3D ACOUSTIC SOURCE LOCALIZATION IN THE SPHERICAL HARMONIC DOMAIN BASED ON OPTIMIZED GRID SEARCH

Sina Hafezi, Alastair H. Moore and Patrick A. Naylor

Department of Electrical and Electronic Engineering, Imperial College London, UK
{s.hafezi14, alastair.h.moore, p.naylor}@imperial.ac.uk

ABSTRACT

An approach for 3D source localization using a spherical microphone array is proposed that gives improved accuracy compared to intensity-based methods. First order spherical harmonics are first used to obtain an initial approximate localization result and then the initial result is improved based on an optimized grid search in the local vicinity using the method of least squares and high-order spherical harmonics. We show that this approach outperforms the first-order approach and shows strong robustness to reverberation and noise. The worst average error of 3 degrees was found in our experiments in the presence of realistic reverberation and noise.

Index Terms— spherical microphone arrays, localization, direction-of-arrival estimation, spherical harmonic, intensity vector

1. INTRODUCTION

Recently the use of Spherical Microphone Arrays (SMAs) has had growing interest due to their ability to capture sound fields in three dimensions [1, 2, 3, 4, 5]. In this paper we address source localization using SMAs, which has a diverse range of applications in acoustic signal processing such as spatial filtering, source separation, dereverberation, source tracking and acoustic environment mapping.

Source localization methods are widely studied and are generally categorised into three main groups, namely subspace-based methods (ESPRIT, MUSIC) [6, 7, 8], steered response power (SRP) [9], and intensity-based methods [10]. The methods that are based on the first two categories [11, 12, 13] are computationally expensive due to the need for an exhaustive search. On the other hand, the intensity-based methods using pseudo-intensity vectors (PIVs) are fast to compute and have good localization accuracy for a single source in relatively dry (low reverberation) environments [10]. However, as with most localization algorithms, as the level of reverberation and the number of sound sources increase, localization accuracy is reduced [14].

This paper is structured as follow: Section 2 briefly reviews the background theory of spherical harmonics and Section 3 introduces PIVs computed using zeroth and first-order eigenbeams. Section 4 presents our approach using high-order eigenbeams to enhance the localization accuracy, Section 5 presents a combination of PIV and SRP as a baseline for comparison, and finally in Section 6 we compare the methods and evaluate their accuracy and stability.

2. SPHERICAL HARMONICS

Let $p\left(k, r, \Omega\right)$ denote the sound pressure field at a point $(r, \Omega) = (r, \theta, \varphi)$ in spherical coordinates with range r, inclination θ and azimuth φ, where k is the wavenumber. The Spherical Harmonic Transform (SHT) of this field is given by [15, p 192]:

$$p_{lm}(k, r) = \int_{\Omega \in S^2} p(k, r, \Omega) Y_{lm}^{*}\left(\Omega\right) d\Omega, \tag{1}$$

where $\int_{\Omega \in S^2} d\Omega = \int_0^{2\pi} \int_0^{\pi} \sin\left(\theta\right) d\theta d\varphi$, and $(.)^*$ denotes the complex conjugate.

The spherical harmonics $Y_{lm}\left(\Omega\right)$ of order l and degree m (satisfying $|m| \leq l$) are given by [15, p 190]:

$$Y_{lm}\left(\Omega\right) = \sqrt{\frac{(2l+1)}{4\pi} \frac{(l-m)!}{(l+m)!}} P_{lm}\left(\cos\left(\theta\right)\right) e^{im\varphi}, \tag{2}$$

where P_{lm} is the associated Legendre function and $i^2 = -1$.

The sound pressure field can be reconstructed using the inverse SHT [16]:

$$p(k, r, \Omega) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} p_{lm}(k, r) Y_{lm}(\Omega), \tag{3}$$

where the coefficients p_{lm} are often called eigenbeams.

