2019 American Control Conference (ACC)
Philadelphia, PA, USA, July 10-12, 2019

Tracking a Sound Source with Unknown Dynamics Using Bearing-Only Measurements Based on A Priori Information*

Masoud Jahromi Shirazi[1] and Nicole Abaid[1]

Abstract— **The problem of sound source localization has attracted the interest of researchers from different disciplines ranging from biology to robotics and navigation. It is in essence an estimation problem trying to estimate the location of the sound source using the information available to sound receivers. It is common practice to design Bayesian estimators based on a dynamic model of the system. Nevertheless, in some practical situations, such a dynamic model may not be available in the case of a moving sound source and instead, some a priori information about the sound source may be known. This paper considers a case study of designing an estimator using available a priori information, along with measurement signals received from a bearing-only sensor, to track a moving sound source in two dimensions.**

I. INTRODUCTION

Sound source localization, which is also referred to as passive sonar, considers the challenge of estimating the location of a source of an incoming sound. Based on the wide range of research areas interested in the problem of sound source localization, it is hard to find the origin of this problem in the literature. Perhaps one of the first studies on this problem is a set of experiments performed by Rayleigh [1] to find out how humans can localize the source of a sound. This is still an active field of research in psychoacoustics, and summaries of new developments in this area can be found in [2],[3]. Biologists are also interested in sound localization since some animals such as bats and whales use sound to perceive their environment [4]. Although sound localization originated as a research question in biology, it finds many applications in engineering, especially in robotics. The more recent advances in sound source localization and its application in robotics can be found in [5] and [6].

The problem of sound source localization involves many challenges. Information about the location of the sound source is usually extracted through analysing the magnitude and the time of arrival of the sound to different sensors. Therefore, the first challenge is detecting the sound and estimating the delay in the reception time across receivers. Many signal processing techniques, such as generalized cross-correlation and eigenvalue decomposition along with different models for sound propagation and reverberation, are implemented to achieve this goal. A comprehensive review of these approaches is presented in [7].

*This work was support by the National Science Foundation under grant #CMMI-1751498.

[1]The authors are with the Engineering Mechanics Program, Virginia Tech, 495 Old Turner Street, Blacksburg, VA 24061, USA mshirazi@vt.edu, nabaid@vt.edu

After the sound signals are received and processed, the next step is extracting the direction of the incoming sound and the distance it traveled using different acoustic cues. Among these cues, the difference of the arrival time of the sound to each receiver, usually referred as Time Difference of Arrival (TDoA), is commonly used to estimate the sound direction [8], [9]. Finding sound direction is more challenging when only two receivers are used, which mimics human localization (binaural localization). The complication in this case is due to geometrical symmetry called cone of confusion, or front-back confusion in the two dimensional case. This confusion can be resolved by breaking the symmetry using a head motion [10], or taking into account the acoustic shadow created by head defining a head model often referred as the Head Related Transfer Function (HRTF) [11], [12].

Range estimation, on the other hand, is a more challenging task in sound source localization and still is an active research area. In some approaches, the redundancy of several receivers is exploited to estimate the range of the sound source [8]. Another work estimates range using consecutive measurements as a head with a binaural sensor turns [13]. In indoor situations, some localization approaches take advantage of the reverberant sound energy to estimate the range [14]. The idea is that, while the energy of the sound signal received directly from the source depends on the distance between the source and sensor, the reverberated sound energy is independent of this distance. The work presented in [15] is another example of using environmental effects on the sound signal for localization by exploiting sound reflection and diffraction in an indoor environment.

Another approach in range estimation is to take advantage of sensor/source mobility by implementing some type of Bayesian estimation. In this approach, the localization problem essentially is a target tracking problem using a bearing-only sensor. The work represented in [16] employs a multiple mode Kalman filter to mitigate the effect of noise and uncertainty in tracking a moving sound source, but the authors assume the direct measurement of its location. They also assume three hypotheses for motion of the source–static, constant velocity, and constant acceleration–and based on Bayes' formula, estimate the probability of any of these assumptions. In [17], particle filtering is implemented to localize a moving sound source as the sensor is moving. In this work, the filter is designed assuming the sound source is stationary and then investigates what type of sensor motion will reduce the localization error in the case of a moving sound source. The work presented in [18] considers localization of an intermittent moving sound source using a

978-1-5386-7926-5/$31.00 ©2019 AACC

mixture Kalman filter. The location and speed of the sound source, as well as its activity status, are included in the state vector and therefore, a dynamic model is assumed for the motion of sound source and its activity.

