EXTENDED KALMAN FILTER WITH PROBABILISTIC DATA ASSOCIATION FOR MULTIPLE NON-CONCURRENT SPEAKER LOCALIZATION IN REVERBERANT ENVIRONMENTS

Soumitro Chakrabarty[1], *Konrad Kowalczyk*[1,2], *Maja Taseska*[1] *and Emanuël A. P. Habets*[1,2]

[1]International Audio Laboratories Erlangen*
[2]Fraunhofer Institute for Integrated Circuits (IIS)
Am Wolfsmantel 33, 91058 Erlangen, Germany

ABSTRACT

Acoustic source localization and tracking (ASLT) in reverberant environments is a challenging task due to the multi-path propagation of acoustic waves. ASLT is often based on the use of a Kalman filter or a particle filter, with time-difference-of-arrival (TDOA) estimates used as measurements. In this work, we aim to track non-concurrent speakers by applying an extended Kalman filter (EKF) with probabilistic data association (PDA) that takes into account multiple measurements simultaneously. By using PDA, the inaccuracy of the measurements caused by room reflections and noise is explicitly taken into account. Unlike in typical approaches where the measurements consist of broadband TDOA estimates, the measurements in the proposed approach consist of multiple narrowband direction-of-arrival (DOA) estimates obtained from distributed microphone arrays. Experimental results demonstrate that incorporating PDA and using properly selected narrowband DOA estimates leads to a better tracking performance, as compared to the standard EKF with a single narrowband or broadband measurement.

Index Terms— acoustic source localization, source tracking, extended Kalman filter, recursive Bayesian filter

1. INTRODUCTION

Many audio and multimedia applications utilize the information about the position of a sound source. Source localization can be used for signal extraction using, e.g., beamforming [1] or to steer a camera towards the speaker [2]. The aim of acoustic source localization and tracking (ASLT) is to estimate the source position within an acoustic environment using signals acquired by distributed microphones. Typically, steered response power [3] or time-differences-of-arrival (TDOAs) estimated using microphone pairs [4, 5] are used as measurements. These measurements are then used within a probabilistic approach [6–9], to localize and track the source.

In practice, a state-space model can be used to describe the source movement. For a linear state-space model with Gaussian state distributions, the optimum Bayesian solution can be obtained using a Kalman filter [10]. When the relation between the state (i.e., the position) and the measurement is non-linear such as the relation between the source position and the estimated TDOAs or directions-of-arrival (DOAs), non-linear Bayesian filters [11] like the extended Kalman filter (EKF), unscented Kalman filter (UKF), and particle filters (PF) can be used.

An EKF that uses a single broadband measurement at each time instant was applied for ASLT in [6, 7]. However, its performance was shown to degrade considerably in reverberant environments. Approaches utilizing multiple measurements were presented in [8] and [9], where a multiple-hypothesis model for ASLT was developed within a particle filtering framework. In [9], the measurements for the PF were selected based on a combination of hard- and soft-decision approaches, for which multiple EKFs were used, which led to very high computational complexity. Furthermore, in both [8] and [9], the probability of each measurement candidate having originated from the active source was assumed to be constant when making soft decisions. A soft-decision approach, known as probabilistic data association (PDA) [12], utilizes all measurement candidates simultaneously by assigning an appropriate posterior probability to each candidate based on the current measurements and the state-space model. Here it was assumed that a single source moves along a continuous trajectory such that the candidates are distributed within a close proximity of the previously estimated source position. A similar PDA based approach to ASLT was presented in [13], in which the multiple candidate measurements are randomly selected peaks in the cross correlation function for pairs of microphones.

In this paper, we apply PDA within an EKF framework in order to increase robustness to reverberation and noise. Here the EKF is preferred over the UKF due to the simplicity of the Taylor series approximation of the non-linear relation between the source position and the DOAs. Unlike conventional approaches to ASLT that are based on broadband TDOA estimates [6, 8, 9], our measurements comprise of multiple narrowband DOA estimates which are selected based on the magnitude squared coherence between microphone signals and are obtained using distributed microphone arrays. In contrast to [12], we aim to track non-concurrent speakers, i.e., we consider the possibility of sharp changes in the trajectory of the target source. In addition, an estimate of the broadband speech presence probability (SPP) is incorporated into the update equations of the EKF to handle non-continuous speech.