The sound pressure field on the surface of a SMA with radius r_a can be considered as $p(k, r_a, \Omega)$, which depends on the array properties, e.g. radius and configuration (open or

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 609465

rigid sphere). This dependency is captured by the frequency-dependent mode strength $b_l(kr_a)$ and is used to form the mode-strength-compensated eigenbeams $a_{lm}(k)$ such that

$$a_{lm}(k) = \frac{p_{lm}(k, r_a)}{b_l(kr_a)}. \quad (4)$$

For a rigid SMA as used in our experimental study, with radius r_a, the mode strength $b_l(kr_a)$ is given by [15, p 228]:

$$b_l(kr_a) = 4\pi i^l \left[j_l(kr_a) - \frac{j_l^{'}(kr_a)}{h_l^{(2)'}(kr_a)} h_l^{(2)}(kr_a) \right], \quad (5)$$

where j_l is the spherical Bessel function of order l, $h_l^{(2)}$ is the spherical Hankel function of the second kind and of order l, and $(.)^{'}$ denotes the first derivative with respect to argument.

3. PSEUDO-INTENSITY VECTORS

In acoustics, sound intensity is a measure of the flow of sound energy through a surface per unit area, in a direction perpendicular to this surface. The intensity vector I, which defines the magnitude and the direction of the energy flow can be determined by calculating the flow of sound energy through the three unit surfaces perpendicular to the Cartesian axes as [17]:

$$\mathbf{I}(k) = \frac{1}{2} \Re \{ q^*.\mathbf{v} \}, \quad (6)$$

where q is the sound pressure, $\mathbf{v} = [v_x, v_y, v_z]^T$ is the particle velocity in the Cartesian directions (with a dipole directivity pattern), and $\Re\{.\}$ denotes the real part of a complex number.

PIVs are conceptually similar to intensity vectors. They are calculated using the zero- and the first-order eigenbeams as [10]

$$\mathbf{I}(k) = \frac{1}{2} \Re \left\{ a_{00}(k)^*. \begin{bmatrix} D_x(k) \\ D_y(k) \\ D_z(k) \end{bmatrix} \right\}, \quad (7)$$

where

$$D_\nu(k) = \sum_{m=-1}^{1} Y_{1m}(\phi_\nu) a_{1m}(k), \quad \nu \in \{x, y, z\} \quad (8)$$

are dipoles steered in the negative direction of Cartesian axes, given by $\phi_x = (\pi/2, \pi)$, $\phi_y = (\pi/2, -\pi/2)$ and $\phi_z = (\pi, 0)$.

A unit vector $\mathbf{u}(k)$ in the direction-of-arrival (DOA) is given by

$$\mathbf{u}(k) = -\frac{\mathbf{I}(k)}{\|\mathbf{I}(k)\|}, \quad (9)$$

where $\|.\|$ indicates a vector's ℓ_2-norm.

4. PROPOSED METHOD

PIV-based localization [10] uses only up to the first order eigenbeams and ignores the higher order harmonics which also carry spatial information. In this section we propose a method for source localization that exploits higher order spherical harmonics, based on single Plane Wave Decomposition (PWD).

Consider a plane wave $S(k) = \alpha(k)e^{i\beta(k)}$ with amplitude $\alpha(k)$, phase at origin $\beta(k)$ and DOA $\Omega_u = (\theta_u, \varphi_u)$. The SHT of this plane wave is given by

$$a_{lm}(k) = S(k) Y_{lm}^*(\Omega_u) + n_{lm}(k), \quad (10)$$

where $n_{lm}(k)$ is a residual due to noise and reverberation.

Writing (10) for SHT orders $l > 1$, results in an overdetermined system. For such a system comprising linear equations, the method of least squares provides the well-known least-squares-optimal solution. As in our case the equations are non-linear and complex, one approach is to use the grid search method over a possible set of data and find the approximate solution which best satisfies the equations.

A grid search method is computationally expensive especially when it is performed for a large number of variables. We can reduce the computation by performing the grid search over a relatively small set of data in the expected vicinity of the solution.

