

# **UCLA**

## **Papers**

### **Title**

Particle Filtering Approach to Localization and Tracking of a Moving Acoustic Source in a Reverberant Room

### **Permalink**

<https://escholarship.org/uc/item/9xc0f566>

### **Authors**

Chen, Chiao-En  
Ali, Andreas M  
Lorenzelli, F.  
et al.

### **Publication Date**

2006-05-14

### **DOI**

10.1109/ICASSP.2006.1661102

Peer reviewed

# PARTICLE FILTERING APPROACH TO LOCALIZATION AND TRACKING OF A MOVING ACOUSTIC SOURCE IN A REVERBERANT ROOM

C. E. Chen\*, H. Wang†, A. Ali\*, F. Lorenzelli\*, R.E. Hudson\*, and K. Yao\*

Electrical Engineering Department\*, Computer Science Department†, UCLA  
Los Angeles, CA 90095, USA

## ABSTRACT

We propose a novel algorithm employing particle filters for acoustic source tracking in a reverberant environment. By incorporating the likelihood function computed through Approximate Maximum-Likelihood (AML) method, the proposed algorithm is applicable to wideband sources and can be implemented for multiple sources tracking. Both computer simulation and experimental results show the effectiveness of the proposed algorithm.

## 1. INTRODUCTION

There is great interest recently in applications such as camera steering for teleconferencing and acoustic beamforming. A critical technique is performing acoustic source tracking using microphone arrays. Traditional approach to this problem is to divide the recorded waveforms into frames, and perform source localization algorithms on a frame-by-frame basis. Tracking acoustic sources in reverberation environments is challenging for traditional approaches since the echoes from scatterers can generate spurious peaks in the likelihood function and lead to false location estimates.

Recently, particle filters have been proposed by researchers to address this problem. Particle filtering is a sequential Monte Carlo methodology where the idea is to recursively compute some probability density through importance sampling and approximation with some discrete random measures. It has not drawn much attention until the last decade due to the high computational complexity. Thanks to the rapid advances in semiconductor technology, particle filtering has now become an active area and is considered a promising approach in overcoming the inefficiency of traditional methods.

The intuition behind is that as the speaker moves around in the reverberant environment, only the peak of the likelihood function due to the true source location conforms to some temporal consistency, while the spurious peaks do not. By formulating the tracking problem as a sequential state estimation problem, the particle filtering approach takes not only

current but all the past observations into consideration, and therefore has better immunity to reverberations.

AML location estimator is a near-field, wideband source localization algorithm [1]. It has been shown to yield superior performance over other techniques such as wideband MUSIC and two-step least-squares method. In this work, we follow the framework proposed by Ward [2] and introduce the AML algorithm to eliminate the difficulties caused by reverberations. The paper is organized as follows. Section 2 reviews the near-field AML algorithm. In section 3, we give a short background introduction to particle filtering technique. In section 4 we introduce the sequential AML source tracking algorithm. Results from computer simulations and real audio measurement are presented in section 5.

## 2. NEAR-FIELD AML ALGORITHM

AML algorithm for wideband source localization extends the classical narrowband maximum likelihood DOA estimator by combining the AML metric results of each subband. In tracking a moving speaker the signal strength at each microphone varies as a function of source locations, and therefore is considered as a near-field localization problem. For a randomly distributed array of  $P$  sensors, the data collected by the  $p$ th sensor at time  $n$  can be given by

$$x_p(n) = \sum_{m=1}^M a_p^{(m)} S^{(m)}(n - t_p^{(m)}) + w_p(n), \quad (1)$$

for  $n = 0, \dots, N - 1$ ,  $p = 1, \dots, P$ , and  $m = 1, \dots, M$ , where  $S^{(m)}$  is the  $m$ th source signal,  $t_p^{(m)}$  is the time delay of the  $m$ th source to the  $p$ th sensor, and  $w_p$  is modeled as zero mean white Gaussian noise with variance  $\sigma^2$ .  $a_p^{(m)}$  is the signal gain and the time delay can be computed by  $t_p^{(m)} = \|\mathbf{r}_{s_m} - \mathbf{r}_p\|/v$ , where  $\mathbf{r}_{s_m}$  is the location of the  $m$ th source,  $\mathbf{r}_p$  is the location of the  $p$ th sensor, and  $v$  is the speed of propagation.