The remainder of the paper is organized as follows. The signal and state-space models are formulated in Section 2. In Section 3, the method for selecting measurement candidates is explained. In Section 4 the EKF with PDA is presented along with the computation of the posterior probability density function. Simulation results are presented in Section 5, followed by concluding remarks in Section 5.

2. PROBLEM FORMULATION

2.1. Signal Model

Consider M distributed microphone arrays, each consisting of L

* A joint institution of the Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits (IIS).

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

microphones. The signal captured at the l-th microphone of the m-th array can be written in the STFT domain as

$$Z_l^{(m)}(n,k) = \sum_{j=0}^{J} Z_{l,j}^{(m)}(n,k) + Z_{l,\text{rev}}^{(m)}(n,k) + Z_{l,v}^{(m)}(n,k), \quad (1)$$

where $Z_{l,0}^{(m)}(n,k)$, $Z_{l,j}^{(m)}(n,k)$, $Z_{l,\text{rev}}^{(m)}(n,k)$, and $Z_{l,v}^{(m)}(n,k)$ denote the complex spectral coefficients of the direct-path signal, the early reflection signals up to order J, the late reverberation and the microphone self-noise at the l-th microphone, respectively. The array index, and the time and frequency indices are denoted by m, n and k, respectively.

2.2. State-Space Model

Considering a state-space approach, the aim is to estimate the position of the target source, represented by the *state variable*, in a 2D Cartesian space at each time step, using measurements taken at each microphone array. Let us represent the state variable as $\mathbf{x}_n = [x_n\ y_n]^T \in \mathbb{R}^2$. Then, two models are required, namely a model describing the state evolution (the process model) and a model that relates the noisy measurements to the state (the measurement model). The process model assumed here is a *random walk* model [6], which is given by

$$\mathbf{x}_n = \mathbf{x}_{n-1} + \mathbf{q}_n, \quad (2)$$

where $\mathbf{q}_n$ is the process noise and the associated covariance matrix is given by $\mathbf{Q} = q_0\,\mathbf{I}$, where q_0 denotes a prior variance that describes the uncertainty in the source motion. The measurement model can be written as

$$\boldsymbol{\theta}_n = \mathbf{h}(\mathbf{x}_n) + \mathbf{g}_n, \quad (3)$$

where $\boldsymbol{\theta}_n = \left[\theta_n^{(1)}, \theta_n^{(2)}, \ldots \theta_n^{(M)}\right]^T$ is the measurement vector that contains the DOAs of M microphone arrays and $\mathbf{g}_n$ is the measurement noise with covariance matrix assumed to be

$$\mathbf{G}_n = \text{diag}\left(\sigma_{n,1}^2, \sigma_{n,2}^2 \ldots, \sigma_{n,M}^2\right), \quad (4)$$

where $\sigma_{n,m}^2$ denote the variance of the DOA estimates for the m-th array, which depends on the DOA estimation method, the acoustic environment and the relative source-array position. The elements of $\mathbf{h}(\mathbf{x}_n) = [\text{h}_1(\mathbf{x}_n), \ldots \text{h}_M(\mathbf{x}_n)]^T$ relate the DOA measurement at m-th array to the state variable $\mathbf{x}_n$ by

$$\text{h}_m(\mathbf{x}_n) = \arctan\left(\frac{r_{y_n}^{(m)}}{r_{x_n}^{(m)}}\right), \quad (5)$$

where $r_{x_n}^{(m)} = x_n - d_x^{(m)}$ and $r_{y_n}^{(m)} = y_n - d_y^{(m)}$ denote the displacement of the state variable from the center of the m-th array denoted as $\mathbf{d}^m = [d_x^{(m)} d_y^{(m)}]^T$ in x and y directions, respectively.