As is briefly mentioned above, the sound localization algorithm based on Bayesian estimation assumes some dynamic model for the sound source to build transition probability. However, such a dynamic model may not always be available. Instead it may be possible to use some a priori information about the motion of the sound source. This information may come from our knowledge about the source. As an example, a beacon for guiding a group of vehicles using passive sonar may be designed to move with constant speed or generate sound with a specific pattern. This information can also be available through measurement, such as estimating speed based on the Doppler shift effect. This paper considers a case study to investigate the possibility of designing an estimator by exploiting a priori information combined with measurement data from a bearing-only sensor to perform sound localization.

II. MODELING

Consider the problem of localization of a sound source using a bearing-only sensor. As an example, a microphone array can be used as a bearing-only sensor that can measure the bearing angle of the incoming sound but not the range. The dynamics of the sound source is unknown but we know that the source is moving in the horizontal plane with constant speed v and it emits sound pulses at known intervals. Therefore, the measurement available to the sensor consists of the bearing angle of the incoming sound and the time the signal is received. To completely localize the sound source, an estimation algorithm is needed to extract the range of the sound source with respect to the sensor.

Although a dynamic model of the sound source is not available, it is possible to build up the dynamics and a measurement model using the existing information about its motion and the sensor measurement. To this end, the measurement vector is divided into two disjoint parts. The first part, which is referred to z^p, is used to build a dynamic model to predict the next state of the system. However, the second part of measurement vector, z^c, is used to build a measurement model to correct the prediction of states.

The time the sound reaches the sensor is related to the distance between the sound source and the sensor. When the time interval between pulses is known, it can be used to extract some information about the range of the source. If Δ is the time interval of the sound source and the k^{th} signal is received at time t_k when the sound source is at range r_k, the next signal is expected at time t_{k+1} when the source is at range r_{k+1} and the following relation holds:

$$r_{k+1} = r_k + c\,(t_{k+1} - t_k - \Delta) + \nu_k, \tag{1}$$

where c is the speed of sound, and ν_k is the process noise, which is a zero-mean random variable independent from the states of the system which models uncertainties and measurement noise on reception time. In this model, it is assumed that the speed of the sound source is negligible compared to the speed of the sound, and therefore, when the signal is received, the sound source is approximately at the same location as when it emits the sound.

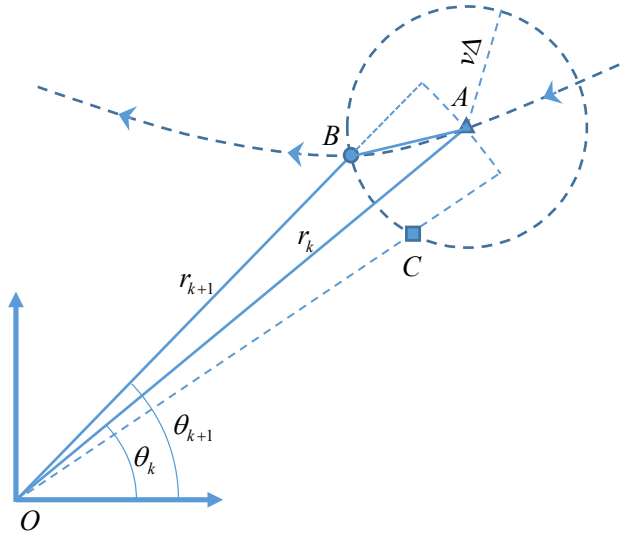

Fig. 1. A schematic of the sound source moving in the plane with constant speed.

Fig. 1 shows a schematic of the motion of the sound source. The sensor is located at the origin and the localization goal is to estimate the range r. The sound source is moving on the dashed line in the direction of the arrows. When the sound source is at the point A at time step k, it emits a sound and then moves along the path to reach the point B at time step *k+1*, when it generates another sound. Since the speed of the source and the time interval between pulses are known, the distance the sound source has traveled during this time can be approximated by $v\Delta$, assuming the path to be linear during this time period. Therefore, the location of point B is located on the circle with radius $v\Delta$ centered at A. Also, the range of B is known by (1), which reduces the possible location of the sound source at time step *k+1* to two points labeled as B and C in Fig. 1. Using the law of cosines in the triangle *OAB* gives us

$$(v\Delta)^2 = r_{k+1}^2 + r_k^2 - 2r_{k+1}r_k \cos\left(\theta_{k+1} - \theta_k\right). \tag{2}$$

where θ_k is the bearing of the sound source with respect to the sensor at time step k. Solving (2) for θ_{k+1} and then disturbing the result with a zero-mean independent noise ω_k leads to

$$\theta_{k+1} = \theta_k \pm \cos^{-1}\left(\frac{r_{k+1}^2 + r_k^2 - v^2\Delta^2}{2r_{k+1}r_k}\right) + \omega_k. \tag{3}$$

This equation shows the existence of two possible solutions for the triangulation as mentioned earlier (points B and C in Fig. 1). However, this confusion can be resolved when θ_{k+1} is measured by the sensor by choosing the sign that leads to closer value to the measured angle.

