

# TRACKING OF TWO ACOUSTIC SOURCES IN REVERBERANT ENVIRONMENTS USING A PARTICLE SWARM OPTIMIZER

F.Antonacci D.Riva A.Sarti M.Tagliasacchi S.Tubaro  
Dipartimento di Elettronica ed Informazione, Politecnico di Milano  
Piazza Leonardo da Vinci, 32, 20133 Milano (Italy)

## Abstract

*In this paper we consider the problem of tracking multiple acoustic sources in reverberant environments. The solution that we propose is based on the combination of two techniques. A blind source separation (BSS) method known as TRINICON [5] is applied to the signals acquired by the microphone arrays. The TRINICON de-mixing filters are used to obtain the Time Differences of Arrival (TDOAs), which are related to the source location through a nonlinear function. A particle filter is then applied in order to localize the sources. Particles move according to a swarm-like dynamics, which significantly reduces the number of particles involved with respect to traditional particle filter. We discuss results for the case of two sources and four microphone pairs. In addition, we propose a method, based on detecting source inactivity, which overcomes the ambiguities that intrinsically arise when only two microphone pairs are used. Experimental results demonstrate that the average localization error on a variety of pseudo-random trajectories is around 40cm when the  $T_{60}$  reverberation time is 0.6s.*

## 1 Introduction

The problem of tracking wide-band acoustic sources is relevant in several applications, including audio surveillance. For example localization of acoustic events can drive a steerable camera. This application is useful in video surveillance and video conference systems. When working in closed environments, traditional localization algorithms fail for the presence of reverberations, resulting in a wrong steering of the camera. In the literature, several works address the problem of localizing and tracking a single acoustic source. A survey on this topic is presented in [2]. The authors propose to combine TDOAs obtained with Generalized Cross Correlation (GCC) and Adaptive Eigenvalue Decomposition (AED) with particle filtering [6].

The main contribution of this paper consists in extending the work in [2] to take into account two moving acoustic sources in reverberating environments. In addition, we

consider the case that part of the sources might be inactive during some time periods. A common way to localize multiple acoustic sources is to use a separation technique as a pre-processing phase. In the proposed system, we use the TRINICON algorithm [5], which claims to achieve separation for the case of non-instantaneous mixing of multiple sources. The TRINICON algorithm is applied to the problem of source localization in [4], where the TDOAs are estimated by identifying the positions of the extrema of the de-mixing filters. Since the proposed work partially relies on the ideas presented in [4], we briefly summarize this approach in Section 2.

The solution of the localization problem in a 2D space cannot be fully determined when only one microphone pair (thus one TDOA estimate) is used [4]. For the single source localization problem, TDOAs estimated using more than one microphone pair can be efficiently combined together. When more acoustic sources need to be localized, an ambiguity arises: each microphone pair estimates one TDOA for each source, but it is not obvious how to determine the correspondence between TDOAs of the same source at different microphone pairs. In [3] we showed that at least three microphone pairs must be used to localize two sources unless we can rely on some a-priori information.

The de-mixing filters estimated with the TRINICON algorithm are not always reliable when one of the sources is inactive. In [8] a pause detection technique based on the unbalancing of the power spectra at the output of the de-mixing filters is presented. This pause detection algorithm enables to figure out the reliability of the TDOA estimate based on the output of the de-mixing filters. In this paper we propose an extension of the technique proposed in [8]. We define a global pause detection function that can be computed in real-time. When a pause occurs, a pause detection index is computed for each microphone pair. The a-priori information about the activity of the sources can be efficiently exploited to solve the aforementioned ambiguity problem. In fact, when one of the sources is in silence, the TDOAs corresponding to the other source are correctly identified and a label is assigned to each TDOA which informs us about the source it refers to. Recently,

a novel approach to solve the ambiguity problem was presented [1], based on the maximization of cross-correlation between the outputs of the de-mixing filters at different microphone pairs. Though effective, this method cannot be applied in presence of pauses and it relies on the fact that the sources can be effectively separated.