In this work, we propose to use multiple narrowband DOA estimates per array within the state-space model described above. Due to the presence of reverberation and noise, not all DOA estimates correspond to the active source. Therefore, we propose to select only $I_n^{(m)}$ DOA estimates at time n, which we refer to as candidates. The candidates for the m-th array can be written in vector notation as $\tilde{\boldsymbol{\theta}}_n^{(m)} = [\tilde{\theta}_n^{(m,1)}, \tilde{\theta}_n^{(m,2)} \ldots \tilde{\theta}_n^{(m,I_n^{(m)})}]^T$.

In Section 3, we describe how the candidates are selected, and in Section 4 we describe the EKF for single and multiple candidates.

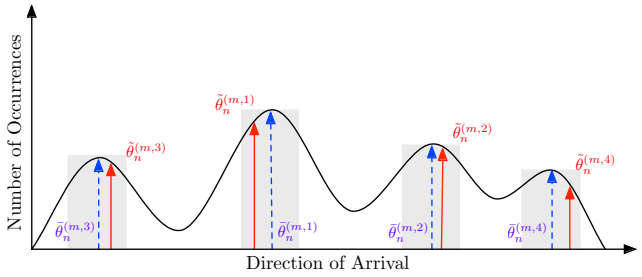

Fig. 1. Illustrative example of MSC-based measurement selection. The narrowband DOAs $\tilde{\theta}_n^{(m,i)}$ corresponding to the highest MSC, within the DOA range δ (depicted by the shaded area) around each local maximum $\bar{\theta}_n^{(m,i)}$ are selected as the measurement candidates.

3. CANDIDATE SELECTION

We aim at selecting $I_n^{(m)}$ candidates from the estimated narrowband DOAs at time step n and array m, where the DOA at frequency bin k is denoted by $\widehat{\theta}_n^{(m)}(k)$. As a direct sound results in high magnitude squared coherence (MSC) between the two microphone signals, we propose to use only the DOA estimates that are characterized by a high MSC. The MSC between the 1-st and L-th microphone signals of the m-th array can be calculated using

$$\gamma_{1L}^{(m)}(n,k) = \frac{\left|\phi_{1L}^{(m)}(n,k)\right|^2}{\phi_{11}^{(m)}(n,k)\phi_{LL}^{(m)}(n,k)}, \quad (6)$$

where $\phi_{1L}^{(m)}(n,k)$ is the cross power spectral density (PSD) and $\phi_{11}^{(m)}(n,k)$ and $\phi_{LL}^{(m)}(n,k)$ are the auto PSDs.

To account for the uncertainty in the estimates and to obtain sufficiently diverse candidates, we search for local maxima in the histogram of all narrowband DOA estimates $\widehat{\theta}_n^{(m)}(n,k)$, which are denoted by $\bar{\boldsymbol{\theta}}_n^{(m)} = [\bar{\theta}_n^{(m,1)}, \ldots, \bar{\theta}_n^{(m,I_n^{(m)})}]^T$. The frequency indices of the candidates are obtained by selecting the narrowband DOA estimate in the vicinity of each local maximum $\bar{\theta}_n^{(m,i)}$ for $i \in \{1, \ldots, I_n^{(m)}\}$ with the highest MSC, i.e.,

$$c_i^{(m)} = \arg\max_{k'} \gamma_{1L}^{(m)}(n,k') \ \text{ s.t. } \ |\bar{\theta}_n^{(m,i)} - \widehat{\theta}_n^{(m)}(k')| \le \delta, \quad (7)$$

where δ denotes the DOA range around the local maximum. This DOA selection procedure is illustrated in Fig. 1. Finally, the candidates are given by

$$\tilde{\boldsymbol{\theta}}_n^{(m)} = \left[\widehat{\theta}_n^{(m)}\left(c_1^{(m)}\right), \ldots, \widehat{\theta}_n^{(m)}\left(c_{I_n^{(m)}}^{(m)}\right)\right]^T. \quad (8)$$

The DOAs with the highest MSCs are selected as candidates since they are likely to correspond to strong directional sounds. Therefore, by using the proposed method we exclude the DOAs that most likely correspond to clutter due to noise and reverberation i.e., the DOAs with low MSC. In the following, we show how to incorporate these candidates into the EKF using PDA.

