

STOCHASTIC PARTICLE FILTERING: A FAST SRP-PHAT SINGLE SOURCE LOCALIZATION ALGORITHM

Hoang Do, Harvey F. Silverman

LEMS

Division of Engineering

Box D, Brown University, Providence, RI 02912, USA

{hdo,hfs}@lems.brown.edu

ABSTRACT

Computational cost has been an issue for the proven robust source localization algorithm, steered response power (SRP) using the phase transform (SRP-PHAT). Some proposed computation reduction algorithms degrade under high noise and reverberant conditions. Some require at least 10% the cost of a full SRP-PHAT grid-search. In ICASSP 2007, we introduced a robust, low-cost global optimization technique, stochastic region contraction (SRC). In this paper, we present another algorithm, stochastic particle filtering (SPF), which uses SRC's initialization and is a kind of Importance Sampling technique. In this paper, the SRP is computed using a modification to the conventional PHAT, namely β -PHAT. Extensive experiments using real data and simulated data are shown. The results indicate that, while maintaining the desirable accuracy of the full search, this method reduces the cost to about half the cost of SRC (0.03% the cost of full search), thus making SRP-PHAT more practical for real-time applications.

Index Terms – Microphones, acoustic radiators, position measurement, particle filters.

1. INTRODUCTION

Locationing algorithms using steered beamforming, such as SRP-PHAT have been shown to be more robust than two-stage, TDOA-based and high-resolution spectral estimation-based algorithms under high noise and reverberant conditions [1, 2]. However, SRP-PHAT's tremendously high computational cost prevents it from practical implementation. This cost mainly lies in the search for the global extremum in a non-linear, multi-local extremum SRP space.

There have been a few approaches to solve the large computational cost of the search process, such as: hierarchical search [3], spherical intersection [4], and inverse mapping of the time-delays [5]. Unfortunately, the hierarchical search performance degrades quickly when the reverberation is high (at 430ms) [6]. The spherical intersection method requires that the source has to be in the near-field, which limits this solution from far-field cases. The time-delay inverse-mapping method stands the test of high reverberations and is applicable in all cases. However, the time-delay inverse-mapping and the spherical intersection methods still take up to about 10% the cost of the full grid-search. When the search volume is significantly large, the achieved savings might not be enough for real-time use. Recently, we introduced SRC [7] at ICASSP 2007 to cut the search cost to about 0.06%–0.1% of the full grid-search's. This method exploits the probabilistic relationship between the volume of the global maximum V_{peak} and

the search volume. Instead of evaluating every grid-points as in the full-search, it iterates only on a small subset of highly probable points with a substantially small probability ($\leq 0.1\%$) of missing the peak. The search volume is then contracted smaller and smaller until the global peak is captured.

In this paper, we present a new method, stochastic particle filtering (SPF), to find the global extremum using the initialization of SRC. In contrast to many particle filter algorithms[8, 9], where the initial set of particles (or 3-D points) is chosen randomly or uniformly without considering the probabilistic relation between the global maximum and the initial search volume, SPF uses the probabilistic approach of SRC to select the initial set of highly probable points (**importance sampling** step) that guarantees to capture the global peak with a negligible missing probability at a much faster convergence rate. In addition, we introduce a new **resampling** algorithm in SPF to avoid the *degeneracy* or also referred as *impoverishment* problem of particle filters. Furthermore, a modification of PHAT, namely β -PHAT [10] is used to compute the SRP functional. The β -PHAT has been shown empirically to smoothen the SRP surface better than the conventional one. Thorough experiments in a real room under adverse conditions as well as simulations were done to test the algorithm. The results show that SPF outperforms all other existing, fast global searching methods in computational saving. It costs only about 0.03% the cost of a full grid-search, while maintaining the same desirable accuracy. In a real-time system, we would like to make an estimate every 51.2ms, implying a 1.27GF (GigaFlops) machine for SRP-PHAT using SPF versus 1.24TF (TeraFlops) for SRP-PHAT using the full-search. This real-time performance can be achieved with today's powerful processors.