Although a model of the sound source motion is not available, using the known information about the sound

source augmented with some measurement information (i.e. time), this method enables us to build up a dynamic model for the sound source as well as a measurement model of the remaining measurement vector (i.e. bearing angle) as expressed in (1) and (3), respectively. Now it is possible to apply a model-based estimation technique to find the range. To this end, in the following section, an estimator based on minimum mean squared error (MMSE) is presented. We comment that the strategy for the design of the estimator follows analogous logic to the classical proof of an extended Kalman filter (EKF), but must take into account our novel formulation that uses parts of the measurement for both prediction and correction [19].

III. LINEAR MMSE ESTIMATION ALGORITHM

The dynamic model described in (1) can be expressed in the following format:

$$x_{k+1} = f\left(x_k, \nu_k; z_k^p, z_{k+1}^p\right), \tag{4}$$

where x_k and ν_k are state and process noise vectors. Also, z_k^p is the part of the measurement vector that is used for state prediction at time step k. It is important to note here that, since these measurements are available through the sensor, they are treated as a deterministic parameter while the state and process noise are stochastic variables.

The measurement model described in (3) can be expressed as

$$z_{k+1}^c = h\left(x_{k+1}, x_k; z_k^c\right) + \omega_k, \tag{5}$$

where z_k^c is the portion of the measurement vector used for correction of the predicted state at time step k.

The goal of this section is to design an iterative estimator to find the range of the sound source. In other words, assuming that the state estimate at time step k, $\hat{x}_k$, is known and that all the measurements z_k^p, z_{k+1}^p, z_k^c, and z_{k+1}^c are available, it is desired to find an expression for $\hat{x}_{k+1}$.

Using the Taylor expansion about point $(\hat{x}_k, 0)$, one can linearize the dynamic model (4) to approximate x_{k+1} by

$$x_{k+1} = f\left(\hat{x}_k, 0; z_k^p, z_{k+1}^p\right) + F_k\left(x_k - \hat{x}_k\right) + \Gamma_k \nu_k, \tag{6}$$

where $F_k = \left.\frac{\partial f}{\partial x_k}\right|_{(\hat{x}_k,0)}$ and $\Gamma_k = \left.\frac{\partial f}{\partial \nu_k}\right|_{(\hat{x}_k,0)}$. The MMSE estimator returns the expected value of the posterior as the estimation. Therefore, the goal is to calculate

$$\hat{x}_{k+1} = E\left[x_{k+1} \middle| Z^{k+1}\right], \tag{7}$$

where Z^k is a notation to show all the measurements taken up to and including time step k, and $E[\cdot]$ is the expected value. Taking expected value from the linearized dynamics (6), and noting that ν_k is a zero-mean random variable and is independent from the states, one can find the expected value of the prior, or the predicted state as follows:

$$\bar{x}_{k+1} = E\left[x_{k+1} \middle| Z^k, z_{k+1}^p\right] = f\left(\hat{x}_k, 0; z_k^p, z_{k+1}^p\right). \tag{8}$$

The covariance of the prior distribution is defined as

$$\bar{P}_{xx} = E\left[\left(x_{k+1} - \bar{x}_{k+1}\right)\left(x_{k+1} - \bar{x}_{k+1}\right)^T \middle| Z^k\right]. \tag{9}$$

Substituting x_{k+1} and $\bar{x}_{k+1}$ from (6) and (8) into (9) yields

$$\bar{P}_{xx} = F_k P_{xx} F_k^T + \Gamma_k Q \Gamma_k^T \tag{10}$$

where P_{xx} is the covariance of the predicted state and Q is the covariance of the process noise. One can use the predicted state and the measurement model in (5) to find the predicted measurement at the next time step. The measurement model can be linearized about $(\bar{x}_{k+1}, \hat{x_k})$ to approximate z_{k+1}^c by

$$\begin{aligned} z_{k+1}^c = {} & h\left(\bar{x}_{k+1}, \hat{x}_k; z_k^c\right) \\ & + \bar{H}_k\left(x_{k+1} - \bar{x}_{k+1}\right) + H_k\left(x_k - \hat{x}_k\right) + \omega_k, \end{aligned} \tag{11}$$

where $\bar{H}_k = \left.\frac{\partial h}{\partial x_{k+1}}\right|_{(\bar{x}_{k+1}, \hat{x}_k)}$ and $H_k = \left.\frac{\partial h}{\partial x_k}\right|_{(\bar{x}_{k+1}, \hat{x}_k)}$. Therefore, the predicted measurement can be found as the expected value of z_{k+1}^c. Using (11), and noting that the measurement noise is zero-mean and independent from the states, the expected measurement can be found as

$$\bar{z}_{k+1}^c = E\left[z_{k+1}^c \middle| Z^k\right] = h\left(\bar{x}_{k+1}, \hat{x_k}; z_k^c\right). \tag{12}$$