In our method, the initial approximate DOA is first calculated using the PIV method and then a grid search optimization procedure is performed across a search window spanning only the vicinity of the initial DOA estimate using higher order eigenbeams. This second step is more computationally expensive than the PIV-based computation in the first step, but this cost is mitigated by restricting the second step only to the expected solution vicinity as indicated by the first step.

4.1. DOA optimization using high order harmonics

Our DOA estimation method using higher-order will next be described. Substituting (2) into (10) for $l = 0$ in a noise-free case gives

$$S(k) = \sqrt{4\pi} a_{00}(k). \quad (11)$$

Substituting (11) into rearranged (10), for an arbitrary look direction Ω, we have the direction-dependant error $n_{lm}(k, \Omega)$:

$$n_{lm}(k, \Omega) = a_{lm}(k) - \sqrt{4\pi} a_{00}(k) Y_{lm}^*(\Omega). \quad (12)$$

We define a cost function $\Psi(k, \Omega)$ as the Root-Mean-Square (RMS) of errors $n_{lm}(k)$ for $l = \{0, \ldots, L\}$ and $m = \{-l, \ldots, l\}$:

$$\Psi(k, \Omega) = \sqrt{\frac{1}{(L+1)^2} \sum_{l=0}^{L} \sum_{m=-l}^{l} | \, n_{lm}(k, \Omega) \, |^2}, \quad (13)$$

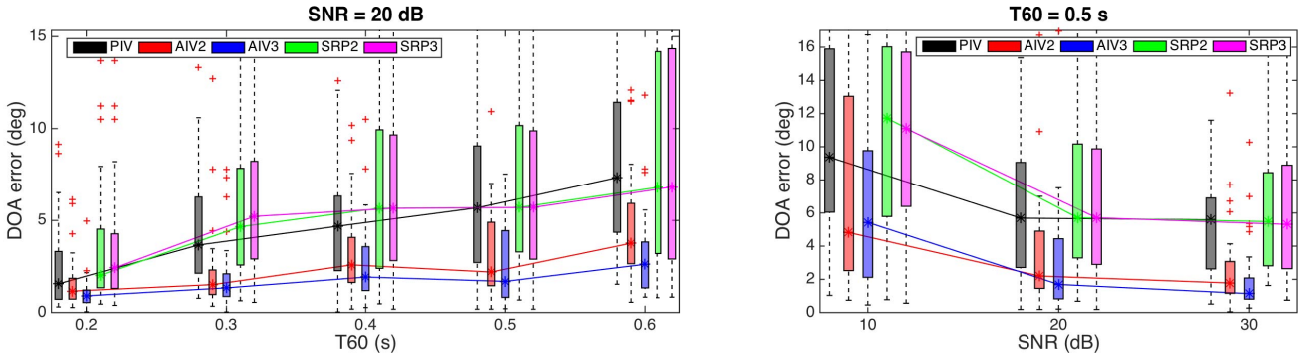

Fig. 1. The DOA errors for the PIV (black), second-order AIV (red), third-order AIV (blue), second-order SRP (green) and third-order SRP (magenta) as a function of reverberation time (a) and sensor noise level (b). The boxes show the median as an asterix, upper and lower quartiles, and the whiskers extend to 1.5 times the interquartile range.

where L is the maximum spherical harmonic order considered in the optimization.