The tracking algorithm described in [2] is based on a state-space formulation and uses a particle filtering approach. Recently the particle filtering (PF) algorithms have gained a great deal of attention as they provide a solution for the problem of state estimation in the nonlinear, multimodal and non-Gaussian case. For the problem at hand, the PF algorithm approximates the state PDF by sampling it at relevant points. In this paper we discuss some modifications to the approach presented in [3], in order to take into consideration the specificity of the acoustic source localization problem. The source dynamics in [3] is described by the Langevin model [2], in which the velocity of the particles is randomly modified between successive iterations: for this reason, only some particles will follow the actual source dynamics. As a consequence, in order to account for the correct source motion, a large number of particles must be used. In [9] and [7] a different dynamic model is used. The main idea is that at each iteration of the PF, the particle that best explains the observations is assigned the role of "master". All the other particles will change their velocity in order to follow the master with a certain momentum. In this paper, we will show how this "swarm intelligence" can be exploited to track sources, but with a small number of particles. If one of the sources is inactive, its position is extrapolated from the last observations.

The rest of this paper is organized as follows: in Section 2 an overview on localization task with TRINICON is presented. Section 3 illustrates the pause detection technique, compared to the one presented in [8]. Section 4 focuses on implementation of PF with Swarm Intelligence. Finally Section 5 discusses some experimental results.

## 2 Localization with Trinicon

In the BSS literature, data is usually modeled as a non-instantaneous mixture: the signal received by each of the  $P$  microphones  $x_p$  is described as a sum of delayed and filtered replica of the source signals  $s_q$ :

$$x_p(n) = \sum_{q=1}^Q \sum_{k=0}^{M-1} h_{pq}(k) s_q(n-k), \quad (1)$$

where  $Q$  is the number of active acoustic sources and  $h_{pq}(k)$ ,  $k = 0, \dots, M-1$  denotes the coefficients of the finite impulse response (FIR) from the  $q$ -th source to the  $p$ -th microphone. In the following, we assume that the number of microphones equals the number of sources ( $Q = P$ ).

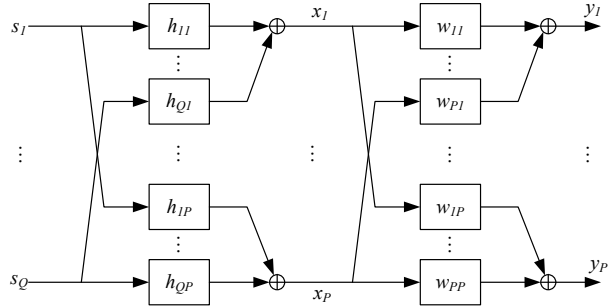


Figure 1: Block diagram of BSS MIMO model

The goal of BSS is to find a de-mixing system where the output signals  $y_q(n)$ ,  $q = 1, \dots, Q$  are described by:

$$y_q(n) = \sum_{p=1}^P \sum_{k=0}^{L-1} w_{pq}(k) x_p(n-k), \quad (2)$$

where  $w_{pq}(k)$  is the de-mixing filter weighing the  $p$ -th sensor contribution to the  $q$ -th output signal. The overall block diagram of BSS MIMO model is illustrated in Figure 1. On the left-hand side we see the mixing filters ( $h_{pq}$ ) and on the right-hand sides are depicted the de-mixing filters ( $w_{pq}$ ).

The fundamental assumption of TRINICON algorithm is that sources are non-Gaussian and statistically independent. With this hypothesis, the adaptive estimation process of de-mixing filters converges to the correct solution when the overall probability density function of the outputs can be factored out in the product of the marginal PDFs. Once de-mixing filters have been estimated, they can be used to retrieve the TDOAs of the active sources according to equations (3) and (4). For the case of two sources and two microphones according to the following equation (for the case  $Q = P = 2$ ), we obtain:

$$\hat{\tau}_1 = (\arg \max_n |w_{12}(n)| - \arg \max_n |w_{22}(n)|) f_s^{-1}, \quad (3)$$

$$\hat{\tau}_2 = (\arg \max_n |w_{11}(n)| - \arg \max_n |w_{21}(n)|) f_s^{-1}, \quad (4)$$

where  $f_s$  is the sampling frequency. From equations (3) and (4) we can appreciate that the information contained in de-mixing filters is only partially exploited to determine the TDOAs: we need only the position of global maxima/minima to achieve source localization.

