiple signal classification (EB-MUSIC) algorithm [12]. Apart from its advantages when used with EB-MUSIC, DPD test can be used as a preprocessing stage for any DOA estimation method. DPD test operates on the windowed Fourier transform of the SHD coefficients of the sound field. Therefore, the discussion that follows assumes the SHD coefficients $\mathbf{p}_{nm}(\tau, \kappa)$ are available up to a given order, N.

DPD test [12] is based on the observation that, the spatial correlation matrix has unit rank for time-frequency bins where only a single direct component exists. While the real spatial correlation matrix cannot be exactly known, its expectation in a given time-frequency bin can be used as an approximation. This can be calculated by averaging the observed spatial correlation matrices over time and frequency, such that:

$$\tilde{\mathbf{R}}_{\mathbf{p}}(\tau, \kappa) = \frac{1}{J_{tot}} \sum_{j_\tau=-J_\tau}^{J_\tau} \sum_{j_\kappa=-J_\kappa}^{J_\kappa} \mathbf{R}_{\mathbf{p}}(\tau + j_\tau, \kappa + j_\kappa) \tag{8}$$

where $J_{tot} = (2J_\tau + 1)(2J_\kappa + 1)$ and the spatial correlation matrix $\mathbf{R}_{\mathbf{p}}(\tau, \kappa) = \mathbf{p_{nm}}(\tau, \kappa) \mathbf{p_{nm}}^H(\tau, \kappa)$ is evaluated at the time-frequency bin (τ, κ). Time-frequency averaging across $2J_\tau + 1$ time and $2J_\kappa + 1$ bins acts as the expectation. Notice that $\mathbf{R}_{\mathbf{p}}(\tau + j_\tau, \kappa + j_\kappa)$ is by itself unit rank, but the time-frequency averaged spatial correlation matrix need not be, unless the SHD coefficient vectors, $\{\mathbf{p_{nm}}^H(\tau, \kappa)\}$ in the time-frequency averaging window are scaled copies of each other. The latter case would only occur if there is a single source in the time-frequency averaging window. However, there will always be multiple components (e.g. sources and coherent reflections) in a real sound field, and the spatial correlation matrix will be full rank.

In order to differentiate the case where the direct sound is dominant for cases where the spatial correlation matrix is numerically calculated to be full rank, the DPD test involves the identification of time-frequency bins where the *effective rank* of the time-frequency averaged spatial correlation matrix is unity. This is achieved via obtaining the singular value decomposition (SVD) of the averaged spatial correlation matrix and calculating the ratio of the two largest singular values. In other words time-frequency bins for which the effective rank is unity are selected such that:

$$\mathscr{P}_{DPDtest} = \left\{ (\tau, \kappa) : \text{erank}\left(\tilde{\mathbf{R}}_{\mathbf{p}}(\tau, \kappa) \right) = 1 \right\} \tag{9}$$

where,

$$\mathscr{E} = \text{erank}\left(\tilde{\mathbf{R}}_{\mathbf{p}}(\tau, \kappa) \right) = 1 \quad \text{if} \quad \mathscr{R} = \frac{\sigma_1(\tau, \kappa)}{\sigma_2(\tau, \kappa)} > T_{\text{DPD}}, \tag{10}$$

Here, $\sigma_1(\tau, \kappa)$ and $\sigma_2(\tau, \kappa)$ are the largest and second-largest singular values of $\tilde{\mathbf{R}}_{\mathbf{p}}(\tau, \kappa)$ and $T_{\text{DPD}} \gg 1$ is a threshold. Notice that the ratio of the singular values given in (10) can never be less than unity.