In the frequency domain, the array signal model is given by

$$\mathbf{X}(k) = \mathbf{D}(k)\mathbf{S}(k) + \boldsymbol{\eta}(k), \quad (2)$$

This work was partially supported by NSF CENS program, NSF grant EF-0410438, AROD-MURI PSU Contract 50126, and ST Microelectronics, Inc.

for  $k = 0, \dots, N-1$ , where the array data spectrum, the steering matrix, source spectrum and noise spectrum are given by  $\mathbf{X}(k) = [X_1(k), \dots, X_R(k)]^T$ ,  $\mathbf{D}(k) = [\mathbf{d}^{(1)}(k), \dots, \mathbf{d}^{(M)}(k)]$ ,  $\mathbf{S}(k) = [S^{(1)}(k), \dots, S^{(M)}(k)]^T$  and  $\eta(k)$  respectively. The steering vector can be expressed as

$$\mathbf{d}^{(m)}(k) = [d_1^{(m)}(k), \dots, d_P^{(m)}(k)]^T,$$

where each element is given by  $d_p^{(m)}(k) = a_p^{(m)} e^{-j2\pi k t_p^{(m)}/N}$ ,

Under the zero mean iid Gaussian assumption of the noise, the maximum likelihood estimation of the source locations and source signals are given by maximizing the log-likelihood function,  $l(\Theta)$  given by

$$\hat{\Theta} = \max_{\Theta} l(\Theta) = \max_{\Theta} \left\{ - \sum_{k=1}^{N/2} \|\mathbf{X}(k) - \mathbf{D}(k)\mathbf{S}(k)\|^2 \right\}, \quad (3)$$

where  $\Theta = [\tilde{\mathbf{r}}_s^T, \mathbf{S}^{(1)T}, \dots, \mathbf{S}^{(M)T}]^T$  and  $\tilde{\mathbf{r}}_s = [r_{s_1}^T, \dots, r_{s_M}^T]^T$ . By the technique of separating variables, the AML source locations estimate can be obtained by maximizing the ML metric  $J(\tilde{\mathbf{r}}_s)$  given by

$$J(\tilde{\mathbf{r}}_s) = \sum_{k=1}^{N/2} \|\mathbf{P}(k, \tilde{\mathbf{r}}_s)\mathbf{X}(k)\|^2. \quad (4)$$

where  $\mathbf{P}(k, \tilde{\mathbf{r}}_s) = \mathbf{D}(k)(\mathbf{D}(k)^H\mathbf{D}(k))^{-1}\mathbf{D}(k)^H$ . For more information about the AML refer to references [1, 3, 4].

### 3. SEQUENTIAL BAYESIAN ESTIMATION USING PARTICLE FILTERS

Tracking of an acoustic source can be formulated as a sequential state estimation problem under the Bayesian framework. Define the source state and the observation at time  $t$  as  $\alpha_t$ , and  $y_t$ , and assume they follow the state dynamic equation (5) and the observation equation (6).

$$\alpha_t = f_t(\alpha_{t-1}, w_t), \quad (5)$$

$$y_t = g_t(\alpha_t, v_t) \quad (6)$$

Let  $y_{1:t} = [y_1, \dots, y_t]$  denote the concatenation of all measurements up to time  $t$ , and assume states follow a first order Markov process, and observations are independent given the states. Then the posterior density  $p(\alpha_t|y_t)$  can be computed recursively through the following equations.

$$p(\alpha_k|y_{1:k}) = \frac{p(y_k|\alpha_k)p(\alpha_k|y_{1:k-1})}{p(y_k|y_{1:k})}, \quad (7)$$

$$p(y_k|y_{1:k-1}) = \int p(y_k|\alpha_k)p(\alpha_k|y_{1:k-1})d\alpha_k, \quad (8)$$

where  $p(\alpha_k|y_{1:k-1}) = \int p(\alpha_k|\alpha_{k-1})p(\alpha_{k-1}|y_{1:k-1})d\alpha_{k-1}$  is the prior and  $p(y_k|\alpha_k)$  is the likelihood function.