Each TDOA determines a locus of potential positions consistent with the observations. The locus is an hyperbola, but it can be confused with a straight line when the distance of the source from the microphones is much larger than the distance between the microphones (far-field). In this case, there is a one-to-one mapping between the TDOA and the direction of arrival (DOA) of the source signal. The triangulation of the DOAs obtained from different microphone pairs can be used to identify the source position.

When multiple acoustic sources are active, a permutation problem arises. In fact, using the TRINICON algorithm, the  $q$ th output of the de-mixing stage,  $y_q(n)$ , can be mapped to any of the  $Q$  original source signals. Furthermore, when more than one microphone pair is used, such a mapping is generally different for each pair. If triangulation of DOAs is used, at least three microphone pairs are needed in order to correctly localize the sources in a 2D space [3]. In the following, we illustrate how the presence of pauses in the source activity can be used to solve the ambiguity problem.

### 3 Source Inactivity Detection

Detection of pauses is carried out through the analysis of power spectra of the signals produced at the output to the de-mixing stage, i.e.  $y_q(n)$ ,  $q = 0, \dots, Q$ . In [8] source inactivity detection is performed in the Fourier domain, since the goal is to design a post-processing de-noising filter in the same domain. Therefore, for each frequency bin, one source is considered inactive when its power is below a fraction of the power of the other source, for the same frequency bin. This technique is effective when separation is at least partially achieved, i.e. the cross-talk between the outputs  $y_q(n)$  is limited. We have already observed in a previous work [3] that source localization can be performed without source separation being fully achieved, since the former depends only on the location of the minima/maxima of the de-mixing filters, while the latter requires convergence of all the filter taps. Therefore, the system presented in this paper does not assume full source separation, thus making the approach in [8] impractical. On the other hand, we notice that if one of the two sources is inactive, the power of one of the two outputs diminishes considerably. Therefore, we propose to keep track of the history of the output power, in the range of frequencies actually covered by the signal spectra. If we detect a sudden decrease of one of the output power, we declare one of the sources to be inactive. We perform the same check at each microphone pair. As a side effect, since only one output will be active for each pair, we are able to resolve the ambiguity problem, by assigning the current active output channels to the same source. Figure 2 depicts an example of pause detection. One of the two signals is inactive for part of the time. As can be seen, the pause detection flag well identifies when the source is inactive. Due to the fact that the pause detection is based on the history of the short-time output power, a delay on the decision is innate in the algorithm: in the example in Figure 2 it amounts to 0.4s.

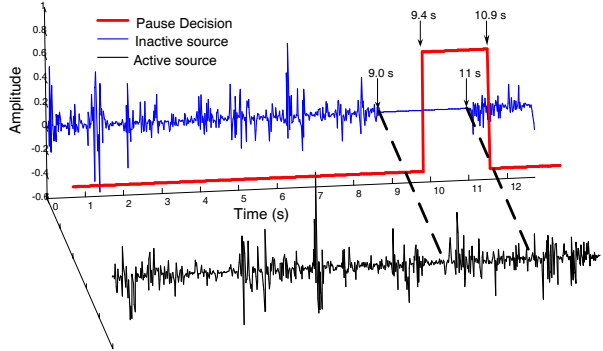


Figure 2: Source signals and inactivity flag

### 4 Localization through Swarm Particle Filters

In presence of reverberations the TDOAs measurements are noisy, due to the presence of outliers, which cause localization errors. Outliers removal is accomplished by a particle filter. The state space information associated with each particle at time  $t$  is described by a vector  $\alpha(t)$  containing the source position and velocity in the 2D space:

