

PASSIVE ACOUSTIC TRACKING OF WHALES IN 3-D

Junsu Jang, Florian Meyer, Eric R. Snyder, Sean M. Wiggins,
Simone Baumann-Pickering, and John A. Hildebrand

Scripps Institution of Oceanography, University of California San Diego, La Jolla, USA

ABSTRACT

Passive acoustic monitoring (PAM) is a nonintrusive approach to studying behaviors of vocalizing marine organisms underwater that otherwise would remain unexplored. In this paper, we propose a data processing chain that can detect and track multiple whales in 3-D from passively recorded underwater acoustic signals. In particular, time-difference-of-arrival (TDOA) measurements of echolocation clicks are extracted from a volumetric hydrophone array's acoustic data by using a noise-whitening cross-correlation. For multi-target tracking, the TDOA measurements are then processed by a Bayesian inference engine consisting of two stages that is based on the sum-product algorithm (SPA). Particle flow is embedded in the SPA to make tracking computationally feasible in the considered nonlinear and high-dimensional scenario. The capability to track multiple whales without human intervention is demonstrated in scenarios with simulated and real data.

Index Terms— Array processing, Bayesian estimation, high-dimensional estimation, multi-target tracking.

1. INTRODUCTION

Bioacoustic signals from marine organisms provide valuable information for understanding and surveying marine ecosystems. Passive acoustic monitoring (PAM) is a nonintrusive and efficient method for acquiring bioacoustic signals that can potentially be used to detect, localize, and track vocalizing marine organisms. Among these organisms, odontocetes (toothed whales) generate directional acoustical pulses called echolocation clicks for foraging and navigating underwater [1]. Echolocation clicks acquired by volumetric hydrophone arrays can be used to compute TDOA measurements of potential whale locations. However, tracking of echolocating whales is challenged by the presence of noise, echoes, and simultaneously vocalized signals by other marine organisms, which give rise to false positive TDOA measurements. Moreover, long periods of missed detections can occur due to the aspect dependence of echolocation click transmissions with respect to the receivers. Therefore, most approaches for tracking whales from acoustic signals rely on a human operator or heuristics to combine the correct TDOA measurements from the same source and initialize tracks accordingly [2–7].

Bayesian tracking of multiple sources in 3-D based on TDOA measurements is challenging because of the highly nonlinear measurement model and the high-dimensional object states. Existing Bayesian multi-target tracking (MTT) methods for marine mammals tracking include [8] and [9]. In [8], multi-hypothesis tracking (MHT) is applied to tracking echolocating whales in 3-D. MHT,

This research was supported in parts by the National Science Foundation (NSF) under CAREER Award N2146261, by the Office of Naval Research (ONR) under YIP Award N00014-15-1-2587, and by the Cooperative Ecosystems Study Unit under Cooperative Agreement N62473-18-2-0016 for the U.S. Navy, U.S. Pacific Fleet.

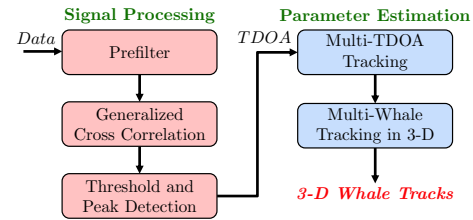


Fig. 1: Block diagram of the proposed data processing chain for tracking whales in 3-D. The signal processing module extracts time-difference-of-arrival (TDOA) measurements of echolocation clicks. The parameter estimation module then estimates tracks from TDOA measurements in two stages. The first stage relies on a linear model and performs tracking in the TDOA domain for each hydrophone pair in parallel. The second stage relies on a high-dimensional nonlinear model and forms tracks in 3-D.

however, is limited to linear measurement models or mildly nonlinear measurement models and faces high computational complexity and memory requirements [10]. In [9], a Gaussian mixture probability hypothesis density (GM-PHD) filter [11] is extended to incorporate amplitude information for tracking whales in the TDOA domain. The GM-PHD is similarly limited to measurement models that are linear or mildly nonlinear and is thus unsuitable for tracking in 3-D from TDOA measurements. Additionally, localization and tracking in 2-D based on TDOA measurements using Bayesian MTT methods have been introduced in [12, 13].

