BeBeC-2026-D02

ASYNCHRONOUS DISTRIBUTED ACOUSTIC ARRAY-BASED 3D LOCALIZATION OF A DRONE

Yonghyun Kim[1], Ookjin Jung[1], Sooho Bae[1] and Youngkey Kim[1]

[1]SM Instruments Co., Ltd.

Daejeon, Republic of Korea

yhkim@smins.co.kr

ABSTRACT

We address 3D localization of a drone using its acoustic emissions and asynchronous distributed acoustic arrays. The problem is formulated as a filter-based acoustic SLAM framework, in which the source trajectory is estimated jointly with the positions, orientations, and clock offsets of distributed microphone arrays. In this setting, acoustic observations are recorded at reception time but physically correspond to past source states due to propagation delay, which introduces structural bias if the measurements are evaluated directly at the reception-time state. To address this issue, the proposed estimator performs time-aligned measurement prediction. Each DOA or TDOA observation is associated with the source state that generated it, and the resulting residuals are fused within an error-state estimator. This enables joint refinement of the source trajectory, array geometry, and clock offsets under asynchronous operation. In simulation, a reception-time formulation yields source-position errors on the order of 10–40 m, while the proposed formulation reduces the error to approximately 1–3 m and improves array orientation and clock-offset estimates. These results demonstrate that SLAM-based joint estimation with consistent measurement-time handling enables consistent localization using asynchronous distributed acoustic arrays.

1 INTRODUCTION

Accurate drone localization is important for surveillance, airspace monitoring, and autonomous navigation. Passive acoustic sensing is attractive because it is low-power and robust to lighting conditions and occlusion, although its spatial resolution is generally lower than that of LiDAR or vision [1].

In distributed acoustic sensing, microphone arrays may operate without a common clock, and their array geometry may be only approximately known. Acoustic propagation delay further complicates the problem because an observation recorded at reception time

corresponds to a past source state. If this reception-emission mismatch is ignored, it introduces structural bias into the estimation, especially over long ranges or during fast source motion.

We formulate this problem as acoustic SLAM [2]. The drone is represented by the source state, while each ground array is treated as a landmark with position, orientation, and clock offset. DOA measurements constrain the bearing from each array to the source, whereas TDOA measurements constrain relative range and inter-array clock offsets [3, 4]. Their joint use enables simultaneous estimation of the source trajectory, array geometry, and clock offsets.

Related acoustic SLAM and microphone-array calibration methods have jointly estimated source positions, array geometry, and asynchronous clock parameters [6, 7, 8, 9, 10]. Recent batch-SLAM approaches, in particular, have addressed joint calibration of multiple asynchronous microphone arrays and sound source localization, including observability analysis and initialization from weakly known or unknown states [9]. These works provide an important foundation for asynchronous acoustic SLAM and calibration.

This paper addresses a complementary modelling issue. Existing formulations commonly associate acoustic measurements with source states indexed by the event or reception timestamp. For a dynamically evolving acoustic source, however, the source state at emission time corresponds to an earlier instant due to propagation delay. This mismatch can introduce structural bias even when the SLAM state includes array geometry and clock offsets. Therefore, this work focuses on time-aligned prediction within a warm-started filter-based SLAM formulation, rather than on fully uninitialized global calibration or FIM-based observability analysis.

The remainder of this paper is organized as follows. Section 2 presents the system model. Section 3 describes the proposed estimator. Section 4 evaluates the method through numerical simulations.

2 SYSTEM MODEL

2.1 System Overview

We consider multiple asynchronous microphone arrays observing a common acoustic source. Each array provides observations with its own reception timestamp, and these timestamps are not synchronized across arrays.

The source state at time t is defined by position $\mathbf{P}(t)$, velocity $\mathbf{V}(t)$, and angular velocity $\mathbf{W}(t)$, all in $\mathbb{R}^3$. Each array i is characterized by position $\mathbf{T}_i$, orientation $\mathbf{R}_i$ in SO(3), and clock offset β_i. The speed of sound is denoted by c. The source may be static or dynamically evolving, while the array landmarks are assumed static over the estimation window.