$$\alpha(t) = [X(t), Y(t), \dot{X}(t), \dot{Y}(t)]. \quad (5)$$

The technique presented in this paper distinguishes two operating modes: in the initialization phase a localization technique based on triangulation is used, while in a second stage two particle filters are applied separately, each tracking one acoustic source. The reason for introducing these two phases is that is that the TRINICON algorithm requires a finite amount of time to reach convergence. During this period, thereby measurements might be unreliable, thus hindering the accuracy of the localization. On the other side, when reliable measurements become available, the a-posteriori PDF of the state exhibits two sharp peaks. Therefore the PDF can be efficiently sampled by a limited number of particles. The convergence of the few particles around the estimated location is obtained with an efficient swarm intelligence model. In the following, we introduce the key concepts of the modified particle filter with swarm dynamics.

The proposed tracking algorithm is summarized below:

1. **Initialization:** two swarm particle filters (one for each source) are instantiated only when TDOA estimations become stable. Particles are initialized in a random position in proximity of the two geometric triangulations of the DOAs. In points 1 and 2 the apex  $g$  will denote the instance of the swarm particle filter which we are referring to.

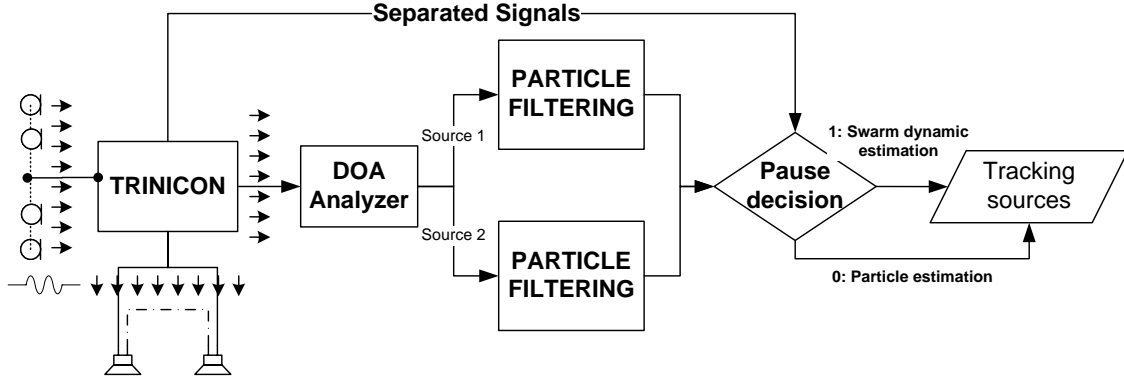


Figure 3: Overall description of the proposed system

2. **Dynamic model evolution:** Each particle  $\alpha_i^{\{g\}}(t)$  is shifted according to the following dynamic model:

$$\dot{X}_i^{\{g\}}(t) = I\dot{X}_i^{\{g\}}(t-1) + c_p\beta(P_i^{\{g\}} - X_i^{\{g\}}(t-1)) + c_s\beta(P^{\{g\}} - X_i^{\{g\}}(t-1)); \quad (6)$$

$$X_i^{\{g\}}(t) = X_i^{\{g\}}(t-1) + \dot{X}_i^{\{g\}}(t), \quad (7)$$

where  $P_i^{\{g\}}$  is the best past position of the particle  $i$ , i.e. the position for which the particle achieved the highest likelihood in the whole past history.  $P^{\{g\}}$  is the position of the most likely particle at the current iteration,  $\beta$  is a random real value sampled from a uniform distribution in the range  $[0, 1]$ . The inertia coefficient  $I$  determines the reliability of the last particle movement. Analyzing equations (6) and (7) we can observe that each particle is updated following a “private” memory of the particle history with weight  $c_p$  (the first line of equation (6)) and a social behavior with weight  $c_s$  (the second line of equation (6)). When a pause is detected, the dynamical model is different: particles are shifted extrapolating their position based on the last reliable TDOAs observations.

