INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Reverberation-Robust One-Bit TDOA Based Moving Source Localization for Automatic Camera Steering

Harshavardhan Sundar[1], Gokul Deepak M.[1], T. V. Sreenivas[2], and Chandra Sekhar Seelamantula[1]

[1]Department of Electrical Engineering,
[2]Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India
{harshas123,mgokuldeepak}@gmail.com, tvsree@ece.iisc.ernet.in, chandra.sekhar@ieee.org

Abstract

We address the problem of moving acoustic source localization and automatic camera steering using one-bit measurement of the time-difference of arrival (TDOA) between two microphones in a given array. Given that the camera has a finite field of view (FoV), an algorithm with a coarse estimate of the source location would suffice for the purpose. We use a microphone array and develop an algorithm to obtain a coarse estimate of the source using only one-bit information of the TDOA, the sign of it, to be precise. One advantage of the one-bit approach is that the computational complexity is lower, which aids in real-time adaptation and localization of the moving source. We carried out experiments in a reverberant enclosure with a 60 dB reverberation time of 600 ms (RT60 = 600 ms). We analyzed the performance of the proposed approach using a circular microphone array. We report comparisons with a point source localization-based automatic camera steering algorithm proposed in the literature. The proposed algorithm turned out to be more accurate in terms of always having the moving speaker within the field of view.

Index Terms: Sign of TDOA, coarse localization, error correcting codes, Hamming distance, automatic camera steering, acoustic source localization.

1. Introduction

Acoustic source localization plays a key role in technologies such as sound rendering for telepresence applications [1], speech enhancement for mobile phone applications [2], automatic camera steering [3, 4], etc.. Automatic camera steering finds attractive applications in the context of lecture recording, round-table meetings, intrusion detection for surveillance, etc..

Automatic camera steering towards an acoustic source requires the acoustic source to be within the field of view (FOV) of the camera. Previous approaches to automatic camera steering have focused on estimating a point location of the source [3] or on estimating the direction of arrival (DOA) of the source [5]. Reverberation is a detrimental factor for any source localization algorithm and the point estimates or the DOA estimates are typically poorer under reverberation, thus requiring temporal smoothing of the location estimates for smooth tracking of the acoustic source [6] which is required for smooth steering of the camera. Since typical cameras have a finite field of view (FOV), we believe that it would be appropriate to have a coarse location estimate of the source as opposed to a point location, as is typically given by most localization algorithms [3, 7, 8]. Although direction of arrival (DOA) is certainly useful for camera steering, a precise DOA may not be required as the FOV of a camera is typically about 30 degrees at nominal levels of zoom. Therefore, we believe that a reverberation-robust coarse location estimate of the source would be appropriate for camera steering.

1.1. This Paper

We propose to estimate a coarse location of the source (source region) and use this for automatically steering a camera so that the entire source region is within the field of view of the camera. The proposed approach is computationally simpler than the previous techniques proposed for acoustic-source-localization-based camera steering [3] and can be realized with a low footprint using logic gates given the TDOA estimates. Interestingly, we show that relying on a coarse location, as discussed in Sec. 2 and 3.1, is more robust to reverberation than relying on point estimates as far as automatic camera steering is concerned. The contributions of this paper are: (i) Estimating coarse localization of a source using only the sign of TDOAs (Sec. 2); (ii) Transforming the problem of estimating the coarse location of the source into a detection framework (Sec. 3.1); (iii) The use of Hamming distance to reduce computational complexity and to incorporate reverberation robustness to the proposed approach; and (iv) Experimental validation of the proposed approach in a real reverberant enclosure (RT60 = 600 ms), for automatic camera steering (Sec. 4.2).