To address these challenges, we propose a data processing chain that extracts TDOA measurements of echolocation clicks acquired by volumetric hydrophone arrays and tracks whales in 3-D. First, TDOA measurements are computed based on a generalized cross-correlation (GCC) [14] that whitens the instrument noise. Then, two stages of an MTT method based on factor graphs and the sum-product algorithm (SPA) [15] are applied. Whales are first tracked individually for each hydrophone pair in the TDOA domain to close gaps of missed detections and reject false positive measurements. A second tracking stage fuses the resulting TDOA estimates in the 3-D domain. To cope with the nonlinear measurement model and high-dimensional state space, particle flow [16] is embedded in the SPA [17]. Simulation and real data application results show that the proposed data processing chain successfully tracks whales from their echolocation clicks in a fully automated and tractable manner.

2. THE PROPOSED DATA PROCESSING CHAIN

The data processing chain for detecting and tracking echolocating whales is in two parts (Fig. 1). First, the signal processing stage detects echolocation clicks and computes potential TDOA measurements from pairs of hydrophones. These measurements are used in the parameter estimation stage, where the whales are first tracked in the TDOA domain and then in the 3-D domain.

2.1. GCC-WIN

The time delay of a coherent signal is often detected and estimated using a GCC [14]. The cross-power spectral density (PSD) between two signals is normalized with a frequency weighting optimal for time-delay estimation for the given signals. We adopt GCC to whiten the instrument noise, thereby enhancing the TDOA detection rate in the presence of a coherent instrument noise signal. We refer to this method as GCC-WIN.

Consider a spatially separated receiver pair (s_1, s_2) , which forms a TDOA sensor $s \in \{1, \dots, n_s\}$ with n_s being the number of pairs of receivers. A GCC of length n_g samples is performed at each discrete tracking time step. The discrete received signals at time n from a remote source that generated a signal $\chi_s[n]$ in the presence of noise, $n_s[n]$, for sensor s are given by

$$y_{s_1}[n] = \chi_{s_1}[n] + n_{s_1}[n] \quad (1)$$

$$y_{s_2}[n] = \alpha \chi_{s_1}[n + d] + n_{s_2}[n] \quad (2)$$

where $\chi_{s_1}[n]$, $n_{s_1}[n]$, and $n_{s_2}[n]$ are real, stationary, and ergodic random processes, α is a scaling factor, and d is the TDOA.

The GCC is performed in the frequency domain, where the discrete frequency is represented by l . Let the discrete Fourier transform pairs of $y_{s_1}[n]$ and $y_{s_2}[n]$ be $Y_{s_1}[l]$ and $Y_{s_2}[l]$, respectively, and the cross-PSD estimate of the received signals be $R_s[l] = Y_{s_1}[l] Y_{s_2}^*[l]$. The corresponding GCC as a function of discrete time delay, m , is given by [14]

$$\hat{h}_s[m] = \frac{1}{n_g} \sum_{l=0}^{n_g-1} \psi[l] R_s[l] e^{j2\pi ml/N} \quad (3)$$

with $\psi[l]$ being the frequency weighting. In GCC-WIN, the frequency weighting is $\psi_s[l] = 1/(G_{s_1,s_1}[l] G_{s_2,s_2}[l])^{1/2}$, where $G_{s_1,s_1}[l]$ and $G_{s_2,s_2}[l]$ are the auto PSD estimates of the noise at the respective receivers. In this work, the instrument introduced a coherent noise signal with time-varying but periodic statistics that dominated the background noise. Hence, a sequence of time-varying noise PSD was estimated from spectrograms of the noise signals and applied as the frequency weighting.