3. **Weight assignment:** at each time instant, a new weight  $w_i^{\{g\}}(t)$  is assigned according to the likelihood of the particle given the observed measurements. Every pair of microphones provides TDOA measurements according to equations (3) and (4). A likelihood function  $F_m(\alpha(t))$  is computed for each microphone array as in [2]. Under the hypothesis of statistically independent measurements, the particle weights are computed according to the global likelihood function:

$$w_i^{\{g\}}(t) = F(\alpha_i^{\{g\}}(t)) = \prod_{m=1}^M F_m(\alpha_i^{\{g\}}(t)). \quad (8)$$

Particles belonging to each swarm particle filter are then normalized according to the following constraint:

$$\sum_{i=1}^{N_s} w_i^{\{g\}}(t) = 1. \quad (9)$$

4. **Localization:** The estimated source locations and velocities correspond to the centroids of the two swarm particle filters:

$$f_p^{\{g\}} = \sum_{i=1}^{N_s} w_i^{\{g\}}(t)\alpha_i^{\{g\}}(t) \quad (10)$$

## 5 Experimental results

Figure 3 illustrates the block diagram of the proposed system. As stated in Section 4, two particle filters are instantiated, one for each source. The pause detection block decides if TDOA measurements are reliable or not.

This section is divided into two parts: first, the effectiveness of the algorithm is shown with a simple piece-wise linear trajectory, then, tracking of more generic pseudo-random trajectories is addressed. In both cases the original signals are simultaneous speech male segments sampled at  $f_s = 44.1\text{KHz}$ . In order to collect ground truth data, impulse responses from each source to each microphone are simulated using a fast beam tracing algorithm every  $0.125\text{s}$  along the source path. Simulations are carried out in a room of size  $5\text{m} \times 5\text{m} \times 2.7\text{m}$ . The reverberation time ranges from  $0.11\text{s}$  to  $0.61\text{s}$ . In order to test the tracking capabilities with silent sources, we can arbitrarily set pauses in any of the source signals.

Figure 4 depicts an example of tracking with two active sources in mildly reverberating conditions ( $T_{60} = 0.2\text{s}$ ) when sources move along piece-wise linear trajectories.

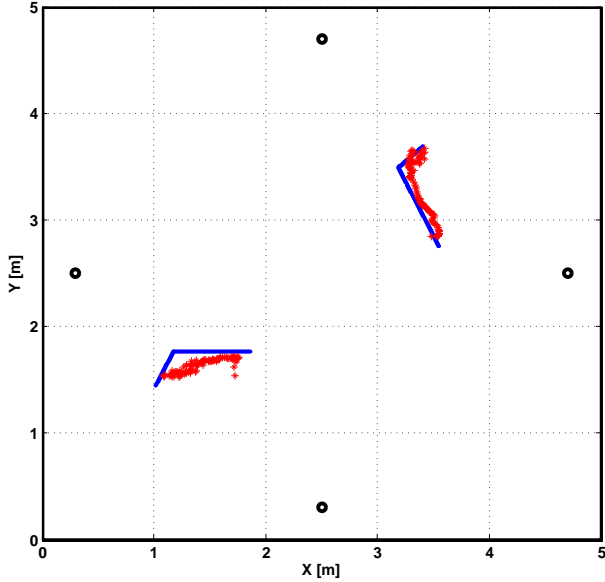


Figure 4: Trajectory followed by sources and corresponding localizations with the swarm particle filter (two active sources)

The continuous lines represent the ground truth data. The number of particles tracking each source is set to 25 for each instance of the swarm particle filter.

In Figure 5 the same experiment is repeated, but when one of the two sources is kept inactive at the time instants indicated in Figure 2. The pause occurs in proximity of the change of direction of source 2, therefore the extrapolation of the trajectory based on the last TDOAs observations introduces a noticeable error.