Using the linear MMSE estimator, the state estimate and its covariance can be found as [19]

$$\hat{x}_{k+1} = \bar{x}_{k+1} + \bar{P}_{xz}\bar{P}_{zz}^{-1}\left(z_{k+1}^c - \bar{z}_{k+1}^c\right), \tag{13}$$

$$P_{xx} = \bar{P}_{xx} - \bar{P}_{xz}\bar{P}_{zz}^{-1}\bar{P}_{xz}^T, \tag{14}$$

where $\bar{P}_{xz}$ is the covariance between predicted state and predicted measurement and $\bar{P}_{zz}$ is the predicted measurement covariance. To complete an iteration step, one needs to calculate $\bar{P}_{xz}$ and $\bar{P}_{zz}$. Using the definition of the covariance,

$$\bar{P}_{xz} = E\left[\left(x_{k+1} - \bar{x}_{k+1}\right)\left(z_{k+1}^c - \bar{z}_{k+1}^c\right)^T \middle| Z^k, z_{k+1}^p\right]. \tag{15}$$

Substituting z_{k+1}^c and $\bar{z}_{k+1}^c$ respectively from (11), and (12) leads to

$$\begin{aligned} \bar{P}_{xz} = {} & E\left[\left(x_{k+1} - \bar{x}_{k+1}\right)\left(x_{k+1} - \bar{x}_{k+1}\right)^T \bar{H}_k^T \middle| Z^k, z_{k+1}^p\right] \\ & + E\left[\left(x_{k+1} - \bar{x}_{k+1}\right)\left(x_k - \hat{x}_k\right)^T H_k^T \middle| Z^k, z_{k+1}^p\right] \\ & + E\left[\left(x_{k+1} - \bar{x}_{k+1}\right)\omega_k^T \middle| Z^k, z_{k+1}^p\right]. \end{aligned} \tag{16}$$

The last term in (16) vanishes since the measurement noise is assumed to be zero-mean and independent of the state variables. Substituting x_{k+1} and $\bar{x}_{k+1}$ from (6) and (8) into the second term in (16), leads to the following:

$$\bar{P}_{xz} = \bar{P}_{xx}\bar{H}_k^T + F_k P_{xx} H_k^T. \tag{17}$$

Similarly, one can calculate $\bar{P}_{zz}$ as follows

$$\begin{aligned} \bar{P}_{zz} = {} & E\left[\left(z_{k+1}^c - \bar{z}_{k+1}^c\right)\left(z_{k+1}^c - \bar{z}_{k+1}^c\right)^T \middle| Z^k, z_{k+1}^p\right] \\ = {} & E\left[\bar{H}_k\left(x_{k+1} - \bar{x}_{k+1}\right)\left(x_{k+1} - \bar{x}_{k+1}\right)^T \bar{H}_k^T \middle| Z^k, z_{k+1}^p\right] \\ & + E\left[\bar{H}_k F_k\left(x_k - \hat{x}_k\right)\left(x_k - \hat{x}_k\right)^T H_k^T \middle| Z^k, z_{k+1}^p\right] \\ & + E\left[H_k\left(x_k - \hat{x}_k\right)\left(x_k - \hat{x}_k\right)^T \bar{H}_k^T F_k^T \middle| Z^k, z_{k+1}^p\right] \\ & + E\left[H_k\left(x_k - \hat{x}_k\right)\left(x_k - \hat{x}_k\right)^T H_k^T \middle| Z^k, z_{k+1}^p\right] \\ & + E\left[\omega\omega^T \middle| Z^k, z_{k+1}^p\right] \\ = {} & \bar{H}_k\bar{P}_{xx}\bar{H}_k^T + \bar{H}_k F_k P_{xx} H_k^T + H_k P_{xx}\bar{H}_k^T F_k^T \\ & + H_k P_{xx} H_k^T + R, \end{aligned} \tag{18}$$

where R is the covariance of the measurement noise. Again, the fact that the process and measurement noises are zero-mean and independent from state variables is used to simplify the expression for $\bar{P}_{zz}$. Now it is possible to calculate the estimated state and its covariance for the next time step using (13) and (14), respectively. A summary of one iteration of the algorithm is presented in table I.

