



Experimental tracking of an ultrasonic source with unknown dynamics using a stereo sensor

Aidan J. Bradley^a, Masoud Jahromi Shirazi^b, Nicole Abaid^{c,*}

^a Engineering Mechanics Program, Virginia Tech, Blacksburg 24061, VA, USA

^b Halliburton, 3000 North Sam Houston Pkwy East, Houston 77032, TX, USA

^c Department of Mathematics, Virginia Tech, Blacksburg 24061, VA, USA

ARTICLE INFO

Article history:

Received 6 April 2023

Received in revised form 25 May 2023

Accepted 18 June 2023

Keywords:

Stereo sensing

Bioinspiration

Linear MMSE estimation

Sound source localization

ABSTRACT

Sound source localization (SSL) is the ability to successfully estimate the bearing and distance of a sound in space relative to the sensing position and pose. SSL as a topic of interest for engineers often revolves around the ability of robots to track other robots, human voices, or other acoustic objects. Common approaches to this goal frequently use large arrays, computationally intensive and complex machine learning methods, or require known dynamic models of a system which may not always be available. In this work we seek to experimentally verify a solution to SSL using a minimal amount of inexpensive equipment on a two microphone, or stereo, sensing platform. A previously developed Bayesian estimator allows for localization of an emitter using easily available *a priori* information and timing data received from the sensor platform. Our results show that our approach is accurate for the tested paths and that the estimator can correct itself when dynamic assumptions are broken for short times due to hardware and software limitations.

© 2023 Elsevier Ltd. All rights reserved.

1. Introduction

The ability to know where a sound originates in space, known as sound source localization (SSL), has interested biologists for centuries and roboticists for decades [27,34,22,15,26]. A large contemporary portion of this research in both biological and robotic sensing revolves around the phenomenon of ultrasonic echolocation in bats [38,16,31], which is characterized not just by their high frequency sensing but by their abilities of Doppler shift compensation [28], multiple target acquisition and tracking [12], and the use of passive sensing to avoid jamming [6]. The feats of these small mammals make them a unique source of inspiration, as we continue to push the boundaries of compact sonar and ultrasonic sensing strategies for robotic systems.

Sonar or ultrasonic sensing systems are categorized as active or passive, distinguished by the ultrasonic source they use to obtain information from their environment [32,33]. Active systems use projectors to transmit acoustic or ultrasonic pulses that reflect off surrounding objects to transducers, which translate the data to be used to localize the source of the reflections. The main advantages of active systems are that the shape and transmission time of the acoustic pulse are known to the sender, allowing for ranging

and echo classification to be more easily accomplished. Passive systems rely on discriminating the transmissions or radiated noise of a target of interest from background and self noise, losing the advantages of prior knowledge of the signal. Passive systems do have advantages in that they sense their surroundings without revealing their position and they tend to be less directional than active systems. While there has been more research into stereo passive SSL in recent years, [4,39,14], there are still many interesting questions left to answer. The system used in this paper is a stereo, passive, ultrasonic sensor, enabling us to explore a challenge not previously investigated, by removing the ability to use triangulation for SSL from multiple sensors.

For SSL to be successful, two main questions must be answered:

1. What is the bearing of the source with respect to the sensor?
2. What distance is the source from the sensor?

The bearing of ultrasonic sources, or direction of arrival (DOA), estimation is the easier of the questions to answer via interaural time difference (ITD) [19,4,13], or complex interaural level difference (ILD) [24,25]. For single ultrasonic sources, a reliable and relatively simple method of DOA estimation is the use of generalized cross correlation with phase transform weighting prefilters (GCC-PHAT) where the time of maximum correlation is the delay between the sensors [18]. This process can be used in a stereo

* Corresponding author.

setup but does not solve the confounding factor of the so-called 'cone of confusion' [34], which arises due to the geometry of the stereo array creating identical time delays between the semicircles separated by the array plane. This can be solved by breaking the symmetry of the array through rotation or using an artificial head in conjunction with a head related transfer function (HRTF) [8,23,4,39,35,14].

Distance estimation has been solved with a number of solutions including: triangulation using time difference of arrival (TDOA) measurements from three or more microphones, direct-to-reverberant ratio estimation techniques, and learning based approaches [26]. These techniques work successfully but can have limits dependent on the size of the equipment, *a priori* information about the operating environment and dynamics of emitter if motion is involved, or available computing power.

By repeatedly performing SSL, its functionality as a way to track a dynamic emitter becomes apparent. One of the most popular estimation techniques used for localization and tracking is Kalman filtering (or its nonlinear counterpart extended Kalman filtering, EKF). In [36], a three-dimensional human tracking system is created using fused data from a monocular camera and ultrasonic sensor in an EKF. Recent work in [9] uses an EKF with time delay of arrival estimation to show a minimum of 4 distributed agents, connected via a communication network, are needed create a generically observable localization process where each agent can accurately track the target. An EKF is also employed in [39] to track multiple sources using ITD and head rotation data. Many other tracking methods and analyses can be found in the review papers [2,10]. We fill a gap in the research by showing that it is possible to localize an ultrasonic source with an EKF-style algorithm, using a static stereo sensor, and without the need for the previously discussed HRTFs or machine learning algorithms.

The work presented in this paper is a proof-of-concept experimental investigation into performing SSL with minimal *a priori* information and an inexpensive stereo sensor. Here, we use a custom sensor to generate data used by an EKF-style estimation algorithm presented in [30]. This paper builds upon our previous work in [5] in three main ways:

- We introduce an improved sensor platform which contains a hardware filter and noise-protected design.
- The ability to measure the time between received pulses has been added to the sensor, allowing us to use the full algorithm developed in [30], which was lacking in [5].
- A new circular experiment platform allows us to hold to the necessary assumption of constant sound source speed for longer periods of time compared to [5] and observe how the sensor and algorithm perform.

To enable the estimation of range we assume that the source has known constant speed v and emission interval Δ (here after referred to as inter-pulse time (IPT)). These conditions may be achievable in a situation where multiple, cooperating, echolocating robots are implemented. Given the assumptions above, we show that it is possible to localize a source in the horizontal plane with unknown dynamics using our stereo sensing system.

2. Dynamic model

Toward performing this estimation, we note that the travel time of signals reaching the sensor is related to the velocity of the source relative to the sensor due to the Doppler effect. When the emitter is moving at a fixed distance from the sensor, the measured inter-pulse time (IPT_z) is the same as the true IPT; if it is moving towards or away from the sensor, IPT_z decreases or increases

respectively. By assuming the IPT, Δ , is constant and known, it is possible calculate distance information from the source. By recording the time between two sensed pulses Δ_k and having a current estimate of the source distance x_k , at time step k , we can use the following relation to estimate the distance at the next time step:

$$x_{k+1} = x_k + c(\Delta_k - \Delta) + v_k, \quad (1)$$

where c is the speed of sound in air (taken as 343 m/s) and v_k is the process noise, assumed to be a zero mean random variable that is independent from the states of the system with realizations generated independently at each time step.

