Towards Informative Path Planning for Acoustic SLAM

Christine Evers, Alastair H. Moore, and Patrick A. Naylor
Imperial College London
Department of Electrical and Electronic Engineering Exhibition Road, London, SW7 2AZ, United Kingdom
Email: c.evers@imperial.ac.uk

Zusammenfassung

Acoustic scene mapping is a challenging task as microphone arrays can often localize sound sources only in terms of their directions. Spatial diversity can be exploited constructively to infer source-sensor range when using microphone arrays installed on moving platforms, such as robots. As the absolute location of a moving robot is often unknown in practice, Acoustic Simultaneous Localization And Mapping (a-SLAM) is required in order to localize the moving robot's positions and jointly map the sound sources. Using a novel a-SLAM approach, this paper investigates the impact of the choice of robot paths on source mapping accuracy. Simulation results demonstrate that a-SLAM performance can be improved by informatively planning robot paths.

Introduction

Robot audition addresses the processing of audio and acoustic signals obtained from microphone installed on the body of robots [1]. Applications are often targeted towards human-robot interaction in realistic acoustic conditions, including interference, noise, and reverberation [2]. Situational awareness of the surrounding acoustic environment is therefore crucial for intuitive interaction of a robot with human talkers.

Acoustic scene mapping aims to construct a three-dimensional representation of the acoustic environment surrounding a microphone array. Using sound source localization, estimates of the instantaneous source Directions-of-Arrival (DoAs) can be obtained. In practice, the distance – or range – between the sensor and each source is often unmeasured. However, the range is required in order to reconstruct the three-dimensional Cartesian source positions within the acoustic scene map.

For robot audition, the source positions can be triangulated from the source DoAs obtained at consecutive positions along the path of a moving robot. Spatial diversity of the robot can therefore be exploited constructively for sound source mapping. However, in realistic environments, reverberation and noise often results in DoA estimation errors. Moreover, early reflections can lead to spurious DoA estimates, considered as clutter, as well as missing source detections.

Multi-source tracking approaches utilize temporal models of the source dynamics in order to infer the Cartesian source trajectories from the instantaneous DoAs. Source tracking can therefore be used to obtain smoothed estimates of the source positions, extrapolate sources through periods of missing detections, and identify incoherent clutter DoAs.

Naturally, the estimated DoAs are relative to the origin of the microphone array. In order to reconstruct absolute source positions from the relative DoAs, the robot path must therefore be known. However, in practice, the executed robot motion diverges from the reported robot motion corresponding to the robot's internal belief of its position [3].

Nevertheless, the acoustic scene map can be used to identify the robot position and orientation that optimally aligns the DoAs with the estimated source positions. Robot localization and source mapping are therefore jointly dependent and should be solved simultaneously, leading to the concept of Acoustic Simultaneous Localization and Mapping (a-SLAM) [4].

We previously proposed a novel and robust a-SLAM approach in [4], that addresses the challenges of DoA errors, clutter, as well as missing detections in reverberant and noisy environments. The novel contribution of this paper is the investigation of the contribution of the choice of robot path on accuracy of the estimated acoustic scene map. Using simulation results, we will demonstrate that a-SLAM performance can be improved by informatively planning robot paths.

System Model

The three-dimensional Cartesian position of a single sound source is given by the vector, $\mathbf{s}_{t,n} \triangleq \left[\hat{x}_{t,n}, \hat{y}_{t,n}, \hat{z}_{t,n}\right]^T$, specified relative to the robot position, $\mathbf{r}_t$, where $n = 1, \ldots, N_t$ and N_t is the number of sources at time t. In this paper, the sources are assumed static with subtle movements due to head and body rotations, such that,

$$\mathbf{s}_{t,n} = \mathbf{s}_{t-1,n} + \mathbf{n}_{t,n}, \qquad \mathbf{n}_{t,n} \sim \mathcal{N}\left(\mathbf{0}_{3\times 1}, \mathbf{Q}\right), \tag{1}$$

where $\mathbf{n}_{t,n}$ is the process noise with covariance $\mathbf{Q}$.