In general, no close form solution exists for (7). Particle filtering recursively computes a set of random measures to represent an unknown probability distribution. The random measures are composed of weighted particles, where the particles are samples of the unknown states from the state space, and the particle weights are computed by using Bayes theory described in this section. As the number of particles becomes very large, it forms an equivalent representation of the true posterior density. Given these particles, statistical estimates of the system state can be calculated. For more information about particle filtering, please refer to [5, 6].

## 4. TRACKING A MOVING ACOUSTIC SOURCE IN A REVERBERANT ROOM

To implement a generic particle filtering algorithm for acoustic source tracking, one needs to determine two kinds of models: 1)source dynamic model; 2)likelihood function. In this paper, we use the well-known Langevin's model as our source dynamic model, and apply the AML algorithm to compute the likelihood function.

### 4.1. Source dynamic model

Several possible source dynamic models have been adopted in the literature [7, 8] to model the dynamics of a person's movement. One popular choice is the Langevin's model which has been shown to work well in practice. Movements in X and Y coordinates are assumed independent in this model. Define the source state as  $\alpha = [\alpha_x, \alpha_y, v_x, v_y]$ , where  $(\alpha_x, \alpha_y)$  is the source location and  $(v_x, v_y)$  is the source velocity. Then the source motion in each coordinate (here we only show the x-coordinate for example), is described by:

$$\begin{bmatrix} \alpha_x(t) \\ v_x(t) \end{bmatrix} = \begin{bmatrix} 1 & NT_s \\ 0 & a_x \end{bmatrix} \begin{bmatrix} \alpha_x(t-1) \\ v_x(t-1) \end{bmatrix} + \begin{bmatrix} 0 \\ b_x \end{bmatrix} F_x, \quad (9)$$

$$a_x = e^{-\beta_x NT_s}, \quad (10)$$

$$b_x = v_x \sqrt{1 - a_x^2}, \quad (11)$$

where  $F_x = \mathcal{N}(0, 1)$  is a Gaussian random variable,  $N$  is number of samples per frame in the localization algorithm, and  $v_x$  is the velocity.

### 4.2. Likelihood function of AML algorithm

The AML estimator described in (4) in general has no close form solutions. One needs to perform exhaustive grid search to compute the ML metric  $J(\tilde{\mathbf{r}}_s)$  over the area of interest. The likelihood function we propose here belongs to the category of so-called "pseudo-likelihood function" of a "direct localization function" in Ward's work [2]. But instead of taking  $r$ -th power of the localization function (here, the  $J$  function) as

in [2], it makes more sense in our application to formulate the likelihood function,  $\mathcal{L}(\hat{\mathbf{r}}_s)$  in exponential form, since AML is derived under Gaussian noise assumption. We define the likelihood function as  $\mathcal{L}(\hat{\mathbf{r}}_s) \propto \exp\{CJ(\hat{\mathbf{r}}_s)\}$ , where  $C$  is a parameter one chooses to shape the likelihood function.

We would like to highlight the fact that the AML algorithm has the capability to localize multiple sources. By incorporating the likelihood model derived from the AML algorithm into the particle filtering framework, the tracking algorithm derived in this work can be directly implemented to track multiple sources.

### 4.3. Implementation of proposed algorithm

The procedure described below summarizes the implementation of our algorithm.

Form an initial set of particles  $\{\alpha_0^{(i)}, i = 1 : K\}$  with weights  $\{w_0^{(i)}, i = 1 : K\}$ . Then, as each new frame of data is received:

1. Resample the particles from the previous frame  $\{\alpha_{t-1}^{(i)}, i = 1 : K\}$  according to their weights  $\{w_{t-1}^{(i)}, i = 1 : K\}$  to form the re-sampled set of particles  $\{\tilde{\alpha}_{t-1}^{(i)}, i = 1 : K\}$  that have uniform weights  $\{\tilde{w}_{t-1}^{(i)} = 1/K, i = 1 : K\}$ ;
2. Predict the new set of particles  $\{\alpha_t^{(i)}, i = 1 : K\}$  by propagating the re-sampled set of particles using the state dynamic equation (Eq. 9).
3. Compute the AML metric of the  $t$ th frame through Eq. 4.
4. Compute  $\mathcal{L}(\mathbf{r}_s)$  through Eq. ?? and assign new weighting as  $w_t^{(i)} = \mathcal{L}(\mathbf{r}_{s,t}^{(i)}) / \sum_{i=1}^K \mathcal{L}(\mathbf{r}_{s,t}^{(i)})$
5. Compute the current target location estimate  $\hat{\mathbf{r}}_{s,t}$  as the expectation of  $\mathbf{r}_{s,t}$  with respect to the posterior target state distribution,