4. EXTENDED KALMAN FILTER WITH PROBABILISTIC DATA ASSOCIATION

4.1. Extended Kalman Filter

First, we consider a single candidate per array, i.e., $I_n^{(m)} = 1$ and $\theta_n^{(m)} = \tilde{\theta}_n^{(m,1)}$. Even though $\text{h}_m(\mathbf{x}_n)$ in (5) is a non-linear function,

a local linearization may be a good approximation in practical applications. The EKF can then be used to estimate the source position as a mean of the posterior pdf of the state given the measurements. The *prediction* and *update* steps for the EKF are given by [11]:

$$\mathbf{x}_{n|n-1} = \widehat{\mathbf{x}}_{n-1}, \tag{9}$$

$$\mathbf{P}_{n|n-1} = \mathbf{P}_{n-1|n-1} + \mathbf{Q}, \tag{10}$$

$$\mathbf{S}_n = \mathbf{H}(\mathbf{x}_{n|n-1})\,\mathbf{P}_{n|n-1}\,\mathbf{H}^T(\mathbf{x}_{n|n-1}) + \mathbf{G}_n, \tag{11}$$

$$\mathbf{K}_n = \mathbf{P}_{n|n-1}\,\mathbf{H}^T(\mathbf{x}_{n|n-1})\,\mathbf{S}_n^{-1}, \tag{12}$$

$$\widehat{\mathbf{x}}_n = \mathbf{x}_{n|n-1} + \mathbf{K}_n\,\mathbf{v}_n, \tag{13}$$

$$\mathbf{P}_{n|n} = \mathbf{P}_{n|n-1} - \mathbf{K}_n\,\mathbf{H}(\mathbf{x}_{n|n-1})\,\mathbf{P}_{n|n-1}, \tag{14}$$

where $\widehat{\mathbf{x}}_{n-1}$ is the state estimate from the previous time step, and $\mathbf{x}_{n|n-1}$ and $\mathbf{P}_{n|n-1}$ are the predicted mean and covariance of the state at time step n, respectively. The Kalman gain is denoted by $\mathbf{K}_n$ and $\mathbf{S}_n$ is the measurement prediction covariance. The innovation vector $\mathbf{v}_n$ consists of M elements where each element is given by $\mathrm{v}_n^{(m)} \triangleq \theta_n^{(m)} - \mathrm{h}_m(\mathbf{x}_{n|n-1})$. The Jacobian matrix of the measurement function $\mathbf{h}(\mathbf{x}_{n|n-1})$ is given by $\mathbf{H}(\mathbf{x}_{n|n-1}) = \left[\mathbf{H}_1(\mathbf{x}_{n|n-1}), \ldots, \mathbf{H}_M(\mathbf{x}_{n|n-1})\right]^T$ which is an $M \times 2$ matrix with elements $\mathbf{H}_m(\mathbf{x}_{n|n-1}) = \left[\frac{\partial \mathrm{h}_m(\mathbf{x}_{n|n-1})}{\partial x_n}, \frac{\partial \mathrm{h}_m(\mathbf{x}_{n|n-1})}{\partial y_n}\right]^T$. For the measurement function given by (5), the entries of $\mathbf{H}_m(\mathbf{x}_n)$ are given by

$$\frac{\partial \mathrm{h}_m(\mathbf{x}_n)}{\partial x_n} = -\frac{r_{y_n}^{(m)}}{\left(r_{x_n}^{(m)}\right)^2 + \left(r_{y_n}^{(m)}\right)^2}, \tag{15}$$

$$\frac{\partial \mathrm{h}_m(\mathbf{x}_n)}{\partial y_n} = \frac{r_{x_n}^{(m)}}{\left(r_{x_n}^{(m)}\right)^2 + \left(r_{y_n}^{(m)}\right)^2}. \tag{16}$$

The current state estimate and the associated variance at time step n are obtained in the update step according to (13) and (14), respectively. In the following, we describe how multiple narrowband DOA measurements can be used and how the broadband speech presence probability (SPP) can be incorporated in the EKF update equations.