A schematic of the experimental setup with source motion is shown in Fig. 1. The source is moving around the sensor at speed v . At time step k , the source is located at A and emits a pulse; it continues to travel along the path for time Δ to location B where emits a second pulse. Using the triangle OAB formed by the two emission locations and sensor position, the rule of cosines can be used to find:

$$(v\Delta)^2 = x_{k+1}^2 + x_k^2 - 2x_{k+1}x_k \cos(\theta_{k+1} - \theta_k). \quad (2)$$

where θ_k is the bearing of the source with respect to the direction of the sensor at time step k . Manipulating this equation allows us to solve explicitly for θ_{k+1} and define a measurement model as:

$$\theta_{k+1} = \theta_k \pm \cos^{-1} \left(\frac{x_{k+1}^2 + x_k^2 - v^2 \Delta^2}{2x_{k+1}x_k} \right) + \omega_k, \quad (3)$$

where ω_k is the measurement noise, assumed to be a zero mean random variable that is independent from the system states and generated independently at each time step. It should be noted there is a sign uncertainty present in (3) which can be resolved by choosing whichever sign minimizes the error between the outcome of the measurement model (3) and the measured bearing angle at $k+1$ [30].

As the sensor only measures ITD, we convert to a bearing angle measurement by:

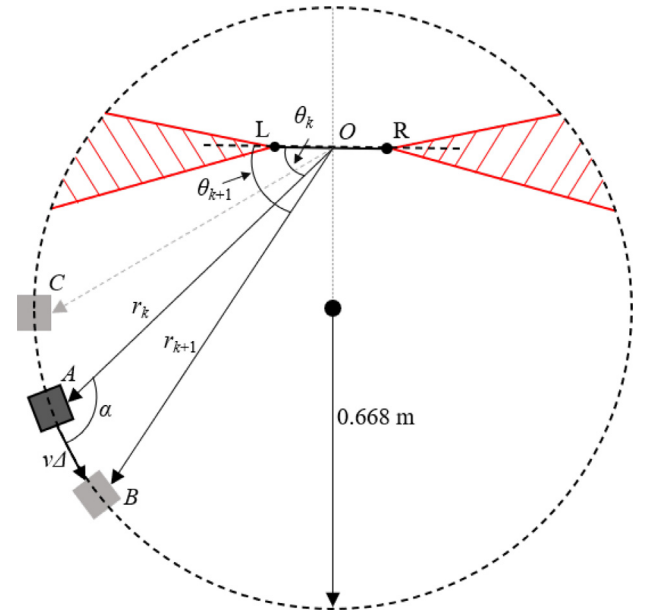


Fig. 1. A schematic of the noncentral experiment setup with a radius 0.668 m. The current position of the emitter is represented as point A at time step k with B and C being potential positions for the emitter at time step $k+1$. Position O is the origin of the array with the left and right microphones being respectively represented by L and R. The 0.1692 radian areas where the sensor is blind are shown by the striped area.

$$\theta_k = \cos^{-1} \left(\frac{c\tau_0}{f_s d} \right). \quad (4)$$

where τ_0 is the ITD of the source in samples, f_s is sampling frequency, and d is the inter-microphone distance [13]. This is developed from the assumption that the incoming signal is a plane wave, allowing for the use of the geometry between the microphones and the wave to calculate the bearing of the signal.

The above calculations are built on the assumption that the sensor is operating in the far field of the emitter, meaning the signal can be assumed to be traveling as a plane wave. As discussed in [26] the assumption of far-field operation “greatly simplifies the mapping procedure between feature and location”, though one must take care to not enter the near field while using a strictly far-field model due to the change from planar to circular sound waves decreasing SSL performance [1]. According to [17], for plane piston transducers, the far field region can be approximated by $r \gg a^2/\lambda$, where r is the distance between the source and the sensor, a is the radius of the source's transducer, and λ is the wavelength of the signal. For our experiments, $a^2/\lambda = 0.019^2/0.007 = 0.0516$ m and the minimum distance between the source and sensor happens when the sensor is non-central with the emitter directly behind the sensor, where $r \approx 0.25$ m. At just under five times greater than the requirement, the sensor should always operate under the far field assumption.

3. Methods

3.1. Stereo microphone array

The sensor used for these experiments is an updated version of the sensor found in [5]. Data collection is done with a Teensy microcontroller (Teensy 4.0, PJRC, USA) in conjunction with two Dodotronic Momimic analog ultrasonic microphones (Momimic Analog Microphone, Dodotronic, Italy), see Fig. 2. To help decrease noise and unwanted signals, both microphone outputs are connected to a 4th order low-pass Chebyshev filter with cutoff frequencies above our Nyquist frequency of 100 kHz and an RC high pass filter to reduce effects of noise below ultrasonic frequencies.

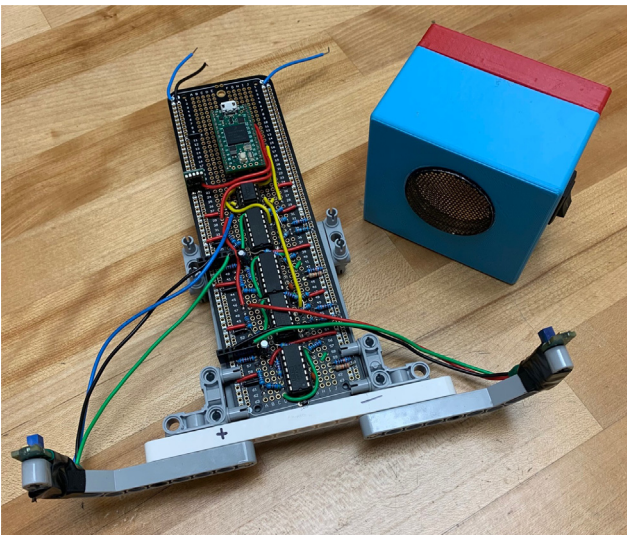


Fig. 2. The sensor, with microphones connected to stalks at bottom of photo, and emitter, in the blue and red box in upper right, used during testing.

The inspiration for this sensor comes from the frequency ranges and shapes of the calls of echolocating bats [28]. The Teensy's ARM Cortex M7 CPU can operate both of its analog-to-digital converters at sub $2 \mu\text{s}$ speeds while also measuring time between pulses down to 0.0001 s, a feat the sensor had not yet been programmed to do in [5]. This allows measurements of signals with frequencies beyond 20–60 kHz (most commonly used by insectivorous bats that hunt in mid air [11]) and detection of distance changes at reasonably low speeds. It further enables us to remove the assumption of a quasi-static sensor as done in [5].

As the sensor is stereo, all measurements will suffer from front-back ambiguity (i.e. cone of confusion [34]), but it is well documented that this ambiguity can be resolved by rotating the array and accounting for this rotation in the dynamic model of the system, see for example [14,39,23]. As this sensor will eventually be mounted on a robotic platform with all the associated complexity, we chose to keep the sensor static and post process the data accordingly.

3.2. Linear MMSE estimation algorithm

A full derivation of the algorithm can be found in [30]; here we will briefly discuss the main ideas about how we can linearize the problem to make use of a linear minimum mean square error (MMSE) estimator. To initialize the algorithm at time step $k = 0$, we define a rough initial estimate of the range (state) using:

$$x_0 = \frac{v\Delta}{|d\theta|} \quad (5)$$

and the initial state error covariance as $P_0 = v\Delta$, as done in [30]. To simplify notation we will express the dynamic model (1) as:

$$x_{k+1} = f(x_k, v_k; z_{k+1}^p), \quad (6)$$

where x_k is the stochastic state vector, v_k is the process noise vector, and z_{k+1}^p is IPT_z, a deterministic value. The measurement model (4) will be expressed as:

$$z_{k+1}^c = h(x_{k+1}, x_k; z_k^c) + \omega_k, \quad (7)$$

where z_k^c is the measured bearing used for the correction of the predicted state, and ω_k is the zero mean, independent measurement noise at time step k .

