SVD-PHAT: A FAST SOUND SOURCE LOCALIZATION METHOD

François Grondin, James Glass

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
{fgrondin,glass}@mit.edu

ABSTRACT

This paper introduces a new localization method called SVD-PHAT. The SVD-PHAT method relies on Singular Value Decomposition of the SRP-PHAT projection matrix. A k-d tree is also proposed to speed up the search for the most likely direction of arrival of sound. We show that this method performs as accurately as SRP-PHAT, while reducing significantly the amount of computation required.

Index Terms— Sound Source Localization, SRP-PHAT, SVD-PHAT, Direction of Arrival

1. INTRODUCTION

Distant speech processing is a challenging task, as the target sound source is usually corrupted by noise from the environment and is degraded by reverberation [1]. Beamforming methods are often used as a preprocessing step to enhance the corrupted speech signal using multiple microphones. Many beamforming methods, such as Delay and Sum (DS), Geometric Source Separation (GSS) [2] and Minimum Variance Distortionless Response (MVDR) [3], require the target source direction of arrival (DOA). Sound source localization consists in estimating this DOA, and often relies on Multiple Signal Classification (MUSIC) [4] or Steered-Response Power Phase Transform (SRP-PHAT) [5] methods.

MUSIC is based on Standard Eigenvalue Decomposition (SEVD-MUSIC), and was initially used for narrowband signals, then adapted to broadband signals to make localization robust to additive noise [6]. The latter method however assumes that the target signal is more powerful than noise. To cope with this limitation, MUSIC based on Generalized Eigenvalue Decomposition (GEVD-MUSIC) handles scenarios when noise is more powerful than the signal of interest [7]. Alternatively, MUSIC based on Generalized Singular Value Decomposition (SVD-MUSIC), reduces the computational load of GEVD-MUSIC and improves DOA estimation accuracy [8]. However, all MUSIC-based methods require performing online eigenvalue or singular value decompositions, which involve a significant amount of computations, and make real-time processing more challenging.

SRP-PHAT computes the Generalized Cross-Correlation with Phase Transform (GCC-PHAT) between each pair of microphones [9]. The exact SRP-PHAT solution involves fractional Time-Difference of Arrival (TDOA) delays, and requires a significant amount of computation. The Fast Fourier Transform (FFT) is thus often used to speed up the computation of GCC-PHAT, which makes this method appealing for real-time applications [10, 11]. However, using the FFT restricts the transform to discrete TDOA values, which reduces localization accuracy. Interpolation [12, 13, 14], fractional delay estimation [15] and fractional Fourier transform [16] attempt to overcome the FFT discretization drawback. Moreover, searching for sound source involves a significant amount of computations when scanning the 3D-space. Stochastic region contraction [17], hierarchical search [18, 19, 20] and vectorization [21] are proposed to speed up scanning, but are usually restricted to a 2D surface.

In this paper, we propose a new method inspired from the original SRP-PHAT approach, called SVD-PHAT. The objective is to reduce the amount of computations typically involved in the exact SRP-PHAT, while preserving its accuracy. The proposed technique relies on SVD to generate a transform related to the matrix geometry that maps the initial observations to a smaller subspace. This subspace is then searched with a k-d tree, which returns the estimated DOA.

2. SRP-PHAT

SRP-PHAT relies on the TDOA estimation for all pairs of microphones (for an array with M microphones, there are $P = M(M-1)/2$ possible pairs). The TDOA (in sec) corresponds to the difference between the distance from the source $\mathbf{s}_q \in \mathbb{R}^3$ to microphone i at position $\mathbf{r}_i \in \mathbb{R}^3$, and the distance between the same source and another microphone j at position $\mathbf{r}_j \in \mathbb{R}^3$, divided by the speed of sound in air $c \in \mathbb{R}^+$ (in m/sec). Since all signals are discretized in time, it is also convenient to express the TDOA in terms of samples by adding

This work was supported in part by the Toyota Research Institute and by the Fonds de recherche du Québec - Nature et technologies.

the sample rate ($f_S \in \mathbb{R}^+$) in the expression, as shown in (1).

$$\tau_{q,i,j} = \frac{f_S}{c} \left(\|\mathbf{s}_q - \mathbf{r}_i\|_2 - \|\mathbf{s}_q - \mathbf{r}_j\|_2 \right) \quad (1)$$

where $\|\dots\|_2$ stands for the Euclidean norm.