2. STEERED RESPONSE POWER USING THE BETA - PHASE TRANSFORM (SRP- β -PHAT)

Using the signal model derived in [7], the estimate of the (single) source location for a time frame n is given by,

$$\hat{\mathbf{d}}_s^{(n)} = \underset{\mathbf{d}}{\operatorname{argmax}} P(\mathbf{d})^{(n)}. \quad (1)$$

where \mathbf{d} is the 3-D spatial vector of a point in the space, and $P(\mathbf{d})^{(n)}$ denotes the steered response power (SRP) of point \mathbf{d} , and is defined as,

$$P(\mathbf{d})^{(n)} \equiv \sum_{k=1}^M \sum_{l=k+1}^M \int_{-\infty}^{\infty} \Psi_{kl}(\omega) X_k(\omega) X_l^*(\omega) e^{j\omega(\tau(\mathbf{d},l) - \tau(\mathbf{d},k))} d\omega. \quad (2)$$

The conventional phase transform (PHAT) is an especially effective weighting of a generalized cross-correlation (GCC) [11] for finding a TDOA from speech signals in highly-reverberant environment. The PHAT weighting is,

$$\Psi_{kl}(\omega) \equiv \frac{1}{|X_k(\omega)X_l^*(\omega)|}. \quad (3)$$

Recently, a modification of PHAT, namely, β -PHAT [10] has been shown empirically to be more robust, i.e., smoothing the SRP-PHAT surface and reducing the noise better than the conventional PHAT. It is defined as,

$$\beta - \Psi_{kl}(\omega) \equiv \frac{1}{(|X_k(\omega)X_l^*(\omega)|)^\beta}. \quad (4)$$

Here, β varies between 0 and 1 (with 0 being no weighting applied to the GCC, and 1 being the conventional PHAT). In this work, we chose $\beta = 0.8$ since it was shown in [10] and in our testing to give the best performance. In this case, β -PHAT does not completely remove the magnitude of the spectrum as in the conventional PHAT but preserves a fraction of it. This results in an enhanced, smoother surface of the SRP, where a large number of the local maxima corresponding to the noise are reduced significantly, see, e.g. Fig. 1. The calculation of any *particu-*

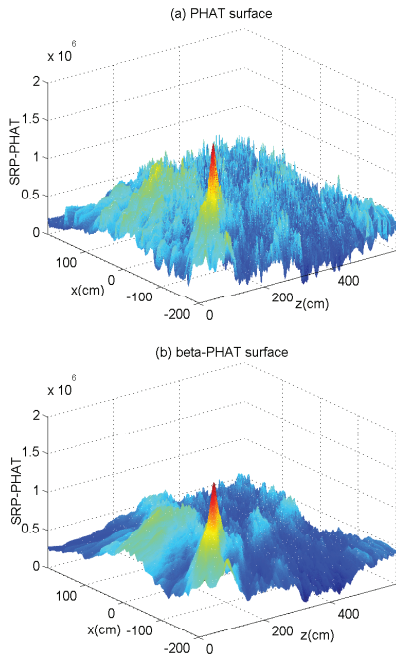


Figure 1: SRP plots of (a) a conventional PHAT and (b) a β -PHAT ($\beta = 0.8$) for a horizontal slice through (approximately) the height of the talker in the same frame.

lar point of $P(\mathbf{d})^{(n)}$ will be called a functional evaluation (fe). For the SRP-PHAT functional (for simplicity, we will implicitly refer to SRP- β -PHAT as SRP-PHAT in the rest of the paper), we want to determine a point-source location in the room that gives the maximum value of $P(\mathbf{d})^{(n)}$. In the next section, we introduce a new method, stochastic particle filtering (SPF) that evaluates a substantially small number of fe's compared to the conventional full grid-search in finding the SRP-PHAT global maximum.

3. STOCHASTIC PARTICLE FILTERING (SPF)

Stochastic particle filtering (SPF) uses a Importance Sampling technique [12] with the initialization of stochastic region contraction SRC.