The linear MMSE algorithm estimates the state at the next time step as in [3]:

$$\hat{x}_{k+1} = \bar{x}_{k+1} + \bar{P}_{xz} \bar{P}_{zz}^{-1} (z_{k+1}^c - \bar{z}_{k+1}^c), \quad (8)$$

$$P_{xx} = \bar{P}_{xx} - \bar{P}_{xz} \bar{P}_{zz}^{-1} \bar{P}_{xz}^T, \quad (9)$$

where $\hat{x}_{k+1}$ denotes the estimated state and is equal to the expected value of the posterior distribution of x , $\bar{x}_{k+1}$ denotes the predicted state and is equal to the expected value of the prior distribution of x , z_{k+1}^c is the measured bearing angle, and $\bar{z}_{k+1}^c$ is the predicted measurement. The covariance of the predicted state, predicted measurement, and between the predicted state and measurement are represented as $\bar{P}_{xx}$, $\bar{P}_{zz}$, and $\bar{P}_{xz}$, respectively. Finally, P_{xx} shows the covariance of the estimated state.

To calculate all the unknown terms of Eqs. (8) and (9), one can perform a Taylor expansion of Eqs. (6) at the point $(\hat{x}_k, 0)$:

$$x_{k+1} = f(\hat{x}_k, 0; z_{k+1}^p) + F_k(x_k - \hat{x}_k) + \Gamma_k v_k. \quad (10)$$

where F_k and Γ_k are the partial derivatives of f with respect to x_k and v_k , respectively, evaluated at the point $(\hat{x}_k, 0)$. Then, assuming independent zero-mean process and measurement noises, the predicted state and its covariance can be calculated using this linearized model prior to measurement z_{k+1}^c as:

$$\bar{x}_{k+1} = E[\bar{x}_{k+1}|Z^k, z_{k+1}^p] = f(\hat{x}_k, 0; z_{k+1}^p), \quad (11)$$

$$\bar{P}_{xx} = E[(x_{k+1} - \bar{x}_{k+1})(x_{k+1} - \bar{x}_{k+1})^T | Z^k, z_{k+1}^p] = F_k P_{xx} F_k^T + \Gamma_k Q_k \Gamma_k^T, \quad (12)$$

where Z^k denotes all the measurements received before time step $k+1$ and Q_k is the process noise covariance. Similarly, linearizing Eq. (7) about point $(\bar{x}_{k+1}, \hat{x}_k)$ yields:

$$z_{k+1}^c = h(\bar{x}_{k+1}, \hat{x}_k; z_k^c) + \bar{H}_k(x_{k+1} - \bar{x}_{k+1}) + H_k(x_k - \hat{x}_k) + \omega_k \quad (13)$$

where H_k and $\bar{H}_k$ are respectively the partial derivatives of h about $(h_{k+1}, \hat{x}_k)$ with respect to x_{k+1} and x_k . Finally, taking the expected value of Eq. (13) and using the definition of the covariance, we can find the following expressions:

$$\bar{z}_{k+1}^c = h(\bar{x}_{k+1}, \hat{x}_k; z_k^c), \quad (14)$$

$$\bar{P}_{xz} = \bar{P}_{xx} \bar{H}_k^T + F_k P_{xx} H_k^T, \quad (15)$$

$$\bar{P}_{zz} = \bar{H}_k \bar{P}_{xx} \bar{H}_k^T + \bar{H}_k F_k P_{xx} H_k^T + H_k P_{xx} H_k^T + R_k, \quad (16)$$

where R_k denotes the covariance of the measurement noise ω_k . One iteration step is summarized in Algorithm 1.

An interesting confounding factor, caused by the physical relationship between the distance between the microphones and the maximum ITD measurable, created sections of space where the assumptions of the algorithm appear to not be followed in the collected data, see the red striped areas in Fig. 1. These blind spots force data that should be closer to 0 or π radians to appear at the

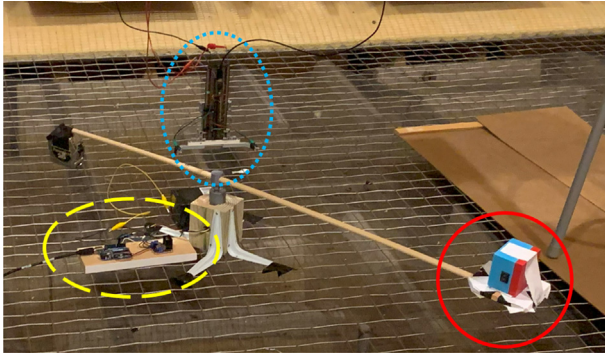


Fig. 3. Experimental setup in anechoic chamber. The emitter on the rotary base is circled in solid red, the hanging sensor is indicated by a dotted blue circle, and the Arduino controlling the motor is indicated by a dashed yellow circle.

maximum or minimum values the sensor can represent, in turn causing the algorithm to believe the source is at a different bearing than it really is. The result of these artifacts is errors in distance estimation at these extremes. With a maximum measurable of ITD of ± 104 samples, a sampling frequency of 200 kHz, and a microphone separation of 0.179 m, the sensing gap on each side of the sensor can be found by doubling the value of (4), producing a blind spot of 0.1692 radians (or 9.6931°). To combat incurred blind spot errors, we implemented a solution that ignores the correction step of the algorithm when the sensor measures data inside blind spots referred to as threshold ϕ . As seen in Fig. 4, a larger ϕ lessens the amount of data points that are considered to be useful information. This decrease in effective data is offset by the ability to avoid larger tracking errors introduced by said data. It is important to note that these data points are not removed from the data pool, instead their corresponding range update is based solely on (1); we refer to these data points as rejected.

Algorithm 1 Linear MMSE Iteration

INPUT:

$$(\hat{x}_k, P_{xx}, Z_k, z_{k+1}, \phi)$$

PREDICT:

$$\bar{x}_{k+1} = f(\hat{x}_k, 0; z_{k+1}^p)$$

$$F_k = \frac{\partial f}{\partial x_k} |_{(\hat{x}_k, 0)}, \Gamma_k = \frac{\partial f}{\partial v_k} |_{(\hat{x}_k, 0)}$$

$$\bar{P}_{xx} = F_k P_{xx} F_k^T + \Gamma_k Q_k \Gamma_k^T$$

CORRECT:

$$\bar{z}_{k+1}^c = h(\bar{x}_{k+1}, \hat{x}_k; z_k^c)$$

$$\bar{H}_k = \frac{\partial h}{\partial x_{k+1}} |_{(\bar{x}_{k+1}, \hat{x}_k)}, H_k = \frac{\partial h}{\partial x_k} |_{(\bar{x}_{k+1}, \hat{x}_k)}$$

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

$$\bar{P}_{zz} = \bar{H}_k \bar{P}_{xx} \bar{H}_k^T + \bar{H}_k F_k P_{xx} H_k^T + H_k P_{xx} H_k^T + R_k$$