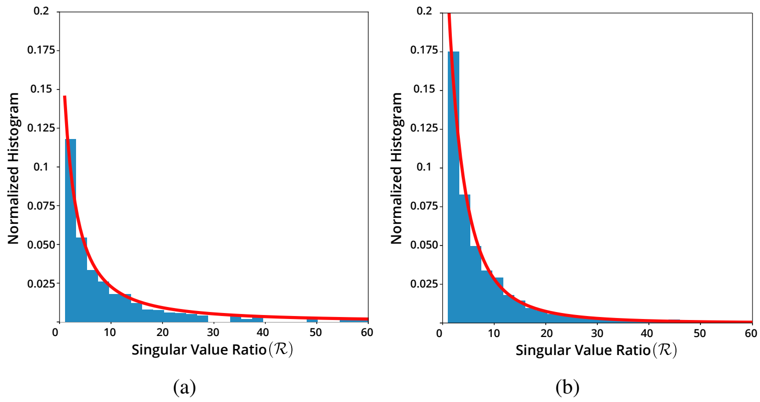

Figure 1. Normalised histograms of $\mathscr{R}$ for sources at a distance of (a) 0.5 m and (b) 2.6 m. The fitted generalized Pareto distributions are also shown with red curves.

The selection of the DPD threshold is typically made in an *ad hoc* manner by selecting a satisfactory value which which will eliminate a majority of time-frequency bins and at the same time provide good DOA estimates subject to the capabilities of the selected DOA estimation method [10, 17]. Therefore, DPD threshold selection has a profound effect on the trade-off between the accuracy and computational cost of DOA estimation.

The data-driven threshold selection approach we propose in this paper is based on the observation that the singular value ratios of spatial correlation matrices come from heavy right-tailed probability distributions. These distributions were observed to change with the reverberation time, direct-to-reverberant (D/R) ratio, as well as the number and types of sources to be localized.

A flexible probability distribution that can be used to approximate such distibutions is the generalised Pareto distribution (GPD). GPD is a three-parameter probability distribution that is effective when used to model the tails of other probability distributions [15]. Cumulative distribution function (CDF) of GPD [3] is given as:

$$F_{\zeta,\sigma,\mu}(x) = 1 - [1 - \zeta(x-\mu)/\sigma]^{1/\zeta} \tag{11}$$

Note that $F_{\zeta,\sigma,1}(1) = 0$. Probability density function (PDF) of GPD is given as:

$$f_{\zeta,\sigma,\mu}(x) = \frac{1}{\sigma}\left[1 - \frac{\zeta(x-\mu)}{\sigma}\right]^{(1-\zeta)/\zeta} \tag{12}$$

where $\zeta \leq 0$ is the shape parameter, $\sigma > 0$ is the scale parameter, and $\mu > 0$ is the location parameter.

In reverberant environments, the probability of observing a high ratio of singular values is low. Similarly, for most realistic cases involving speech or music signals, the time-frequency representation is sparse, resulting in a majority of time-frequency bins which have a low singular value ratio. If we assume that effective ranks calculated for each time-frequency bin as in (10) come from a GPD with a location parameter, $\mu = 1$, the selection of the DPD threshold can be made based on the probability of observing a ratio greater than a desired ratio threshold, T_{DPD} such that:

$$P[\mathscr{R} > T_{\text{DPD}}] = 1 - \int_1^{T_{\text{DPD}}} f_{\zeta,\sigma,1}(x)dx = 1 - F_{\zeta,\sigma,1}(T_{\text{DPD}}) + F_{\zeta,\sigma,1}(1) \tag{13}$$

where $P[\cdot]$ represents the probability. In other words, if the scale and shape parameters can be estimated from the singular value ratios of time-frequency bins, the threshold for the DPD test can be selected as the minimum value which satisfies:

$$P[\mathscr{R} > T_{\text{DPD}}] = 1 - F_{\zeta,\sigma,1}(T_{\text{DPD}}) < \widehat{P} \tag{14}$$

The threshold to be used in the DPD test in order to guarantee that the ratio for the selected bins have the probability of occurrence, $\widehat{P}$ should then be selected to satisfy:

$$T_{\text{DPD}} > \sigma(1-\widehat{P}^{\zeta})/\zeta \tag{15}$$

The scale and shape parameters of the underlying distribution can be trivially obtained via maximum likelihood estimation [6, 2]. This way it becomes possible to select the DPD threshold in a deterministic way. The number of selected bins can also be approximately controlled, i.e. for K time-frequency bins that are analysed, the the expected number of bins selected this way will be approximately $K\hat{P}$.