While much work has been done on particle filtering used for tracking acoustic sources [13, 14], which usually requires information from previous frames to update the tracker, the SPF method presented in this paper is a real-time SRP-PHAT-based source localization implementation using only a single frame of data. Many particle filtering algorithms [8, 9], choose the initial set of particles (or points) randomly or uniformly without regarding a probabilistic measure of converging to the global maximum. On the other hand, SPF selects the initial set of particles, J_0 , in such a way that it guarantees a particle in the volume surrounding the global maximum, V_{peak} , is captured with a missing probability of less than 0.1%. This is the same initialization step to get J_0 particles in SRC [7]. If V_0 is the original search volume, we can estimate the number of particles needed to ensure that the probability of missing V_{peak} altogether is less than a given percent. The probability of a random particle hitting V_{peak} in the initial search volume V_0 is,

$$\mathbf{P}(\text{hit } V_{peak}) = \frac{V_{peak}}{V_0} \quad (5)$$

Hence, the probability of a random point missing V_{peak} is,

$$\mathbf{P}(\text{miss } V_{peak}) = 1 - \frac{V_{peak}}{V_0} \quad (6)$$

The event of throwing a random particle is independent from one to another. Therefore, the probability of throwing J_0 random particles missing V_{peak} is,

$$\mathbf{P}(J_0\text{-misses}) = (1 - \frac{V_{peak}}{V_0})^{J_0} \quad (7)$$

Taking the logarithm of both sides, separating J_0 to one side, we have,

$$J_0 = \frac{\log \mathbf{P}(J_0\text{-misses})}{\log(1 - \frac{V_{peak}}{V_0})} \quad (8)$$

From this relationship between the probability of throwing J_0 random points missing V_{peak} and the ratio between V_{peak} and V_0 , we can determine how many random particles, J_0 needed to throw to ensure that the missing probability, $\mathbf{P}(J_0\text{-misses})$ is *negligible* (substantially small in a realistic sense). In addition, a new, robust resampling technique is introduced to avoid the *degeneracy* problem of particle filters. The SPF searching algorithm for a frame f consists of two steps: Importance sampling and Resampling as follows,

(i) Importance Sampling:

- 1) Initialize: $k = 0$. For $p = 1, \dots, J_k$, randomly sample $\mathbf{d}(p) \sim \mathbb{U}[\vec{B}_k]$, where \mathbb{U} denotes the uniform distribution, and \vec{B}_k defines the 3-D boundary of the search volume V_k .

- 2) Evaluate to get the set W of the J_k importance weights:

$$W = w(\mathbf{d}(p)) = P(\mathbf{d}(p))^{(f)} \quad (9)$$

- 3) Select a set N_k containing the best $N \ll J_k$ out of the J_k particles:

$$\mathbf{d}(n) = \underset{\mathbf{d}(p)}{\operatorname{argmax}}(W), n = 1, \dots, N \quad (10)$$

- 4) Calculate the standard-deviation, $\sigma(k)$ of the N particles $\mathbf{d}(n)$.

- 5) Normalize the importance weights:

$$SW_k = \sum_{i=1}^N w(\mathbf{d}(n)) \quad (11)$$

$$w(\mathbf{d}(n)) = \frac{w(\mathbf{d}(n))}{SW_k}, n = 1, \dots, N \quad (12)$$

- 6) Number of functional evaluations (fe's):

$$\Phi_k = J_k \quad (13)$$

- (ii) **Resampling:** IF $\sigma(k) \geq 0.05m$:

- 1) Sort the particles $\mathbf{d}(n)$, $n = 1, \dots, N$ according to their weights in descending order.

- 2) Calculate the number of replications (a binary value, either 0 or 1, since all the weights are normalized from Eq. 12), $i(k, n)$ for each particle $\mathbf{d}(n)$:
For $n = 1 \rightarrow N$:

$$i(k, n) = \lfloor w(\mathbf{d}(n)) \times N \rfloor; n = 1, \dots, N. \quad (14)$$

- 3) Calculate the number of particles that are eliminated, $e(k)$:

$$e(k) = |\cdot| \ni i(k, n) = 0 \quad (15)$$

where $|\cdot|$ denotes cardinality of the set.