The localized DoA due to source $\mathbf{s}_{t,n}$ is defined as $\boldsymbol{\omega}_{t,m} \triangleq \left[\theta_{t,m}, \phi_{t,m}\right]^T$ where $m \in \{1, \ldots, M_t\}$ and where M_t is the number of DoAs at t. Each DoA constains the inclination $\theta = \cos^{-1}\left(z/\sqrt{x^2+y^2+z^2}\right)$ and azimuth $\phi = \arctan\left(y/x\right)$ and is modelled as

$$\boldsymbol{\omega}_{t,m} = g(\mathbf{s}_{t,n}) + \mathbf{m}_{t,m}, \qquad \mathbf{m}_{t,m} \sim \mathcal{N}\left(\mathbf{0}_{2\times 1}, \mathbf{R}\right) \tag{2}$$

where $g(\mathbf{s}_{t,n})$ is the Cartesian-to-spherical transformation, and where $\mathbf{m}_{t,m}$ is the measurement noise with covariance, $\mathbf{R}$.

The multi-source state contains all N_t sources and can hence be expressed as the random finite set [5], $\mathbf{S}_t \triangleq$

$\{\mathbf{s}_{t,n}\}_{n=1}^{N_t}$. The multi-source state can be modelled to explicitly account for source initialisation, survival between time steps, and termination, such that

$$\mathbf{S}_t = \left[\bigcup_{n=1}^{N_{t-1}} P(\mathbf{s}_{t-1,i})\right] \cup B_t, \tag{3}$$

where B_t is a birth process, and $P(\mathbf{s}_{t-1,i}) = \mathbf{s}_{t-1,j}$ if $\mathbf{s}_{t-1,i}$ persists between $t-1$ to t, and $P(\mathbf{s}_{t-1,i}) = \emptyset$ otherwise.

In order to model clutter and missing detections, the multi-source measurement process is expressed as

$$\mathbf{\Omega}_t = \left[\bigcup_{n=1}^{N_t} D(\mathbf{s}_{t,n})\right] \cup C_t, \tag{4}$$

where C_t is the clutter process and $D(\mathbf{s}_{t,n}) = \boldsymbol{\omega}_{t,m}$ if $\mathbf{s}_{t,n}$ is detected and $D(\mathbf{s}_{t,n}) = \emptyset$ otherwise.

Similar to (1), the robot position, $\mathbf{p}_t = \begin{bmatrix} x_{t,r} & y_{t,r} & v_{t,r} \end{bmatrix}^T$, contains the Cartesian position $(x_{t,r}, y_{t,r})$ and speed, $v_{t,r}$, and is modelled as

$$\mathbf{p}_t = \mathbf{F}_{t,r}\,\mathbf{p}_{t-1} + \mathbf{v}_{t,\mathbf{p}}, \qquad \mathbf{v}_{t,\mathbf{p}} \sim \mathcal{N}\left(\mathbf{0}_{3\times 1},\, \mathbf{\Sigma}_{t,\mathbf{v}}\right) \tag{5}$$

where $\mathbf{v}_{t,\mathbf{p}}$ is the process noise with covariance $\mathbf{\Sigma}_{t,\mathbf{v}}$ and the height, $z_{t,r}$, is constant and known. The dynamical model, $\mathbf{F}_t$, is general but expressed as a constant velocity model in this paper, such that

$$\mathbf{F}_{t,r} = \begin{bmatrix} 1 & 0 & \Delta_T \sin\gamma_{t,r} \\ 0 & 1 & \Delta_T \cos\gamma_{t,r} \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

where Δ_T is the time delay between $t-1$ and t and $\gamma_{t,r}$ is the robot orientation, modelled as a random walk,

$$\gamma_{t,r} = \gamma_{t-1,r} + v_{t,\gamma}, \qquad v_{t,\gamma} \sim \mathcal{N}\left(0,\, \sigma^2_{v_{t,\gamma}}\right). \tag{7}$$

for process noise $v_{t,\gamma}$ with variance $\sigma^2_{v_{t,\gamma}}$. The unknown robot state is therefore defined as $\mathbf{r}_t \triangleq \begin{bmatrix} \mathbf{p}_t^T & \gamma_{t,r} \end{bmatrix}^T$.