In most microphone array configurations, the array aperture is small compared to the distance between the source and the array, such that the farfield assumption holds. In this case, (1) can be formulated as in (2).

$$\tau_{q,i,j} = \frac{f_S}{c} \left(\mathbf{r}_j - \mathbf{r}_i \right) \cdot \frac{\mathbf{s}_q}{\|\mathbf{s}_q\|_2} \quad (2)$$

Let $x_m[n]$ be the signal of microphone m in the time domain. The expression $X_m^l[k] \in \mathbb{C}$ is obtained with a Short Time Fourier Transform (STFT) with a sine window, where $N \in \mathbb{N}$ and $\Delta N \in \mathbb{N}$ stand for the frame and hop sizes in samples, respectively, and $k \in \mathbb{N} \cap [0, N/2]$ and $l \in \mathbb{N}$ stand for the frequency bin and frame indexes, respectively. For clarity, the frame index l is omitted in this paper without loss of generality. The normalized cross-spectrum for each pair of microphones (i, j) (where $i \neq j$) corresponds to the expression $X_{i,j}[k] \in \mathbb{C}$ in (3). The operators $\{\dots\}^*$ and $|\dots|$ stand for the complex conjugate and the absolute value, respectively.

$$X_{i,j}[k] = \frac{X_i[k] X_j[k]^*}{|X_i[k]||X_j[k]|} \quad (3)$$

In the frequency domain, the TDOA $\tau_{q,i,j}$ leads to the coefficient $W_{q,i,j}[k] \in \mathbb{C}$ in (4) according to SRP-PHAT beamforming.

$$W_{q,i,j}[k] = \exp\left(2\pi\sqrt{-1}k\tau_{q,i,j}/N\right) \quad (4)$$

For each potential source position located at $\mathbf{s}_q$, SRP-PHAT returns an energy value expressed by $Y_q \in \mathbb{R}$, where $\Re\{\dots\}$ extracts the real part.

$$Y_q = \Re\left\{ \sum_{i=1}^{M} \sum_{j=(i+1)}^{M} \sum_{k=0}^{N/2} W_{q,i,j}[k] X_{i,j}[k] \right\} \quad (5)$$

The estimated direction of arrival (DOA) of sound corresponds to the position denoted by $\mathbf{s}_{\bar{q}}$, where $\bar{q}$ is obtained in (6). Moreover, the scalar $Y_{\bar{q}}$ is often used to discriminate a valid sound source from backgroud noise.

$$\bar{q} = \arg\max_q \{Y_q\} \quad (6)$$

Computing Y_q for $q \in \mathbb{N} \cap [1, Q]$ as in (6) involves a complexity order of $\mathcal{O}(QPN)$, and searching for the best potential source results in (6) leads to a $\mathcal{O}(Q)$ search. When the number of points to scan (Q) gets large, the SRP-PHAT involves numerous computations, which makes the method less suitable for real-time applications. The proposed SVD-PHAT method described in the next section aims to alleviate this limitation.

3. SVD-PHAT

To define the SVD-PHAT method, it is convenient to start from SRP-PHAT expressed in matrix form. We define the vector $\mathbf{X} \in \mathbb{C}^{P(N/2+1)\times 1}$ in (7), which concatenates all normalized cross-spectra previously introduced in (3).

$$\mathbf{X} = \begin{bmatrix} X_{1,2}[0] & X_{1,2}[1] & \cdots & X_{M-1,M}[N/2] \end{bmatrix}^T \quad (7)$$

Similarly, the matrix $\mathbf{W} \in \mathbb{C}^{Q\times P(N/2+1)}$ holds all the SRP-PHAT coefficients:

$$\mathbf{W} = \begin{bmatrix} W_{1,1,2}[0] & W_{1,1,2}[1] & \cdots & W_{1,M-1,M}[N/2] \\ \vdots & \vdots & \ddots & \vdots \\ W_{Q,1,2}[0] & W_{Q,1,2}[1] & \cdots & W_{Q,M-1,M}[N/2] \end{bmatrix} \quad (8)$$

Finally, the vector $\mathbf{Y} \in \mathbb{C}^{Q\times 1}$ stores the SRP-PHAT energy for all Q potential sources and is obtained from the following matrix multiplication:

$$\mathbf{Y} = \begin{bmatrix} Y_1 & \dots & Y_Q \end{bmatrix}^T = \Re\{\mathbf{W}\mathbf{X}\} \quad (9)$$

As mentionned for SRP-PHAT, this matrix multiplication is computationally expensive when there are numerous potential source positions to scan. To cope with this limitation, we propose to perform Singular Value Decomposition on the matrix $\mathbf{W}$, where $\mathbf{U} \in \mathbb{C}^{Q\times K}$, $\mathbf{S} \in \mathbb{C}^{K\times K}$ and $\mathbf{V} \in \mathbb{C}^{P(N/2+1)\times K}$, as shown in (10).

$$\mathbf{W} \approx \mathbf{U}\mathbf{S}\mathbf{V}^H \quad (10)$$

where $\{\dots\}^H$ stands for the Hermitian operator.