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

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

CHECK:

$$\text{if } (z_{k+1}^c < \phi) | (\hat{x}_{k+1} < 0)$$

$$\hat{x}_{k+1} = \bar{x}_{k+1}$$

$$P_{xx} = \bar{P}_{xx}$$

else

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

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

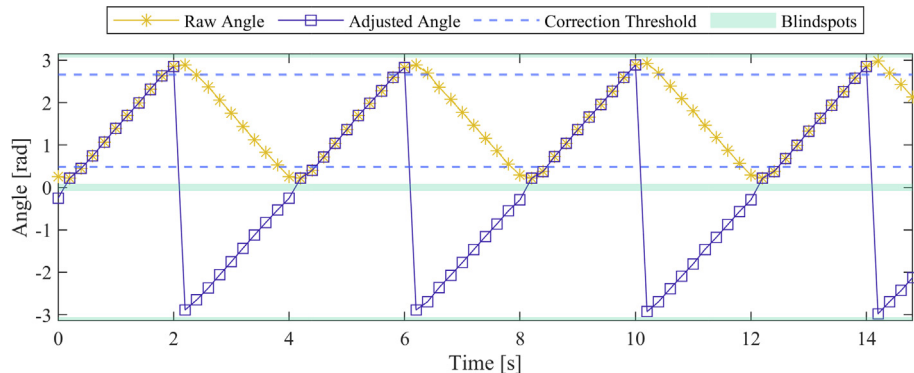


Fig. 4. Raw measured bearing angle (yellow starred line ranging from $[0.0846, 3.057]$ rad) compared to adjusted bearing angle (purple squared line ranging between ± 3.057 rad). Blind spots are illustrated by the shaded areas covering $[\pm\pi, \pm 3.057]$ rad and $[-0.0846, 0.0846]$ rad. The dashed lines represent the cutoff of ϕ for accepting the position update of the algorithm that the raw angles are compared against. Increasing ϕ brings these lines closer together, meaning the distance correction of any raw data points outside of range is ignored.

3.3. Experimental setup

To properly test the performance of the sensor with minimal interference, data was gathered in a full anechoic chamber, see Fig. 3. The emitter built for the experiments used a SensComp SMT 6500 Ranging Module to produce a 50 kHz signal output to a SensComp Series 600 Instrument Grade Ultrasonic Transducer. This emitter has a minimum transmission sensitivity of 110 dB at its operating frequency and a maximum operating range of 10.7 m [29] meaning the signal received by the sensor at any point of the emitter's path, which ranges between 0.25 m - 1.25 m, will be well above any noise present in the anechoic chamber ensuring accurate measurements. Emission timing was controlled by an Arduino Micro, allowing for easy changes to IPT between experimental trials. The emitter was attached to a 0.668 m rotating arm, with constant angular speed set via an Arduino Uno running a PI controller. A circular path was chosen as it provided numerous benefits:

- It allowed for the speed assumption to be met within a compact space.
- It provided a known trajectory to discern from tracking results.
- We were able to observe the behavior of the sensor and algorithm for longer periods of time without breaking assumptions as was done on the linear track in [5].

The sensor was hung over this setup, ensuring that the incoming signal was not interrupted by the base of the rotating arm.

Three parameters – IPT, speed, and sensor location – were varied across nine runs of the experiment. IPT was varied between 0.15 and 0.2 s, providing a way to investigate tracking performance at different measurement update periods. The speed of emitter rotation was varied between 12, 15, and 18 rotations per minute as the increased speed of the source means IPT_z can reach a greater range of values (e.g. pure linear motion towards the center will result in a smaller IPT_z the faster the emitter is moving). This greater range of values may make it easier to estimate changes in distance, using (1), when the motion of a source is close to radial about the sensor due to the smaller difference between IPT_z and IPT. Multiple combinations of both parameters were tested with the sensor at a central and noncentral location, allowing for the investigation of two movement patterns. A centralized sensor position produces the simplest trajectory for tracking as both change in bearing angle and IPT_z should stay constant, other than the effects of noise and imperfect centering between each sample. Noncentral positioning causes these measurements to vary, creating a model closer to arbitrary movement that upholds the assumptions of the algorithm. See Table 1 for the tested parameter combinations.

To initialize the tracking algorithm, statistics for the measurement and process noises must be calculated and tuned. Measurement noise was calculated by averaging standard deviations for both bearing angle and IPT_z for twenty repetitions of measurements at six positions around the sensor. Regarding process noise, an important difference between [5] and this work is that the distance model (6) does not assume a pseudo-stationary source, which implies zero-mean uncorrelated Gaussian process noise would no longer have an effect on the filter performance. To choose

the process noise covariance, we used the knowledge of the speed of the source v and IPT Δ to estimate the change in distance of the source as $dr = v\Delta \cos(\alpha)$. Here, α is the angle between the path of the source and the line segment from source to sensor at time step k , see Fig. 1. If we have no knowledge of the next position of the source and assume it is performing a random walk then α becomes a random variable sampled from a uniform distribution with zero mean and standard deviation $v\Delta/\sqrt{2}$. As this is a practical implementation of the system, it can be assumed a smooth path is followed compared to a true random walk. Therefore, we choose to tune the standard deviation further to be $v\Delta/3\sqrt{2}$, making the process noise covariance:

$$Q = \left(\frac{v\Delta}{3\sqrt{2}} \right)^2 \quad (17)$$

3.4. Data processing

When a signal of interest is detected by the Teensy, it sends the IPT_z along with 20 ms of data from both microphones, starting at the time the signal was first detected, over serial USB to a laptop running a script to adjust and analyze the signals. The data is split into left and right channels, normalized, then an absolute voltage level is used to determine the start point of the signal in each microphone from which the ITD is calculated. While it would be more rigorous to use a form of weighted cross correlation such as the generalized cross-correlation phase transform algorithm found in [18] to determine ITD, a minimum voltage level and sample time difference proved to be accurate enough for our needs, as we are in an anechoic chamber with the emitter pointed generally towards the sensor at all times. The bearing estimate is calculated using (4) which produces an angle estimate that is bounded between $[0, \pi]$. Due to our sensor position being fixed and the algorithm expecting bearing data to correspond to a full circle, the slope of the data is used to adjust the bearing range to $[-\pi, \pi]$ where decreasing bearing data is made negative, allowing us to simulate what the data would look like if the sensor did rotate. An example of this adjustment is shown in Fig. 4. This data was combined with the IPT_z as the measurement matrix used for tracking at each time step. While data collection and processing was performed online, tracking at each time step was performed offline to simplify analysis. Timing measurements were performed for both data adjustment and algorithm calculations and it was found that the scripts are fast enough to be implemented online in future versions of the sensor.

To evaluate performance of the sensor and algorithm in localizing the source, we use the measured true path to calculate the root mean square error in the estimated range over the length of each trial. We define the metric as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{k=1}^N (\hat{x}_k - x_k)^2} \quad (18)$$

Table 1
Experiment Parameters.

Sensor location	Central	Noncentral
Interpulse time [s]	0.15, 0.2	0.15
Speed [rpm]	12, 15, 18	12, 15, 18

Table 2
Process Noise Covariance (Q).

Speed [rpm]	IPT [s]	Q
12	0.2	1.6×10^{-3}
15	0.2	2.4×10^{-3}
18	0.2	3.5×10^{-3}
12	0.15	8.81×10^{-4}
15	0.15	1.4×10^{-3}
18	0.15	2.0×10^{-3}

Table 3

Experimental Results: Position of sensor is either central (C) or noncentral (NC) in the first column with speed and inter-pulse time of the emitter shown in the next two columns respectively. In the following columns the root mean square error (RMSE) of the predicted distance to the true distance is shown next to the percentage of updates rejected for each run. The degree values represent the threshold ϕ , centered around 90° and 270° from the front of the sensor, where the distance update was rejected.

