Speaker Localization with Moving Microphone Arrays

(Invited Paper)

Yuval Dorfan, Christine Evers∗, Sharon Gannot and Patrick A. Naylor∗
Faculty of Engineering, Bar-Ilan University, Ramat-Gan, 5290002, Israel
dorfany@gmail.com; sharon.gannot@biu.ac.il
∗Department of Electrical and Electronic Engineering, Imperial College, London, SW7 2AZ, UK
c.evers@imperial.ac.uk; p.naylor@imperial.ac.uk

***Abstract*—Speaker localization algorithms often assume static location for all sensors. This assumption simplifies the models used, since all acoustic transfer functions are linear time invariant. In many applications this assumption is not valid. In this paper we address the localization challenge with moving microphone arrays. We propose two algorithms to find the speaker position. The first approach is a batch algorithm based on the maximum likelihood criterion, optimized via expectation-maximization iterations. The second approach is a particle filter for sequential Bayesian estimation. The performance of both approaches is evaluated and compared for simulated reverberant audio data from a microphone array with two sensors.**

I. Introduction

Localization using static sensors has been dealt with theoretically and practically for various signal processing applications including passive or active radio detection and ranging (RADAR). Passive sonar using moving hydrophones has been dealt for years [1]. In particular, approaches for acoustic sensors deal with the specific challenges of reverberation, see, e.g., [2]–[5]. In general, most sound source localization approaches in the literature utilize either time of arrival (TOA) [6], time difference of arrivals (TDOAs) or direction of arrivals (DOAs) as measurements of the source in order to reconstruct the Cartesian source position. A common assumption is that the acoustic sensor is stationary and that its position is known. Nonetheless, spatial diversity of an acoustic sensor installed on a moving platform could be exploited for improved inference of the source position. Moving microphone arrays are particularly suitable for the field of robot audition [7], where microphone arrays can be installed in the limbs and head of an autonomous robot.

The movement of microphone arrays is particularly useful in situations where the sensor moves faster than the sources. In this case, the displacement of the sensors over time can be interpreted as a synthetic widening of the array aperture in space. This interpretation was first implemented for synthetic aperture radar (SAR) [8]. We implement this principle for localizing the coordinates of a source with a single pair of microphones.

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

In this paper, we address the challenge of sound source localization from a moving platform by considering and comparing two philosophically different approaches.

A maximum likelihood (ML) technique is presented that iteratively estimates the source positions from a batch of measurements. ML estimation procedures for localization [9], [10] are usually characterized by high computational complexity and by the nonexistence of closed-form solutions. The iterative expectation-maximization (EM) [11] or recursive EM (REM) procedures can be applied to maximize the likelihood instead. The first version of the incremental distributed expectation-maximization (IDEM) algorithm [12] solves a static localization problem. In this paper it is used to incorporate a dynamic problem.

In practice, applications such as robot audition often require near real-time processing, such that sound sources must be localized from short frames of audio data. We decided to adapt a Bayesian technique to sequentially estimate the source positions from on-line data. Bayesian estimation [3], [13] considers not only the likelihood of the desired random variables, but also incorporates prior information by modeling belief about the dynamics of the sources. The posterior probability density function (p.d.f.) is therefore maximized instead of the likelihood. In this paper, the prior imposes the static location of the source. The resulting Maximum *a posteriori* (MAP) estimator can therefore be considered as a penalized ML approach. In this paper, a particle filter is proposed for sequential sound source localization.

For both approaches, we assume that only a single, static source is localized and that the trajectory and positions of the microphones are known *a priori*.

The remainder of this paper is organized as follows. The general model of the problem is in Section II. A description of the dynamic IDEM is given in Section III. The Bayesian approach using a particle filter is presented in IV. Section V is dedicated to simulation results. Conclusions are drawn in Section VI.

II. Signal Model

In the following section models for the source and sensor dynamics are defined. Furthermore, the measurement model

of the source is presented which will be used for the sound source localization algorithms in Sections III and IV.