This structure requires joint estimation of the source trajectory and static array parameters. The clock offset β_i represents array clock asynchrony and couples timing with geometry.

2.2 Source Motion Model

To evaluate the source state at the emission time, we use a short-term constant-turn kinematic model:

$$d\mathbf{P}/dt = \mathbf{V}, \quad d\mathbf{V}/dt = \mathbf{W} \times \mathbf{V}, \quad d\mathbf{W}/dt = 0 \tag{1}$$

Let $\boldsymbol{\Omega} = \mathrm{skew}(\mathbf{W})$. Instead of relying on a small-time approximation, the source state is propagated using the closed-form solution

$$\mathbf{V}(t_k + \Delta t) = \exp(\mathbf{W}\Delta t) \cdot \mathbf{V}_k \tag{2}$$
$$\mathbf{P}(t_k + \Delta t) = \mathbf{P}_k + f(\Delta t, \mathbf{W}) \cdot \mathbf{V}_k \tag{3}$$

where

$$f(\Delta \mathrm{t}, \mathbf{W}) = \int_0^{\Delta t} \exp(\boldsymbol{\Omega} \cdot \tau) d\tau \tag{4}$$

For nonzero omega, this integral can be written as

$$f(\Delta, \mathbf{W}) = \mathbf{I} \cdot \Delta t + (1 - \cos\theta)/\|\mathbf{W}\|^2 \cdot \boldsymbol{\Omega} + (\theta - \sin\theta)/\|\mathbf{W}\|^3 \cdot \boldsymbol{\Omega}^2 \tag{5}$$
$$\theta = \|\mathbf{W}\|\Delta t$$

This closed-form propagation is used to interpolate the source state between reception and emission times, where the time interval may be larger than a single filter step.

2.3 Measurement Model

The DOA measurement is modelled as a unit direction vector in the local coordinate frame of array i:

$$\mathbf{z}_i^{doa} = \mathbf{R}_i^T \cdot \rho\big(\mathbf{P}(t_{e,i}) - \mathbf{T}_i\big) + \mathbf{n}_i^{doa} \tag{6}$$

where $\rho(\mathbf{x}) = \mathbf{x}/\|\mathbf{x}\|$

The TDOA measurement between arrays i and j is

$$z_{ij}^{tdoa} = \big\|\mathbf{P}(t_{e,j}) - \mathbf{T}_j\big\|/c - \big\|\mathbf{P}(t_{e,i}) - \mathbf{T}_i\big\|/c + \beta_j - \beta_i + n_{ij}^{tdoa} \tag{7}$$

Measurement noise can be modelled as zero-mean Gaussian with known covariance.

2.4 Reception–Emission Time Relationship

For array i, reception time and emission time satisfy

$$t_{r,i} = t_{e,i} + \big\|\mathbf{P}(t_{e,j}) - \mathbf{T}_j\big\|/c + \beta_i \tag{8}$$

Thus, the measurement time is coupled with the source state, array geometry, and clock offset. Unlike standard SLAM formulations where the measurement timestamp is assumed to be aligned with the state, asynchronous acoustic SLAM must account for the delay between emission and reception.

If prediction is performed directly at $t_{r,i}$, the approximate position bias is

$$\Delta\mathbf{P} \cong \mathbf{V} \cdot \big(t_{r,i} - t_{e,i}\big) \tag{9}$$

This bias increases with propagation delay and source velocity, and it can propagate into array orientation and clock-offset estimates. The proposed estimator therefore predicts measurements at the corresponding source state at emission time rather than directly at the reception-time state.

3 PROPOSED ESTIMATOR

3.1 Acoustic SLAM Formulation

The problem is formulated as filter-based acoustic SLAM. The source trajectory is the dynamic state, while the distributed microphone arrays are static landmarks. Each landmark includes position, orientation, and clock offset, representing both geometric and clock calibration parameters.