The reported robot motion is defined as $\mathbf{y}_t \triangleq \begin{bmatrix} z_{t,v} & z_{t,\gamma} \end{bmatrix}^T$, containing the robot velocity and orientation, $z_{t,v}$ and $z_{t,\gamma}$ respectively. In order to account for the errors between the reported and executed robot motion, the robot reports are modelled as

$$z_{t,v} = \mathbf{h}\,\mathbf{p}_t + w_{t,v} \qquad w_{t,v} \sim \mathcal{N}\left(0,\, \sigma^2_{w_{t,v}}\right) \tag{8a}$$

$$z_{t,\gamma} = \gamma_t + w_{t,\gamma} \qquad w_{t,\gamma} \sim \mathcal{N}\left(0,\, \sigma^2_{w_{t,\gamma}}\right) \tag{8b}$$

where $w_{t,v}$ and $w_{t,\gamma}$ are the speed and orientation noise with variances $\sigma^2_{w_{t,v}}$ and $\sigma^2_{w_{t,\gamma}}$ respectively, and $\mathbf{h} \triangleq \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$.

Proposed a-SLAM Approach

The a-SLAM approach used in this paper was originally proposed in [4] and is summarized in this section for completeness of this paper.

In order to fully describe the statistics of the unknown process, $\mathbf{X}_t \triangleq (\mathbf{r}_t, \mathbf{S}_t)$, its posterior Probability Density Function (pdf) should be estimated and propagated in time. For multi-source mapping, however, the pdf is numerically intractable. Rather than propagating the pdf, the posterior can be approximated by its first-order moment, the Probability Hypothesis Density (PHD) [5]. The joint PHD can be factorized into the PHD of the robot state, $\lambda(\mathbf{r}_t|\mathbf{y}_{1:t})$, and the multi-source PHD, $\lambda(\mathbf{s}_t|\mathbf{r}_t, \mathbf{\Omega}_{1:t})$, conditional on the robot position, such that,

$$\lambda(\mathbf{x}_t|\mathbf{\Omega}_{1:t}, \mathbf{y}_{1:t}) = \lambda(\mathbf{r}_t|\mathbf{y}_{1:t})\,\lambda(\mathbf{s}_t|\mathbf{r}_t, \mathbf{\Omega}_{1:t}). \tag{9}$$

Source Mapping

The source PHD accounts for the contribution of undetected, detected, and newborn sources and is given by:

$$\lambda(\mathbf{s}_t|\mathbf{r}_t, \mathbf{\Omega}_{1:t}) = (1-p_d)\lambda(\mathbf{s}_t|\mathbf{r}_t, \mathbf{\Omega}_{1:t-1}) \tag{10}$$

$$+\, p_d\,\lambda_s(\mathbf{s}_t|\mathbf{r}_t, \mathbf{\Omega}_{1:t}) + \lambda_b(\mathbf{s}_t|\mathbf{r}_t, \mathbf{\Omega}_t) \tag{11}$$

where p_d is the probability of source detection, $\lambda(\mathbf{s}_t|\mathbf{r}_t, \mathbf{\Omega}_{1:t-1})$ is the predicted PHD of undetected sources, $\lambda_b(\mathbf{s}_t|\mathbf{r}_t, \mathbf{\Omega}_t)$ is the PHD of newborn sources, and $\lambda_s(\mathbf{s}_t|\mathbf{r}_t, \mathbf{\Omega}_{1:t})$, is the PHD of surviving sources.

For the Gaussian source model in (1), $\lambda_s(\mathbf{s}_t|\mathbf{r}_t, \mathbf{\Omega}_{1:t-1})$ can be modelled as a Gaussian Mixture Model (GMM) [6] of J_{t-1} components and Gaussian Mixture (GM) weights, $w_{t-1}^{(j)}$ for all $j \in \{1, \ldots, J_{t-1}\}$, such that

$$\lambda(\mathbf{s}_t|\mathbf{r}_t, \mathbf{\Omega}_{1:t-1}) = \sum_{j=1}^{J_{t-1}} w_{t-1}^{(j)}\,\mathcal{N}\left(\mathbf{s}_t \,\middle|\, \mathbf{m}_{t|t-1}^{(j)},\, \mathbf{\Sigma}_{t|t-1}^{(j)}\right), \tag{12}$$

where the GM mean and covariance terms, $\mathbf{m}_{t|t-1}^{(j)}$ and $\mathbf{\Sigma}_{t|t-1}^{(j)}$, are given by the Kalman Filter (KF) prediction equation [7].