The last step to finish the design of the estimator is to assume initial state and initial covariance. Similar to the EKF, this estimator operates based on linearization of the dynamic and measurement equations about the best estimate of the previous state. Therefore, the performance of the estimator strongly depends on how close the initial state estimation is to reality. This initial state and its covariance can be set using some a priori information about the initial state or through measurement. Here, we used a simple approach to get a rough estimate of source location. The change in bearing angle, $d\theta$, is measured by finding the difference between two consecutive bearing measurements. Then, assuming that the sound source is moving on a circular path, one can estimate the initial position to be

$$x_0 = \frac{v\Delta}{|d\theta|}, \tag{19}$$

and its variance is set to be equal to $v\Delta$, the amount that the sound source is moving between two pulses.

Finally, it is important to note a limitation of this approach. As it was mentioned before, the general idea of this localization approach is to use the measured time of arrival of the sound to predict its source's next location while using the bearing angle to correct the prediction, compensate the effect of noise, and improve localization confidence. However, when the sound source is moving radially toward or away from the sensor, there is no change in bearing angle and therefore implementing the correction step is out of the question. In fact, due to the existence of noise in the bearing angle measurement, the same problem occurs when the change in angle is small compared to the standard deviation of the bearing angle measurement noise. In these situations, the measured bearing angle is dominantly noise and, if it is used for correction step, it may lead to divergence of the algorithm. Therefore, we used a metric similar to signal-to-noise ratio to evaluate the information in the measurement and skip the correction step when the signal is dominated by noise. If the change in bearing angle is shown by $d\theta$, we define this ratio ρ, measured in dB, as

$$\rho = 20 \log \frac{|d\theta|}{\sigma_\theta}, \tag{20}$$

where σ_θ is the standard deviation of the bearing angle measurement noise. This metric defines a bound around zero within which we assume the measurement is dominated by noise.

TABLE I

AN ITERATION OF THE ESTIMATION ALGORITHM

Inputs: $\hat{x}_k, P_{xx}, z^p_k, z^p_{k+1}, z^c_k, z^c_{k+1}$
Outputs: $\hat{x}_{k+1}, P_{xx}$
Prediction:
$\bar{x}_{k+1} = f\left(\hat{x}_k, 0; z^p_k, z^p_{k+1}\right)$
$F_k = \frac{\partial f}{\partial x_k}\Big\vert_{(\hat{x}_k,0)}$, $\Gamma_k = \frac{\partial f}{\partial x_k}\Big\vert_{(\hat{x}_k,0)}$
$\bar{P}_{xx} = F_k P_{xx} F_k^T + \Gamma_k Q \Gamma_k^T$
Correction:
$\bar{z}^c_{k+1} = h\left(\bar{x}_{k+1}, \hat{x_k}; z^c_k\right)$
$H_k = \frac{\partial h}{\partial x_k}\Big\vert_{(\bar{x}_{k+1},\hat{x}_k)}$, $\bar{H}_k = \frac{\partial h}{\partial x_{k+1}}\Big\vert_{(\bar{x}_{k+1},\hat{x}_k)}$
$\bar{P}_{xz} = \bar{P}_{xx}\bar{H}_k^T + F_k P_{xx} H_k^T$.
$\bar{P}_{zz} = \bar{H}_k \bar{P}_{xx} \bar{H}_k^T + \bar{H}_k F_k P_{xx} H_k^T + H_k P_{xx} \bar{H}_k^T F_k^T + H_k P_{xx} H_k^T + R$
$\hat{x}_{k+1} = \bar{x}_{k+1} + \bar{P}_{xz}\bar{P}_{zz}^{-1}\left(z^c_{k+1} - \bar{z}^c_{k+1}\right)$
$P_{xx} = \bar{P}_{xx} - \bar{P}_{xz}\bar{P}_{zz}^{-1}\bar{P}_{xz}^T$
$k = k+1$ and repeat

IV. SIMULATION RESULTS

To examine the performance of the proposed localization algorithm, numerical simulation is used. The sound source is moving with constant speed of 1 m/s in x-y plane. The bearing-only sensor is located at the origin and the bearing angle is measured from positive x-axis. The value of pulse intervals is set to be $\Delta = 0.2$ s and the variance of process noise is assumed to be $Q = 4 \times 10^{-4}$ m^2. The standard deviation of the measurement noise in sensing the bearing angle and time of arrival are respectively set to $\sigma_\theta = 0.005$ rad, and $\sigma_t = 0.1$ ms. Therefore the measurement noise covariance matrix is equal to $R = \mathrm{diag}\left[\left(\sigma_\theta^2, \sigma_t^2\right)\right]$, where $\mathrm{diag}\left[v\right]$ is a diagonal matrix in which the elements of vector v are located on the main diagonal.