2.2. Review of Graph-Based MTT

At the discrete time k , we assume there are i_k targets. However, since i_k is unknown and time-varying, we introduce potential target (PT) states, which are denoted by index $j \in \{1, \dots, j_k\}$ with j_k being the maximum number of PTs given the available measurements at time k . The existence of the PT j is represented by a binary variable, $r_k^{(j)} \in \{0, 1\}$, where $r_k^{(j)} = 1$ if and only if PT j exists. The state of PT j is then denoted as $\mathbf{y}_k^{(j)} = [\mathbf{x}_k^{(j)T} r_k^{(j)}]^T$, where $\mathbf{x}_k^{(j)}$ is the position and motion-related parameters of the target.

We are interested in estimating the state of PTs, $\mathbf{x}_k^{(j)}$, if they exist, based on the measurements up to time k . Let there be n_s sensors, and each sensor $s \in \{1, \dots, n_s\}$ generates $m_{k,s}$ measurements $\mathbf{z}_{k,s} = [\mathbf{z}_{k,s}^{(1)}, \dots, \mathbf{z}_{k,s}^{(m_{k,s})}]^T$. There can be at most one measurement at each sensor from each target, and the rest are false positives (FPs). The probability of detecting a measurement originating from PT j at sensor s is modeled as $p_d^{(s)}(\mathbf{x}_k^{(j)})$ and follows a Bernoulli distribution. The m^{th} measurement, where $m \in \{1, \dots, m_{k,s}\}$, is distributed according to $f(\mathbf{z}_{k,s}^{(m)} | \mathbf{x}_k^{(j)})$. The number of FPs is Poisson distributed with a mean $\mu_{\text{fp}}^{(s)}$. Each FP is iid according to pdf $f_{\text{fp}}^{(s)}(\mathbf{z}_{k,s}^{(m)})$ and independent of the measurements from the targets.

Two types of PTs are possible: new PTs and legacy PTs. The new PTs have generated the measurement for the first time, while

the legacy PTs have made at least one measurement at a previous time step or sensor. A PT j at time $k-1$ continues to exist at time k with survival probability $p_{\text{su}}(\mathbf{x}_k^{(j)})$, and all PTs at time $k-1$ become legacy PTs at time k . States of the new and legacy PTs are denoted by $\bar{\mathbf{y}}_{k,s}^{(j)} = [\bar{\mathbf{x}}_{k,s}^{(j)T} \bar{r}_{k,s}^{(j)}]^T$ and $\mathbf{y}_{k,s}^{(j)} = [\mathbf{x}_{k,s}^{(j)T} r_{k,s}^{(j)}]^T$, respectively. Then, $\mathbf{y}_k = [\bar{\mathbf{y}}_k \mathbf{y}_k]^T$ is the vector that consists of all PT state at time k , where $\bar{\mathbf{y}}_k$ and $\mathbf{y}_k$ denote the vectors that consist of all new and legacy PT states, respectively.

A latent random vector $\mathbf{a}_{k,s} = [a_{k,s}^{(1)}, \dots, a_{k,s}^{(j_k)}]^T$ is used to model the unknown association between measurements and targets [18]. The random variable $a_{k,s}^{(j)}$ is equal to $m \in \{1, \dots, m_{k,s}\}$ if PT j generates measurement m and 0 otherwise. The indicator function $\psi(\mathbf{a}_{k,s}) \in \{0, 1\}$ ensures that at most one measurement is generated by a target at every time step, where $\psi(\mathbf{a}_{k,s}) = 0$ if $\exists j, j' \in \{1, \dots, n_t\}$ such that $j \neq j'$ and $a_{k,s}^{(j)} = a_{k,s}^{(j')} \neq 0$.

Let $\mathbf{y}_k = [\bar{\mathbf{y}}_k^{(1)}, \dots, \bar{\mathbf{y}}_k^{(j_k)}]^T$, $\mathbf{a}_k = [a_{k,1}, \dots, a_{k,n_s}]^T$, and $\mathbf{z}_k = [\mathbf{z}_{k,1}, \dots, \mathbf{z}_{k,n_s}]^T$ be the joint vectors. The joint posterior pdf of $\mathbf{y}_{1:k}$ and $\mathbf{a}_{1:k}$ given $\mathbf{z}_{1:k}$ is given by [15]