- 4) Find the set of R replicated particles:

$$\tilde{X}_k = \tilde{\mathbf{d}}(r) = \mathbf{d}(n) \times \mathbf{1}(i(k, n)) \quad (16)$$

where $r = 1, \dots, R \leq N$, and $\mathbf{1}$ denotes a vector having all elements equal to 1. This creates $i(k, n)$ copies of $\mathbf{d}(n)$ for each n .

- 5) Find the set of replicated (de-normalized) weights:

$$\tilde{W}_k = w(\tilde{\mathbf{d}}(r)) = w(\mathbf{d}(n)) \times \mathbf{1}(i(k, n)) \times SW_k \quad (17)$$

- 6) Adding $e(k)$ new particles, which are randomly selected within 0.1m-deviation of the "top weighted" $e(k)$ original particles:

$$\hat{X}_k = \hat{\mathbf{d}}(a) = \mathbf{d}(a) \pm \gamma \times 0.1 \quad (18)$$

where $a = 1, \dots, e(k)$, and γ is a random number in $[0, 1]$.

- 7) Evaluate the importance weights of the $e(k)$ newly added particles:

$$\hat{W}_k = w(\hat{\mathbf{d}}(a)) = P(\hat{\mathbf{d}}(a))^{(f)} \quad (19)$$

- 8) Create the new set of particles and weights for next iteration:

$$X_{k+1} = [\tilde{X}_k; \hat{X}_k] \quad (20)$$

$$W_{k+1} = [\tilde{W}_k; \hat{W}_k] \quad (21)$$

- 9) Update the normalizing factor (weight sum):

$$SW_{k+1} = \sum_{n=1}^N (W_{k+1}) \quad (22)$$

- 10) Normalize the importance weights:

$$W_{k+1} = \frac{W_{k+1}}{SW_{k+1}} \quad (23)$$

- 11) Calculate the new standard-deviation, $\sigma(k+1)$ of the N particles $\mathbf{d}(n)$ in the set X_{k+1} .

- 12) Update the number of fe's:

$$\Phi_{k+1} = \Phi_k + e(k) \quad (24)$$

- 13) If $\Phi_{k+1} \geq \Phi_T$: Exit and discard the estimate.

- 14) Iterate: $k = k + 1$.

ELSE: Output the final location estimate, \bar{x} and number of fe's, $\bar{\Phi}$:

$$\bar{x} = \sum_{n=1}^N (\mathbf{d}(n) \times w(\mathbf{d}(n))) \quad (25)$$

$$\bar{\Phi} = \Phi_k \quad (26)$$

Note: In our case, for $V_0 = 400cm \times 100cm \times 600cm = 24 \times 10^6 cm^3$, and typically for a low SNR case, $V_{peak} \approx 12 \times 10^4 cm^3$, hence from Eq. 8, $J_0 = 3000$ and $N = 100$ will err by missing V_{peak} less than **0.1%** of the time, which is virtually zero. Although the selection of J_0 seems a bit ad-hoc, absolute exactness is not required in this stage. A rough estimate of J_0 is sufficient, and the resulting missing probability is guaranteed to be in a negligible range. The threshold for the number of fe's is $\Phi_T = 20,000$.

4. EXPERIMENTAL EVALUATION

The new method was evaluated with experiments using real data in a real room with high noise and reverberations, as well as with simulated data (to have full control over different high levels of reverberations). The cost of the pre-processing stage (signal processing and cubic-splines interpolation) is the same for grid-search, SRC, coarse-to-fine region contraction (CFRC)[15], SPF, and is equal to 46.6 mo/f (*million operations per frame*)[7, 15]. These methods only differ in the search stage, and the search cost, hence the total computational cost depends on the number of fe's (each fe costs 2588 ops) evaluated in this stage. As shown in [7], the cost of a full grid-search is 63,113 mo/f (note that the cost of the pre-processing stage is tiny compared to this), implying a *1.28TF* machine for 51.2ms/estimate! SPF's cost is significantly less, and will be shown experimentally since it varies (within a small order of fe's) with the source locations in the room.