4.2. Probabilistic Data Association

Given the measurements of M arrays, the innovation vector in (13) is given by $\mathbf{v}_n = [\mathrm{v}_n^{(1)}, \mathrm{v}_n^{(2)} \ldots \mathrm{v}_n^{(M)}]^T$, where the m-th element denotes the innovation term of the m-th array. To take all candidates into account, the m-th innovation term is computed as a weighted sum of the candidate innovations, i.e.,

$$\mathrm{v}_n^{(m)} \triangleq \sum_{i=1}^{I_n^{(m)}} \beta_n^{(m,i)} \mathrm{v}_n^{(m,i)}, \tag{17}$$

where $\beta_n^{(m,i)}$ denotes the weight of the i-th candidate and $\mathrm{v}_n^{(m,i)} \triangleq \tilde{\theta}_n^{(m,i)} - \mathrm{h}_m(\mathbf{x}_{n|n-1})$ denotes the innovation of the i-th candidate. As in [14], the weights can be defined as an *a posteriori* probability as described in Section 4.3.

To make the algorithm more robust, the update equations of the standard EKF are modified to take into account the presence of speech. The modified update equation for the state estimate is given by

$$\widehat{\mathbf{x}}_n = \mathbf{x}_{n|n-1} + \Pr\left(\mathcal{H}_1(n)|Z_l^{(m)}\right)\mathbf{K}_n\,\mathbf{v}_n, \tag{18}$$

where $\Pr(\mathcal{H}_1(n)|Z_l^{(m)})$ denotes the broadband SPP computed as in [15], and $Z_l^{(m)}$ is the l-th microphone signal acquired by the m-th array. In this work we used $l = 1$ and $m = 1$ to compute the SPP.

Note that for an EKF based on a single measurement, we can rewrite (14) as $\mathbf{P}_{n|n} = \mathbf{P}_{n|n-1} - \mathbf{K}_n\mathbf{S}_n\mathbf{K}_n^{\boldsymbol{T}}$ [16]. Following [14], we add the term $\mathbf{K}_n\mathbf{F}_n\mathbf{K}_n^{\boldsymbol{T}}$ to the standard EKF variance update equation to account for the multiple candidates, where $\mathbf{F}_n = \mathrm{diag}(f_n^{(1)}, \ldots, f_n^{(M)})$ is an $M \times M$ matrix with elements

$$f_n^{(m)} \triangleq \left[\sum_{i=1}^{I_n^{(m)}} \beta_n^{(m,i)} \left(\mathrm{v}_n^{(m,i)}\right)^2 - \left(\mathrm{v}_n^{(m)}\right)^2\right]. \tag{19}$$

Incorporating also the broadband SPP, the state covariance $\mathbf{P}_{n|n-1}$ can be updated using

$$\mathbf{P}_{n|n} = \mathbf{P}_{n|n-1} + \Pr\left(\mathcal{H}_1(n)|Z_l^{(m)}\right)\mathbf{K}_n(\mathbf{F}_n - \mathbf{S}_n)\mathbf{K}_n^{\boldsymbol{T}}. \tag{20}$$

With a proper choice of $\beta_n^{(m,i)}$ for each candidate, $\mathbf{F}_n$ accounts for the effect of measurements that do not correspond to the true source position by increasing the state covariance $\mathbf{P}_{n|n}$ in (20), thereby improving on the algorithm robustness against reverberation and noise.