2. Coarse localization using sign of TDOAs

Consider an array of N microphones placed at known locations. The position vector (p.v.) of the i^{th} microphone (M_i) is denoted by $\vec{p}_i = [x_i, y_i, z_i]$; $1 \leq i \leq N$. Let the unknown source location be denoted by $\vec{p} = [x, y, z]$. Let d_i denote the distance of the source from the i^{th} microphone. The TDOA of the source w.r.t. the i^{th} and the j^{th}microphones is given by:

$$\tau_{ij} \triangleq \frac{1}{v}\left(d_i - d_j\right) = \frac{1}{v}\left(||\vec{p} - \vec{p}_i|| - ||\vec{p} - \vec{p}_j||\right), \qquad (1)$$

where v is the speed of sound in the medium within the enclosure. If $\tau_{ij} < 0$ then $d_i < d_j$, leading to the conclusion that the source is closer to M_i and hence lies in the half-space containing M_i. Let H_{ij} represent the plane of support of this half-space. H_{ij} passes through the mid-point of the line segment joining M_i and M_j with the line joining the two microphones as the normal. Thus, H_{ij} can be expressed in point-normal form as: $H_{ij} = \left\{\vec{p} : \underline{P}_{ij}\vec{p} = q_{ij}\right\}$, where $\underline{P}_{ij} = \vec{p}_i - \vec{p}_j$, and $q_{ij} = \frac{1}{2}\left(\vec{p}_i - \vec{p}_j\right).\left(\vec{p}_i + \vec{p}_j\right)$. Figure 1 shows an illustrative enclosure with microphones M_1 and M_2. The plane H_{21} divides $\mathbb{R}^3$ into two half-spaces given by: $\mathcal{R}_{21}^{(1)} \triangleq \{\vec{p} : \underline{P}_{21}.\vec{p} > q_{21}\}$ and $\mathcal{R}_{21}^{(2)} \triangleq \{\vec{p} : \underline{P}_{21}.\vec{p} \leq q_{21}\}$. The source region $\mathcal{R}_{21}$, can be specified as: $\mathcal{R}_{21} \triangleq \{\vec{p} : \left(\underline{P}_{21}.\vec{p} \leq q_{21}\right).\text{sgn}(\tau_{21})\}$, where $\text{sgn}(\cdot)$ denotes the signum function. We consider the following definition of the signum function: $\text{sgn}\left(x\right) = 1,\ \forall\ x > 0$ and $\text{sgn}\left(x\right) = 0,\ \forall\ x \leq 0$. Thus, $\text{sgn}(\tau_{21})$ decides whether $\vec{p}$ lies in $\mathcal{R}_{21}^{(1)}$ or $\mathcal{R}_{21}^{(2)}$. This information reduces the ambiguity of the source location by half. Thus, the sign of TDOA has important

  http://dx.doi.org/10.21437/Interspeech.2016-575

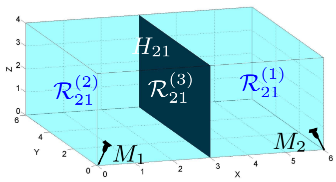
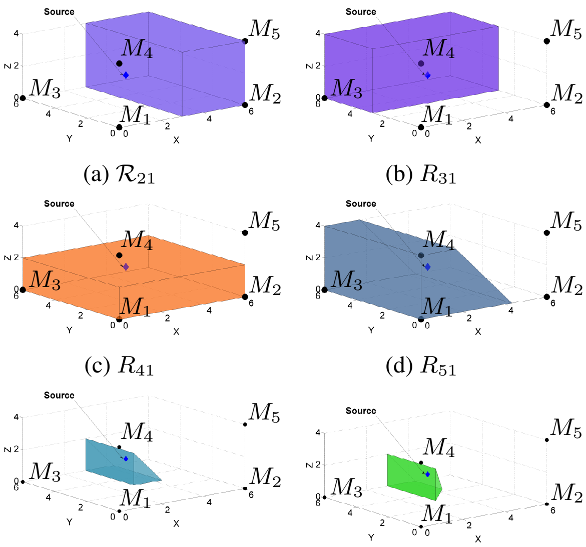


Figure 1: *Hyperplane (H_{21}) dividing an enclosure into the regions-* $\mathcal{R}_{21}^{(1)} = \{\vec{p} : \underline{P}_{21}.\vec{p} > q_{21}\}$, $\mathcal{R}_{21}^{(2)} = \{\vec{p} : \underline{P}_{21}.\vec{p} < q_{21}\}$ *and* $\mathcal{R}_{21}^{(3)} = \{\vec{p} : \underline{P}_{21}.\vec{p} = q_{21}\}$.