A. Stationary source dynamics

The source position, $\mathbf{p}_t \triangleq \left[x_s(t), y_s(t)\right]^T$, is defined as the absolute two-dimensional Cartesian position of the source within the room. In this paper, a static source is assumed, i.e.,

$$\mathbf{p}_0 = \cdots = \mathbf{p}_t = \mathbf{p}. \tag{1}$$

B. Measurement model

The microphone array used in this paper consists of one microphone pair with two-dimensional Cartesian positions, $\mathbf{p}^m(t) \triangleq \left[x^m(t), y^m(t)\right]^T$ for $m = 1, 2$, and where the platform moves with speed, $v(t)$. The localization procedure starts with a pair-wise relative phase ratio (PRP) extraction [2]:

$$\phi(t,k) \triangleq \frac{z^2(t,k)|z^1(t,k)|}{z^1(t,k)|z^2(t,k)|}, \tag{2}$$

where $z^m(t,k)$ is the short-time Fourier transform (STFT) of the mth microphone signal. The time and frequency indices are $t = 1, \ldots, T$ and $k = 0, \ldots, K-1$, respectively.

These PRPs are induced by the TDOA, which can be defined as:

$$\tau(\mathbf{p}, \mathbf{p}^1(t), \mathbf{p}^2(t)) \triangleq \frac{||\mathbf{p} - \mathbf{p}^2(t)|| - ||\mathbf{p} - \mathbf{p}^1(t)||}{c}, \tag{3}$$

where $||\cdot||$ denotes the Euclidean norm, and c is the sound velocity.

We model the PRPs using a Gaussian mixture model (GMM):

$$\phi(t,k) \sim \sum_{\mathbf{p}} \psi_{\mathbf{p}} \mathcal{N}^c\left(\phi(t,k); \tilde{\phi}^k(\mathbf{p},t), \sigma^2\right), \tag{4}$$

where $\psi_{\mathbf{p}}$ is the probability that the speaker emitting at time t and frequency k is located at position $\mathbf{p}$. $\mathcal{N}^c(\cdot;\cdot,\cdot)$ denotes the complex-Gaussian p.d.f. with variance σ^2. In our model, the variance is fixed and chosen empirically. The mean of each Gaussian can be calculated in advance on a grid of all possible locations:

$$\tilde{\phi}^k(\mathbf{p},t) \triangleq \exp\left(-j\frac{2\pi k \tau(\mathbf{p}, \mathbf{p}^1(t), \mathbf{p}^2(t))}{KT_s}\right) \tag{5}$$

$\forall \mathbf{p} \in \mathcal{P}$, where T_s denotes the sampling period and $\mathcal{P}$ being the set of all possible locations.

III. The EM localization algorithm

Based on [14] an EM localization algorithm has been suggested in [2] with a vector of PRP measurements. In [15] IDEM has been presented for the same localization problem. Following [15] and [2], we present an algorithm for the moving microphones case.

A. Maximum likelihood for dynamic localization

The joint p.d.f. of the PRPs in (2), assuming independence along time and frequency indexes, is given by:

$$f(\mathbf{\Phi} = \boldsymbol{\phi}; \boldsymbol{\psi}) = \prod_{t,k} \sum_{\mathbf{p}} \psi_{\mathbf{p}} \mathcal{N}^c\left(\phi(t,k); \tilde{\phi}^k(\mathbf{p},t), \sigma^2\right), \tag{6}$$

where $\boldsymbol{\psi} = \text{vec}_{\mathbf{p}}\left(\psi_{\mathbf{p}}\right)$ and $\boldsymbol{\phi} = \text{vec}_{t,k}\left(\phi(t,k)\right)$.