Two different paths are considered for the sound source. In the first simulation, the sound source starts from point (1,1) and moves on a circle around the origin in the counter clockwise direction. Fig. 2 shows the actual (blue dashed line) and tracked (solid red line) paths of the sound source. Time series for the range estimation error, its covariance, and the change in measured bearing angle of sound source in this case are shown in Fig. 3. The red dashed lines show the bound when ρ is set to be 24 dB for simulation purposes. In practice, this range should be chosen based on the sensor accuracy and bearing angle estimation approach.

In the second simulation, the sound source is moving on a straight line. The simulation parameters are the same as mentioned in the first paragraph of this section. In this case, the sound source starts from (2,1) and moves horizontally to the left. The actual path and the estimated one are shown in Fig. 4. The range estimation error, estimation covariance, and the measured change in bearing angle for this case is shown in Fig. 5. Again, the red dashed lines show the bound when ρ is set to be 24 dB.

In order to investigate the robustness of the designed estimator, Monte Carlo simulations are performed for 100 repetition. The mean value of the range estimation error$\pm$ one standard deviation for both paths are shown in Fig. 6.

V. DISCUSSION

The algorithm developed in this work is the first of its kind, to the best of our knowledge, that uses part of

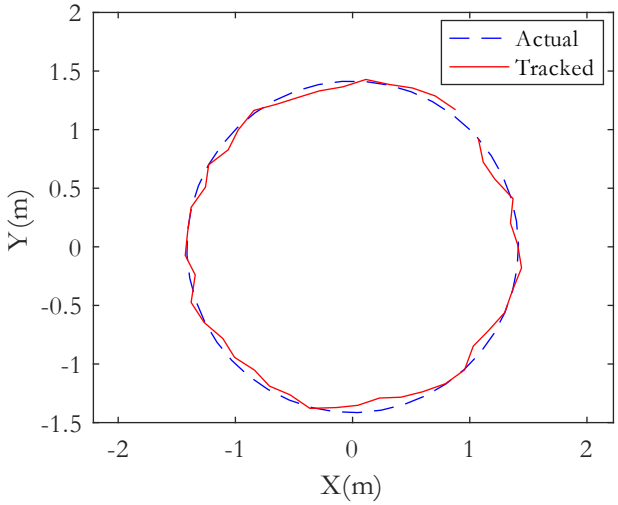

Fig. 2. The actual path and the path tracked by the algorithm when the sound source moves on a circular, counter clockwise path around the sensor.

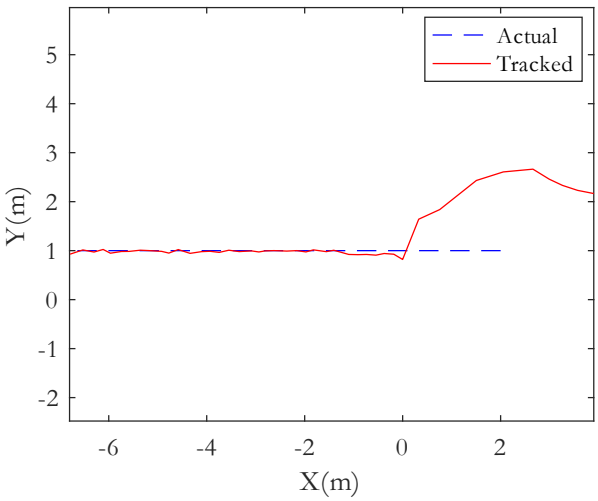

Fig. 4. The actual path and the path tracked by the algorithm when the sound source moves on a linear path.

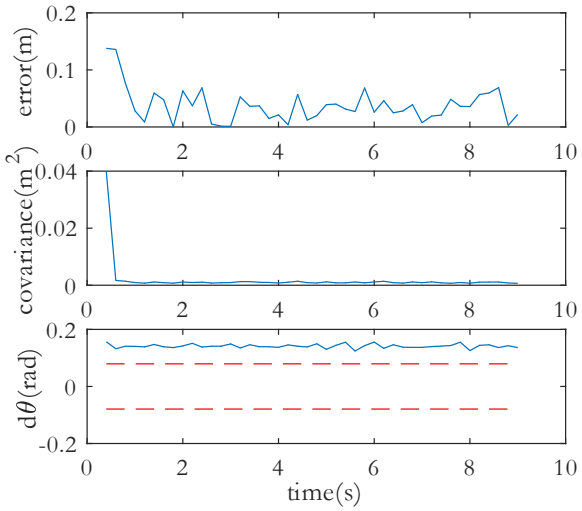

Fig. 3. The range estimation error (top), its covariance (middle), and the change in measured bearing angle (bottom) of the sound source moving on a circular path around the sensor. The dashed red lines in the bottom plot show the bound when ρ is set to be 24 dB.