The simulation described in Figure 5 has been repeated setting different reverberation times, in order to compare the following localization techniques: 1) proposed swarm particle filter (SPF); 2) traditional particle filter (PF); 3) triangulation of DOAs. As for the latter, we identify the source position as the point that minimizes the average square distance from the DOAs. The localization error is defined as the root mean square error between the ground truth data and the estimated location. One of the two sources has been kept inactive as in Figure 2. The results of these simulations are illustrated in Figure 6. We observe that the swarm particle filter outperforms the other techniques in reverberant environments ( $T_{60} > 0.55s$ ), keeping the average localization error below 0.25m. For mildly reverberating environments, traditional PF and SPF give similar performance. Nevertheless, the SPF achieves the same localization error but at a fraction of the computational cost of PF. In fact, by exploiting the swarm dynamics, the number of particles is equal to 50 for SPF (25 for each source), versus 300 particles for PF.

In order to test the tracking capabilities of the swarm par-

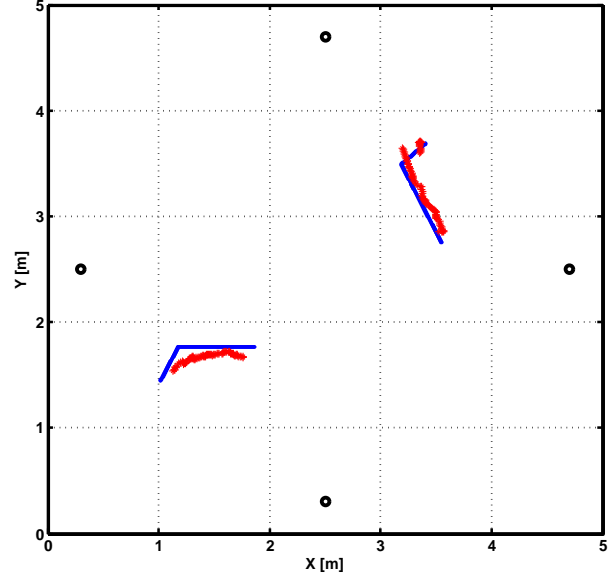


Figure 5: Trajectory followed by sources and corresponding localizations with swarm particle filter (source 2 inactive from 9.0s to 11.0s)

ticle filter, the same experiment conducted in Figure 5 has been repeated by testing pseudo-random non linear trajectories. The positions of the sources follow the Langevin dynamical model: every 0.125s the sources positions and velocities have been updated adding a random noise. Fifteen random trajectories have been generated: a sample trajectory is shown in Figure 7. Each trajectory has been tested with T60 from 0.11s to 0.61s, as in experiment of Figure 6. From the observation of trajectories in Figure 7 we can expect that the effectiveness of the localization algorithm decreases with respect to previous experiments. The RMS localization error is depicted in Figure 8: we notice that the localization error is greater than with linear trajectories: the RMS error of Swarm Particle Filter in mildly reverberating conditions increases from 0.15m to 0.25m. A similar consideration holds for traditional Particle Filtering. On the other hand, the performance of triangulation dramatically decreases with pseudo-random trajectories: the RMS error is almost always greater than 0.5m, confirming that triangulation is not a viable solution to the problem of tracking moving sources.

## 6 Conclusions

This paper presents an efficient multiple acoustic source localization algorithm stemming from the combination of a BSS technique and a particle filter with swarm intelligence. Furthermore, a solution for the removal of ambiguity problem is provided using a source inactivity detection algo-

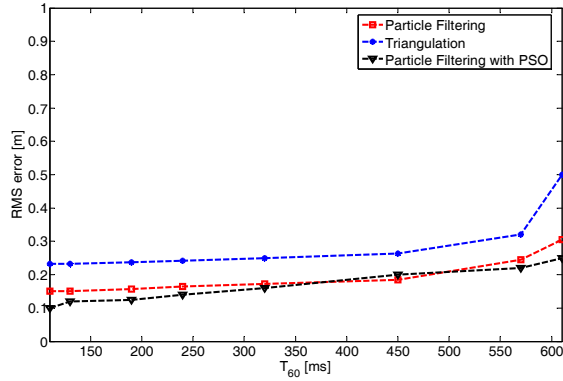


Figure 6: Average localization error of different algorithms vs. reverberation time on linear trajectories