(a) $\mathcal{R}_{21}$ (b) R_{31}

(c) R_{41} (d) R_{51}

(e) $\bigcap_{i=2}^{5} R_{i1}$ (f) R_S

Figure 2: *(Color Online): N=5 microphones (black circles) example in an enclosure of dimension* 6 *m* ×6 *m* ×4 *m. (a-d) Source region from individual TDOAs . (e) Source region using* $N-1$ *TDOAs. (f) Source region using all the TDOAs.*

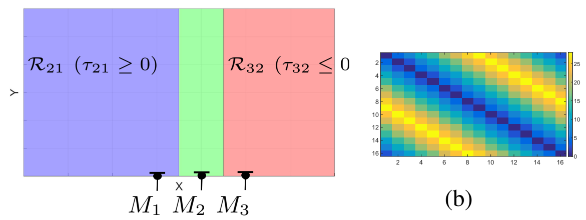
source location information that can be determined from each pair of microphone signals. The source region $\mathcal{R}_{ij}$ obtained using the microphones M_i, M_j and the TDOA - τ_{ij} is given as:

$$\mathcal{R}_{ij} \triangleq \left\{\vec{p} : \left(\underline{P}_{ij}.\vec{p} \le q_{ij}\right).\text{sgn}(\tau_{ij})\right\}. \tag{2}$$

Since $\mathcal{R}_{ij} = R_{ji}$, we choose only $\binom{N}{2}$ regions out of the possible N^2 regions, by considering $j < i$. We refer to the regions $\{\mathcal{R}_{ij}; 1 \le i, j, \le N; j < i\}$ as 1-bit TDOA regions, as these regions are completely specified by the 1-bit sign information of TDOAs and the microphone locations. Since the source is included in all these 1-bit TDOA regions $\mathcal{R}_{ij}$; $1 \le i, j \le N; j < i$, independently determined, it must lie within the intersection of these regions. Considering all pairs of microphones, the final source region (R_S) is obtained as:

$$R_S \triangleq R_E \bigcap_{\substack{1 \le i,j \le N \\ j < i}} \mathcal{R}_{ij}, \tag{3}$$

where R_E denotes the enclosure. Since each 1-bit TDOA region formed by (2) is a convex polyhedron, the intersection of these polyhedra is also a convex polyhedron [9], which is expressed as the linear matrix inequality: $R_S = \{\vec{p} : P.\vec{p} \le \underline{q}\}$.



(a) (b)

Figure 3: (a) Intersection of $\mathcal{R}_{21}$ and $\mathcal{R}_{32}$ is a null set. (b) Hamming distance matrix of the codewords corresponding to the valid regions shown in Figure 4(a).

The above set of inequalities is reduced to a minimal set using an algorithm presented in [10] for which a MATLAB implementation available in [11] is used. Figure 2 shows an example of an acoustic enclosure with five microphones and with the regions obtained using individual TDOAs (given by 2) in Figure 2(a)-2(d). The region obtained using $(N-1)$ TDOAs, by intersecting the regions shown in Figure (2(a))-(2(d)) is shown in Figure 2(e) and the region obtained using all the TDOAs (given by (3)) is shown in Figure 2(f). We observe that the source region obtained using all TDOAs is much smaller than using the individual TDOAs alone or a subset of $(N-1)$ TDOAs.

3. Coarse Localization Based Automatic Camera Steering

The main challenge in automatic camera steering is to track a moving source. We address this problem by processing the microphone signals in non-overlapping frames of duration T_W s. In each frame, the coarse source location (R_S) is estimated and the camera is automatically steered towards R_S in such a way that the entire region falls within the field of view of the camera. In the following section, we show that we can obviate the need for estimating the source regions in every frame using the method described in Sec. 2.

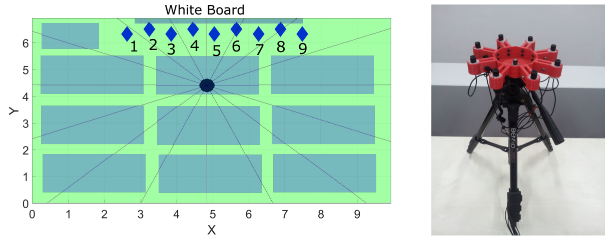
3.1. Precomputing source regions and relation to error correcting codes