The parameter $K \in \mathbb{N} \cap]0, K_{max}]$, where the upper bound $K_{max} = \max\{Q, P(N/2+1)\}$, is chosen to ensure accurate reconstruction of $\mathbf{W}$, according to the condition in (11), where the user-defined parameter δ is a small positive value that models the tolerable reconstruction error. The operator $\mathrm{Tr}\{\dots\}$ stands for the trace of the matrix.

$$\mathrm{Tr}\left\{\mathbf{S}\mathbf{S}^T\right\} \geq (1-\delta)\,\mathrm{Tr}\left\{\mathbf{W}\mathbf{W}^H\right\} \quad (11)$$

The vector $\mathbf{Z} \in \mathbb{C}^{K\times 1}$ results from the projection of the observations $\mathbf{X}$ in the K-dimensions subspace:

$$\mathbf{Z} = \mathbf{V}^H\mathbf{X} \quad (12)$$

The matrix $\mathbf{D} \in \mathbb{C}^{Q\times K}$ is obtained in (13) and can be decomposed in a set of Q vectors $\mathbf{D}_q \in \mathbb{C}^{1\times K}$:

$$\mathbf{D} = \mathbf{U}\mathbf{S} = \begin{bmatrix} \mathbf{D}_1^T & \mathbf{D}_2^T & \dots & \mathbf{D}_Q^T \end{bmatrix}^T \quad (13)$$

The index of the most likely DOA obtained in (6) now corresponds to:

$$\bar{q} = \arg\max_q \{\Re\{\mathbf{D}_q \cdot \mathbf{Z}^H\}\} \quad (14)$$

One way to find the correct value of q in (14) consists in computing every Y_q for $q \in \mathbb{N} \cap [1, Q]$, and then finding the index q that leads to the maximum value, which obviously involves a significant amount of computations, as the complexity order is linear ($\mathcal{O}(Q)$). It is therefore relevant to look for an alternate cost function that would allow a more efficient search. For all values of q, the expressions $\|\mathbf{D}_q\|_2$ are almost identical (when the reconstruction meets condition in (11) for a small value of δ) , but do not necessary equal to 1. We thus define the new vectors $\hat{\mathbf{D}}_q = \mathbf{D}_q / \|\mathbf{D}_q\|_2$ and the normalized vector $\hat{\mathbf{Z}} = \mathbf{Z}/\|\mathbf{Z}\|_2$. With $\|\hat{\mathbf{D}}_q\|_2^2 = 1$ and $\|\hat{\mathbf{Z}}\|_2^2 = 1$, the dot product can therefore be expressed as follows:

$$\Re\{\mathbf{D}_q \cdot \mathbf{Z}^H\} = 1 - \frac{1}{2}\|\hat{\mathbf{D}}_q - \hat{\mathbf{Z}}^H\|_2^2 \tag{15}$$

and thus maximizing (14) now corresponds to minimizing $\|\hat{\mathbf{D}}_q - \hat{\mathbf{Z}}^H\|_2^2$. This minimization can be done by computing $\|\hat{\mathbf{D}}_q - \hat{\mathbf{Z}}^H\|_2^2$ for all values of q and finding q that leads to the minimum value, but this brings us back to the linear complexity order $\mathcal{O}(Q)$ as in (14). Fortunately, the new formulation based on sum of squares becomes a nearest neighbor search problem, which can be solved efficiently using a k-d tree [22].

Algorithm 1 summarizes the offline configuration performed prior to processing and the online computations. The real-time performances are independent of the computationally expensive SVD and tree construction since these are done offline. During online processing, computing the vector $\mathbf{Z}$ involves a complexity order of $\mathcal{O}(KPN)$ and the k-d tree search exhibits on average a complexity $\mathcal{O}(\log Q)$ [22].