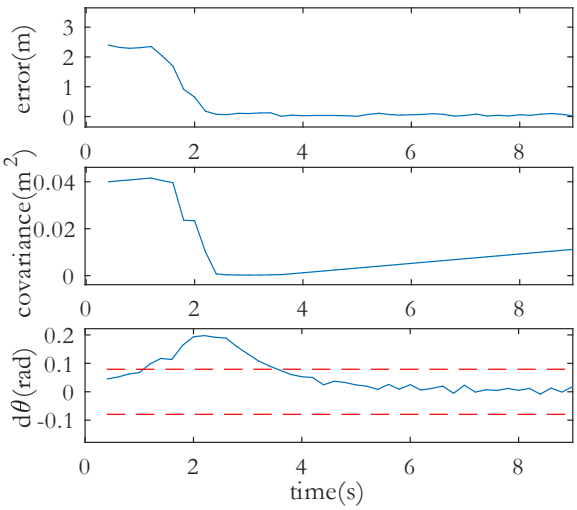

Fig. 5. The range estimation error (top), its covariance (middle), and the change in measured bearing angle (bottom) of the sound source moving on a linear path. The dashed red lines in the bottom plot show the bound when ρ is set to be 24 dB.

a measurement for prediction within a standard Bayesian estimation procedure. As a result, we are able to use bearing-only measurements to track a sound source with unknown dynamics. The simulation results show that the algorithm can estimate the sound source location using the available information and measurement. In Fig. 2, the assumed initial range of the sound source is close to the actual value and the estimator can track the sound source in presence of process and measurement noise. Fig. 3 shows the performance of the estimator. The first plot shows range estimation error is relatively small although there are fluctuations due to the noise in the system. The covariance of estimation drops rapidly, which shows the confidence of the algorithm. This confidence can lead to inconsistency of the estimator mainly due to linearization, as it is well-known and studied for EKF [19]. The last plot shows that, in this example, the change in measured bearing angle is large compared to the noise standard deviation as it is outside the bounds set by ρ and therefore, these measurements are informative.

In the second case study, as it is shown in Fig. 4, the initial range estimation is relatively far from the actual state, but still the algorithm manages to converge. The first plot in Fig. 5 shows the range estimation error reduces to near zero. The estimation covariance decreases rapidly and then starts to increase. This increase in covariance happens when the change in bearing angle is perceived to be corrupted by noise, as the sound source moves away from the sensor at constant speed. Therefore the correction step is skipped and the covariance increases due to noise accumulation. Finally, Monte Carlo simulations show that, given an initial state and noise covariances, the estimation algorithm results are robust over multiple runs.

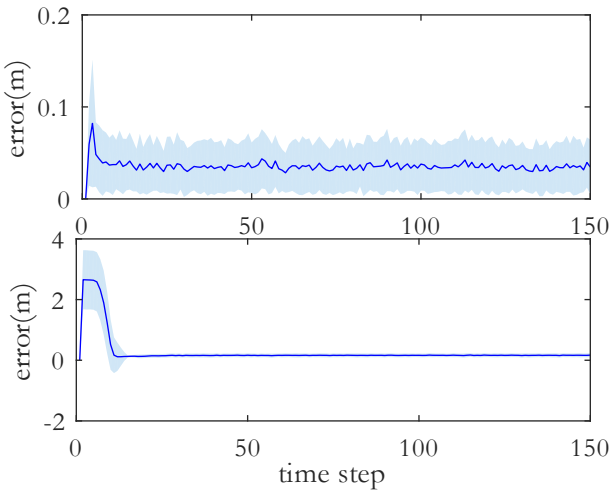

Fig. 6. The mean value of the range estimation error over 100 Monte Carlo simulations for the circular path (above) and the linear path (below). Shaded region shows $\pm$ one standard deviation.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, the tracking problem of a moving sound source using a bearing-only sensor is studied when a dynamic model of the sound source is not available. Using information about its motion as well as some measurement information, a dynamic and measurement model is developed for the sound source. This model is used to predict the next location of the sound source and the remaining measurement information is exploited to correct this estimation. An EKF-type estimator is designed using MMSE estimation and its performance is studied using numerical simulation.