With a single source of sound, a set of N microphones yields $\binom{N}{2}$ TDOAs each corresponding to a pair of microphones. Thus, $\binom{N}{2}$ 1-bit TDOA regions intersect to form R_S, as given by (3). Consider a bit-vector $\underline{\tau}$, which contains the 1-bit TDOA information across all pairs of microphones, that is, $\underline{\tau} \in \{-1, 1\}^{\binom{N}{2}}$. Thus, such a bit-vector represents the coarse location R_S, obtained using the intersection of 1-bit TDOA regions, which are in turn obtained using the elements in $\underline{\tau}$. Since there are $2^{\binom{N}{2}}$ distinct bit-vectors in the space - $\{-1, 1\}^{\binom{N}{2}}$, we deduce that, there can be a maximum of $2^{\binom{N}{2}}$ source regions corresponding to different locations of the source within the enclosure. Interestingly, not all of these $2^{\binom{N}{2}}$ source regions are valid regions, since some of them may turn out to be empty sets. Consider an example of a linear array with three microphones as shown in Figure 3(a). We observe that, if the TDOAs $\tau_{21} \ge 0$ and $\tau_{32} \le 0$, then irrespective of the sign of τ_{31}, the intersection of $\mathcal{R}_{21}$, $\mathcal{R}_{32}$ and $\mathcal{R}_{31}$ is an empty set. In this example, the bit-vector $\underline{\tau} = [\text{sgn}(\tau_{21}), \text{sgn}(\tau_{31}), \text{sgn}(\tau_{32})] = [1, \text{sgn}(\tau_{31}), -1]$ is considered to be an invalid bit-vector, as it corresponds to the source region being an empty set. Let L denote the number of valid regions, which are obtained as the intersection of $\binom{N}{2}$ 1-bit TDOA regions. We refer to the corresponding bit-vectors as codewords. Let $\underline{\tau}^{(\ell)} = \left[\mathrm{sgn}^{(\ell)}(\tau_{21}), \mathrm{sgn}^{(\ell)}(\tau_{31}), \ldots, \mathrm{sgn}^{(\ell)}(\tau_{N,N-1})\right]$ denote the vector of the ℓ^{th} codeword, which represents the ℓ^{th} valid region, denoted as $\mathcal{R}_S^{(\ell)}$, and given by:

$$\mathcal{R}_S^{(\ell)} \triangleq \mathcal{R}_E \bigcap_{\substack{1 \leq i,j \leq N \\ j<i}} \mathcal{R}_{ij}^{(\ell)}, \ \forall\, 1 \leq \ell \leq L, \tag{4}$$

where $\mathcal{R}_{ij}^{(\ell)} \triangleq \left\{\vec{p} : \left(\underline{P}_{ij}.\vec{p} \leq q_{ij}\right).\mathrm{sgn}^{(\ell)}(\tau_{ij})\right\}$. We search for the valid regions $\left\{\mathcal{R}_S^{(\ell)}\right\}_{\ell=1}^{L}$ and the codewords $\boldsymbol{\tau} = \left\{\underline{\tau}^{(1)}, \underline{\tau}^{(2)}, \ldots, \underline{\tau}^{(L)}\right\}$, by a grid search over the enclosure space. Importantly, the valid regions can be precomputed given the microphone positions and the enclosure boundaries. Therefore, we have reduced the problem of estimating the source region in every frame to that of detecting one of L regions containing the source. In real-time processing, from each frame of the microphone signals, we first estimate the TDOA of the source w.r.t. all pairs of microphones. Consider the bit-vector $\hat{\underline{\tau}} = [\mathrm{sgn}(\tau_{21}), \mathrm{sgn}(\tau_{31}), \ldots, \mathrm{sgn}(\tau_{N,N-1})]$, consisting of the sign of TDOAs of the acoustic source w.r.t. all the microphone pairs. The region containing the acoustic source is detected as the region corresponding to the codeword $\in \boldsymbol{\tau}$ that is closest to the bit-vector of the acoustic source ($\hat{\underline{\tau}}$). Since we are computing distances between bit-vectors, we employ the Hamming distance, which gives the number of positions in which the two vectors differ. Thus, a coarse location of the source ($\mathcal{R}_S^*$) is detected as: $\ell^* = \arg\min_{\ell} \mathcal{H}\left(\underline{\tau}^{(\ell)}, \hat{\underline{\tau}}\right)$, $\mathcal{R}_S^* = \mathcal{R}_S^{(\ell^*)}$, where $\mathcal{H}\left(\underline{x}, \underline{y}\right)$ denotes the Hamming distance between the bit-vectors $\underline{x}$ and $\underline{y}$. Thus, in real-time processing, once the TDOA estimates across all microphones are estimated, only L Hamming distances need to be computed to detect the coarse location of the source. This detection strategy using Hamming distances not only reduces the computational complexity of the algorithm but also imparts robustness to reverberation. This is because, reverberation in the enclosure creates phantom sources within the enclosure leading to erroneous TDOA estimates. Errors in TDOA manifests as bit flips in the bit-vector containing the sign of TDOAs. Thus, assigning regions based on least Hamming distances amounts to error correction of a maximum of $\delta - 2$, where δ is the minimum Hamming distance between the codewords in $\boldsymbol{\tau}$. It is worthwhile to note that, for a fixed array of microphones, the codewords and hence the valid regions are fixed. The design of regions such that their corresponding codewords are maximally apart is a more challenging problem involving the design of the microphone array, which will be separately addressed.