Given this cost as a function of the look direction and wavenumber, we then perform a grid search across a grid of discrete look directions $\{\Omega_M\}$ around the initial DOA with the search window size $\triangle\Omega_M = (\triangle\theta_M, \triangle\varphi_M)$. Note that the larger search window size and the higher grid resolution both increase the accuracy of estimation as well as the computational cost. The optimized DOA Ω_s is the direction in which the cost function $\Psi(k, \Omega_s)$ is minimised:

$$\Omega_s(k) = \arg\min_{\Omega} \Psi(k, \Omega), \quad \Omega \in \{\Omega_M\}, \tag{14}$$

which is then converted into Cartesian coordinates to form the optimized DOA unit vector $\mathbf{u}_s(k)$. The Augmented Intensity Vector (AIV) $\mathbf{I}_s(k)$ is formed using (9) and the initial intensity norm $\|\mathbf{I}(k)\|$

$$\mathbf{I_s(k)} = -\mathbf{u_s}(k)\|\mathbf{I}(k)\|. \tag{15}$$

4.2. DOA quantization

The intensity vectors are calculated for each time frame. Since the grid search is intrinsically discrete in the DOA domain, we quantize the estimated directions in azimuth and inclination separately. Each estimated DOA is quantized to the nearest DOA sample. For each DOA sample, the overall intensity is the sum of the intensities of associated quantized DOAs across all frequencies and time frames. This results in a 2-dimensional intensity matrix with row as inclination, column as azimuth, and value as intensity. For a single source, the overall estimated DOA is the azimuth and inclination of the global maximum intensity.

5. STEERED RESPONSE POWER

As a baseline for comparison, we introduce SRP that also uses high order spherical harmonics. In SRP a beam with an arbitrary directivity pattern is steered into different look directions to find the direction with the highest power. The beam pattern depends on the harmonic orders that is used in beamforming. The output of the beamformer steered into an arbitrary look direction Ω is given [18]

$$y(k, \Omega) = \sum_{l=0}^{L_d} \sum_{m=-l}^{l} a_{lm}(k) Y_{lm}(\Omega), \tag{16}$$

where L_d is the maximum harmonic order used in beamforming.

For a single active source, the estimated DOA Ω_d is given

$$\Omega_d = \arg\max_{\Omega} \sum_{k} \mid y(k, \Omega) \mid^2 . \tag{17}$$

In practice, it is not efficient to steer many beams indiscriminately in all directions. Therefore a coarse approximation of DOA can be taken at first, e.g. using PIV, and then a fine grid search can be performed over the area of interest around the initial DOA. For the sake of equal conditions in evaluation, the maximum order of beamforming L_d, the search window, and the initial DOA in SRP is assumed to be the same as in our method AIV.

6. EVALUATION

The aim of our evaluation is to assess the DOA accuracy improvements due to use of high-order eigenbeams, compared to PIV that uses only zeroth and first-order eigenbeams, and SRP which also uses high-order eigenbeams. For each approach, we calculate the DOA error ε (in degree) between the

true DOA unit vector $\mathbf{u}_o$ and the estimated DOA unit vector $\mathbf{u}_s$:

$$\varepsilon = \cos^{-1}\left(\mathbf{u}_o^T \mathbf{u}_s\right). \quad (18)$$

An evaluation of the proposed approach is conducted using simulated data with one active talker at all times. The Acoustic Impulse Responses (AIRs) of a 32-element rigid spherical microphone array were simulated using Spherical Microphone arrays Impulse Response Generator (SMIRgen) [19] based on Allen & Berkley's image method [20]. The array with radius 4.2 cm is placed at (2.52, 1.97, 3.11) m in a 5x4x6 m shoebox room. We consider 40 different source positions at the distance of 1 m from the centre of array with a DOA that is randomly selected from a uniform distribution around the sphere. The source signal consists of an anechoic speech [21] with duration 5 s. A sampling frequency of 8 kHz was used with frame length of 4 ms and 50% overlap of time frame. For each source position, the test was repeated for different level of sensor noise (white Gaussian) with Signal-to-Noise Ratio (SNR) = $\{10, 20, 30\}$ dB and different reverberation time (RT) with $T_{60} = \{0.2, 0.3, 0.4, 0.5, 0.6\}$ s. The Direct-to-Reverberant Ratios (DRRs) of all trials have the mean of $\{4.6, 1.4, -0.9, -2.8, -4.3\}$ dB and the standard deviation of $\{2.3, 1.9, 1.8, 1.8, 1.8\}$ dB respectively for $T_{60} = \{0.2, 0.3, 0.4, 0.5, 0.6\}$ s.