$$\hat{\mathbf{r}}_{s,t} = \sum_{i=1}^K w_t^{(i)} \mathbf{r}_{s,t}^{(i)}. \quad (12)$$

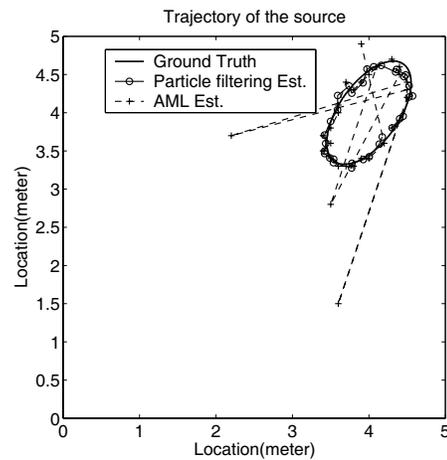
6. Store the particles and their respective weights  $\{\alpha_t^{(i)}, w_t^{(i)} = 1/K, i = 1 : K\}$ , back to step 1, when the next frame of data arrives.

## 5. SIMULATION AND EXPERIMENTAL RESULTS

In this section, we present results of tracking an acoustic source in a reverberant room. Frame-by-frame AML localization algorithm and particle filtering approach are performed for comparison. Both simulation and experimental results show the performance improvement offered by the particle filtering approach.

### 5.1. Simulations results

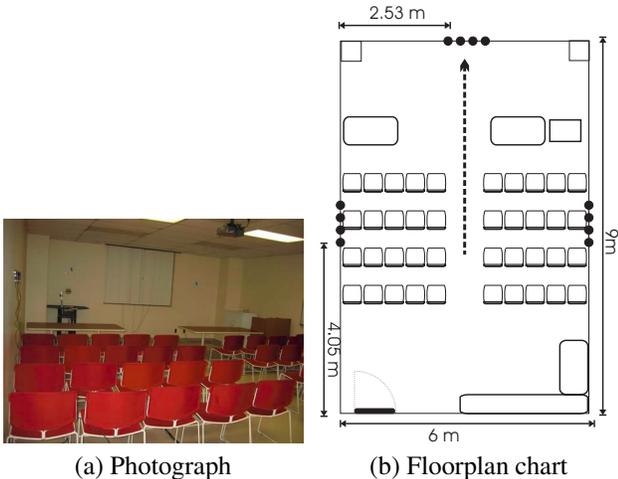
Simulation is conducted in a 5 meter by 5 meter reverberant room. The impulse response of the room reverberation is simulated using the image method [9], up to fourth-order reflections with reflectivity=0.7. We also assume the signal gain level is proportional to the inverse of the distance of signal/image to the microphone, and is known in the AML algorithm. A sound file of a male speaker is used as the audio source, with 16 bit resolution and 8kHz sampling rate. Four microphones are placed at the corners of the simulated 2D room, and the AML algorithm is performed frame by frame (1024 samples/frame). As we can see from Fig. 1, the reverberation itself can cause false localization even in a noise-free environment. Simulation result shows the performance improvement of particle filtering approach using  $K=100$  particles compared to the conventional frame-by-frame AML source localization under reverberation.



**Fig. 1.** Simulation results of tracking an acoustic source in a 2D square room.  $K=100$  in the particle filtering approach.