At each frame, DOA and TDOA observations are fused within an error-state estimator. Because these measurements are affected by acoustic propagation delay, they are associated with the corresponding source state at emission time during prediction rather than evaluated directly at the reception time.

The state vector at time step k is

$$\mathbf{x}_k = [\mathbf{P}_k \quad \mathbf{V}_k \quad \mathbf{W}_k \quad \{\mathbf{T}_i \quad \mathbf{R}_i \quad \beta_i\}_{i=1}^{N}] \tag{10}$$

The measurement vector is

$$\mathbf{z}_k = \left[\mathbf{z}_i^{doa} \quad z_{ij}^{tdoa}\right] \tag{11}$$

This formulation allows the estimator to refine both the source trajectory and the distributed array configuration using the same acoustic observations.

3.2 Time-Aligned Prediction

For each asynchronous observation, the corresponding source state is determined by solving the time-of-flight relationship in Section 2.4. Given the current SLAM state estimate, this implicit relationship is solved within a bounded interval to ensure physical validity.

The temporal association is initialized using the geometric propagation delay computed from the current source and array estimates, and is refined iteratively until the time-of-flight equation is satisfied. Once $t_{e,i}$ is obtained, the source state is interpolated using the motion model, and the predicted measurement is computed.

The residual is

$$\mathbf{r}_k = \mathbf{z}_k - h(\mathbf{x}_k, t_e) \tag{12}$$

where $h(\mathbf{x}_k, t_e)$ denotes the DOA/TDOA measurement function. This time-aligned prediction ensures that the SLAM update uses the source state that generated the measurement rather than the later reception-time state.

3.3 Estimation Update

The estimator follows an error-state Extended Kalman Filter (ESKF) formulation [5]. The nominal state is kept on the original state space, while the update is computed in a local error-state space:

$$\delta\mathbf{x}_k = [\delta\mathbf{P}_k \quad \delta\mathbf{V}_k \quad \delta\mathbf{W}_k \quad \{\delta\mathbf{T}_k \quad \delta\boldsymbol{\theta}_k \quad \delta\beta_k\}_{i=1}^{N}] \tag{13}$$

The term $\delta\boldsymbol{\theta}_i$ denotes the minimal perturbation of the array rotation. Euclidean components are updated additively, while rotations are updated on SO(3):

$$\mathbf{R}_i \leftarrow \mathbf{R}_i \cdot \exp\big(\mathrm{skew}(\delta\boldsymbol{\theta}_i)\big) \tag{14}$$

Linearization accounts for both direct error-state dependence and indirect dependence through measurement time. Since the residual is defined as Eq. (12), the error-state Jacobian is written as

$$d\mathbf{r}/d\delta\mathbf{x} = -(dh/d\delta\mathbf{x} - dh/dt_e \cdot dt_e/d\delta\mathbf{x}) \tag{15}$$

In this expression, t_e denotes the measurement-specific emission time for a DOA residual, or the set of emission times involved in a TDOA residual. The term $dt_e/d\delta\mathbf{x}$ is obtained by implicitly differentiating the time-of-flight constraint. It describes how perturbations in the source state, array geometry, and clock offset change the emission time associated with the measurement.

This indirect term is important in asynchronous acoustic SLAM because clock offsets, array geometry, and source motion jointly determine the temporal association between a reception-time observation and the corresponding source state.

3.4 Initialization

The proposed estimator assumes a coarse initial array layout, which is commonly available in distributed acoustic deployments from manual placement, deployment metadata, or external positioning sources such as GPS. This warm start is used only as an initial landmark estimate; the array poses and clock offsets are still refined during the SLAM update. The method therefore does not assume perfectly known array poses or synchronized clocks.

This initialization strategy reflects the focus of this paper. Rather than addressing fully uninitialized global map construction, we consider the practical case in which a rough deployment map is available and examine how asynchronous acoustic measurements should be handled consistently within SLAM. Under this setting, the warm-started array geometry reduces the search space without removing the need for joint estimation.

The source position and velocity can be initialized using simple geometric methods, such as DOA intersection and finite differences, while clock offsets can be initialized from TDOA residuals. These initial values provide a feasible starting point for the error-state estimator and are subsequently refined using the acoustic measurements.