The ML estimate of the source locations is given by:

$$\hat{\boldsymbol{\psi}} = \underset{\boldsymbol{\psi}}{\text{argmax}}[\log f(\mathbf{\Phi} = \boldsymbol{\phi}; \boldsymbol{\psi}) \quad \text{s.t.} \sum_{\mathbf{p} \in \mathcal{P}} \psi_{\mathbf{p}} = 1 \text{ and } 0 < \psi_{\mathbf{p}} < 1]. \tag{7}$$

B. Hidden variables

The hidden variables, $y(t,k,\mathbf{p})$ are defined as the association of each measurement with a source at position $\mathbf{p}$. Let $\mathbf{y} = \text{vec}_{t,k,\mathbf{p}}\left(y(t,k,\mathbf{p})\right)$ be the vector concatenation of the hidden variables. The p.d.f. of $\mathbf{y}$ is given by:

$$f(\mathbf{Y} = \mathbf{y}; \boldsymbol{\psi}) = \prod_{t,k} \sum_{\mathbf{p}} \psi_{\mathbf{p}} y\left(t,k,\mathbf{p}\right). \tag{8}$$

Given the hidden variables, the p.d.f. of the observations is:

$$f(\mathbf{\Phi} = \boldsymbol{\phi}|\mathbf{y}; \boldsymbol{\psi}) = \prod_{t,k} \sum_{\mathbf{p}} y\left(t,k,\mathbf{p}\right) \times \mathcal{N}^c\left(\phi(t,k); \tilde{\phi}^k(\mathbf{p},t), \sigma^2\right). \tag{9}$$

The p.d.f. of the *complete data* can be deduced from (8)-(9):

$$f(\mathbf{\Phi} = \boldsymbol{\phi}, \mathbf{Y} = \mathbf{y}; \boldsymbol{\psi}) = \prod_{t,k} \sum_{\mathbf{p}} \psi_{\mathbf{p}} y\left(t,k,\mathbf{p}\right) \times \mathcal{N}^c\left(\phi(t,k), \tilde{\phi}^k\left(\mathbf{p},t\right), \sigma^2\right). \tag{10}$$

C. The IDEM algorithm

The original IDEM algorithm is capable of detecting the number of active sources (including the detection of no activity) and their locations for static scenarios. The IDEM is applied here for moving sensors. Thanks to the dynamics of the sensors, we can use here only a single node.

The *E-step* can be stated as:

$$Q(\boldsymbol{\psi}|\hat{\boldsymbol{\psi}}^{(\ell-1)}) \triangleq E\left\{\log\left(f(\mathbf{\Phi} = \boldsymbol{\phi}, \boldsymbol{Y} = \mathbf{y}; \boldsymbol{\psi})\right)|\boldsymbol{\phi}; \hat{\boldsymbol{\psi}}^{(\ell-1)}\right\} \tag{11}$$

$$= \sum_{t,k,\mathbf{p}} E\left\{y(t,k,\mathbf{p})|\phi(t,k); \hat{\boldsymbol{\psi}}^{(\ell-1)}\right\} \cdot \left[\log\psi_{\mathbf{p}} + \log\mathcal{N}^c(\phi(t,k); \tilde{\phi}^k(\mathbf{p},t), \sigma^2)\right],$$

which, in our case, simplifies to:

$$\upsilon^{(\ell)}\left(t,k,\mathbf{p}\right) \triangleq E\left\{y\left(t,k,\mathbf{p}\right)|\phi(t,k); \hat{\boldsymbol{\psi}}^{(\ell-1)}\right\} = \frac{\hat{\psi}_{\mathbf{p}}^{(\ell-1)} \mathcal{N}^c\left(\phi(t,k); \tilde{\phi}^k\left(\mathbf{p},t\right), \sigma^2\right)}{\sum_{\mathbf{p}'} \hat{\psi}_{\mathbf{p}'}^{(\ell-1)} \mathcal{N}^c\left(\phi(t,k); \tilde{\phi}^k\left(\mathbf{p}',t\right), \sigma^2\right)}. \tag{12}$$

The IDEM applies the E-step, followed by the M-step, as summarized in Algorithm 1.