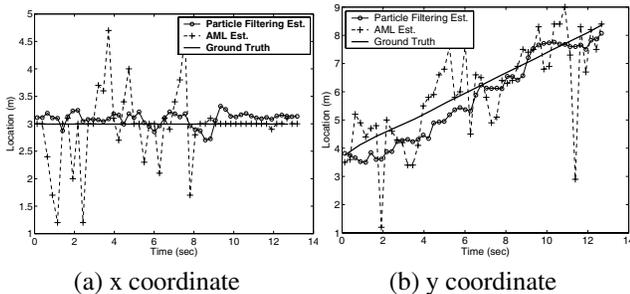
### 5.2. Experimental results

In this subsection, we present the tracking results of a moving speaker in a real reverberant room. The experiment is performed in a rectangular conference room (Fig. 2a) of dimension 6m x 9m. We intentionally retained the chairs, tables and podium in the experiment to make the environment more realistic. Twelve microphones are used in this experiment, and the recordings are synchronized by two Presonus Firepod 8 channel 94 bit/96 kb firewire-based recording systems. The microphones are placed 1.6 meter from the ground, which are approximately the same height as the voice source and placed over paddings to avoid vibrations from the building before attaching to the walls. A person starts from the middle of the room and speaks while moving in the y-direction. The source trajectory and the floorplan of the room is shown in Fig. 2b.



**Fig. 2.** Photograph and floorplan chart of the conference room. Dotted line in b) indicates the moving trajectory of the speaker. The microphone spacing in each subarray is 0.3 meter.

The recorded data is first downsampled from 44.1 kHz to 16kHz, and the AML algorithm is applied at a frame rate of 2048 samples/frame. In the previous derivation of the AML algorithm, we always assume the signal gain level is known, which is hard to obtain in practice. To avoid this problem, we normalize the power of the 12 recorded waveforms to make them of equal power within each frame before processing, and set  $a_p^{(m)} = 1$  in the AML algorithm. Another advantage of the AML lies in the fact that we can match our algorithm to frequencies of human vocalization by simply choosing corresponding frequency bins in computing the AML metric. In this work, we apply a spectral mask (from 250 Hz to 2kHz), and choose 60 frequency bins that have highest power inside this mask. Tracking results of the particle filtering approach in Fig. 3 uses  $K=200$  particles. The frame-by-frame AML algorithm yields poor location estimates although it does follow the right trend of true trajectory. Proposed particle filtering approach significantly reduces localization error as is predicted by the computer simulation.



**Fig. 3.** Tracking results of an acoustic source in a reverberant room.  $K=200$  in the particle filtering approach.

## 6. CONCLUSION

In this paper, we proposed a novel algorithm for wideband source tracking in a reverberant environment. Both computer simulations and measurement results from a real conference room setting demonstrate the effectiveness of the proposed algorithm. The proposed algorithm is general and can be extended for multiple sources tracking.

## 7. REFERENCES

- [1] J. Chen, R.E. Hudson, and K. Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," *IEEE Trans. Signal Processing*, vol. 50, no. 8, pp. 1843–1854, August 2002.
- [2] D. Ward, E. Lehmann, and R. Williamson, "Particle filtering algorithms for tracking an acoustic source in reverberant environment," *IEEE Trans. Speech Audio Proc.*, vol. 11, no. 6, pp. 826–836, Nov 2003.
- [3] J. Chen, K. Yao, and R.E. Hudson, "Source localization and beamforming," *IEEE Signal Processing Magazine*, vol. 19, pp. 30–39, March 2002.
- [4] J. Chen, L. Yip, J. Elson, H. Wang, D. Maniezzo, R.E. Hudson, K. Yao, and D. Estrin, "Coherent acoustic array processing and localization on wireless sensor network," *Proc. of the IEEE*, no. 8, pp. 1154–1162, August 2003.
- [5] J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, P.M. Djuric, J.H. Kotecha, and J. Miguez, "Particle filtering," *IEEE Signal Processing Magazine*, vol. 20, pp. 19–38, September 2003.
- [6] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking," *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 174–188, February 2001.
- [7] X. Sheng and Y.H. Hu, "Sequential acoustic energy based source localization using particle filter in a distributed sensor network," in *Proc. ICASSP'04*, Canada, May 2004.
- [8] J. Vermaak and A. Blake, "Nonlinear filtering for speaker tracking in noisy and reverberant environments," in *Proc. ICASSP'01*, Salt Lake City, UT, May 2001.
- [9] J.B. Allen and D.A. Berkeley, "Image method for efficiently simulating small-room acoustics," *JASC*, vol. 65, pp. 943–950, 1979.