4 SIMULATION STUDIES

4.1 Setup

We evaluate the proposed method in a controlled simulation designed to demonstrate SLAM-based localization using asynchronous distributed acoustic arrays. Four microphone arrays are placed in a linear configuration with 10 m inter-array spacing and initially uncertain positions and orientations. Each array is modelled using the microphone configuration of the SeeSV-S206 acoustic camera, configured as shown in Fig. 1, consisting of 96 microphones sampled at 25.6 kHz. The arrays operate asynchronously, each with an independent constant clock offset.

The acoustic source follows a smooth curved trajectory around the array configuration, with non-zero angular velocity. This setting induces non-negligible propagation delays, so that each observation is associated with a source state earlier than its reception timestamp.

The initial array poses and clock offsets are perturbed from their true values and then jointly estimated with the source trajectory. We compare two prediction strategies: (i) a reception-time model that evaluates measurements at the reception timestamp, and (ii) the proposed time-aligned model that evaluates each measurement at the corresponding source state at emission time. Both methods use the same SLAM estimator and differ only in how measurement time is handled during prediction.

Synthetic DOA and TDOA observations are generated from the simulated source and array configuration, while ground truth is not provided to the estimator and is used only for performance evaluation. This controlled setup isolates the structural bias caused by temporal misalignment.

4.2 Results

Fig. 2 compares the ground-truth source trajectory with estimates from the reception-time and time-aligned models. The reception-time model exhibits a systematic shift due to temporal misalignment, whereas the time-aligned model closely follows the ground truth. This is consistent with the bias relation in Section 2.4.

Fig. 3 shows the frame-wise RMSE of the source position. The reception-time model maintains errors on the order of 10–40 m after the initial transient, whereas the time-aligned model converges below 5 m and remains near 1–3 m for most frames. This indicates that reception-time prediction introduces a persistent structural error that filtering alone cannot remove.

Fig. 4 shows the RMSE of the estimated array orientations. The reception-time model remains around 0.2–0.5 rad after convergence, while the time-aligned model converges below 0.1 rad. Because array orientation is estimated from the same DOA residuals used for source localization, reducing temporal bias also improves the orientation estimate.

Fig. 5 shows the RMSE of the estimated clock offsets. The reception-time formulation leaves residual errors on the order of several milliseconds, whereas the time-aligned formulation reduces the error to near-zero levels. Consistent measurement-time handling reduces ambiguity between propagation delay and clock offset.

Overall, the results show that the performance gain is not caused by a different filter structure or noise suppression, but by removing structural bias caused by reception-time prediction. Time-aligned prediction enables consistent joint estimation of source trajectory, array orientation, and clock offsets within the same SLAM framework.

Fig. 1. Unit array configuration (SeeSV-S206).