Algorithm 1: Dynamic IDEM localization.

input $z^1(t,k)$, $z^2(t,k)$;.
Calculate $\phi(t,k)$ using (2).
set $\tilde{\phi}^k(\mathbf{p},t)$ using (5).
init $\hat{\psi}_{\mathbf{p}}^{(-1)}$ to uniform p.d.f..
Calculate $\upsilon^{(0)}(t,k,\mathbf{p})$ using (12).
Calculate their mean: $\hat{\psi}_{\mathbf{p}}^{(0)} = \frac{\sum_{t,k}\upsilon^{(0)}(t,k,\mathbf{p})}{T\cdot K}$.
for $\ell = 1$ **to** L **do**
  E-step
  Calculate $\upsilon^{(\ell)}(t,k,\mathbf{p})$ using (12).
  M-step
  Calculate $\hat{\psi}_{\mathbf{p}}^{(\ell)} = \frac{\sum_{t,k}\upsilon^{(\ell)}(t,k,\mathbf{p})}{T\cdot K}$.
end
output $\hat{\psi}_{\mathbf{p}}^{(L)}, \upsilon^{(L)}(t,k,\mathbf{p})$.

IV. Bayesian filter

As discussed in the previous section, ML estimation infers knowledge about the source position from the observations only (see (7)). As only knowledge about the measured data is taken into account, ML estimators are based on purely objective observations. Prior belief about the source position can also be utilized when considering a Bayesian framework.

Under the Bayesian paradigm the desired source position, $\mathbf{p}$, is considered as a state. Estimates of $\mathbf{p}$ can hence be obtained by construction of the posterior p.d.f. of the states, $f_t(\mathbf{p}|\boldsymbol{\phi}_{1:t})$, given the PRPs, $\boldsymbol{\phi}_{1:t} \triangleq \left[\boldsymbol{\phi}_1^T,\ldots,\boldsymbol{\phi}_t^T\right]^T$ where $\boldsymbol{\phi}_t \triangleq \left[\phi(t,1),\ldots,\phi(t,K)\right]^T$ which is related to the likelihood, $f(\boldsymbol{\phi}_t|\mathbf{p})$, via Bayes's theorem:

$$f_t(\mathbf{p}|\boldsymbol{\phi}_{1:t}) = \frac{f(\boldsymbol{\phi}_t|\mathbf{p})\, f_{t|t-1}(\mathbf{p}|\boldsymbol{\phi}_{1:t-1})}{f(\boldsymbol{\phi}_t)}, \tag{13}$$

where the instantaneous likelihood, $f\left(\boldsymbol{\phi}_t\middle|\,\mathbf{p}\right)$ is modelled similar to (6) by assuming independence of PRPs in time and frequency:

$$f\left(\boldsymbol{\phi}_t\middle|\,\mathbf{p}\right) = \prod_{k=1}^{K}\mathcal{N}^c\left(\phi(t,k);\tilde{\phi}^k(\mathbf{p},t),\,\sigma^2\right), \tag{14}$$

where σ^2 is the measurement noise variance. Furthermore, $f_{t|t-1}(\mathbf{p}|\boldsymbol{\phi}_{1:t-1})$ in (13) is the predicted p.d.f. given by:

$$f_{t|t-1}(\mathbf{p}|\boldsymbol{\phi}_{1:t-1}) = \int_{\mathbb{R}^2} f(\mathbf{p})\, f_{t-1}(\mathbf{p}|\boldsymbol{\phi}_{1:t-1})d\mathbf{p}, \tag{15}$$

where $f(\mathbf{p})$ is the prior p.d.f. capturing the static nature of the source and $f_{t-1}(\mathbf{p}|\boldsymbol{\phi}_{1:t-1})$ is the posterior p.d.f. at time $t-1$.

To sequentially obtain the optimal value of $\mathbf{p}$ at each time, t, MAP estimates can be evaluated by maximization with respect to the variables of interest, i.e.,

$$\hat{\mathbf{p}} \triangleq \underset{\mathbf{p}}{\operatorname{argmax}}\, f(\mathbf{p}|\boldsymbol{\phi}_{1:t}). \tag{16}$$