Position	Speed [rpm]	IPT [s]	RMSE(0°)	Reject%(0°)	RMSE(30°)	Reject%(30°)	RMSE(50°)	Reject%(50°)
C	12	0.20	0.14	0	0.14	12.93	0.09	27.21
C	15	0.20	0.33	1.78	0.31	9.47	0.08	26.63
C	18	0.20	0.19	1.02	0.17	9.18	0.11	26.53
C	12	0.15	0.78	1.00	2.82	10.0	1.01	26.00
C	15	0.15	0.14	0	0.12	12.27	0.08	26.38
C	18	0.15	0.2	0	0.17	8.96	0.11	26.87
NC	12	0.15	0.17	1.00	0.16	9.50	0.15	24.5
NC	15	0.15	0.17	0.61	0.15	10.91	0.14	24.85
NC	18	0.15	0.18	0.61	0.18	12.27	0.17	25.15

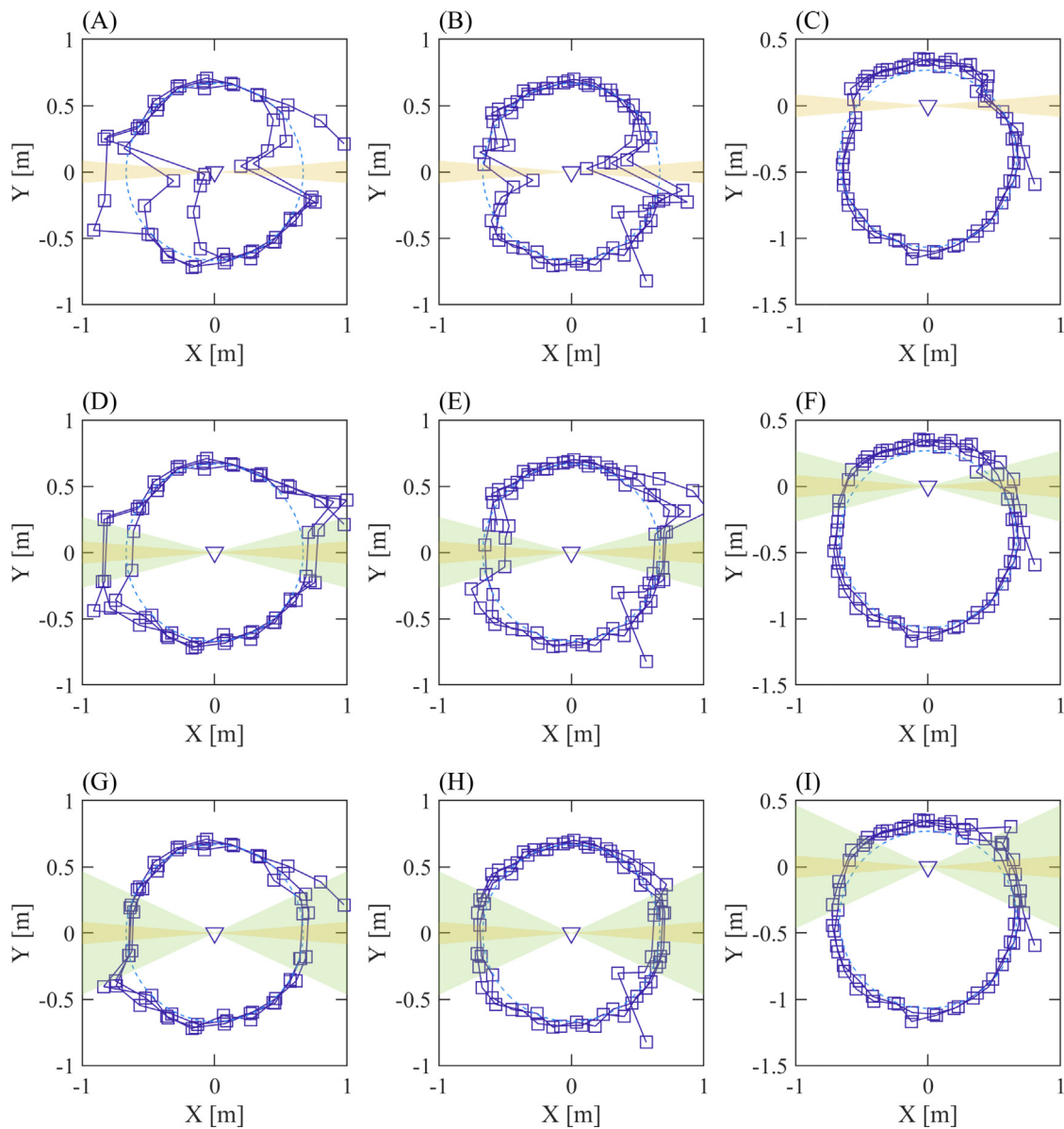


Fig. 5. Tracked paths of the emitter rotating at 15 rpm for central tracking at IPT = 0.2 s (A,D,G) and IPT = 0.15 s (B,E,H) and noncentral tracking at IPT = 0.15 s (C,F,I). The sensor's position and direction are represented by the triangle, blue squares are the estimated positions, and the light blue dashed line represents the true path of the emitter. The yellow shaded region is used to illustrate the $\sim 10^\circ$ blind spots of the sensor while the green shaded region represents the extended area of ϕ which increases along each column [0° , 30° , 50°].

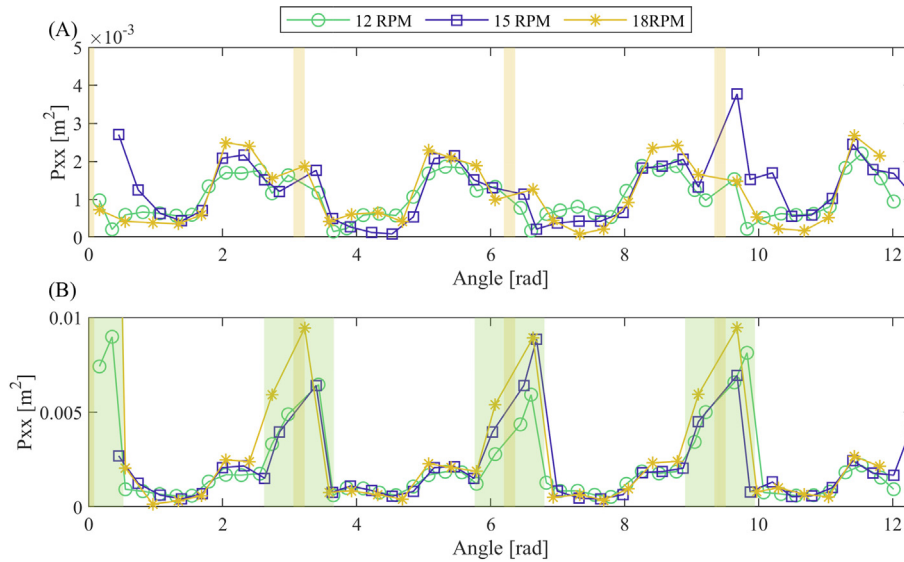


Fig. 6. Unwrapped state error covariance for trials at 12, 15, 18 rpm with IPT = 0.2 s for two rotations. (A) shows raw covariance for each iteration and (B) shows the associated covariance when the extended blind spots are introduced, at each measured angle.

where $\hat{x}_k$ is the estimated range, x_k is the true range and N is the total number of samples.

