INTERNATIONAL AUDIO LABORATORIES ERLANGEN

A joint institution of Fraunhofer IIS and Universität Erlangen-Nürnberg

AUDIO LABS

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

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

[1]International Audio Laboratories, Erlangen, Germany [2]Fraunhofer Institute for Integrated Circuits (IIS), Erlangen, Germany

E-Mail: soumitro.chakrabarty@audiolabs-erlangen.de

1. Introduction

- Aim - Estimate the source position at each time step within an acoustic environment using signals acquired by distributed microphones
- Conventional Method - Use broadband time-difference-of-arrival (TDOA) estimates as measurements within a state-space model based probabilistic framework
- Proposed Method - Use narrowband direction-of-arrival (DOA) estimates selected based on magnitude squared coherence between microphone signals as measurements within an extended kalman filter (EKF) framework with probabilistic data association (PDA)

2. Problem Formulation

Signal model in TF domain

direct sound + early reflections   late reverberation   noise signal   array index

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

time index   frequency index   reflection order   microphone index

State-space model

$$\mathbf{x}_n = [x_n \; y_n]^T = \mathbf{x}_{n-1} + \mathbf{q}_n$$

state variable   process noise

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

measurements   measurement noise

Measurements

- Multiple narrowband DOA estimates per array
- Measurement for m-th array

$$\tilde{\boldsymbol{\theta}}_n^{(m)} = [\tilde{\theta}_n^{(m,1)}, \tilde{\theta}_n^{(m,2)} \dots \tilde{\theta}_n^{(m,I_n^{(m)})}]^T$$

$I_n^{(m)}$- number of candidates

Measurement function

$$h_m(\mathbf{x}_n) = \arctan\left(\frac{r_{y_n}^{(m)}}{r_{x_n}^{(m)}}\right) \qquad r_{y_n}^{(m)} = y_n - d_y^{(m)} \qquad r_{x_n}^{(m)} = x_n - d_x^{(m)} \qquad \mathbf{d}^m = [d_x^{(m)} d_y^{(m)}]^T$$

array center

- Given the model, compute the state estimate $\hat{\mathbf{x}}_n$

3. Candidate Selection

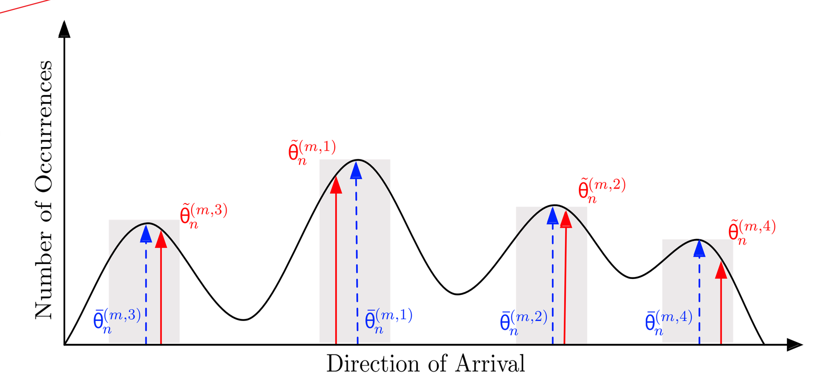
- Candidates selected based on magnitude squared coherence (MSC)
- Search for local maxima in the histogram of all narrowband DOA estimates: $\bar{\boldsymbol{\theta}}_n^{(m)} = [\bar{\theta}_n^{(m,1)}, \dots, \bar{\theta}_n^{(m,I_n^{(m)})}]^T$
- Frequency indices of the candidates obtained by:

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

estimated narrowband DOAs

- Selected candidates:

$$\tilde{\boldsymbol{\theta}}_n^{(m)} = \left[\hat{\theta}_n^{(m)}\left(c_1^{(m)}\right), \dots, \hat{\theta}_n^{(m)}\left(c_{I_n^{(m)}}^{(m)}\right)\right]^T$$