A. Sequential importance sampling

To impose real-valued source states despite the complex observations, sequential importance sampling [16] is used in this paper. The posterior at $t-1$ is approximated by:

$$f_{t-1}(\mathbf{p}|\boldsymbol{\phi}_{1:t-1}) = \sum_{j=1}^{J_{t-1}} w_{t-1}^{(j)}\,\delta_{\hat{\mathbf{p}}^{(j)}}(\mathbf{p}), \tag{17}$$

where $\delta_{\hat{\mathbf{p}}^{(j)}}(\mathbf{p})$ denotes the Dirac measure of random variable, $\mathbf{p}$, centered on particle $\hat{\mathbf{p}}^{(j)}$. Inserting (17) into (15) yields:

$$f_{t|t-1}(\mathbf{p}|\boldsymbol{\phi}_{1:t-1}) = \int_{\mathbb{R}^2} f(\mathbf{p}) \sum_{j=1}^{J_{t-1}} w_{t-1}^{(j)}\,\delta_{\hat{\mathbf{p}}^{(j)}}(\mathbf{p})\,d\mathbf{p} \tag{18}$$

In order to capture the static nature of the source whilst modelling uncertainty in the particles, $\hat{\mathbf{p}}^{(j)}$, the prior, $f(\mathbf{p})$, is approximated by drawing P importance samples for each particles, $\hat{\mathbf{p}}^{(j)}$, from the proposal distribution,

$$\pi\left(\mathbf{p}|\hat{\mathbf{p}}^{(j)}\right) = \mathcal{N}\left(\mathbf{p};\hat{\mathbf{p}}^{(j)},\,\mathbf{Q}\right) \tag{19}$$

where the covariance, $\mathbf{Q}$, allows for deviations of the new particles via (19), $\hat{\mathbf{p}}^{(j,p)}$, from the old particles, $\hat{\mathbf{p}}^{(j)}$.

Using (18) and (13), the posterior p.d.f. of the states, $\mathbf{p}$, can hence be expressed as

$$f_t(\mathbf{p}|\boldsymbol{\phi}_{1:t}) = \sum_{j=1}^{J_{t-1}}\sum_{p=1}^{P} w_t^{(j,p)}\delta_{\hat{\mathbf{p}}^{(j,p)}}(\mathbf{p}), \tag{20}$$

with weights:

$$w_t^{(j,p)} = \tilde{w}_t^{(j,p)} \Bigg/ \sum_{j=1}^{J_{t-1}}\sum_{p=1}^{P}\tilde{w}_t^{(j,p)}, \tag{21}$$

where the unnormalized weights, $\tilde{w}_t^{(j,p)}$, are defined as:

$$\tilde{w}_t^{(j,p)} \triangleq w_{t-1}^{(j)}\, f(\boldsymbol{\phi}_t|\hat{\mathbf{p}}^{(j,p)}). \tag{22}$$

The point estimate of the source position at each t is extracted as the weighted average of the particles,

$$\tilde{\mathbf{p}} = \sum_{j=1}^{J}\sum_{p=1}^{P} w_t^{(j,p)}\,\hat{\mathbf{p}}^{(j,p)} \Bigg/ \sum_{j=1}^{J}\sum_{p=1}^{P} w_t^{(j,p)}. \tag{23}$$

In order to avoid an explosion of the number of particles, systematic resampling to $J_{\max}$ particles is applied to the particle cloud, $\left\{\hat{\mathbf{p}}^{(j,p)} : j=1,\ldots,J\,;\,p=1,\ldots,P\right\}$, after each recursion [17]. The Bayesian algorithm is summarised in Algorithm 2.

V. Simulation study and performance measures

A. Simulation setup

To evaluate the performance of the algorithms, audio data was generated using the following simulation.

The origin of the microphone array, $\mathbf{p}^0(t) = \left[x^0(t), y^0(t)\right]^T$, is generated using a constant velocity model where

$$\mathbf{p}^0(t) = \mathbf{F}(t)\,\mathbf{p}^0(t-1) + \mathbf{n}_{\mathbf{p}}(t), \tag{24}$$

Algorithm 2: Particle filter source tracker

Input PRPs, $\left\{\tilde{\phi}^k(\mathbf{p},t)\right\}_{k=1}^{K}$
for $j=1$ **to** J **do**
  for $p=1$ **to** P **do**
    Sample $\hat{\mathbf{p}}^{(j,p)}$ from (19)
    Evaluate $\tilde{w}_t^{(j,p)}$ from (22)
  end
end
Normalize weights, $w_t^{(j,p)}$, from (21)
Re-sample $\hat{\mathbf{p}}^{(j,p)}$ [17]
Extract point estimate, $\tilde{\mathbf{p}}$ (23)
Output Cartesian source position, $\tilde{\mathbf{p}}$.