The choice of the number and shape of valid regions is also a matter of aesthetics in video recording. If there are far too many valid regions, the camera keeps on adjusting and fluctuating between adjacent regions although the source maybe stationary. This results in a poor video recording. If there are far too less valid regions, then the camera will be quite stable and consistent, however the source of sound may be outside the field of view for significant amount of time. Since the field of view of a camera is typically conical, we choose a uniform circular array with $N = 8$ microphones, resulting in $2N = 16$ sector-like valid source regions as shown in Figure 4(a). The sectors match the conical field of view of the camera. Such a choice yields codewords whose Hamming distance matrix is shown in Figure 3(b).



(a) (b)

Figure 4: (a) Real enclosure schematic with calibrated source positions (blue diamond), desks (blue rectangles), sector boundaries(black lines); (b) UCA with $N = 8$ microphones.

Distinct codewords have a minimum distance of 3. Thus, with the chosen configuration of microphones, we can correct single bit errors in the sign-of-TDOAs vector to correctly detect the source region.

4. Experiments in a Real Reverberant Enclosure

4.1. Experimental Setup

The block diagram of the proposed approach is as shown in Figure 5. The microphone signals are synchronously sampled at 44.1 kHz and quantized at 24 bits per sample using the M-Audio M-Track Eight [12] multichannel audio interface and acquired into a workstation. The digitized microphone signals are re-sampled to 16 kHz. The TDOAs across all pairs of microphone signals are estimated using generalized cross correlation with phase transform (GCC-PHAT) [13] technique. We now compute θ_{l^*}, the DOA of the centroid of $\mathcal{R}_S^{(\ell^*)}$. This angle is transmitted to a Raspberry Pi [14] over a TCP/IP socket which steers the pan-tilt-zoom (PTZ) camera (Minrray VHD-A910) towards θ_{l^*} using pre-recorded IR commands with the Linux infra-red remote control (LIRC) package. We set the field of view of the camera to about 30 deg such that a human speaker is captured well within the video frames. The experiments reported in this paper are done in a computing lab of size 9.936 m $\times$6.919 m $\times$4.142 m. The 60 dB reverberation time (RT60) of the enclosure was found to be approximately 600 ms. A schematic of the enclosure, along with the $2N = 16$ valid regions obtained (regions between the black lines) is shown in Figure 4(a). A uniform circular array (UCA) of radius 10 cm is used, as shown in Figure 4(b). In order to analyze the performance of the proposed method, nine locations have been calibrated within the enclosure as shown in Figure 4(a).

A sound source is placed on each of these locations for a duration of 2 minutes. Speech data from 4 speakers (2 male and 2 female) from the Dialect 8 (DR 8) of the TIMIT database [15] is played out through a mobile phone speaker as it has a small aperture approximating a point source. With a frame length of 1 s, we obtain a set of 120 audio frames. We use a frame energy-based voice activity detector (VAD) to identify the frames containing speech. Only speech frames are processed to obtain coarse location estimates.