4. Results

The motion of the emitter is illustrated in Fig. 4. Starting just behind the left microphone of the sensor, the slope of the data increases as the emitter moves around the backside of the sensor to the right microphone; likewise as the slope decreases the emitter is moving in front from right to left. The emitter moves on the track with constant speed and emits a 50 kHz tone with a pulse interval of either 0.2 s or 0.15 s, see Table 1. The inter-microphone distance of the sensor is 0.179 m and it is facing downward and orthogonal to the emitter's plane of motion. The process noise covariance calculated from (17) depends on speed and IPT and can be found in Table 2 while measurement noise covariance was estimated to be $1.21 \times 10^{-4} \text{ rad}^2$ for bearing measurements and $1.163 \times 10^{-7} \text{ s}^2$ for the IPT. For this section, we will graphically follow runs of the experiment when the emitter was traveling at 15 rpm. Relevant information for all runs is found in Table 3.

Two features are common between all runs of the experiment. The first, seen across the first row of Fig. 5, is that the jumps in distance estimation away from the true path occur near the blind spots of the sensor. These errors are lessened and shifted by the introduction of ϕ , a phenomenon most easily visible down the first column of Fig. 5. The second is the cyclical nature of the covariance which can be seen in Fig. 6. The covariance begins to settle and hits a low point a sample before the emitter enters the sensor's blind spot and then jumps up just after the blind spot as the algorithm works to correct itself.

Corresponding covariance time series for all runs shown in Fig. 5 can be found in the appendix.

4.1. Central sensor

The central sensor configuration consists of six of the nine runs conducted, see Table 3. Fig. 5(A) shows the tracked path of the emitter with an IPT = 0.2 s and (B) shows the path of the emitter with IPT = 0.15 s. For all but one of the tested cases, setting $\phi = 30^\circ$ on each side produces little quantitative change in RMSE but does create a visible qualitative change in the tracked path, pushing

errors closer to the true path of the emitter. The introduction of $\phi = 50^\circ$ produces a greater reduction in RMSE while also leading to a range update rejection for just over 25% of the points for each trial.

The experimental trial, with a speed of 12 rpm and IPT of 0.15 s, stands out as an anomaly with a raw RMSE of 0.78 m. The RMSE also increases significantly at $\phi = 30^\circ$ to 2.82 m and doesn't drop below its 0° value at the greater $\phi = 50^\circ$. It was found that this error was introduced by an incorrectly estimated bearing measurement soon after the blind spot, see Fig. A.8 in the Appendix.

4.2. Noncentral sensor

The noncentral sensor trials did not experience the drop in RMSE from varying ϕ at the level the centered trials did. Data in the last three rows of Table 3 shows ϕ having only a slight effect on decreasing the RMSE value by at least 0.02 m. These results can be seen qualitatively in the last column of Fig. 5 where even the raw tracking of (C) does not have the large errors of (A) and (B) due to the blind spots.

5. Discussion

From these results, we find that the stereo sensor measures bearing and ITP_z accurately enough to track the source and correct error, even with blind spots that disrupt the assumptions of the algorithm. It is known that the performance of linear MMSE filters depends on the accuracy and confidence of the initial guess [3]. If the initial guess is highly inaccurate, the algorithm may never converge or at best converge to an incorrect value. For all above trials, we set the initial estimated range to be 1 m for clarity of plotting, though tests showed that the algorithm can still converge with a initial range ten times the actual value, see Fig. A.9 in the Appendix. In the noncentral case, introduced errors such as those described above seem to have less of an effect on the accuracy of the algorithm. We interpret that this is due to the area of the blind spots being related to an angle instead of a length of area that cannot be observed, see Fig. 1. Due to this, the closer the source is to the sensor the less time it spends operating in the blind spot areas, thus reducing the amount of time we are operating outside of our assumptions.

The estimation covariance never fully settles as the algorithm converges and its confidence seems to be tied to behavior around blind spots, see Fig. 6. As the EKF is known to have problems showing overconfidence in estimations [3], the low covariance after the blind spot in Fig. 6(A) could be due to the interruption of our assumptions, causing the algorithm to believe it is more accurate than really is and have to compensate when the sensor is once again following the assumptions. Looking at Fig. 6(B), we see the effects of ϕ on the state error covariance. As we are only using the prediction of our next covariance in this area, the covariance continues to rise as long as the data received is within the bounds of ϕ . Once the sensor is outside ϕ the covariance corrects back to expected levels.

An interesting question arises when comparing tracking performance to the speed and IPT relationship. When the source is on a rotating path and the sensor is placed in the center, it becomes possible to calculate a ratio of emissions per rotation (or per some unit length). While the majority of results follow the same trends, it may be possible that there exists a ratio that produces optimal algorithm performance that could translate to arbitrary path tracking.

The cause of the bearing estimation error of trial 4 in Table 3 is not immediately clear. Some possible sources of the error include an undefined echo from the equipment setup or simply an error in our time delay estimation calculation. This error does introduce a need for a more robust time delay estimation calculation for later experiments such as the GCC-PHAT technique described in the introduction [18].

Comparing this work to other stereo sensors doing SSL in the literature, we see that our sensor shows a decrease in accuracy due to its static nature. In [13], the stereo array is able to both rotate 360 degrees and translate slightly, from which distance for static sound sources can be estimated exploiting the dynamics of the sensor motion. This results in errors on the order of one centimeter. In [21], a stereo sensor equipped on a mobile robot platform was able to accurately track a mobile sound source with estimation errors on the order of one centimeter as well. It is important to note that, due to the lateral blind spots of the array, the RMSE is inflated more than it would be if this was accounted for by a redesign in the hardware or by allowing the array to rotate. This can be seen in the average decrease of the RMSE in Table 3 by the introduction of the update rejection step, bringing our results closer to those observed in [13,21].

6. Conclusion

In this work, we performed an improved and more accurate investigation of passive stereo SSL of a source with unknown dynamics presented in [5], including a demonstrative experiment with novel hardware. We made use of minimal inexpensive hardware and computer power while still maintaining accuracy in measurements and estimation. Our presented results show that SSL of this form is possible and that there are a number of important hardware characteristics to be aware of. By changing the spacing

of our microphones, it should be possible to effectively negate the sensor's lateral blind spots and mounting it on a mobile platform will remove the need for post processing of bearing data.

In the future, to allow for a relaxation of the assumptions necessary for the current tracking problem we plan to update the sensor to a phased array design similar to what is described in [20], though in a uniform line array format. By equipping mobile robotic platforms with these sensor arrays and an ultrasonic emitter, we plan to investigate the feasibility of combining typical active sensing and localization with passive tracking capabilities similar to what is described in this paper. We expect that these instrumented robots will allow investigation into more applicable challenges such as simultaneous localization and mapping and other multi-agent robotic tasks [7,37].

CRediT authorship contribution statement

Aidan J. Bradley: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft. **Masoud Jahromi Shirazi:** Conceptualization, Methodology, Software, Writing - review & editing. **Nicole Abaid:** Conceptualization, Data curation, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Data availability

The data and code used in this paper can be found at <https://doi.org/10.7294/22227883>

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was funded by the National Science Foundation under the Graduate Research Fellowship Program to A.J.B. and under grant 1751498. The authors would like to thank the Department of Mechanical Engineering at Virginia Tech for use of the anechoic chamber. The authors would also like to thank Darcy Davidson and Eighdi Aung for their repeated help in experimental set up and break down.

Appendix A. Appendix

Fig. A.7 shows the state error covariance values which correspond to the tracked paths shown in Fig. 5. Fig. A.8 shows the large error introduced by the incorrect bearing measurement in trial four of the experiments, with data shown in row four of Table 3. Fig. A.9 is included to show that with a very large error in the estimated initial range the algorithm can still converge to track the true path of the emitter.