4.1. Experiments using real data

The system (a room with a $T_{60} = 450ms$ and a focal volume $V_{room} = 4m \times 1m \times 6m$) that we used in our experiments has been described in [7]. The source was an Advent AV009 speaker playing an 8-second clean-speech recording (wav file) at 5 different locations facing the 24 microphones in the room as shown in Fig. 2. A recording of different native American English talkers was played at each location. Hence, there were 5 recordings of 5 different talkers (1 female and 4 males) played at 5 locations. By varying the source locations and talkers, versatility of SPF were tested both for different SNR situations and talker's characteristics, such as: pitches, tones, etc. Frames of 102.4ms, advancing each 25.6ms,

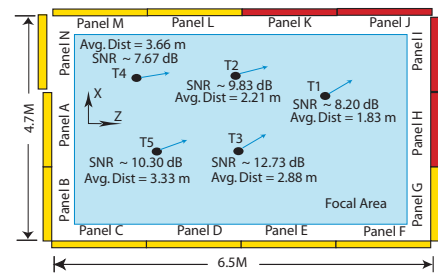


Figure 2: Top view of the array, showing source locations and panels. This experiment used 24 microphones of the 128 on panels H, I, J, K. The arrows indicate the orientation of the talker and the SNR's are for background noise only.

and a sampling rate of 20 KHz were the conditions for testing. We selected “speech frames” by hand-labeling frames from the close-talking data that have speech in at least 90% of the framelength. They total to about 62% of all 1500 frames. On the “speech” frames, performance of the full grid-search was compared to the measured locations of the source. On the frames where the full grid-search gave correct estimates, denoted as “good” frames (~ 916 frames), performance of SPF, SRC-I, and CFRC[15] were compared **relatively to the performance of the grid search**, i.e., performance is listed as 100% if SPF, SRC-I, & CFRC achieved the global maximum everywhere the grid-search did. Results are given in Table 1 for accuracy and the average number of fe’s used for SPF, SRC-I, and CFRC on “good” frames, and for grid-search on all “speech” frames. A location estimate was considered an error if it was either off by more than 5cm in x or z or 10cm in y , the vertical dimension.

Algorithm	T1 8.2dB		T2 9.83dB		T3 12.73dB		T4 7.67dB		T5 10.30dB	
	% Corr.	# fe’s	% Corr.	# fe’s	% Corr.	# fe’s	% Corr.	# fe’s	% Corr.	# fe’s
Grid Search	98.54	2.4×10^7	98.17	2.4×10^7	97.96	2.4×10^7	71.57	2.4×10^7	89.96	2.4×10^7
SPF	99	8,074	100	7,317	100	7,144	100	7,064	100	7,770
SRC - I	99	11,649	100	10,300	100	10,884	100	10,913	100	10,748
CFRC	98	12,171	100	11,031	99	11,380	100	11,510	100	12,808

Table 1: Performance and cost of SRP-PHAT using a full grid search over all speech frames; SPF, SRC, and CFRC over good frames for 5 different locations.

4.2. Evaluation using simulated data

We utilized the Brown Acoustic Simulator (BAS)[16] to simulate a fixed, single source facing 24 microphones in the room described in section 4.1. Uncorrelated white Gaussian noise was added to the microphone signals, making SNR at the source 12dB. We tested SPF under 4 adverse room’s conditions, i.e., T_{60} =315, 400, 500 and 585ms. Table 2 shows the performance (% correct) and cost (number of fe’s and % cost respectively) of SPF relative to those of the full search on 300 frames where the full search gave correct estimates, i.e., “good” frames.