Fig. 1 shows the normalised histograms and the fitted generalized Pareto distributions for two sources at the different distances of 0.5 m and 2.6 m. It may be observed that moving the source away from the RSMA results in a higher number of bins with lower $\mathscr{R}$ since the D/R ratio decreases. The parameters, (ζ, σ, μ) of the fitted GPD are $(-1.7241, 10.1377, 1)$ and $(-0.53987, 6.14617, 1)$.

4 EVALUATION

We evaluated the DPD threshold selection method using emulated sound recordings. The details of this evaluation is given in this section.

4.1 Sound scene emulations

We used two sets of anechoic audio signals for the evaluations: music and speech. Music signals consisted of 4 seconds of anechoic recordings of Mahler's Symphony Nr. 1, Mvt. 4 [14] at the sampling rate of $f_S = 48$ kHz. The four tracks where violins are playing in unison are selected. Speech signals consisted of the first 4 seconds of energy normalized dry speech signals from two male and two female speakers recorded at METU SPARG audio lab. The sampling rate was $f_S = 48$ kHz.

We used the acoustic impulse responses (AIRs) recorded with Eigenmike em32 microphone array in a classroom with a reverberation time of $T_{60} \approx 1.12$ s [13]. Signals used in the evaluations were obtained by convolving the source signals with measured AIRs.

Three simple scenarios were simulated. In all three scenarios, the four sources were positioned at the drections $(\pi/2, 0)$, $(\pi/2, \pi/2)$, $(\pi/2, \pi)$, and $(\pi/2, 3\pi/2)$. In the first, second and third scenarios, the sources were positioned at a the distances of 0.5 m, 1m and 1.5 m from the RSMA, respectively. These distances correspond to average direct-to-reverberant (D/R) ratios of 10.72, 5.69, and 2.12 dB, for the scenarios 1, 2, and 3, respectively.

4.2 DPD test parameters

Windowed Fourier transform with a 1024-point FFT and a Hamming window with 75% overlap are used in the evaluations. The frequency range of analysis was between 2608 and 5216 Hz which allowed a decomposition of order $N = 4$. Smoothing parameters for the DPD test were $J_\tau = 4$ and $J_\nu = 15$. Only the SHD coefficients up to an order of $N = 3$ were used for calculating the spatial correlation matrices.

Selection of bins using the DPD test was made for five different probability levels, $\widehat{P} = 0.00625$, 0.0125, 0.025, 0.05, and 0.1 after fitting a GPD to the observed data via maximum likelihood estimation as described in [2].

4.3 DOA Estimation Method

We used hierarchical grid refinement (HiGRID) [1] as the DOA estimation method. HiGRID is based on the calculation of a sector averaged power of a plane-wave decomposition steered response and identifies the DOAs of sources by seeking a data-adaptive multresolution steered response representation that minimises spatial entropy. The algorithm provides very accurate DOA estimates on par with other state-of-the-art DOA estimation methods such as EB-MUSIC at a lower computational cost. It is also robust to noise and reverberation, and can localise multiple coherent sources. HiGRID is an output sensitive algorithm which has a lower computational cost for a low source count. Therefore, using the DPD test as a pre-processing stage to identify bins with a

Table 1. DPD test thresholds for different probabilities, $\widehat{P}$.

$\widehat{P}$	0.00625	0.0125	0.025	0.05	0.1
Scenario 1 (Violins)	79.6	53.7	35.6	23.0	14.2
Scenario 2 (Violins)	46.7	34.9	25.5	18.0	12.2
Scenario 3 (Violins)	**34.7**	26.6	20.0	14.5	10.0
Scenario 1 (Speech)	**459.5**	**226.5**	110.9	53.6	25.2
Scenario 2 (Speech)	**94.0**	**58.7**	36.2	21.8	12.6
Scenario 3 (Speech)	**31.6**	**22.7**	16.0	11.0	7.2

Table 2. Number of bins selected for different probabilities, $\widehat{P}$.

$\widehat{P}$	0.00625	0.0125	0.025	0.05	0.1
Scenario 1 (Violins)	379	675	1414	2536	4426
Scenario 2 (Violins)	386	727	1344	2433	4733
Scenario 3 (Violins)	**502**	828	1328	2335	4570
Scenario 1 (Speech)	**55**	**191**	797	2389	5102
Scenario 2 (Speech)	**67**	**287**	1013	2638	5317
Scenario 3 (Speech)	**164**	**519**	1364	2825	5494

single dominant component would reduce its computational cost even further.