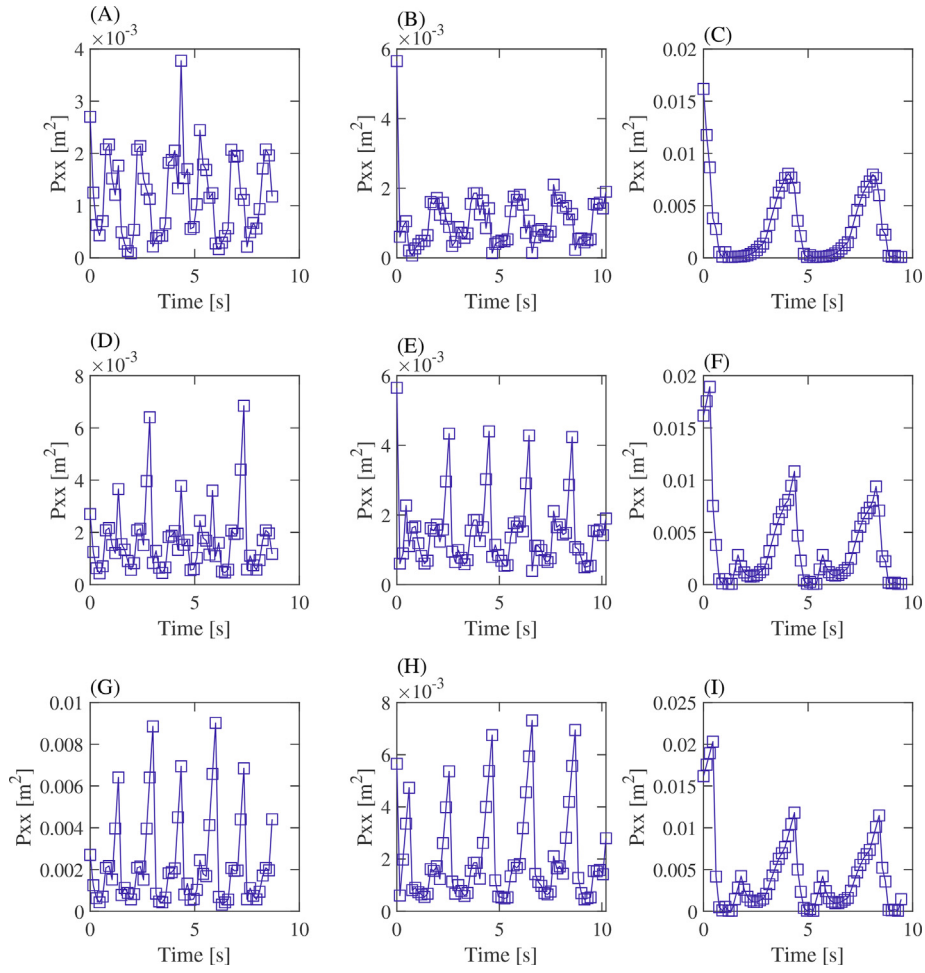


Fig. A.7. State error covariance values corresponding to the tracked paths shown in Fig. 5. For the shown trials the emitter is rotating at 15 rpm with an IPT = 0.2 s (A,D,G) and IPT = 0.15 s (B, E, H) for central tracking and an IPT = 0.15 s (C,F,I) for noncentral tracking.

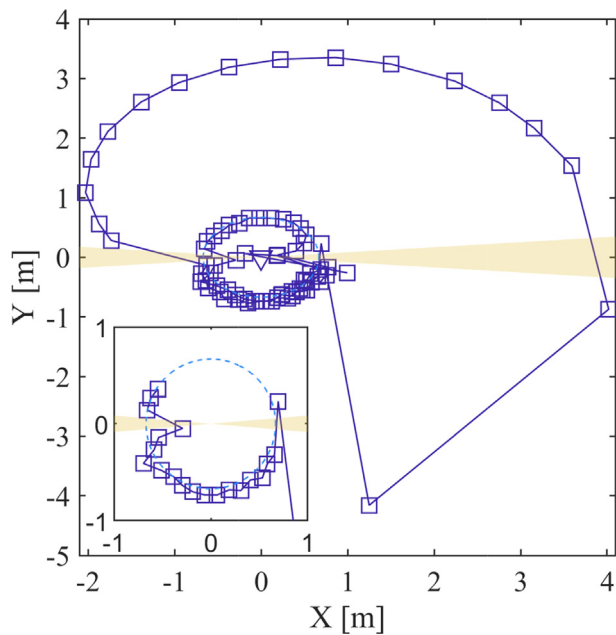


Fig. A.8. The anomaly from the centered experiment trial at 15 rpm with an IPT = 0.15. The zoomed graph insert shows that a tracking point soon after the blind spot of the sensor is what introduced the large error.

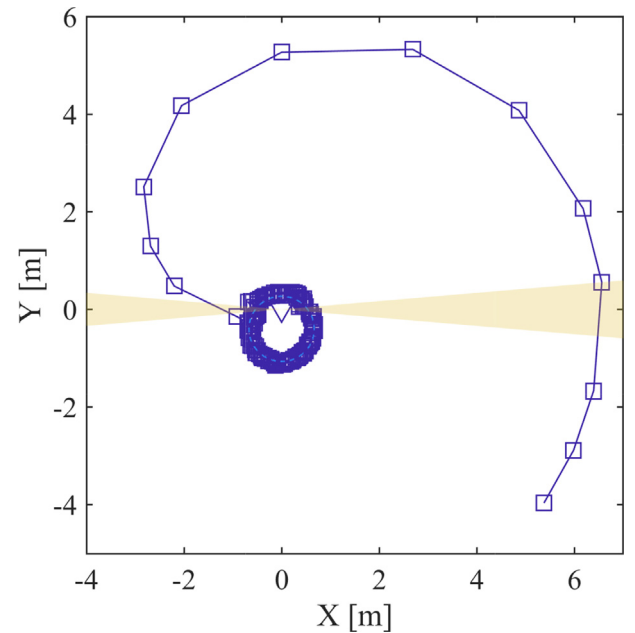


Fig. A.9. An example of the algorithm being able to still converge to the true path from a starting point of 6.68 m (10 times the true radius of rotation).