4.3. Posterior Probability Density Function Computation

Similarly to [14], we define $\beta_n^{(m,i)}$, used in (17) and (19), as the *a posteriori* probability of each candidate measurement having originated from the actual position of the active source, i.e.,

$$\beta_n^{(m,i)} \triangleq \Pr\left(\chi_n^{(m,i)}|\tilde{\boldsymbol{\theta}}_n^{(m)}\right), \quad i = 1 \ldots I_n^{(m)}, \tag{21}$$

where $\chi_n^{(m,i)}$ denotes the event that the i-th measurement originated from the true source position. These events are assumed to be mutually exclusive since at a given point in time, only one measurement candidate can originate from the true source position due to our candidate selection criterion. Using Bayes' rule, (21) can be written as

$$\beta_n^{(m,i)} = \frac{\Pr\left(\tilde{\boldsymbol{\theta}}_n^{(m)}|\chi_n^{(m,i)}\right)\,\Pr\left(\chi_n^{(m,i)}\right)}{\Pr\left(\tilde{\boldsymbol{\theta}}_n^{(m)}\right)}. \tag{22}$$

Assuming clutter measurements are independent and uniformly distributed, the following approximation can be used:

$$\Pr\left(\tilde{\boldsymbol{\theta}}_n^{(m)}|\chi_n^{(m,i)}\right) \sim \Pr\left(\tilde{\theta}_n^{(m,i)}|\chi_n^{(m,i)}\right). \tag{23}$$

Applying marginalization and Bayes' rules to $\Pr(\tilde{\boldsymbol{\theta}}_n^{(m)})$, (22) can be rewritten as

$$\beta_n^{(m,i)} = \frac{\Pr\left(\tilde{\theta}_n^{(m,i)}|\chi_n^{(m,i)}\right)\,\Pr\left(\chi_n^{(m,i)}\right)}{\sum_{i=1}^{I_n^{(m)}} \Pr\left(\tilde{\theta}_n^{(m,i)}|\chi_n^{(m,i)}\right)\,\Pr\left(\chi_n^{(m,i)}\right)}. \tag{24}$$

The probability of each measurement originating from the true source position given past measurements $\Pr(\chi_n^{(m,i)})$ in (24) is assumed to be the same for all candidates and is given by $\Pr(\chi_n^{(m,i)}) = 1/I_n^{(m)}$. In (24), the probability $\Pr(\tilde{\theta}_n^{(m,i)}|\chi_n^{(m,i)})$ can be computed as the likelihood of a measurement candidate given the predicted mean and measurement prediction covariance, which yields

$$\Pr\left(\tilde{\theta}_n^{(m,i)}|\chi_n^{(m,i)}\right) = \mathcal{N}(\tilde{\theta}_n^{(m,i)}; \mathbf{h}_m(\mathbf{x}_{n|n-1}), \mathbf{S}_n). \tag{25}$$

Substituting (25) into (24) we obtain

$$\beta_n^{(m,i)} = \frac{\mathcal{N}(\tilde{\theta}_n^{(m,i)}; \mathbf{h}_m(\mathbf{x}_{n|n-1}), \mathbf{S}_n)}{\sum_{i=1}^{I_n^{(m)}} \mathcal{N}(\tilde{\theta}_n^{(m,i)}; \mathbf{h}_m(\mathbf{x}_{n|n-1}), \mathbf{S}_n)}. \tag{26}$$

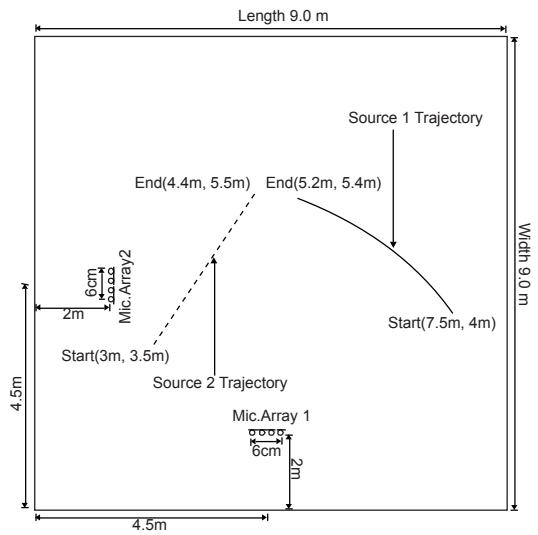

Fig. 2. Simulation setup.