Although the simulation results show the convergence of the algorithm, there are still practical challenges in implementing the algorithm. The most important challenge is the limited range of operation which is related to use of bearing-only sensors. When the sound source is radially moving towards or away from the sensor, or when it is located far enough from the sensor that its bearing angle does not change significantly between two consequent pulses, the change in bearing angle is dominated by noise and implementing the correction update leads to bad estimates of the sound source location. This effect can be mitigated using accurate sensors, i.e. smaller noise covariance, which increases the cost in practical applications.

Another way to improve the performance of the algorithm even in higher noise covariance is to improve estimation of the initial range and its covariance. A more rigorous way to estimate initial state (compared to that used in this work) is to record a fixed amount of samples and use a smoothing algorithm to improve the initial estimation [19]. Since the success or failure of tracking depends on the accuracy of the estimation of the initial state and the parameters defining both the process and measurement noise, a parameter study to understand the dependency of the algorithm's success on these factors is necessary. These improvements are the main direction of our future work on this problem.

REFERENCES

[1] L. Rayleigh, "On our perception of the direction of a source of sound," *Proceedings of the Musical Association*, vol. 2, pp. 75–84, 1875.

[2] A. J. Kolarik, B. C. Moore, P. Zahorik, S. Cirstea, and S. Pardhan, "Auditory distance perception in humans: A review of cues, development, neuronal bases, and effects of sensory loss," *Attention, Perception, & Psychophysics*, vol. 78, no. 2, pp. 373–395, 2016.

[3] G. C. Stecker and F. J. Gallun, *Binaural Hearing, Sound Localization, and Spatial Hearing*. Plural Publishing Inc., 2012, ch. Translational Perspectives in Auditory Neuroscience: Normal Aspects of Hearing, pp. 383–433.

[4] A. Surlykke, P. E. Nachtigall, R. R. Fay, and A. N. Popper, Eds., *Biosonar*. Springer, 2014.

[5] S. Argentieri, P. Danès, and P. Souères, "A survey on sound source localization in robotics: From binaural to array processing methods," *Computer Speech and Language*, vol. 34, p. 87112, 2015.

[6] C. Rascon and I. Meza, "Localization of sound sources in robotics: A review," *Robotics and Autonomous Systems*, vol. 96, pp. 184–210, 2017.

[7] J. Chen, J. Benesty, and Y. A. Huang, "Time delay estimation in room acoustic environments: An overview," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, p. 026503, 2006.

[8] A. Mahajan and M. Walworth, "3D position sensing using the differences in the time-of-flights from a wave source to various receivers," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 1, pp. 91–94, 2001.

[9] Y. Shim, J. Park, and J. Kim, "Relative navigation with passive underwater acoustic sensing," in *12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*. IEEE, 2015, pp. 214–217.

[10] L. Kneip and C. Baumann, "Binaural model for artificial spatial sound localization based on interaural time delays and movements of the interaural axis," *Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 3108–3119, 2008.

[11] B. G. Shinn-Cunningham, S. Santarelli, and N. Kopco, "Tori of confusion: Binaural localization cues for sources within reach of a listener," *Journal of Acoustical Society of America*, vol. 107, pp. 1627–1636, 2000.

[12] S. J. Cho, A. Ovcharenko, and U. p. Chong, "Front-back confusion resolution in 3d sound localization with HRTF databases," in *The 1st International Forum on Strategic Technology*, 2006, pp. 239–243.

[13] D. Tamsett, "Binaural range finding from synthetic aperture computation as the head is turned," *Robotics*, vol. 6, no. 2, p. 10, 2017.

[14] Y.-C. Lu and M. Cooke, "Binaural estimation of sound source distance via the direct-to-reverberant energy ratio for static and moving sources," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1793–1805, 2010.

[15] K. Takami, T. Furukawa, M. Kumon, and L. C. Mak, "Non-field-of-view indoor sound source localization based on reflection and diffraction," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Sept 2015, pp. 59–64.

[16] Z. Liang, X. Ma, and X. Dai, "Robust tracking of moving sound source using multiple model kalman filter," *Applied Acoustics*, vol. 69, no. 12, pp. 1350–1355, 2008.

[17] Y. Lu and M. Cooke, "Motion strategies for binaural localisation of speech sources in azimuth and distance by artificial listeners," *Speech Communication*, vol. 53, no. 5, pp. 622 – 642, 2011.

[18] Q. V. Nguyen, F. Colas, E. Vincent, and F. Charpillet, "Localizing an intermittent and moving sound source using a mobile robot," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 1986–1991.

[19] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: Theory algorithms and software*. John Wiley & Sons, 2004.