Two versions for each AIV and SRP method are used with different optimization orders $L = \{2, 3\}$ in (13). The search window size is $(\triangle\theta_M, \triangle\varphi_M) = (10, 10)$ degrees with the resolution of 1 degree in (14). The estimated DOAs are quantized with the resolution of 1 degree in azimuth and inclination.

Fig. 1 shows the results presented as a distribution of DOA estimation errors. The boxes show the median as the asterix, upper and lower quartiles, and the whiskers extend to 1.5 times the interquartile range based on Monte Carlo simulations. Fig. 1 (a) shows the DOA error as a function of RT for SNR=20 dB. Both of AIVs, compared to PIV and SRP, have noticeable improvement in the median and range of DOA errors for all RTs. Also the median errors for AIVs change smoothly while the PIV and SRPs show a sharp increase of median. Fig. 1 (b) shows the DOA error as a function of sensor noise level for T_{60} of 0.5 s. We can see the significant improvements on the median and the range of errors for both of AIVs in all cases.

Fig. 2 shows the median DOA error across all the trials for each approach as a function of RT for all SNRs. The AIVs, compared to PIV and SRP, have significant improvement for all SNRs and RTs. As the RT increases the improvement, compared to PIV and SRP, becomes more noticeable. AIVs show a better accuracy compared to SRPs although they both use the same eigenbeams and grid search. AIVs, unlike PIV and SRPs, have strong robustness to RT and noise as they change smoothly as RT increases for low and moderate noise level.

Fig. 3 shows the range of DOA estimation error for each approach as a function of RT for all SNRs. For SNR of 20 and 30 dB AIVs, compared to PIV and SRPs, are more stable due to lower range of DOA estimation error.

7. CONCLUSION

We proposed an approach, AIV, to improve the localization accuracy for spherical microphone arrays using high order eigenbeams. We have shown that our approach noticeably outperforms both of the first-order PIV and the high-order SRP method, which uses the same number of eigenbeams. It is also shown that our method has significant robustness to reverberation time and noise. In a highly reverberant environment with realistic level of sensor noise, the second- and the third-order approaches respectively show the average error of 3.7 and 2.6 degrees whereas the first-order PIV, second- and third-order SRP respectively have more than 7.3, 6.8 and 6.8 degrees error. Our approach also shows high stability in DOA estimation as its estimation error varies by a maximum of 3 degrees whereas PIV and SRPs respectively show 7 and 11 degrees variation for moderate noise level. This study also shows that using up to the third-order harmonics in optimization has a slight advantage of 2 degrees higher accuracy over the second-order optimization only on high RTs.

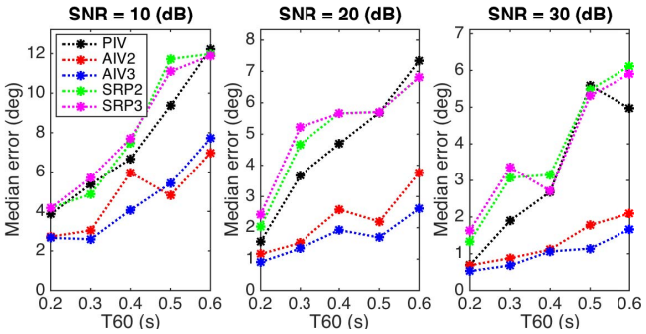

Fig. 2. Median of DOA estimation errors

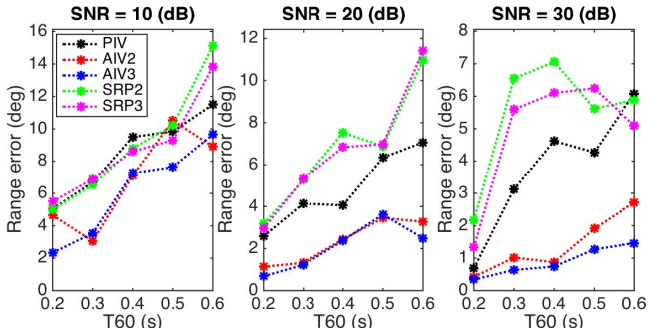

Fig. 3. Range of DOA estimation errors

8. REFERENCES

[1] E. Fisher and B. Rafaely, "The nearfield spherical microphone array," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2008, pp. 5272–5275.