The PHD of surviving components is obtained by updating $\lambda(\mathbf{s}_t|\mathbf{r}_t, \mathbf{\Omega}_{1:t-1})$ with the source DoAs, $\boldsymbol{\omega}_{t,m}$, such that

$$\lambda_s(\mathbf{s}_t|\mathbf{r}_t, \mathbf{\Omega}_{1:t}) = \sum_{m=1}^{M_t}\sum_{j=1}^{J_{t-1}} w_t^{(j,m)}\mathcal{N}\left(\mathbf{s}_t \,\middle|\, \mathbf{m}_t^{(j,m)},\, \mathbf{\Sigma}_t^{(j,m)}\right) \tag{13}$$

where the mean, $\mathbf{m}_t^{(j,m)}$, and covariance, $\mathbf{\Sigma}_t^{(j,m)}$, of each GM component are given by the extended KF equations [4, 7]. The GM weights, $w_t^{(j,m)}$, are given by

$$w_t^{(j,m)} = \frac{w_{t-1}^{(j)}}{\ell(\boldsymbol{\omega}_{t,m}|\mathbf{r}_t)}\,\mathcal{N}\left(\boldsymbol{\omega}_{t,m} \,\middle|\, g\left(\mathbf{m}_{t|t-1}^{(j)}\right),\, \mathbf{S}_t^{(j)}\right), \tag{14}$$

where $\mathbf{S}_t^{(j)}$ is the KF innovation covariance and $g(\cdot)$ denotes the Cartesian-to-spherical transformation. The evidence term, $\ell(\boldsymbol{\omega}_{t,m}|\mathbf{r}_t)$, corresponds to the probability of DoA, $\boldsymbol{\omega}_{t,m}$, being due to either clutter, source birth or source survival, such that

$$\ell(\boldsymbol{\omega}_{t,m}|\mathbf{r}_t) = \kappa_{t,m} + \sum_{j=1}^{J_{t-1}} p_d\, w_{t-1}^{(j)}, \tag{15}$$

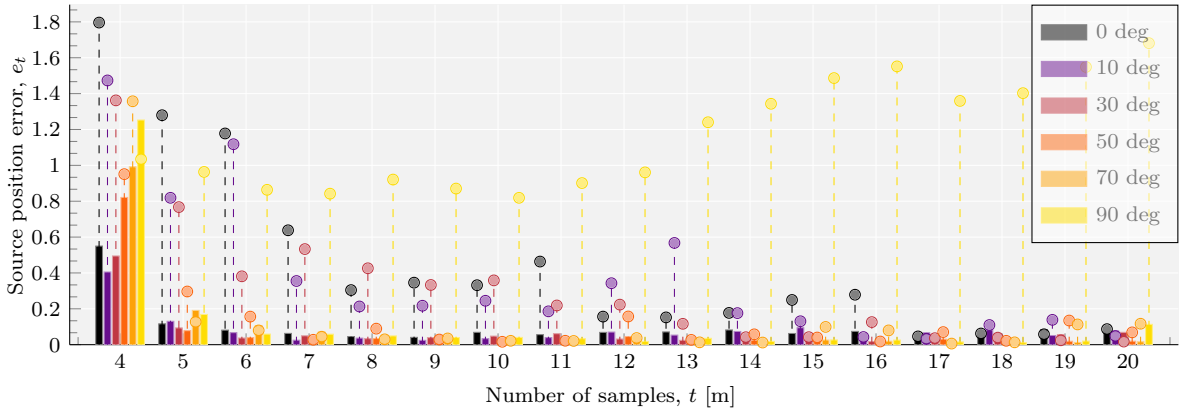

Abbildung 1: Source mapping accuracy for an array at 1.195 m height moving in a straight line with specified orientation from the source. The source is placed at a height of 1.195 m (circles) 1.8 m (bars).

where $\kappa_{t,m} = \lambda_\kappa V \mathcal{U}(\boldsymbol{\omega}_{t,m})$ is the clutter PHD for room volume V with clutter rate, λ_κ, and assuming uniformly distributed clutter DoAs.