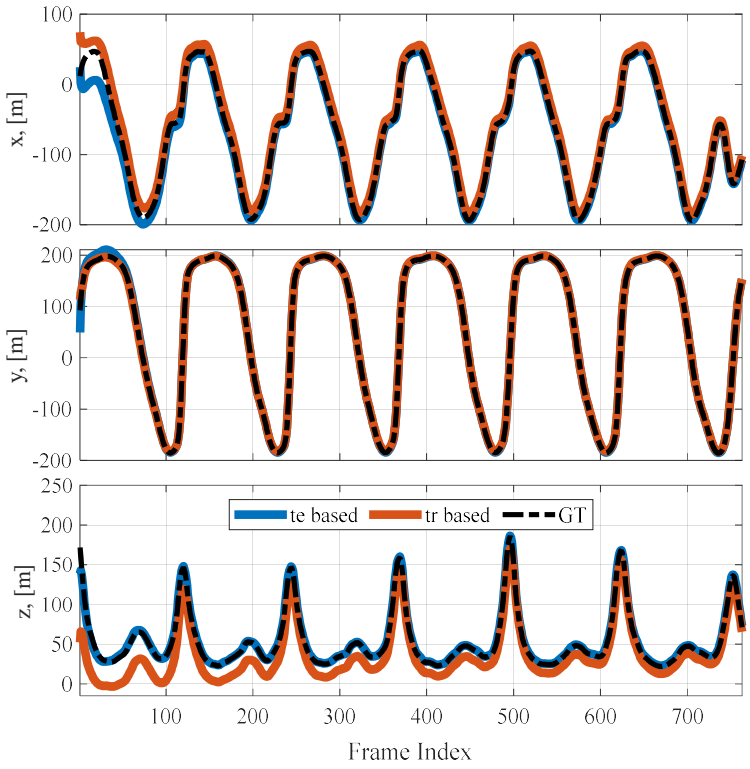

Fig. 2. Source trajectory comparison between t_r-based and t_e-based estimation.

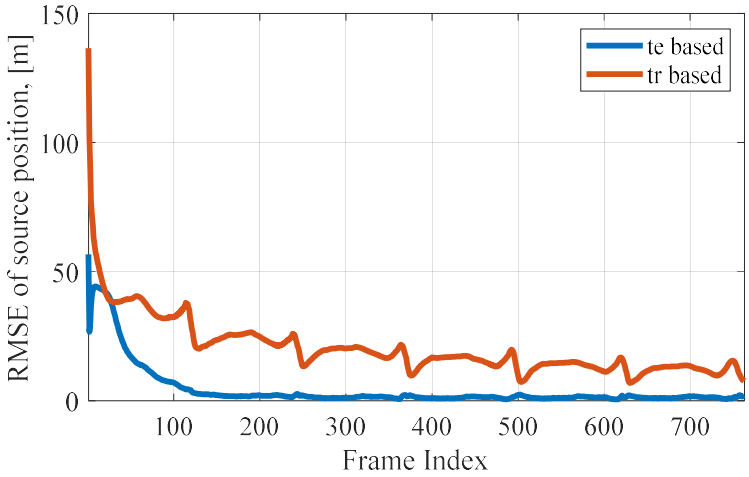

Fig. 3. Frame-wise RMSE of source position for t_r and t_e models.

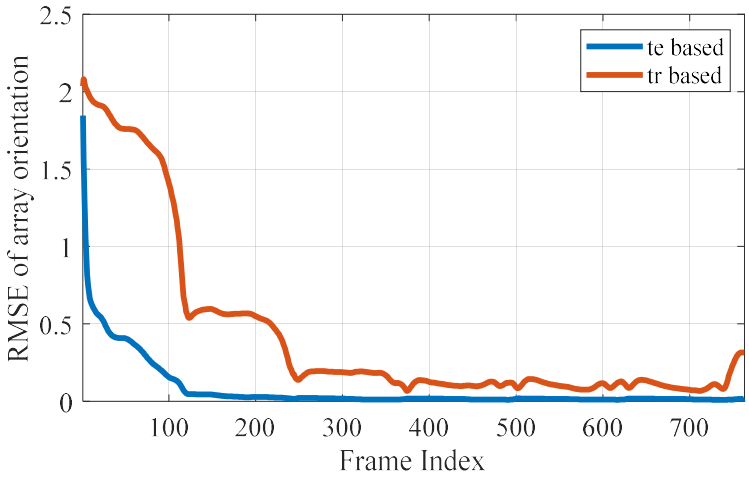

Fig. 4. RMSE of estimated array orientations under t_r and t_e formulations.