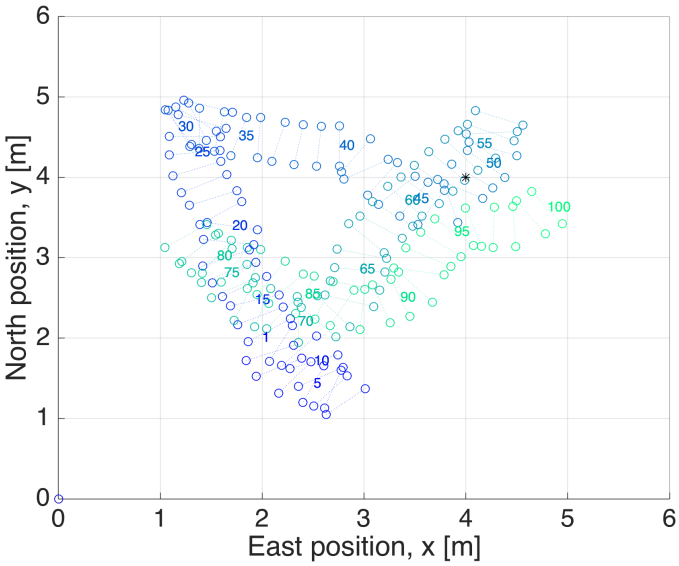

Fig. 1. Scenario of $6 \times 6 \times 2.5$ m room, with source (black asterisk) at $(4, 4, 1.5)$ m and moving sensor with initial position at $(2, 2, 1.5)$ m. The colour code represents a continuum of colors from blue at $t = 1$ to green at $t = 100$.

where $\mathbf{n}_{\mathbf{p}}(t) \sim \mathcal{N}\left(\mathbf{0}_{2\times 1}, \mathbf{\Sigma}^0(t)\right)$ is the process noise with covariance, $\mathbf{\Sigma}^0(t)$. The matrix $\mathbf{F}(t)$ captures the dynamic model, defined in this paper as a constant velocity model given by:

$$\mathbf{F}(t) = \begin{bmatrix} 1 & 0 & \Delta_T\, v(t)\, \sin\gamma(t) \\ 0 & 1 & \Delta_T\, v(t)\, \cos\gamma(t) \end{bmatrix}, \tag{25}$$

where Δ_T is the time delay between $t-1$ and t, $v(t)$ is the velocity the moving platform, and where the array orientation, $\gamma(t)$, is given by the random walk:

$$\gamma(t) = \gamma(t-1) + v_\gamma(t), \quad v_\gamma(t) \sim \mathcal{N}\left(0, \sigma^2_{v_\gamma}(t)\right). \tag{26}$$

The microphone array elements are placed at $\mathbf{r}^{(\{1,2\})} = \left[\pm 0.25, 0, 0\right]^T$ relative to the array center, such that the positions of microphone, $m \in \{1, 2\}$, is given by:

$$\mathbf{p}^m(t) = \mathbf{R}^{-1}(\gamma_t)\, \mathbf{r}^m + \mathbf{p}^0(t), \tag{27}$$

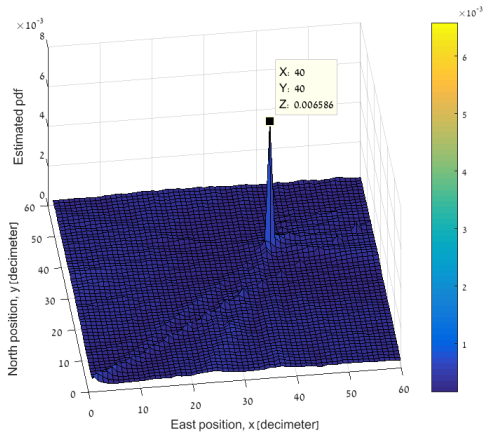

Fig. 2. The estimated posterior p.d.f., where a black circle marks the true source position.