Similar to (12), the birth PHD can be expressed as

$$\lambda_b(\mathbf{s}_t|\mathbf{r}_t, \boldsymbol{\Omega}_t) = \sum_{\ell=1}^{L} w_{b,t}^{(\ell)} \mathcal{N}\left(\mathbf{s}_t \,|\, \mathbf{m}_{b,t}^{(\ell)}, \boldsymbol{\Sigma}_{b,t}^{(\ell)}\right), \qquad (16)$$

where L is the number of newborn GM components, the GM weights are given by $w_{b,0}^{(\ell)} = \frac{N_b}{L}$ where N_b is the expected number of source births per time step, and $\mathbf{m}_{b,t}^{(\ell)}$ and $\boldsymbol{\Sigma}_{b,t}^{(\ell)}$ are the mean and covariance of GM component, $\ell \in \{1, \ldots, L\}$.

Newborn sources are initialized from the measurements. An estimate of the unmeasured range is introduced at initialisation. The range estimate is propagated in time by probabilistic triangulation as previously proposed in [8]. For each DoA in $\{\boldsymbol{\omega}_{t,m}\}_{m=1}^{M_t}$, newborn source states are generated by drawing P random variates $\mathbf{m}_{b,0}^{(p)} \sim \mathcal{N}\left(\left[\boldsymbol{\omega}_{t,m}^T, r_0\right]^T, \boldsymbol{\Sigma}_{b,0}\right)$ for $p = 1, \ldots, P$, such that $L = M_t P$, and where r_0 is the prior range with variance $\sigma_{r_0}^2$, and the covariance is given by $\boldsymbol{\Sigma}_{b,0} \triangleq \text{diag}\left[\mathbf{R}, \sigma_{r_0}^2\right]$.

Robot Localization

Recalling (6), the robot position is non-linearly dependent on the robot orientation, $\gamma_{t,r}$. The non-linearity can be addressed using a Rao-Blackwellized particle filter [9] to estimate $\mathbf{r}_t$ as proposed in [4].

At each time t, P random variates of the orientation, $\hat{\gamma}_t^{(i,p)}$, can be drawn from an importance function, $\pi(\gamma_{t,r}|\hat{\gamma}_{t-1}^{(i)}, z_{t,\gamma})$ [9]. For each of the resulting $I_t = P\, I_{t-1}$ particles, one KF realisation is evaluated to obtain a position particle, $\hat{\mathbf{p}}_t^{(i,p)}$, with covariance, $\boldsymbol{\Psi}_t^{(i,p)}$. The robot PHD hence is given by

$$\lambda(\mathbf{r}_t|\mathbf{Z}_{1:t}) = \sum_{i=1}^{I_{t-1}} \sum_{p=1}^{P} \alpha_t^{(i,p)} \delta_{\hat{\gamma}_t^{(i,p)}}(\gamma_{t,r}) \mathcal{N}(\mathbf{p}_t|\hat{\mathbf{p}}_t^{(i,p)}, \boldsymbol{\Psi}_t^{(i,p)}),$$

where $\delta_{\hat{\gamma}_t^{(i,p)}}(\gamma_{t,r})$ is the Dirac delta function of $\gamma_{t,r}$ evaluated at $\hat{\gamma}_t^{(i,p)}$, and $\alpha_t^{(i,p)}$ are the importance weights at t, given by

$$\alpha_t^{(i,p)} = \frac{\mathcal{L}(\boldsymbol{\Omega}_t|\mathbf{r}_t^{(i,p)})\, \hat{\alpha}_t^{(i,p)}}{\sum_{l=1}^{I_{t-1}} \sum_{m=1}^{P} \mathcal{L}(\boldsymbol{\Omega}_t|\mathbf{r}_t^{(l,m)})\, \hat{\alpha}_t^{(l,m)}}. \qquad (17)$$