4.2. Coarse localization performance

In this section we study the coarse localization performance of the proposed approach presented in Sec. 2. Since the proposed approach yields a region estimate rather than a point estimate, we assess its performance in terms of "Coarse Local-

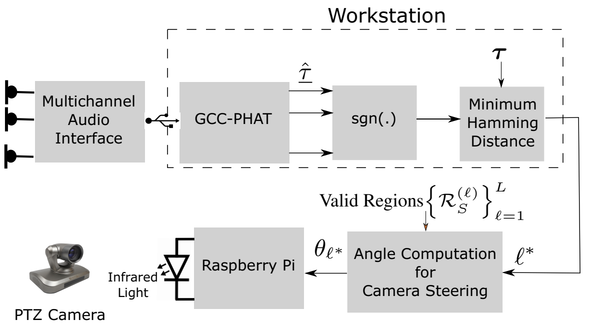
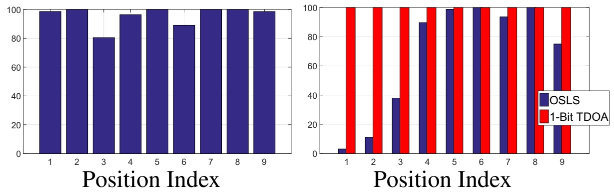


Figure 5: Block diagram of the proposed approach for automatic camera steering.



(a) CLA (%)

(b) % frames within FOV

Figure 6: (a) Coarse localization accuracy (CLA) of the 1-bit TDOA approach. (b) Comparison of the 1-Bit TDOA approach (red bar) with the OSLS approach (blue bar).

ization Accuracy (CLA)" defined as the percentage of speech frames wherein the estimated source region contains the actual source. Since the estimated coarse location of the source is necessarily a convex region (as shown in Sec. 2), we use the dot product of the actual calibrated source position with that of the normals of the faces of the estimated convex region, to determine if the actual source position is within the estimated source region. Figure 6(a) shows a plot of CLA as a function of the position index. We observe that the proposed approach in general has an accuracy closer to 100 % except for position index 3 and 6. Interestingly, position index 3 and 6 are both on, or very close to the border between adjacent sectors. On careful examination, for the case when the source is placed at the position index 3, in 80 % of the frames, the source region is detected correctly and in the remaining 20%, the sector adjacent to the true sector is detected. However, irrespective of which sector gets detected, the source always lies in the field of view of the camera. Similarly, for position 6, those frames in which the source region is not detected correctly, the adjacent sector is detected. We would like to point out that, even without any temporal smoothing of the coarse location estimates, the raw coarse location estimates themselves are consistent without any erratic variations in the region detected.

4.3. Camera steering performance comparison

In this section, we compare the performance of the one-step least-squares (OSLS) [3] technique with the proposed 1-bit TDOA technique. OSLS is a point localization technique which estimates the source location in 3-D. Hence, it requires the microphones to be placed in a non-co-planar arrangement. Therefore, in order to ensure fair comparison with the proposed approach, we use a random array of microphones, for the OSLS technique, placed in the enclosure as shown in Figure 7. We would like to point out that, sign-of-TDOA based region formation is not specific to any array configuration and can be

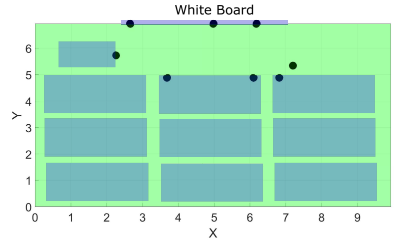


Figure 7: Random placement of microphones used for OSLS technique.

used with a random array of microphones. However, when we use the 1-bit TDOA approach with a random array of microphones, we obtain a large number of valid regions which are narrow. This results in a single speaker occupying more than one valid region. Also, with a large number of regions, the camera could fluctuate between adjacent regions resulting in a poor video recording. Therefore, we use the OSLS technique with a random array and compare it with the proposed approach with a UCA of microphones. Using the OSLS-based approach, the camera is steered such that the point location estimate is in the center of the field of view of the camera (cf. Sec. 4.1). We compare the performance of the OSLS based camera steering with that of the proposed technique by comparing the percentage of audio frames in which the actual source is within the field of view of the camera. We observe that the performance of the OSLS technique is comparable to the proposed approach for position indices $4-8$. However, performance at the position indices 1, 2 and 9 are significantly poorer. The proposed approach performs consistently well at all position indices. Additionally, it is observed that in the OSLS technique there are a number of false alarms which are erratic and hence results in erratic movement of the camera although the source of sound is stationary.