4. Extended Kalman Filter with PDA

Extended Kalman Filter

$$\mathbf{x}_{n|n-1} = \hat{\mathbf{x}}_{n-1},$$

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

process noise covariance

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

measurement noise covariance

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

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

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

- Kalman Gain: $\mathbf{K}_n$
- Measurement prediction covariance: $\mathbf{S}_n$
- Jacobian Matrix: $\mathbf{H}(\mathbf{x}_{n|n-1})$
- Innovation vector: $\mathbf{v}_n$

Probabilistic Data Association Modifications

- Innovation vector for multiple candidates:

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

a posteriori probability

$$\beta_n^{(m,i)} \triangleq \Pr\left(\chi_n^{(m,i)} | \tilde{\boldsymbol{\theta}}_n^{(m)}\right)$$

event: i-th measurement is from true source position

- Computed:

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

- **State estimate**:

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

broadband speech presence probability

- **State covariance**:

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

accounts for measurements ***not*** corresponding to true source position

each element

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

5. Experimental Results

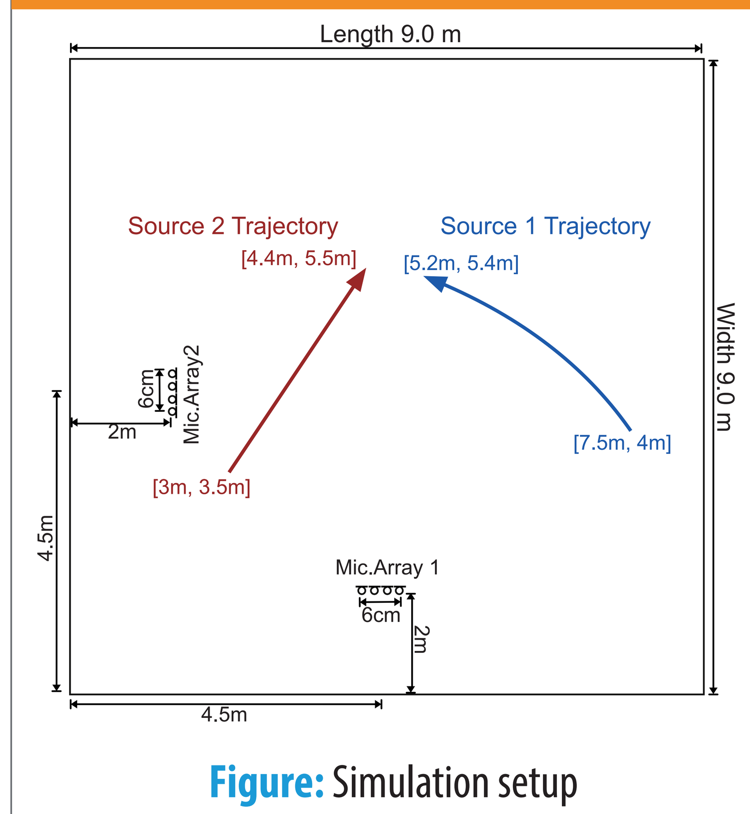


Figure: Simulation setup

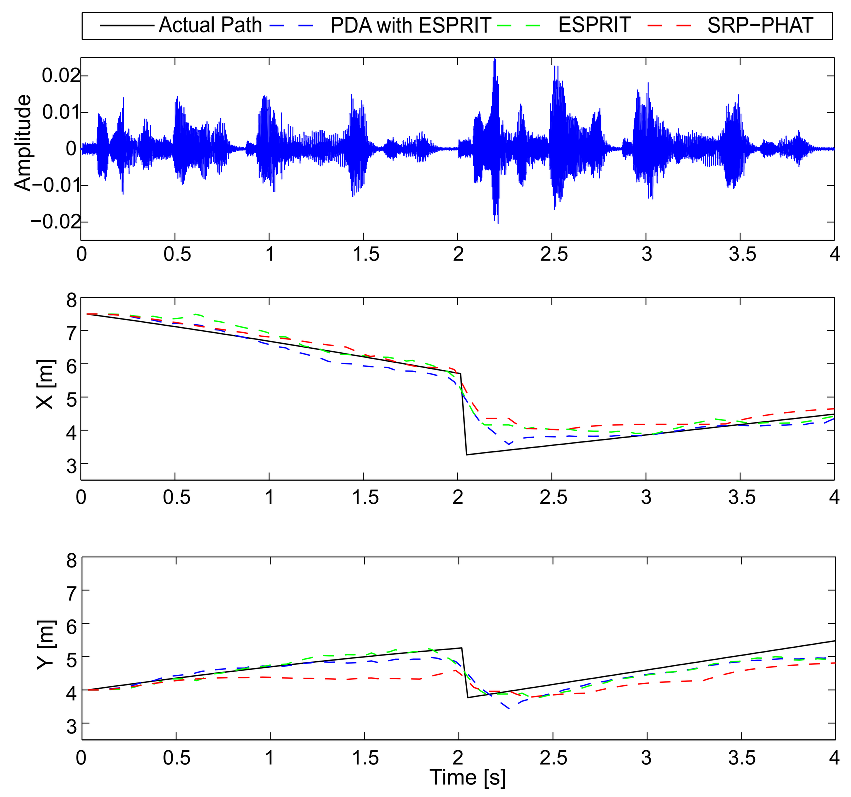


Figure: Microphone signal (top), and tracked position along X and Y dimensions (middle and bottom, respectively) for $RT_{60} = 0.2$s and SNR = 20dB.

Simulation setup

- $F_s = 16$ kHz
- STFT: 32 ms, 50%
- $q_0 = 0.4$
- $\sigma^2 = 0.4$
- DOA estimation method: ESPRIT

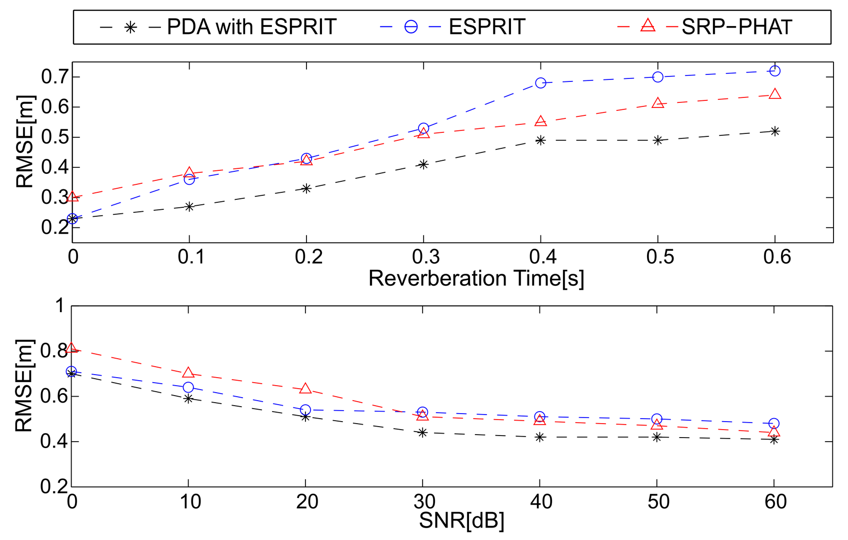


Figure: Position RMSE for varying reverberation times and constant SNR = 30dB (top) and varying SNR levels with a constant $RT_{60} = 0.3$s (bottom).

6. Conclusion

- A method for acoustic source tracking using EKF with PDA was proposed
- Multiple narrowband DOA estimates selected based on the MSC were used as measurement candidates
- The proposed method yields an improved tracking performance over single-candidate based methods