Algorithm 1 SVD-PHAT

Offline:

1: Generate $\mathbf{W}$ from (1), (4), and (8).
2: Perform SVD and obtain $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}$, with K chosen according to condition in (11).
3: Generate the normalized vectors $\hat{\mathbf{D}}_q$ from $\mathbf{D}_q$ in (13).
4: Build a k-d tree for all $\hat{\mathbf{D}}_q$.

Online:

1: Generate $\mathbf{X}$ from the STFT coefficients as in (7).
2: Compute $\mathbf{Z}$ with (12) and generate $\hat{\mathbf{Z}}^H$.
3: Find $\bar{q}$ using the k-d tree search.
4: Find $Y_{\bar{q}}$ with the corresponding row of $\mathbf{W}$ in (9).

4. RESULTS

The parameters for the experiments are summarized in Table 1. The sample rate f_S captures all the frequency content of speech, and the speed of sound c corresponds to typical indoor conditions. The frame size N analyzes segments of 16 msecs, and the hop size ΔN provides a 50% overlap. The potential DOA are represented by equidistant points on a unit sphere generated recursively from a tetrahedron, for a total of 2562 points, as in [11].

Table 1. SVD-PHAT Parameters

f_S	c	N	ΔN	Q
16000	343.0	256	128	2562

We investigate three different microphone array geometries: a 1-D linear array, a 2-D planar array and a 3-D array. The microphones exact xyz-positions are given in cm in Table 2 and the geometries are shown in Fig. 1.

Table 2. Positions (x,y,z) of the microphones in cm

Mic	1-D	2-D	3-D
1	$(0, 0, -5.0)$	$(0, 0, 0)$	$(0, 0, 0)$
2	$(0, 0, -3.3)$	$(5, 0, 0)$	$(-5, 0, 0)$
3	$(0, 0, -1.7)$	$(2.5, 4.3, 0)$	$(5, 0, 0)$
4	$(0, 0, 0)$	$(-2.5, 4.3, 0)$	$(0, -5, 0)$
5	$(0, 0, 1.7)$	$(-5.0, 0, 0)$	$(0, 5, 0)$
6	$(0, 0, 3.3)$	$(-2.5, -4.3, 0)$	$(0, 0, -5)$
7	$(0, 0, 5.0)$	$(2.5, -4.3, 0)$	$(0, 0, 5)$

Simulations are conducted to measure the accuracy of the proposed method. The microphone array and the target source are positionned randomly in a 10m x 10m x 3m rectangular room. For each configuration, the room reverberation is modeled with Room Impulse Responses (RIRs) generated with the image method [23], where the reflection coefficients are sampled randomly in the uniform interval between 0.2 and 0.5. Sound segments from the TIMIT dataset are then convolved with the generated RIRs. Diffuse white noise is added on each channel, for a signal-to-noise ratio (SNR) that varies randomly between 0dB and 30dB. A total of 1000 different configurations are generated for each microphone array.

We vary δ to analyze its impact on the accuracy of localization, measured as Root Mean Square Error (RMSE). For the 1-D linear array, localization can only provide a position on 180° arc. The 3-D position from the 2562 points unit sphere is therefore mapped to an arc:

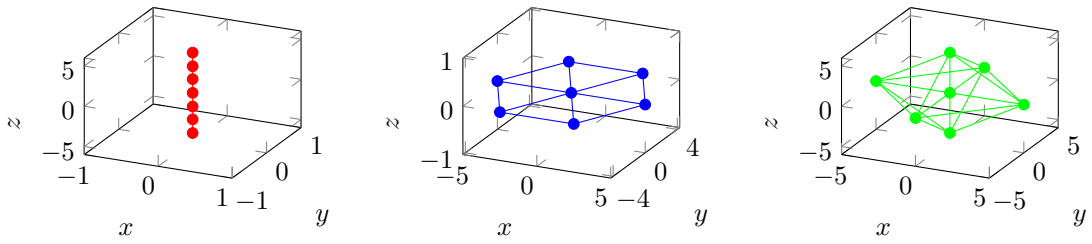
$$f_1(\mathbf{s}) = [\cos(g(\mathbf{s})), 0, \sin(g(\mathbf{s}))] \tag{16}$$

where

$$g(\mathbf{s}) = \operatorname{atan2}\left\{\mathbf{s}|_z, \sqrt{(\mathbf{s}|_x)^2 + (\mathbf{s}|_y)^2}\right\} \tag{17}$$