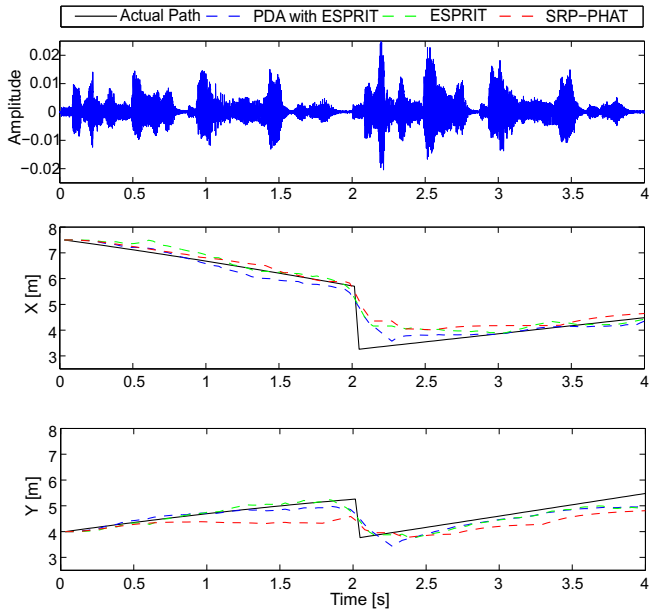

Fig. 3. Microphone signal (top), and tracked position along X and Y dimensions (middle and bottom, respectively) for $T_{60} = 0.2$ s and SNR$= 20$ dB.

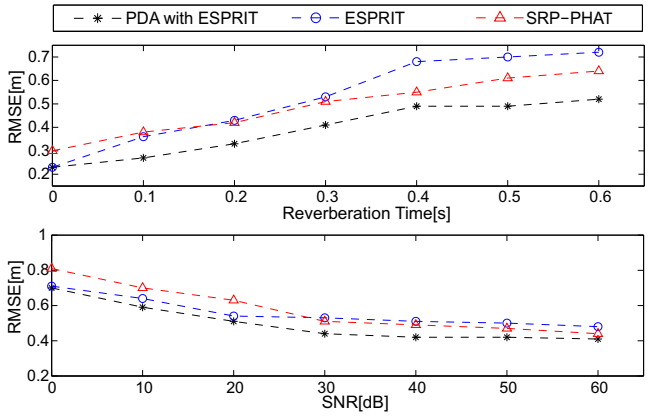

Fig. 4. Position RMSE for varying reverberation times and constant SNR $= 30$ dB (top) and varying SNR levels with a constant $T_{60} = 0.3$ s (bottom).

5. EXPERIMENTAL RESULTS

We simulated a scenario with two non-concurrently active speakers. Each speaker was active during a 2 s segment, continuously moving along the trajectories depicted in Fig. 2. Measurements were obtained using two 4-element uniform linear arrays with 0.02 m inter-microphone spacing. The microphone signals were computed by convolving anechoic speech signals with room impulse responses generated using an image-source method [17], and a white Gaussian noise with a desired signal-to-noise ratio (SNR) between 0 and 60 dB was added. The sampling rate was $F_s = 16$ kHz and a 512-point STFT with 50% overlap was used. The process noise variance was $q_0 = 0.4$ and the measurement variance was kept constant as $\sigma^2 = 0.4$ for all frequencies. The narrowband DOAs were estimated using ESPRIT [18], and an autoregressive averaging with a time constant of 50 ms was applied to smooth the microphone signal PSDs. The number of candidates in PDA was set to five. For comparison, the results of an EKF with a single candidate selected as a maximum in the steered response power phase-transform (SRP-PHAT) function [19], and an EKF with a single candidate with the highest MSC were also presented. The SRP-PHAT is an extension of the generalized cross-correlation phase-transform (GCC-PHAT) [20] for more than 2 microphones, where the GCC-PHAT functions are averaged over all microphone pairs. Note that SPP was also incorporated into the EKF update equations for both single candidate methods for a fair comparison with the proposed multiple candidate approach. As an evaluation measure, root mean square error (RMSE) between the true and estimated positions was used.