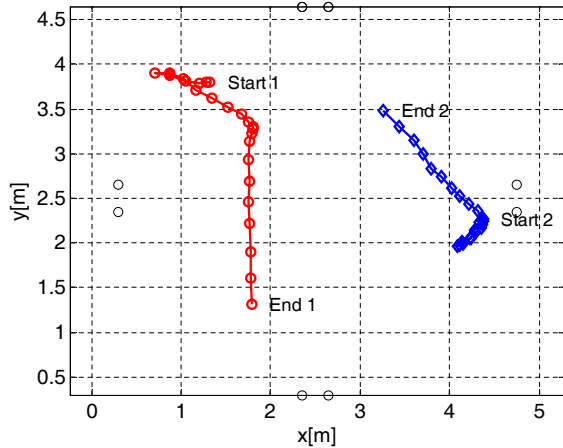


Figure 7: Example of a pseudo-random trajectory

rithm. The use of swarm intelligence allows to decrease the computational complexity while keeping the localization error almost unaltered.

## References

- [1] W.Kellermann A.Lombard, H.Buchner. Multidimensional localization of multiple sound sources using blind adaptive MIMO system identification. In *IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems*, Heidelberg, Germany, Sept. 2006.
- [2] R.C. Williamson D.B. Ward, E.A. Lehmann. Particle filtering algorithms for tracking an acoustic source in a reverberant environment. *IEEE Trans. on Speech and Audio Processing*, 11:826–836, Nov. 2003.
- [3] D.Saiu A.Sarti M.Tagliasacchi S.Tubaro F.Antonacci, D.Riva. Tracking multiple acoustic sources using particle filtering. In *European Signal Processing Conference*, Florence, Italy, Sept. 2006.
- [4] J. Stenglein H. Teutsch W. Kellermann H. Buchner, R. Aichner. Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, volume 3, pages 97–100, Philadelphia, PA, Mar. 2005.
- [5] R. Aichner H. Buchner and W. Kellermann. Trinicon: A versatile framework for multichannel blind signal processing. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, volume 3, pages 898–92, May 2004.
- [6] N.Gordon M.S.Arulampalam, S.Maskell and T.Clapp. A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking. *IEEE Trans. Signal Processing*, 50:174–188, Feb. 2002.
- [7] P. Croene R. Parisi and A. Uncini. Particle swarm localization of acoustic sources in the presence of reverberation. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 4739 – 4742, May 2006.
- [8] H.Buchner W.Kellermann R.Aichner, M.Zourub. Post-processing for convolutive blind source separation. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, volume 5, pages 37–40, May 2006.
- [9] Y. Shi and R.Eberhart. A modified particle swarm optimizer. In *IEEE World Congress On Computational Intelligence*, volume 1, pages 69–73, Anchorage, 1998.

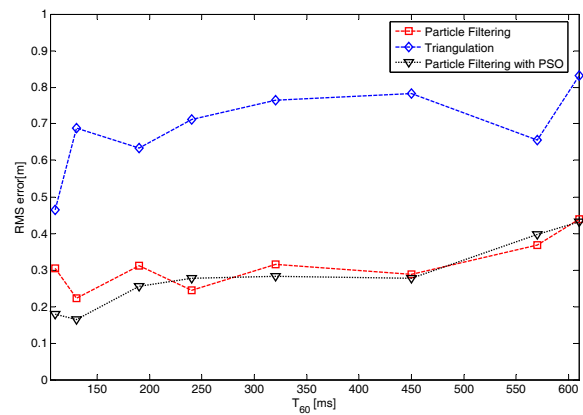


Figure 8: Average localization error vs. reverberation time along pseudo-random trajectories.