$\hat{\alpha}_t^{(i,p)}$ are the un-normalised importance weights given by

$$\hat{\alpha}_t^{(i,p)} = \alpha_{t|t-1}^{(i,p)}\, p(z_{t,\gamma}|\hat{\gamma}_t^{(i,p)})\, p(z_{t,v}|\hat{\mathbf{p}}_{t|t-1}^{(i,p)}), \qquad (18)$$

where $p(z_{t,\gamma}|\hat{\gamma}_t^{(i,p)})$ and $p(z_{t,v}|\hat{\mathbf{p}}_{t|t-1}^{(i,p)})$ are the likelihood terms of the robot orientation and velocity obtained from the KF. The term $\mathcal{L}(\boldsymbol{\Omega}_t|\mathbf{r}_t^{(i,p)})$ in (17) is the multi-source evidence given by

$$\mathcal{L}(\boldsymbol{\Omega}_t|\mathbf{r}_t^{(i,p)}) \triangleq \prod_{m=1}^{M_t} \ell(\boldsymbol{\omega}_{t,m}|\mathbf{r}_t^{(i,p)}), \qquad (19)$$

for single-source likelihood terms, $\ell(\boldsymbol{\omega}_{t,m}|\mathbf{r}_t^{(i,p)})$, in (15).

Results

In order to investigate the impact of the choice of robot path on the mapping accuracy of the proposed a-SLAM approach, the following experiments were conducted by simulation. In Experiment 1, the orientation and height of the robot is investigated, whilst Experiment 2 evaluates the influence of the choice of robot speed.

Experiment 1: Robot path

A single source is located at $(4, 4, 1.195)$ m in a $15 \times 8 \times 2.5$ m^3 room. The robot travels at 20 waypoints in a straight line from starting position $(1, 4, 1.195)$ m at a constant speed of 0.2 m/s and with orientation angles between $[0, 90]$ deg, where 0 deg corresponds to a robot path in the the North direction and 90 deg corresponds to a path in the East direction. Assuming negligible error in the robot reports with $\sigma_{w_{t,v}} = 10^{-9}$ m/s and $\sigma_{w_{t,\gamma}} = 10^{-9}$ rad, the proposed a-SLAM approach is evaluated using 150 GM components. The source position error is evaluated as the Euclidean distance between the true and estimated source position for 50 Monte Carlo (MC) runs.

The resulting source mapping accuracy is plotted in Fig. 1. The mapping accuracy converges to under 5 cm

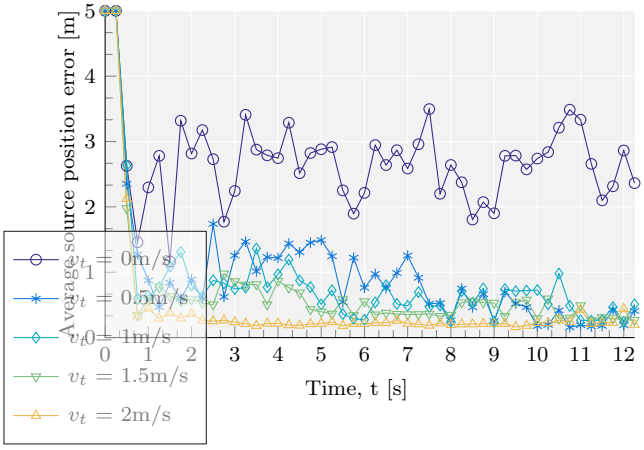

Abbildung 2: Source mapping accuracy for a robot speed between $[0, 2]$ m/s.

after 7 time steps for robot orientation angles of 50 and 70 deg. For an orientation angle of 0 deg, convergence to a comparable accuracy is achieved after 17 time steps. At 90 deg, i.e., facing the source, the estimate position error of the source diverges to 1.8 m. This divergence is due to the fact that the source cannot be triangulated when a robot at the same height heads straight towards the source.