$$\begin{aligned} f(\mathbf{y}_{1:k}, \mathbf{a}_{1:k} | \mathbf{z}_{1:k}) &\propto \prod_{k'=1}^k \left(\prod_{j'=1}^{j_{k'}-1} f(\bar{\mathbf{y}}_{k',s}^{(j')} | \bar{\mathbf{y}}_{k'-1}^{(j')}) \right) \prod_{s=1}^{n_s} \psi(\mathbf{a}_{k',s}) \\ &\quad \times \left(\prod_{j=1}^{j_{k'},s} q(\bar{\mathbf{x}}_{k',s}^{(j)}, \bar{r}_{k',s}^{(j)}, a_{k',s}^{(j)}; \mathbf{z}_{k',s}) \right) \\ &\quad \times \prod_{m=1}^{m_{k',s}} v(\bar{\mathbf{x}}_{k',s}^{(m)}, \bar{r}_{k',s}^{(m)}, a_{k',s}^{(m)}). \end{aligned} \quad (4)$$

Here, $f(\bar{\mathbf{y}}_{k',s}^{(j')} | \bar{\mathbf{y}}_{k'-1}^{(j')})$ represents the motion model of the targets, and the remaining factors represent the measurement model. In particular, the factors $q(\bar{\mathbf{x}}_{k',s}^{(j)}, \bar{r}_{k',s}^{(j)}, a_{k',s}^{(j)}; \mathbf{z}_{k',s})$ and $v(\bar{\mathbf{x}}_{k',s}^{(m)}, \bar{r}_{k',s}^{(m)}, a_{k',s}^{(m)})$ are functions of $p_d^{(s)}(\mathbf{x}_k^{(j)})$, $\mu_{\text{fp}}^{(s)}$, $f_{\text{fp}}^{(s)}(\mathbf{z}_{k,s}^{(m)})$, and $f(\mathbf{z}_{k,s}^{(m)} | \mathbf{x}_k^{(j)})$.

With the joint posterior pdf in (4), one can marginalize and compute the MMSE estimate of the PTs with $r_k(j) = 1$ [19] following

$$\hat{\mathbf{x}}_k^{(j)} \triangleq \int \mathbf{x}_k^{(j)} f(\mathbf{x}_k^{(j)} | r_k^{(j)} = 1, \mathbf{z}_{1:k}) d\mathbf{x}_k^{(j)} \quad (5)$$

where $f(\mathbf{x}_k^{(j)} | r_k^{(j)} = 1, \mathbf{z}_{1:k}) = f(\mathbf{x}_k^{(j)}, r_k^{(j)} = 1 | \mathbf{z}_{1:k}) / p(r_k^{(j)} = 1 | \mathbf{z}_{1:k})$. An efficient marginalization is performed with the framework employing factor graphs and SPA [20]. The complete details on the joint posterior pdf, system model, and the SPA for MTT can be found in [15].

2.3. Particle Flow

A key challenge of performing the SPA for the considered MTT problem is the combination of high-dimensional state space and non-linear measurement model for which a conventional particle implementation of the SPA for MTT based on bootstrap importance sampling [21] is unsuitable. Hence, we make use of invertible particle flow for importance sampling. Particle flow actively migrates particles $\mathbf{x}_k^{(j,i)}$, $i \in \{1, \dots, n_p\}$ drawn from the predicted posterior pdf $f(\mathbf{x}_k^{(j)} | r_k^{(j)} = 1, \mathbf{z}_{1:k-1})$ towards high-likelihood regions [16]. This migration is performed sequentially and at each time step k across pseudo time steps $l \in \{1, \dots, n_f\}$. The particle flow mapping $\mathbf{x}_k^{(j,i)} \rightarrow \bar{\mathbf{x}}_k^{(j,i)}$ from pseudo time step $l = 1$ to $l = n_f$ is proven to be invertible, i.e., there exists a mapping of the particles after the flow to the particle before the flow [22]. By exploiting this invertible mapping, the particle-flow proposal pdf can be evaluated as [22]

$$q(\bar{\mathbf{x}}_k^{(j,i)}|\mathbf{z}) = \frac{f(\mathbf{x}_k^{(j,i)}|r_k^{(j)}=1, \mathbf{z}_{1:k-1})}{\theta^{(j,i)}} \quad (6)$$

where the “mapping factor” $\theta^{(j,i)}$ is computed based on a first-order Taylor series expansion of the TDOA measurement in (10). Performing importance sampling by using the particles after the flow, $\{\mathbf{x}_k^{(j,i)}\}_{i=1}^{n_p}$, and by evaluating the corresponding proposal pdf in (6) is asymptotically optimal and enables MTT in the considered nonlinear and high-dimensional problems [17].