where $\mathbf{R}(\gamma_t)$ is the rotation matrix, defined as:

$$\mathbf{R}(\gamma_t) \triangleq \begin{bmatrix} \cos(\gamma_t) & -\sin(\gamma_t) \\ \sin(\gamma_t) & \cos(\gamma_t) \end{bmatrix}. \tag{28}$$

Using (24) and (27), the trajectory of the microphone center in (24) within a room of size $6 \times 6 \times 2.5$ m^3 was simulated with the initial position at $(2, 2, 1.5)$ m at a speed of 1 m/s with orientation variance of $\sigma^2_{v_\gamma}(t) = 0.1$ rad^2 and process noise covariance, $\mathbf{\Sigma}^c = 10^{-9} \times \mathbf{I}_4$. A single static source was placed at $(4, 4, 1.5)$ m. The scenario is shown in Fig. 1. Using the room impulse response (RIR) generator in [18] the RIRs of 100 time steps at time delays of 0.2 s along the trajectory of the microphone pair were simulated for a reverberation time of 0.3 s. The resulting RIRs were convolved with a 20 s speech signal from a female speaker constructed from the TIMIT database. For localization, the height of the microphones and sources is assumed constant and known, such that the model in Section II can be used.

The input of both algorithms is constructed by the STFT with a rectangular window for each microphone. The results are used to produce the PRPs as described in [2].

B. Results

We present here the results of the two proposed algorithms.

1) IDEM: The IDEM in Alg. 1 is evaluated for $\sigma^2 = 0.1$. The estimated posterior p.d.f. after 4 iterations is plotted in Fig. 2. It can be seen that the position error is zero, when the source is located on the grid. When it is not on the grid the erro is dictated mainly by the grid resolution.

This algorithm is very accurate, but it assumes all samples are used together. As an on-line approach, we have decided to use the particle filter.

2) Particle filter: The particle filter in Alg. 2 is evaluated for $P = 100$ particles for $\mathbf{Q} = 0.01\mathbf{I}$ and with $\sigma^2 = 0.04$ and $J_{\max} = 100$. The filter is initialized by $J_0 = 500$ particles that are uniformly spread within the room region, with at least 1 m

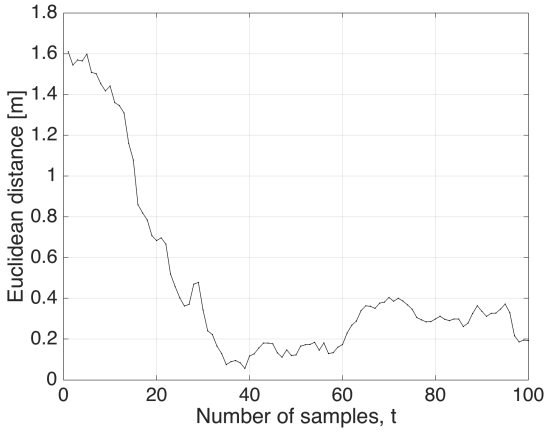

Fig. 3. Euclidean distance between the state estimates, $\tilde{\mathbf{p}}$, and true source positions, $\mathbf{p}$ for the particle filter.

distance from each of the four walls. The importance weights are initialized to $w_0^{(i)} = 1/J_0$, $\forall i \in 1, \ldots, J_0$.

The Euclidean distance between the true source position and the point estimates of the source over time are plotted in Fig. 3. The filter achieves its optimal estimation performance of 5.8 cm at $t = 38$ when the sensor pair is steering towards at a source-sensor distance of 1.5 m (see Fig. 1). The performance degrades to up a Euclidean distance between $30 - 40$ cm between $t \in [60, 98]$ when the sensor is steering away from the source.

VI. Conclusions

In this paper we addressed the challenge of source localization in a reverberant room using a single moving pair of microphones. We adapt the SAR concept from the radio frequency (RF) field. Reverberant audio data was simulated for a microphone array with two sensors and the complex-valued PRPs were extracted as measurements. Two approaches for sound source localization using the PRPs were proposed.

The first approach is a ML approach implemented by an EM algorithm, which processes all data as a batch. Localization accuracy is dictated by the grid resolution. The static nature of the source and the dynamics of the microphones enable accurate results.