As depicted in Fig. 3, the EKF with PDA achieved the highest accuracy between the true and tracked source positions in x and y dimensions. The two EKFs which use a single DOA (estimated using SRP-PHAT or ESPRIT) perform similarly, except for the y dimension. As shown in Fig. 4, the RMSE over the full signal duration increases for increasing reverberation time and noise level for all compared methods. The EKF with PDA exhibits the smallest error, thus being more robust towards reverberation and noise than the techniques that use a single measurement only. Furthermore, SRP-PHAT-based EKF performs better than a single-DOA ESPRIT-based EKF for reverberation times exceeding 0.2 s, while ESPRIT-based EKF outperforms SRP-PHAT-based EKF at low SNRs.

6. CONCLUSIONS

In this paper, a method for acoustic source tracking using an extended Kalman filter with probabilistic data association was presented. Multiple narrowband DOA measurements are selected based on the MSC between the microphone signals, and a broadband speech presence probability is incorporated in the model for tracking of speech sources. The proposed method yields improved tracking performance over standard single-candidate techniques in noisy and reverberant environments.

7. REFERENCES

[1] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[2] Y. Huang, J. Benesty, and G. W. Elko, "Passive acoustic source localization for video camera steering," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Istanbul, Turkey, June 2000, vol. 2, pp. 909–912.

[3] E. A. Lehmann and R. C. Williamson, "Particle filter design using importance sampling for acoustic source localisation and tracking in reverberant environments," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, pp. 017021, 2006.

[4] T. G. Dvorkind and S. Gannot, "Time difference of arrival estimation of speech source in a noisy and reverberant environment," *Signal Processing*, vol. 85, no. 1, pp. 177–204, Jan. 2005.

[5] S. Doclo and M. Moonen, "Robust adaptive time delay estimation for speaker localisation in noisy and reverberant acoustic environments," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 11, pp. 1110–1124, Oct. 2003.

[6] S. Gannot and T. Dvorkind, "Microphone array speaker localizers using spatial-temporal information," *EURASIP Journal on Applied Signal Processing*, vol. 2006, no. 1, pp. 1–17, Jan. 2006.

[7] U. Klee and J. McDonough, "Kalman filtering for acoustic source localization based on time delay of arrival," in *Proc. Joint Workshop on Hands-Free Speech Communication and Microphone Arrays*, Piscataway, New-Jersey, USA, Mar. 2005, vol. C, pp. 5–6.

[8] J. Vermaak and A. Blake, "Nonlinear filtering for speaker tracking in noisy and reverberant environments," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Salt-Lake City, Utah, USA, May 2001, vol. 5, pp. 3021–3024.

[9] A. Levy, S. Gannot, and E. A. P. Habets, "Multiple-hypothesis extended particle filter for acoustic source localization in reverberant environments," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 19, no. 6, pp. 1540–1555, 2011.

[10] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. of the ASME Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

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

[12] Y. Bar-Shalom and E. Tse, "Tracking in a cluttered environment with probabilistic data association," in *Proc. of the Fourth Symp. on Nonlinear Estimation, Theory and its Applications,*, San Diego, California,, Sep. 1973, pp. 13–22.

[13] T. Gehrig and J. McDonough, "Tracking multiple speakers with probabilistic data association filters," in *Multimodal Technologies for Perception of Humans*, Rainer Stiefelhagen and John Garofolo, Eds., vol. 4122 of *Lecture Notes in Computer Science*, pp. 137–150. Springer Berlin Heidelberg, 2007.

[14] Y. Bar-Shalom and E. Tse, "Tracking in a cluttered environment with probabilistic data association," *Automatica*, vol. 11, no. 5, pp. 451–460, Sep. 1975.

[15] M. Brookes, "Voicebox: Speech processing tool for MATLAB," (available online).

[16] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House radar library. Artech House, 2004.

[17] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal Acoust. Soc. of America*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[18] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 984–995, 1989.

[19] M. S. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*, Springer-Verlag, Berlin, Germany, 2001.

[20] J. Chen, J. Benesty, and Y. Huang, "Performance of GCC- and AMDF-based time-delay estimation in practical reverberant environments," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 1, pp. 25–36, 2005.