3. TRACKING OF ECHOLOCATING WHALES

First, MTT is applied in the TDOA domain to each hydrophone pair separately. Next, the tracking algorithm combines the TDOA tracking results from the first stage to track the whales in the 3-D.

3.1. MTT in TDOA Domain

In the TDOA domain, the state of the whale j is denoted as $\mathbf{d}_{k,s}^{(j)} = [d_{k,s}^{(j)} \dot{d}_{k,s}^{(j)}]^T$, where $d_{k,s}^{(j)}$ is the true TDOA and $\dot{d}_{k,s}^{(j)}$ is its rate of change at time step k and sensor s . The state transition model follows a linear constant velocity motion model [18],

$$\mathbf{d}_{k,s}^{(j)} = \begin{bmatrix} 1 & t_m \\ 0 & 1 \end{bmatrix} \begin{bmatrix} d_{k-1,s}^{(j)} \\ \dot{d}_{k-1,s}^{(j)} \end{bmatrix} + \mathbf{u}_{k,s}^{(j)} \quad (7)$$

where t_m is the length of each time step, and the driving noise $\mathbf{u}_{k,s}^{(j)} \in \mathbb{R}^2$ is a zero-mean multivariate Gaussian random vector with a driving noise standard deviation (STD) σ_u and a covariance matrix

$$\Sigma_u = \begin{bmatrix} \frac{t_m^3}{3} & \frac{t_m^2}{2} \\ \frac{t_m^2}{2} & t_m \end{bmatrix} \sigma_u^2. \quad (8)$$

The measurement model is given by $z_{k,s}^{(m)} = d_{k,s}^{(j)} + v_{k,s}^{(m)}$, where $z_{k,s}^{(m)}$ is the TDOA originated from the whale j at sensor s , and $v_{k,s}^{(m)}$ is a zero-mean Gaussian measurement noise with STD σ_v . The pdf of the false positives, $f_{fp}^{(s)}(z_{k,s}^{(m)})$, is uniformly distributed over interval $[-t_s^{\max}, t_s^{\max}]$, where $t_s^{\max}$ is the maximum time delay at sensor s corresponding to an arbitrary hydrophone pair (s_1, s_2) . In particular, given the speed of sound, c , and the positions $\mathbf{q}_{s_1} \in \mathbb{R}^3$ and $\mathbf{q}_{s_2} \in \mathbb{R}^3$ of the hydrophone pair, $t_s^{\max}$ is equal to $\|\mathbf{q}_{s_1} - \mathbf{q}_{s_2}\|/c$.

The SPA-based MTT method (Sec. 2.2) is used since there is measurement-origin uncertainty (MOU), and the number of whales is unknown and time-varying. The tracking results of this stage are sets of TDOA estimates $\hat{d}_{k,s}^{(j)}$, $j \in \{1, \dots, j_{k,s}\}$ for each time step k and sensor s . In the next section, they are used as measurements in the 3-D MTT stage, denoted as $\hat{\mathbf{d}}_{k,s}^{(m)}$, $m \in \{1, \dots, m_{k,s}\}$.

In this work, we also preprocess the TDOA measurements by accumulating and clustering them over a time period to reduce the processing time and be robust against the time-varying nature of the inter-click-interval (ICI) [23] of the echolocation clicks.