In order to facilitate sequential sound source localization from on-line measurements, a particle filter was also proposed. Particle filters are typically aimed at source tracking in highly dynamic scenarios. In this paper, the approach was chosen in order to ensure real-valued source position estimates from the complex PRP measurements. Despite the static nature of the source localization accuracy of up to 5.8 cm can be achieved. Due to the sequential nature of the algorithm, this performance was shown to be dependent on the path of the robot.

In this approach we have decided to compare two modified algorithms from different paradigms, as a first step of addressing the challenge of source localization with a single pair of microphones. Possible extensions include using REM for on-line processing, solving uncertainty in robot trajectory and tracking multiple moving sources.

References

[1] G. C. Carter, "Time delay estimation for passive sonar signal processing," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 29, no. 3, pp. 463–470, 1981.

[2] O. Schwartz and S. Gannot, "Speaker tracking using recursive EM algorithms," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 2, pp. 392–402, 2014.

[3] C. Evers, J. Sheaffer, A. H. Moore, B. Rafaely, and P. A. Naylor, "Bearing-only acoustic tracking of moving speakers for robot audition," in *Proc. IEEE Intl. Conf. Digital Signal Processing (DSP)*, Singapore, Jul. 2015.

[4] A. Plinge, M. H. Hennecke, and G. A. Fink, "Robust neuro-fuzzy speaker localization using a circular microphone array," in *Proc. Intl. Workshop Acoust. Signal Enhancement (IWAENC)*, Tel Aviv, Israel, 2010.

[5] M. H. Hennecke and G. A. Fink, "Towards acoustic self-localization of ad hoc smartphone arrays," in *Proc. Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, Edinburgh, UK, May 2011, pp. 127–132.

[6] T.-K. Le and N. Ono, "Closed-form and near closed-form solutions for toa-based joint source and sensor localization," 2015.

[7] K. Nakadai, G. Ince, K. Nakamura, and H. Nakajima, "Robot audition for dynamic environments," in *Signal Processing, Communication and Computing (ICSPCC), 2012 IEEE International Conference on*, Aug 2012, pp. 125–130.

[8] G. J. Heard and I. Schumacher, "Synthetic aperture matched field approach to acoustic source localization in a shallow-water environment," *Canadian Acoustics*, vol. 26, no. 2, pp. 3–10, 1998.

[9] M. Angjelichinoski, D. Denkovski, V. Atanasovski, and L. Gavrilovska, "SPEAR: Source position estimation for anchor position uncertainty reduction," *IEEE Communications Letters*, vol. 18, no. 4, pp. 560–563, Apr. 2014.

[10] T. Routtenberg and J. Tabrikian, "Non-Bayesian periodic Cramer-Rao bound," *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 1019–1032, Feb. 2013.

[11] S. S. Iyengar, K. G. Boroojeni, and N. Balakrishnan, "Expectation maximization for acoustic source localization," in *Mathematical Theories of Distributed Sensor Networks*. Springer New York, Jan. 2014, pp. 37–54.

[12] Y. Dorfan, G. Hazan, and S. Gannot, "Multiple acoustic sources localization using distributed expectation-maximization algorithm," in *the 4th Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, May 2014, pp. 72–76.

[13] E. A. Lehmann and R. C. Williamson, "Importance sampling particle filter for robust acoustic source localisation and tracking in reverberant environments," in *Proc. Joint Workshop on Hands-Free Speech Communication and Microphone Arrays*, vol. C, New Jersey, USA, Mar. 2005, pp. 7–8.

[14] M. Mandel, R. Weiss, and D. Ellis, "Model-based expectation-maximization source separation and localization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 382–394, 2010.

[15] Y. Dorfan, D. Cherkassky, and S. Gannot, "Speaker localization and seperation using distributed expectation-maximization algorithms for localization of acoustic sources," in *Proc. European Signal Processing Conf. (EUSIPCO)*, 2015.

[16] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science. New York: Springer, 2001.

[17] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb 2002.

[18] E. Habets, "Room impulse response (RIR) generator," http://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator, 2010.