[2] Zhibin Lin and Wei Qingyu, "Localization of multiple acoustic sources using optimal spherical microphone arrays," in *Proc. 9th International Conference on Signal Processing ICSP 2008*, Oct. 2008, pp. 349–352.

[3] Y. Peled and B. Rafaely, "Method for dereverberation and noise reduction using spherical microphone arrays," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2010, pp. 113–116.

[4] H. Sun, S. Yan, and U. P. Svensson, "Robust minimum sidelobe beamforming for spherical microphone arrays," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 4, pp. 1045–1051, May 2011.

[5] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, May 2002, vol. 2, pp. 1781–1784.

[6] R. Roy and T. Kailath, "ESPRIT - estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 984–995, 1989.

[7] H. Teutsch and W. Kellermann, "EB-ESPRIT: 2D localization of multiple wideband acoustic sources using eigen-beams," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2005, vol. 3, pp. iii/89–iii/92.

[8] S. Argentieri and P. Danès, "Broadband variations of the MUSIC high-resolution method for sound source localization in robotics," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Diego, CA, Nov. 2007.

[9] J. P. Dmochowski, J. Benesty, and S. Affes, "A generalized steered response power method for computationally viable source localization," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 8, pp. 2510–2526, Nov. 2007.

[10] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, "3D source localization in the spherical harmonic domain using a pseudointensity vector," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 442–446.

[11] D. Khaykin and B. Rafaely, "Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, Oct. 2009, pp. 221–224.

[12] H. Teutsch and W. Kellermann, "Eigen-beam processing for direction-of-arrival estimation using spherical apertures," in *Proc. Joint Workshop on Hands-Free Speech Communication and Microphone Arrays*, Piscataway, New Jersey, USA, Mar. 2005, IEEE.

[13] H. Teutsch and W. Kellermann, "Detection and localization of multiple wideband acoustic sources based on wavefield decomposition using spherical apertures," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2008, pp. 5276–5279.

[14] C. Evers, A. H. Moore, and P. A. Naylor, "Multiple source localisation in the spherical harmonic domain," in *Proc. Intl. Workshop Acoust. Signal Enhancement (IWAENC)*, Nice, France, July 2014.

[15] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, Academic Press, London, first edition, 1999.

[16] Boaz Rafaely, *Fundamentals of Spherical Array Processing*, Springer Topics in Signal Processing. Springer, Berlin Heidelberg, 2015.

[17] M. J. Crocker and F. Jacobsen, "Sound intensity," in *Handbook of Acoustics*, M. J. Crocker, Ed., chapter 106, pp. 1327–1340. Wiley-Interscience, 1998.

[18] Boaz Rafaely, Yotam Peled, Morag Agmon, Dima Khaykin, and Etan Fisher, "Spherical microphone array beamforming," in *Speech Processing in Modern Communication: Challenges and Perspectives*, Israel Cohen, Jacob Benesty, and Sharon Gannot, Eds., chapter 11. Springer, Jan. 2010.

[19] D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, "Simulating room impulse responses for spherical microphone arrays," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 129–132.

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

[21] John S. Garofolo, Lori F. Lamel, William M. Fisher, Jonathan G. Fiscus, David S. Pallett, Nancy L. Dahlgren, and Victor Zue, "TIMIT acoustic-phonetic continuous speech corpus," Corpus LDC93S1, Linguistic Data Consortium, Philadelphia, 1993.