The experiment is repeated for a source at 1.8 m height. Results for 50 and 70 deg robot orientation are comparable to the source co-located at robot height. The rate of convergence is significantly improved for 0 and 30 deg orientation, where less than 5 cm error are achieved within 6 time steps. Moreover, diversity in height results in convergence for a robot moving straight towards the source at absolute orientation angle of 90 deg.

Therefore, at $50 - 70$ deg robot orientation, i.e., as the robot "grazes" past the source, the azimuth between the source and sensor change significantly between each waypoint along the robot trajectory. In comparison, a robot orientation of 0 deg, i.e., perpendicular to the source, results in modest rate of change in the azimuth between consecutive waypoints. Spatial diversity can be further improved by increasing the difference in height between the source and sensor, such that the rate of change in source inclination is maximised between consecutive waypoints along the robot path.

Experiment 2: Robot speed

The trajectory of a moving robot and two static sources are simulated for 12.5 s with $\Delta_T = 250$ ms, such that $t = 1, \ldots, 50$. The robot trajectory is simulated using (5) with an initial position of $(3, 3, 1.195)$ m, initial orientation of $\gamma_0 = 0$ deg and constant velocity of 1.5 m/s, and for process noise with $\mathbf{\Sigma}_{t,\mathbf{v}} = \text{diag}\left[0.1\ \text{m}^2, 0.1\ \text{m}^2, 10^{-3}\ (\text{m/2})^2\right]$, and $\sigma_{v_{t,\gamma}} = 45$ deg. The reported orientation and velocity of the robot are simulated using (8) for $\sigma_{w_{t,v}} = 10^{-3}$ m/s and $\sigma_{w_{t,\gamma}} = 5$ deg unless otherwise specified in the experimental setup. The sources are initialised at $(1.5, 1.5, 1.7)$ m and $(4.5, 4.5, 1.5)$ m respectively. Small changes due to head and body rotations are simulated using (1) over time with $\mathbf{Q} = \text{diag}\left[0.01\ \text{m}^2, 0.01\ \text{m}^2, 10^{-3}\ (\text{m/s})^2\right]$. The source DoAs are simulated as an oracle localizer using (2) with Root Mean Square (RMS) error of $\mathbf{R}^{1/2} = \text{diag}\,[5, 2]$ deg for 10 MC runs. The source mapping accuracy is evaluated as the Euclidean distance between true and estimated source positions averaged over both sources.

The results are shown in Fig. 2. A static robot at 0 m/s speed results in an average error of 2.75 m between the estimated and true source positions, as the range to static sources cannot be inferred from a static robot. By exploiting spatial diversity of a moving robot, source mapping accuracy of up to 10 cm can be achieved. The results clearly indicate that source mapping accuracy can be improved by increasing the speed of the robot.

Conclusions

a-SLAM is required in order to localize the positions along the path of a moving microphone array and jointly map the surrounding sound sources. Using a novel a-SLAM approach, this paper investigated the impact of the choice of robot paths in terms of the robot orientation, height, and speed on the extent of source mapping accuracy. Simulation results demonstrated that a-SLAM performance can be improved by informatively planning robot paths in order to actively exploit time-varying source-sensor geometries.

Literatur

[1] 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.

[2] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*. Springer, 2010.

[3] L. George and A. Mazel, "Humanoid robot indoor navigation based on 2d bar codes: application to the nao robot," in *IEEE-RAS Intl. Conf. on Humanoid Robots (Humanoids)*, Oct 2013, pp. 329–335.

[4] C. Evers, A. H. Moore, and P. A. Naylor, "Acoustic simultaneous localization and mapping (a-SLAM) of a moving microphone array and its surrounding speakers," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016.

[5] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.

[6] B.-N. Vo and W.-K. Ma, "The Gaussian Mixture probability hypothesis density filter," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.

[7] S. Gannot and A. Yeredor, "The Kalman filter," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Springer-Verlag, 2008, ch. 8, part B.

[8] 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.

[9] A. Doucet, S. Godsill, and C. Andrieu, "On sequential monte carlo sampling methods for bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.