3.2. MTT in 3-D

The whales are tracked in 3-D with the TDOA measurements, where multisensor data association and track initialization are performed [21, 24]. At time k , the state of the whale j is denoted as $\mathbf{p}_k^{(j)} = [p_{k,x}^{(j)} p_{k,y}^{(j)} p_{k,z}^{(j)} \dot{p}_{k,x}^{(j)} \dot{p}_{k,y}^{(j)} \dot{p}_{k,z}^{(j)}]^T$, where $p_{k,x}^{(j)}$, $p_{k,y}^{(j)}$, and $p_{k,z}^{(j)}$ are the position of the whale in the 3-D Cartesian coordinate system and $\dot{p}_{k,x}^{(j)}$, $\dot{p}_{k,y}^{(j)}$ and $\dot{p}_{k,z}^{(j)}$ are the respective velocities. The motion model again follows the linear constant velocity model

$$\mathbf{p}_k^{(j)} = \begin{bmatrix} \mathbf{I}_3 & t_m \mathbf{I}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} \mathbf{p}_{k-1}^{(j)} + \begin{bmatrix} \frac{t_m^2}{2} \mathbf{I}_3 \\ t_m \mathbf{I}_3 \end{bmatrix} \mathbf{w}_k^{(j)} \quad (9)$$

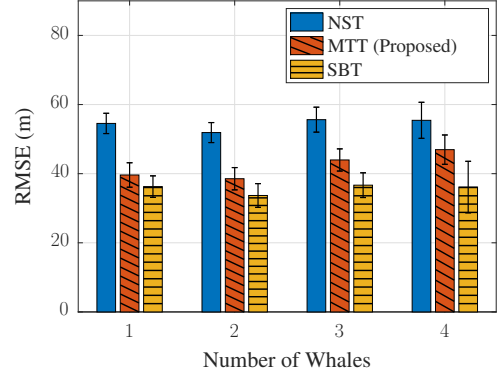


Fig. 2: Average RMSE across 200 Monte Carlo simulations versus the number of simultaneously tracked whales. Three different 3-D whale tracking approaches are considered: a nonsequential tracker (NST), a single Bernoulli tracker (SBT), and the proposed MTT approach (MTT). Unlike (NST) and (SBT) which rely on perfect data association, (MTT) performs data association in the presence of false positive measurements. The error bars denote the 75th percentile of the measured RMSE.

where $\mathbf{w}_k^{(j)} \in \mathbb{R}^3$ is a zero-mean Gaussian driving noise with STD σ_w and covariance $\mathbf{I}_3 \sigma_w^2$.

The TDOA measurement model for the whale j is given by

$$\hat{d}_{k,s}^{(m)} = (\|\mathbf{p}_k^{(j)} - \mathbf{q}_{s_1}\| - \|\mathbf{p}_k^{(j)} - \mathbf{q}_{s_2}\|)/c + b_{k,s}^{(m)} \quad (10)$$

where $b_{k,s}^{(m)} \in \mathbb{R}$ is a zero-mean Gaussian measurement noise with STD σ_b . Since multiple sensors involve acoustic recordings from the same hydrophone, $b_{k,s}^{(m)}$ is not statistically independent across sensors. However, for simplicity, it is assumed independent to facilitate MTT in 3-D. The pdf of the false positives, $f_{fp}^{(s)}(z_{k,s}^{(m)})$, is again uniform on the interval $[-t_s^{\max}, t_s^{\max}]$.

The nonlinear measurement model in (10) is underdetermined and yields a potential whale location on a hyperboloid. To estimate the 3-D location, the TDOAs from multiple sensors have to be fused. Due to the presence of MOU, noise, and an unknown number of whales, no accurate intersection can be found, and thus, reliable state estimation can only be performed sequentially. We again apply the MTT approach from Sec. 2.2 with a particle-based implementation. To avoid particle degeneracy [25] and curse of dimensionality [26], we use particle flow recently proposed in [17].

3.3. Simulation

PAM data often lacks ground truth to verify the tracking outcomes. Thus, the estimated tracks are compared with those from manually annotated approaches, which are subjective and imperfect. To motivate the use of the proposed tracking method, we compare the root-mean-square error (RMSE) of 3-D tracking results from simulated data using three different methods: nonsequential tracking (NST) using a combination of DOAs [6], a single Bernoulli tracker (SBT) [27], and the proposed MTT approach described in Sec. 3.2 (MTT). Note that (MTT) is required to solve the data association problem given the MOU, while (NST) and (SBT) are genie-aided in that the association is known.