5. Conclusions

Motivated by the application of automatic camera steering, we have developed an algorithm to coarsely localize a source. Through real-recordings in a reverberant enclosure, we have shown that the proposed approach is robust to reverberation. Comparisons with point-localization based camera steering clearly shows the advantage of the proposed technique, in terms of consistency and stability of camera movement and robustness to reverberation. From our experiments we have observed that the OSLS technique always yields accurate location estimates when the source is well within the convex hull of the microphones and poorer location estimates when the source is on the border or outside the convex hull of the microphones. In general the proposed method, offers a convenient way to setup an automatic source tracking system in enclosures of different size and shapes, as it involves placement of a small circular array. Placement of a random array is more cumbersome and demands enclosure specific design. The proposed method has a small footprint and very low run-time computational complexity making it ideally suited for a low power single-chip realization.

6. Acknowledgements

The authors would like to thank Mr. Sunil Rudresh and Mr. Anirudhdha Adiga, Department of Electrical Engineering, Indian Institute of Science, Bangalore, for useful technical discussions regarding automatic camera steering.

7. References

[1] F. Keyrouz and K. Diepold, "Binaural source localization and spatial audio reproduction for telepresence applications," *Presence: Teleoper. Virtual Environ.*, vol. 16, no. 5, pp. 509–522, Oct. 2007. [Online]. Available: http://dx.doi.org/10.1162/pres.16.5.509

[2] S. Kim, H. Jeon, and L. Y. Kim, "Source localization for dual speech enhancement technology," in *Advances in Sound Localization*, D. P. Strumillo, Ed. InTech, ch. 9. [Online]. Available: http://www.intechopen.com/books/advances-in-sound-localization/source-localization-for-dual-speech-enhancement-technology

[3] Y. Huang, J. Benesty, and G. W. Elko, "Passive acoustic source localization for video camera steering," in *Proc. Int'l. Conf. Acoust. Speech and Signal Process.*, vol. 2, 2000, pp. 909–912.

[4] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," in *IEEE Int'l Conf. Acoust., Speech, and Signal Process.*, vol. 1, Apr 1997, pp. 187–190.

[5] M. K. Mukul, R. Prasad, M. M. Choudhary, and F. Matsuno, "Steering of camera by stepper motor towards active speaker using microphone array," in *SICE Annual Conference*, Aug 2008, pp. 19–24.

[6] Z. Liang, X. Ma, and X. Dai, "Robust tracking of moving sound source using multiple model Kalman filter," *J. Acoust. Soc. Amer.*, vol. 69, pp. 1350–1355, 2008.

[7] Y. Huang, J. Benesty, and G. W. Elko, "An efficient linear-correction least-squares approach to source localization," in *IEEE Workshop on the Appl. of Signal Process. to Audio and Acoust.*, 2001, pp. 67–70.

[8] J. H. DiBiase, "A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays," Ph.D. dissertation, Brown Univ., Providence, 2000.

[9] B. Grünbaum, *Convex Polytopes*, 2^{nd} ed. Springer, New York, 2003, vol. 221, p. 466.

[10] D. Bremner, K. Fukuda, and A. Marzetta, "Primal-dual methods for vertex and facet enumeration," *Discrete and Computational Geometry*, vol. 20, pp. 333–357, 1998.

[11] J. Matt, "Representing polyhedral convex hulls by vertices or (in)equalities," *Internet: http://www.mathworks.com/matlabcentral/fileexchange/30892-representing-polyhedral-convex-hulls-by-vertices-or-inequalities*, Aug. 2012.

[12] M-Audio, "M-audio M-TRACK EIGHT. Internet: http://www.m-audio.com/products/view/m-track-eight#.VvvoNEeFHbQ," 30^{th} March 2016.

[13] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, 1976. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?&arnumber=1162830

[14] "Raspberry Pi. Internet: https://www.raspberrypi.org/," 17^{th} June 2016.

[15] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, "DARPA TIMIT acoustic phonetic continuous speech corpus CDROM," 1993.