References

- [1] Argentieri, S., Danes, P., Soueres, P., 2006. Modal analysis based beamforming for nearfield or farfield speaker localization in robotics, in: 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, pp. 866–871.
- [2] Argentieri S, Danès P, Souères P. A survey on sound source localization in robotics: From binaural to array processing methods. *Comput Speech Lang* 2015;34:87–112. <https://doi.org/10.1016/j.csl.2015.03.003>.
- [3] Bar-Shalom Y, Rong Li X, Kirubarajan T. *Estimation with Applications to Tracking and Navigation*. New York: John Wiley & Sons; 2001.
- [4] Baumann C, Rogers C, Massen F. Dynamic binaural sound localization based on variations of interaural time delays and system rotations. *J Acoust Soc Am* 2015;138:635–50. <https://doi.org/10.1121/1.4923448>.
- [5] Bradley, A.J., Shirazi, M.J., Abaid, N., 2020. Bearing-only localization of a quasi-static sound source with a binaural microphone array, in: ASME 2020 Dynamic Systems and Control Conference, DSCC 2020, pp. 1–8. doi: 10.1115/DSCC2020-3235.
- [6] Chen C, Wei X, Moss CF. Flying in silence: Echolocating bats cease vocalizing. *PNAS* 2008;105:2–7.
- [7] Chen H, Sun D, Yang J, Chen J. Localization for multirobot formations in indoor environment. *IEEE/ASME Trans Mechatron* 2010;15:561–74. <https://doi.org/10.1109/TMECH.2009.2030584>.
- [8] Cho, S.J., Ovcharenko, A., Chong, U.p., 2006. Front-back confusion resolution in 3d sound localization with HRTF databases, in: 2006 International Forum on Strategic Technology, IEEE, pp. 239–243.
- [9] Ennasr O, Tan X. Time-difference-of-arrival TDOA-based distributed target localization by a robotic network. *IEEE Trans Control Network Syst* 2020;7:1416–27.
- [10] Evers C, Lollmann HW, Mellmann H, Schmidt A, Barfuss H, Naylor PA, Kellermann W. The LOCATA Challenge: Acoustic Source Localization and Tracking. *IEEE/ACM Trans Audio Speech Lang Process* 2020;28:1620–43. <https://doi.org/10.1109/TASLP.2020.2990485>. arXiv:1909.01008.
- [11] Fenton MB, Portfors CV, Rautenbach IL, Waterman JM. Compromises: Sound frequencies used in echolocation by aerial-feeding bats. *Can J Zool* 1998;76:1174–82. <https://doi.org/10.1139/z98-043>.
- [12] Fujioka E, Aihara I, Sumiya M, Aihara K, Hiryu S. Echolocating bats use future-target information for optimal foraging. *Proc Natl Acad Sci USA* 2016;113:4848–52. <https://doi.org/10.1073/pnas.1515091113>.
- [13] Gala D, Lindsay N, Sun L. Realtime active sound source localization for unmanned ground robots using a self-rotational bi-microphone array. *J Intell Robot Syst* 2019;95:935–54.
- [14] Gala D, Lindsay N, Sun L. Multi-Sound-Source Localization Using Machine Learning for Small Autonomous Unmanned Vehicles with a Self-Rotating Bi-Microphone Array. *J Intell Robot Syst: Theory Appl* 2021;103. <https://doi.org/10.1007/s10846-021-01481-4>. arXiv:1804.05111.
- [15] Handzel AA, Krishnaprasad PS. Biomimetic Sound-Source Localization. *IEEE Sens J* 2002;2:607–16.
- [16] Heffner RS, Koay G, Heffner HE. Sound localization in common vampire bats: Acuity and use of the binaural time cue by a small mammal. *J Acoust Soc Am* 2015;137:42–52. <https://doi.org/10.1121/1.4904529>.
- [17] Kino GS. *Acoustic waves: devices, imaging and analog signal processing*. 43 KIN, Prentice Hall; 1987.
- [18] Knapp CH, Carter CG. The generalized correlation method for estimation of time delay. *IEEE Trans Acoust Speech Signal Process* 1976;24:320–7.
- [19] Kuhn GF. Model for the interaural time differences in the azimuthal plane. *J Acoust Soc Am* 1977;62:157–67. <https://doi.org/10.1121/1.381498>.
- [20] Kumar S, Furuhashi H. Characteristics of an ultrasonic phased array transmitter in medium range. *Ultrasonics* 2018;82:331–5. <https://doi.org/10.1016/j.ultras.2017.09.013>. URL: <https://www.sciencedirect.com/science/article/pii/S0041624X17300057>.
- [21] Kumon M, Uozumi S. Binaural localization for a mobile sound source. *J Biomech Sci Eng* 2011;6:26–39.
- [22] Lambert RM. Dynamic theory of sound-source localization. *J Acoust Soc Am* 1974;56:165–71. <https://doi.org/10.1121/1.1903248>.
- [23] Lee S, Park Y, Park YS. Three-Dimensional Sound Source Localization Using Inter-Channel Time Difference Trajectory. *Int J Adv Robot Syst* 2015;12:1–15. <https://doi.org/10.5772/61652>.
- [24] Liu R, Wang Y. Azimuthal source localization using interaural coherence in a robotic dog: modeling and application. *Robotica* 2010;28:1013–20.
- [25] Mansour CB, Koreman E, Laurijssen D, Steckel J, Peremans H, Vanderelst D. Robotic models of obstacle avoidance in bats. In: *ALIFE 2019: The 2019 Conference on Artificial Life*. MIT Press; 2019. p. 463–4.
- [26] Rascon C, Meza I. Localization of sound sources in robotics: A review. *Robot Auton Syst* 2017;96:184–210. <https://doi.org/10.1016/j.robot.2017.07.011>.
- [27] Rayleigh, L., 1875. On Our Perception of the Direction of a Source of Sound. *Proceedings of the Musical Association*, 75–84.
- [28] Schnitzler HU, Moss CF, Denzinger A. From spatial orientation to food acquisition in echolocating bats. *Trends Ecol Evol* 2003;18:386–94. [https://doi.org/10.1016/S0169-5347\(03\)00185-X](https://doi.org/10.1016/S0169-5347(03)00185-X).
- [29] SensComp, 2022. Series 600 Instrument Grade Transducer. SensComp. Livonia, MI.
- [30] Shirazi, M.J., Abaid, N., 2019. Tracking a sound source with unknown dynamics using bearing-only measurements based on a priori information. *Proceedings of the American Control Conference* 2019-July, 4491–4496. doi: 10.23919/acc.2019.8815232.
- [31] Sutlive J, Singh A, Zhang S, Müller R. A biomimetic soft robotic pinna for emulating dynamic reception behavior of horseshoe bats. *Bioinspiration and Biomimetics* 2020;16. <https://doi.org/10.1088/1748-3190/abb73>.
- [32] Urick RJ. *Principles of underwater sound-2*. New York: NY (USA) McGraw-Hill Book; 1975.
- [33] Waite AD. *Sonar for practising engineers*. Wiley; 2002.
- [34] Wallach H. On Sound Localization. *J Acoust Soc Am* 1939;10:270–4. <https://doi.org/10.1121/1.1915985>.
- [35] Wang J, Wang J, Qian K, Xie X, Kuang J. Binaural sound localization based on deep neural network and affinity propagation clustering in mismatched HRTF condition. *Eurasip J Audio, Speech, Music Process* 2020;2020. <https://doi.org/10.1186/s13636-020-0171-v>.
- [36] Wang M, Liu Y, Su D, Liao Y, Shi L, Xu J, Miro JV. Accurate and real-time 3-D tracking for the following robots by fusing vision and ultrasonar information. *IEEE/ASME Trans Mechatron* 2018;23:997–1006.
- [37] Wang Y, Yue Y, Shan M, He L, Wang D. Formation reconstruction and trajectory replanning for multi-uav patrol. *IEEE/ASME Trans Mechatron* 2021;26:719–29. <https://doi.org/10.1109/TMECH.2021.3056099>.
- [38] Wohlgemuth, M.J., Luo, J., Moss, C.F., 2016. Three-dimensional auditory localization in the echolocating bat. *Current Opinion in Neurobiology* 41, 78–86. URL: <https://www.sciencedirect.com/science/article/pii/S0959438816301143>, doi: 10.1016/j.conb.2016.08.002. microcircuit computation and evolution.
- [39] Zhong X, Sun L, Yost W. Active binaural localization of multiple sound sources. *Robot Auton Syst* 2016;85:83–92. <https://doi.org/10.1016/j.robot.2016.07.008>.