Four sets of 200 Monte Carlo simulations of whale tracks are generated and tracked using the approaches (NST), (SBT), and (MTT). Each set of simulations is 85 discrete time steps long and has an increasing number of simultaneously present whales from 1 to 4. Each whale is present for 50 time steps, and a whale is introduced every 10 time steps when multiple whales are simulated.

Hyperparameters	TDOA	3-D
Detection Probability, p_d	0.80	0.80
Survival Probability, p_{su}	0.90	0.99
Mean Number of False Positives, μ_{fp}	10	1
Mean Number of Whale Birth, μ_b	1.0×10^{-4}	1
Measurement Noise STD, σ_u & σ_w	1.0×10^{-5}	3.0×10^{-5}
Driving Noise STD, σ_v & σ_b	1.5×10^{-7}	1.0×10^{-2}
Number of Particles	30,000	100,000
Minimum Track Length	20	5

Table 1: Hyperparameters for tracking whales in TDOA domain and in 3-D.

The whales' initial positions are uniformly distributed on a circle of radius 1000 m on a xy-plane and at a fixed depth of 1000 m. Using the hyperparameters in Table. 1 and the array geometry in Sec. 4, TDOA data from assumed echolocation clicks of those whales are simulated following the model in (10). In (NST) and (SBT), we use the true but noisy TDOAs generated from the corresponding whales. In case of missed detections, (NST) uses the interpolated data, and (SBT) takes the missed detections into account. When the whale is not detected or there is a large error due to false positives or missed detections, the error value of 110 m is applied for (SBT) and (MTT). This penalty is roughly twice the average RMSE of (NST).

Based on the results in Fig. 2, the performance of (MTT) is between those of (SBT) and (NST). Since (NST) does not apply a filter to the interpolated data, the high RMSE is expected. (SBT), however, filters the tracks and also incorporates missed detections into its statistical model, yielding the lowest RMSE. (MTT) employs the Bernoulli filter but faces the data association problem, causing an RMSE between those from (NST) and (SBT).

4. REAL DATA APPLICATION

Echolocation clicks of beaked whales were passively recorded on two high-frequency acoustic recording packages (HARPs) [28]. Each HARP is equipped with four hydrophones that are 1m apart and arranged in a tetrahedral shape (see Fig. 2 in [4]), forming a small-aperture volumetric array. They were deployed off the coast of California at a depth of ~ 1300 m from March 2018 to July 2018. The two arrays were approximately 1km apart, establishing a large-aperture array, and we refer to the arrays located east and west as HARP EE and HARP EW, respectively. The encounters of Cuvier's beaked whales were detected using the long-term spectral average (LTSA) with a software package *Triton* [28].

The details of the tracking setup are as follows. Since the TDOA is computed for every pair of hydrophones in each of the two arrays with four hydrophones, there are $n_s = 12$ TDOA sensors total. TDOA measurements are accumulated over $t_m = 7$ s. We employ the east-north-up (ENU) system, where the x-, y-, and z-axes are positive along the east, north, and up directions. The origin is set between the two arrays and at the sea surface. Additionally, since the beaked whales are detected below the thermocline, we assume that they are in an isovelocity medium with the speed of sound of $c = 1490$ ms⁻¹. The hyperparameters for tracking in TDOA and 3-D domains are shown in Table. 1.