T_{60}	% Corr.	# fe’s	% Cost
315 ms	100	9,331	0.039
465 ms	97	9,018	0.038
510 ms	96	12,254	0.051
585 ms	96	9,809	0.041

Table 2: Performance and cost of SPF relative to those of the full grid-search in 4 adverse reverberant environments

Note: More data and a movie demonstrating how SPF works are available at [16].

5. CONCLUSION

We presented a new global maximum searching method, stochastic particle filtering (SPF) for SRP- β -PHAT, a more robust version of SRP-PHAT for single-source localization. Thorough experiments using real data and simulated data were done to test the new method. Under real, adverse conditions, SPF only requires on average 0.03% the cost of a full grid-search, which is fewer than other existing search methods, such as: spherical intersection (10% of a full search), time-delay inverse mapping (10%), CFRC and SRC (0.06%). While maintaining performance relative to that of a full search, SRP-PHAT using SPF would require only a 1.27 GF (GigaFlop) processor for real-time applications (51.2ms/estimate). This is feasible with a single floating-point DSP today, such as the Analog Devices’ 600MHz TigerSHARC processor (ADSP-TS201S), which can provide up to 1.2 GF.

6. REFERENCES

- [1] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, *Microphone Arrays: Techniques and Applications*, chapter Robust localization in reverberant rooms, pp. 157–180, Springer-Verlag, 2001.
- [2] S. T. Birchfield, “A unifying framework for acoustic localization,” in *Proc. of European Signal Processing Conference (EUSIPCO 2004)*, Vienna, Austria, Sept. 2004, pp. 1127–1130.
- [3] D. N. Zotkin and R. Duraiswami, “Accelerated speech source localization via a hierarchical search of steered response power,” *IEEE Trans. Speech, Audio Process.*, vol. 12, no. 5, pp. 499–508, Sept. 2004.
- [4] J. Peterson and C. Kyriakakis, “Hybrid algorithm for robust, real-time source localization in reverberant environments,” in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, PA, USA, Mar. 2005, vol. 4, pp. 1053–1056.
- [5] J. 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.
- [6] J. Peterson and C. Kyriakakis, “Analysis of fast localization algorithms for acoustical environments,” in *Proc. 39th Asilomar Conf. Signals, Syst., Comput.*, 2005, pp. 1385–1389.
- [7] 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. Acoust. Speech, Signal Process.*, Honolulu, Hawaii, Apr. 2007, vol. 1, pp. 121–124.
- [8] D. B. Ward and R. C. Williamson, “Particle filter beamforming for acoustic source localization in a reverberant environment,” in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, FL, USA, May 2002, vol. 2, pp. 1777–1780.
- [9] J. Vermaak and A. Blake, “Nonlinear filtering for speaker tracking in noisy and reverberant environments,” in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, UT, USA, May 2001, vol. 5, pp. 3021–3024.
- [10] K. D. Donohue, J. Hannemann, and H. G. Dietz, “Performance of phase transform for detecting sound sources with microphone arrays in reverberant and noisy environments,” *Signal Processing*, vol. 87, no. 7, pp. 1677–1691, July 2007.
- [11] C. H. Knapp and G. C. Carter, “The generalized correlation method for estimation of time delay,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.
- [12] D. Crisan and A. Doucet, “A survey of convergence results on particle filtering methods for practitioners,” *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 736–746, Mar. 2002.
- [13] D. B. Ward, E. A. Lehmann, and R. C. Williamson, “Particle filtering algorithms for tracking an acoustic source in a reverberant environment,” *IEEE Trans. Speech, Audio Process.*, vol. 11, no. 6, pp. 826–836, Nov. 2003.
- [14] J. F. Deneault and J. Rouat, “Cohesive particle filtering for sound source localization,” in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, Toulouse, France, May 2006, vol. 4, pp. 829–832.
- [15] H. Do and H. F. Silverman, “A fast microphone array srp-phat source location implementation using coarse-to-fine region contraction (cfrc),” in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA-07)*, New Paltz, NY, Oct. 2007, pp. 295–298.
- [16] Online appendix, “Demonstration movie, real data and simulation,” <http://www.llems.brown.edu/~hdo/spf.html>.