For the 2-D planar array, localization returns a position on a half-sphere, and thus every point is mapped to the positive hemisphere as follows:

$$f_2(\mathbf{s}) = [\mathbf{s}|_x, \mathbf{s}|_y, |\mathbf{s}|_z|] \tag{18}$$



(a) 1-D linear array (b) 2-D planar array (c) 3-D array

Fig. 1. Geometries of the microphone arrays in xyz-coordinates (dimensions are given in cm)

Finally, for the 3-D array, the localization result can span the full 3-D space, such that the mapping function corresponds to identity:

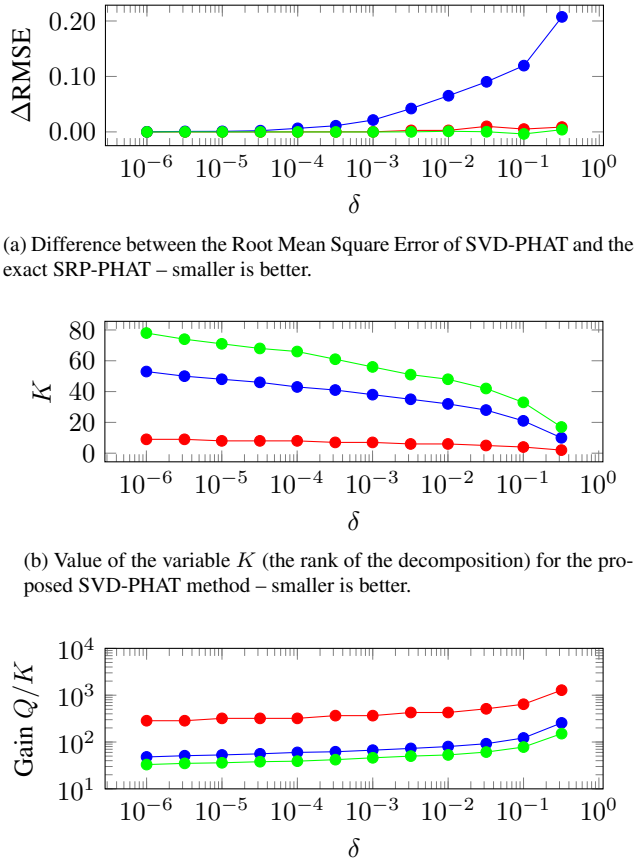
$$f_3(\mathbf{s}) = \mathbf{s} \tag{19}$$

The RMSE between the estimated DOA ($\mathbf{s}_{\bar{q}}$) for all frames for a given room configuration and speech signal is summed and weighted with the energy Y_q, and then compared with the theoretical DOA defined by $\mathbf{s}_0$. The mathematical expression corresponds to (20), where $\alpha = \{1, 2, 3\}$ for 1-D, 2-D and 3-D arrays, respectively.

$$\text{RMSE}_\alpha = \left\| \frac{\sum f_\alpha(\mathbf{s}_{\bar{q}}) Y_{\bar{q}}}{\sum Y_{\bar{q}}} - f_\alpha(\mathbf{s}_0) \right\|_2 \tag{20}$$

Figure 2a shows the difference between the RMSE from SVD-PHAT and SRP-PHAT (denoted as ΔRMSE), with respect to the δ parameter. As expected, when δ increases, the reconstruction error gets significant and this reduces the accuracy of localization for SVD-PHAT. It is interesting to note that the 2-D planar array shows the largest increase in RMSE. Figure 2b shows the value of the K parameter as a function of δ. Note how the K value is smaller when the array spans only one or two dimensions, as expected since the transfer function between DOAs are more correlated. The gain in performance is mostly due to the reduction from a matrix multiplication with Q rows in (9) to a matrix multiplication with K rows in (12). Figure 2c therefore shows the gain Q/K as a function of δ.

It is reasonable to define $\delta = 10^{-5}$ as the RMSE between SRP-PHAT and SVD-PHAT is almost identical. With this configuration, the gain Q/K reaches 320, 53 and 36 for 1-D, 2-D and 3-D arrays, which is considerable, and demonstrates the superiority of SVD-PHAT over SRP-PHAT in terms of computational requirements, while preserving the same accuracy.



(a) Difference between the Root Mean Square Error of SVD-PHAT and the exact SRP-PHAT – smaller is better.

(b) Value of the variable K (the rank of the decomposition) for the proposed SVD-PHAT method – smaller is better.

(c) Performance gain of SVD-PHAT when compared to exact SRP-PHAT method – greater is better.

Fig. 2. Performance of the proposed SVD-PHAT method with respect to the exact SRP-PHAT method. Results are presented for the 1-D linear array (red), the 2-D planar array (blue) and the 3-D array (green).