HiGRID was used on the time-frequency bins selected via the DPD test. The maximum SHD order was $N = 4$, and the maximum resolution level used was 3 which which corresponds to a spherical grid with 768 elements and provides angular resolution of 7.33°. Implementations of HiGRID and DPD in Python 3.6 were used [4].

4.4 Results

Table 1 shows the thresholds calculated for different scenarios at different probability levels. Values shown with bold typeface indicate cases where less than four sources were localised. It may be observed that decreasing $\widehat{P}$ corresponds to a higher DPD test threshold and it is possible to use high DPD thresholds without degrading DOA estimation performance.

Table 2 shows the number of bins selected from a total of 42552 time-frequency bins in each tested case. It may also be observed that the thresholds and the number of selected bins are different for speech and music as well as for different scenarios. The number of selected bins approximately coincide with $\hat{P} \times 100\%$ of the total number of tested bins.

The time-frequency bins that pass the DPD test are processed with HiGRID to determine the source DOAs. Fig. 2 shows the spherical DOA histogram for Scenario 2 with violins using bins selected by setting $\hat{P} = 0.025$. All sources are clearly represented by a distinct peak in the histogram.

Table 3 shows the DOA estimation errors using data-driven DPD test threshold selection and HiGRID for the emulated cases. Except for very low values of $\hat{P}$, all four sources were correctly localised with a very low average DOA estimation error below half of the maximum resolution used in nearly all cases. Another observation is that higher thresholds can be used for closer sources and vice versa. Overall, $\hat{P} = 0.0125$ and $\hat{P} = 0.025$ are good selections for music and speech, respectively, providing low DOA estimation errors at a low computational cost for all the different cases tested. The effect of reverberation time and other acoustical features of the recording venue on threshold selection remains to be investigated.

Table 3. DOA estimation errors in degrees.

$\widehat{P}$	0.00625	0.0125	0.025	0.05	0.1
Scenario 1 (Violins)	3.27°	2.77°	2.24°	1.91°	1.52°
Scenario 2 (Violins)	4.36°	2.92°	2.06°	1.10°	1.38°
Scenario 3 (Violins)	**4.01**°	3.34°	1.56°	0.99°	1.21°
Scenario 1 (Speech)	**8.31**°	**3.31**°	3.53°	3.14°	2.30°
Scenario 2 (Speech)	**0.02**°	**2.41**°	2.61°	2.74°	1.92°
Scenario 3 (Speech)	**1.12**°	**3.32**°	2.87°	1.56°	1.05°

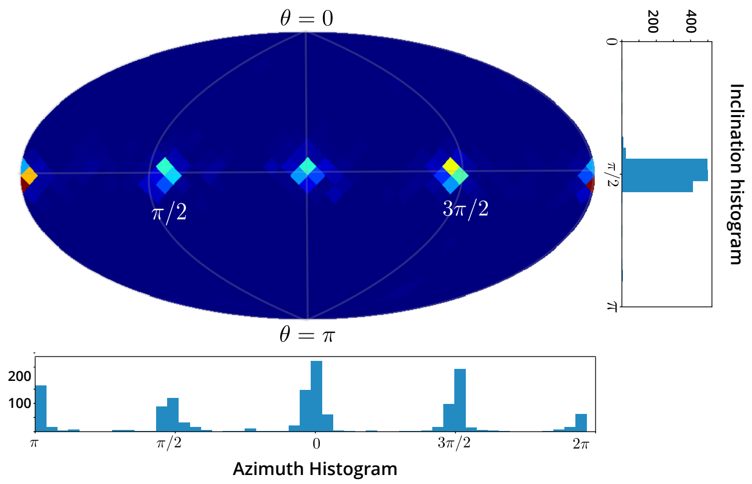

Figure 2. Mollweide projection of DPD-HiGRID DOA histogram showing the peaks of source locations on a spherical grid consisting of 768 pixels.