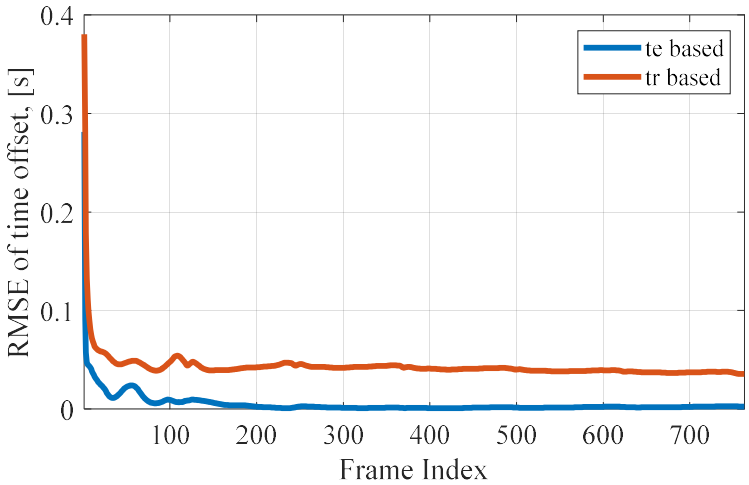

Fig. 5. RMSE of estimated clock offsets for t_r and t_e models.

5 CONCLUSION

We presented a SLAM-based framework for 3D drone localization using asynchronous distributed acoustic arrays. The source trajectory is estimated jointly with the positions, orientations, and clock offsets of the arrays, which is necessary when array geometry and clocks are not perfectly known.

The results show that a SLAM formulation with reception-time prediction is not sufficient for asynchronous acoustic measurements. Propagation delay makes each observation correspond to a past source state, and reception-time prediction introduces structural bias into the SLAM update. The proposed time-aligned prediction associates each observation with the source state that generated it.

Simulation results show that the reception-time formulation produces source-position errors on the order of tens of meters and degrades array orientation and clock-offset estimates. In contrast, the time-aligned formulation reduces the source-position error to a few meters and improves the estimated array parameters. These results confirm that consistent localization with asynchronous distributed acoustic arrays requires both SLAM-based joint estimation and physically consistent treatment of measurement time.

Future work will include validation on real drone acoustic data and extension to larger array networks and multi-source scenarios.

REFERENCES

[1] L. Grumiaux, S. Kitić, L. Girin, and A. Guérin. “A survey of sound source localization.” J. Acoust. Soc. Am., 152, 1079–1098, 2022. doi:10.1121/10.0011809.

[2] H. Durrant-Whyte and T. Bailey. “Simultaneous localization and mapping: Part I.” IEEE Robot. Autom. Mag., 13, 99–110, 2006. doi:10.1109/MRA.2006.1638022.

[3] J. H. DiBiase. “A high-accuracy, low-latency technique for talker localization in reverberant environments.” Ph.D. thesis, Brown University, 2000.

[4] C. Knapp and G. Carter. “The generalized correlation method for estimation of time delay.” IEEE Trans. Acoust., Speech, Signal Process., 24, 320–327, 1976. doi:10.1109/TASSP.1976.1162830.

[5] J. Solà. “Quaternion kinematics for the error-state Kalman filter.” 2017. URL https://arxiv.org/abs/1711.02508.

[6] H. Miura, T. Yoshida, K. Nakamura, and K. Nakadai. “SLAM-based online calibration of asynchronous microphone array for robot audition.” IEEE/RSJ International Conference on Intelligent Robots and Systems, 524–529, 2011.

[7] D. Su, T. Vidal-Calleja, and J. Valls Miro. “Simultaneous asynchronous microphone array calibration and sound source localisation.” IEEE/RSJ International Conference on Intelligent Robots and Systems, 5561–5567, 2015.

[8] K. Sekiguchi, Y. Bando, K. Nakamura, K. Nakadai, K. Itoyama, and K. Yoshii. “Online simultaneous localization and mapping of multiple sound sources and asynchronous microphone arrays.” IEEE/RSJ International Conference on Intelligent Robots and Systems, 1973–1979, 2016.

[9] J. Wang, Y. He, D. Su, K. Itoyama, K. Nakadai, J. Wu, S. Huang, Y. Li, and H. Kong. "SLAM-based joint calibration of multiple asynchronous microphone arrays and sound source localization." IEEE Trans. Robot., 2024. doi:10.1109/TRO.2024.3410456.

[10] C. Zhang, J. Wang, and H. Kong. "Asynchronous microphone array calibration using hybrid TDOA information." 2024. URL https://arxiv.org/abs/2403.05791.