5. CONCLUSION

This paper introduces a new localization method named SVD-PHAT. This technique can perform with the same accuracy as SRP-PHAT, while reducing significantly the amount of computations.

In future work, we will investigate multiple source localization with SVD-PHAT. It would also be interesting to introduce binary time-frequency mask, which could reduce even more the amount of computations. The method could also be extended to deal with speed of sound mismatch, the near-field effect and microphone position uncertainty.

6. REFERENCES

[1] H. Tang, W.-N. Hsu, F. Grondin, and J. Glass, "A study of enhancement, augmentation, and autoencoder methods for domain adaptation in distant speech recognition," in *Proc. INTERSPEECH*, 2018, pp. 2928–2932.

[2] L. Parra and C. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," in *Proc. IEEE Signal Processing Society Workshop*. IEEE, 2001, pp. 273–282.

[3] E. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, "New insights into the MVDR beamformer in room acoustics," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 1, pp. 158, 2010.

[4] R.O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

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

[6] C.T. Ishi, O. Chatot, H. Ishiguro, and N. Hagita, "Evaluation of a MUSIC-based real-time sound localization of multiple sound sources in real noisy environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots & Systems*, 2009, pp. 2027–2032.

[7] K. Nakamura, K. Nakadai, F. Asano, and G. Ince, "Intelligent sound source localization and its application to multimodal human tracking," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots & Systems*, 2011, pp. 143–148.

[8] K. Nakamura, K. Nakadai, and G. Ince, "Real-time super-resolution sound source localization for robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots & Systems*, 2012, pp. 694–699.

[9] M.S. Brandstein and H.F. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms," in *Proc. IEEE Int. Conf. Acoustics, Speech & Signals Process.* IEEE, 1997, vol. 1, pp. 375–378.

[10] F. Grondin, D. Létourneau, F. Ferland, V. Rousseau, and F. Michaud, "The ManyEars open framework," *Auton. Robots*, vol. 34, no. 3, pp. 217–232, 2013.

[11] J.-M. Valin, F. Michaud, and J. Rouat, "Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering," *Rob. Auton. Syst.*, vol. 55, no. 3, pp. 216–228, 2007.

[12] G. Jacovitti and G. Scarano, "Discrete time techniques for time delay estimation," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 525–533, 1993.

[13] M. McCormick and T. Varghese, "An approach to unbiased subsample interpolation for motion tracking," *Ultrasonic imaging*, vol. 35, no. 2, pp. 76–89, 2013.

[14] F. Viola and W.F. Walker, "A spline-based algorithm for continuous time-delay estimation using sampled data," *IEEE Trans Ultrason Ferroelectr Freq Control*, vol. 52, no. 1, pp. 80–93, 2005.

[15] D.L. Maskell and G.S. Woods, "The estimation of subsample time delay of arrival in the discrete-time measurement of phase delay," *IEEE Trans. Instrum. Meas*, vol. 48, no. 6, pp. 1227–1231, 1999.

[16] K.K. Sharma and S.D. Joshi, "Time delay estimation using fractional fourier transform," *EURASIP J. Audio Speech*, vol. 87, no. 5, pp. 853–865, 2007.

[17] H. Do, H.F. Silverman, and Y. Yu, "A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array," in *Proc. IEEE Int. Conf. Acoustics, Speech & Signals Process.*, 2007, vol. 1, pp. 121–124.

[18] D.N. Zotkin and R. Duraiswami, "Accelerated speech source localization via a hierarchical search of steered response power," *IEEE Trans. Audio, Speech, Language Process.*, vol. 12, no. 5, pp. 499–508, 2004.

[19] H. Do and H.F. Silverman, "Stochastic particle filtering: A fast srp-phat single source localization algorithm," in *IEEE Workshop on Applications of Signal Processing to Audio & Acoustics*, 2009, pp. 213–216.

[20] L.O. Nunes, W.A. Martins, M. Lima, L. Biscainho, M. Costa, F. Goncalves, A. Said, and B. Lee, "A steered-response power algorithm employing hierarchical search for acoustic source localization using microphone arrays," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5171–5183, 2014.

[21] B. Lee and T. Kalker, "A vectorized method for computationally efficient srp-phat sound source localization," in *Int. Workshop on Acoustic Echo & Noise Control*, 2010.

[22] J.L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, pp. 509–517, 1975.

[23] 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, 1979.