5 CONCLUSIONS

We proposed a data-driven threshold selection method for the direct path dominance test in this paper. The method is based on the observations that the singular value ratios used in the DPD test come from a heavy right-tailed distribution which can be approximated well by using the generalised Pareto distribution, and also that time-frequency bins that contain a single source only are rare events. These observations allow the selection of a DPD test threshold using the probability of occurence, effectively allowing the control of approximate number bins to be processed. We presented an evaluation of the proposed method for speech signals and noise. The evaluations revealed that DPD test thresholds that are much higher than those reported in literature can be used without degrading the performance of DOA estimation.

ACKNOWLEDGEMENTS

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).

REFERENCES

[1] M. B. Çöteli, O. Olgun, and H. Hacıhabiboğlu. Multiple sound source localization with steered response power density and hierarchical grid refinement. *IEEE/ACM Trans. on Audio, Speech and Lang. Process.*, 26(11):2215 – 2229, November 2018.

[2] V. Choulakian and M. A. Stephens. Goodness-of-fit tests for the generalized pareto distribution. *Technometrics*, 43(4):478–484, November 2001.

[3] S. Coles, J. Bawa, L. Trenner, and P. Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.

[4] H. Hacıhabiboğlu. Hiearchical Grid Refinement (HiGRID): DOA Estimation using Rigid Spherical Microphone Arrays (Python Code and Data), https://doi.org/10.5281/zenodo.2636368, Apr. 2019.

[5] S. Hafezi, A. H. Moore, and P. A. Naylor. Multiple source localization in the spherical harmonic domain using augmented intensity vectors based on grid search. In *Proc. 24th European Signal Process. Conf. (EUSIPCO 2016)*, pages 602–606, Budapest, Hungary, August 2016.

[6] J. R. M. Hosking and J. R. Wallis. Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, 29(3):339–349, March 1987.

[7] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor. 3D source localization in the spherical harmonic domain using a pseudointensity vector. *Proc. 18th European Signal Process. Conf. (EUSIPCO 2010)*, pages 442–446, August 2010.

[8] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor. *Theory and Applications of Spherical Microphone Array Processing*, volume 9 of *Springer Topics in Signal Processing*. Springer, Aug. 2016.

[9] Z. Li, R. Duraiswami, E. Grassi, and L. S. Davis. Flexible layout and optimal cancellation of the orthonormality error for spherical microphone arrays. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages iv–iv, May 2004.

[10] A. Moore, C. Evers, P. A. Naylor, D. L. Alon, and B. Rafaely. Direction of arrival estimation using pseudo-intensity vectors with direct-path dominance test. In *Proc. 23rd European Signal Process. Conf. (EUSIPCO-15)*, pages 2296–3000, Nice, France, August 2015.

[11] 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.*, 25(1):178–192, January 2017.

[12] 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. on Audio, Speech and Language Process*, 10:1494–1505, October 2013.

[13] O. Olgun and H. Hacıhabiboğlu. METU SPARG Eigenmike em32 Acoustic Impulse Response Dataset v0.1.0, https://doi.org/10.5281/zenodo.2635758, Apr. 2019.

[14] J. Pätynen, V. Pulkki, and T. Lokki. Anechoic recording system for symphony orchestra. *Acta Acust. united with Acust.*, 94:856–865, November 2008.

[15] J. Pickands III. Statistical inference using extreme order statistics. *Ann. Statistics*, 3(1):119–131, January 1975.

[16] B. Rafaely. *Fundamentals of Spherical Array Processing*. Springer-Verlag, 2015.

[17] B. Rafaely and D. Kolossa. Speaker localization in reverberant rooms based on direct path dominance test statistics. In *2017 IEEE Int. Conf. on Acoust. Speech and Signal Process. (ICASSP'17)*, pages 6120–6124, New Orleans, USA, March 2017.

[18] 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.*, 131(4):2828–2840, April 2012.

[19] H. Teutsch and W. Kellermann. Detection and localization of multiple wideband acoustic sources based on wavefield decomposition using spherical apertures. In *Proc. IEEE Int. Conf. on Acoust. Speech and Signal Process. (ICASSP-08)*, pages 5276–5279, Las Vegas, NV, USA, Mar. 31 - Apr. 4 2008.