Two whales were detected and tracked from the acoustic data recorded on July 1st, 2018. Their tracks in the TDOA and 3-D domains are compared to the results using (NST) in [6] with the hand-annotated data (Fig. 3). The diving behaviors of the two whales

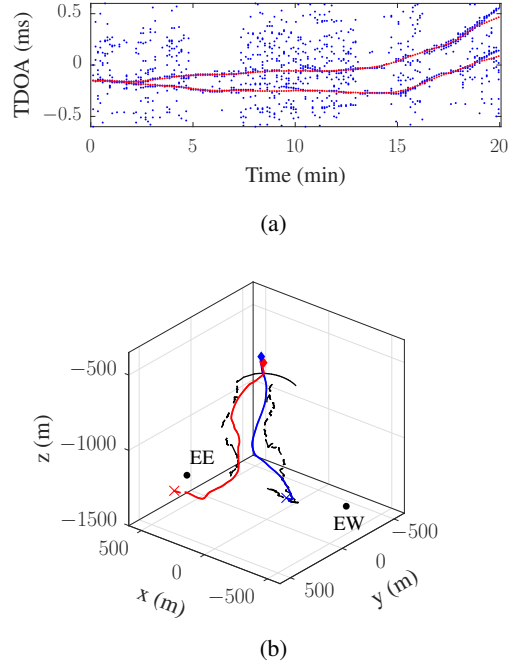


Fig. 3: MTT results using real acoustic data collected on the coast of Southern California. Two whales are simultaneously dive into deeper waters. In (a), TDOA measurements (blue dots) and resultant TDOA tracks (red dots) from one sensor of HARP EE are shown. In (b), MTT results (blue and red lines) are compared with results that rely on hand annotation (dashed lines). Diamonds and crosses indicate the start and end of tracks.

were detected at relatively shallow depths of approximately 450 m. The proposed MTT approach was able to track the two whales that started close to each other. By using GCC-WIN, more echolocation clicks were detected such that longer 3-D tracks (approximately by five more minutes) were generated.

In addition, the accumulated runtime of the data processing chain is measured to be approximately 39 min using a MacBook Pro with an Apple M1 Pro chip and 32 GB memory. Of the 39 min, the TDOA tracking across all sensors and the 3-D tracking took 5 and 20 min, respectively.

5. CONCLUSION AND FUTURE WORK

We propose an algorithmic solution for automatically detecting and tracking whales in 3-D based on their echolocation clicks. Acoustic measurements are provided by a volumetric hydrophone array. For each pair of hydrophones, TDOA measurements of echolocation clicks are extracted using a noise-whitening cross-correlation. Tracking is performed in two stages based on an MTT framework using factor graphs and the SPA. In the second stage, particle flow is embedded into the SPA to make tracking computationally feasible in the considered nonlinear and high-dimensional scenario. The presented results based on synthetic and real data demonstrate that the proposed processing chain can detect and track multiple whales in a fully automated way. To enhance tracking performance, especially in scenarios with closely spaced whales, ICIs could be incorporated as random parameters to be estimated. In addition, future work could extend 3-D tracking to an interacting multiple models (IMM) approach where the best model for whale motion is automatically selected during runtime. Finally, the proposed data processing chain will facilitate scientific studies of echolocating whales based on PAM.

6. REFERENCES

- [1] Z. M. X. Walter, *Passive Acoustic Monitoring of Cetaceans*, Cambridge University Press, Cambridge, UK, 2011.
- [2] C. O. Tiemann, M. B. Porter, and L. N. Frazer, "Localization of marine mammals near Hawaii using an acoustic propagation model," *J. Acoust. Soc. Am.*, vol. 115, no. 6, pp. 2834–2843, 2004.
- [3] E.-M. Nosal, "Methods for tracking multiple marine mammals with wide-baseline passive acoustic arrays," *J. Acoust. Soc. Am.*, vol. 134, no. 3, pp. 2383–2392, 2013.
- [4] M. Gassmann, S. M. Wiggins, and J. A. Hildebrand, "Three-dimensional tracking of Cuvier's beaked whales' echolocation sounds using nested hydrophone arrays," *J. Acoust. Soc. Am.*, vol. 138, no. 4, pp. 2483–2494, 2015.
- [5] J. Macaulay, J. Gordon, D. Gillespie, C. Malinka, and S. Northridge, "Passive acoustic methods for fine-scale tracking of harbour porpoises in tidal rapids," *J. Acoust. Soc. Am.*, vol. 141, no. 2, pp. 1120–1132, 2017.
- [6] S. M. Wiggins, B. J. Thayre, J. S. Trickey, S. Baumann-Pickering, and J. A. Hildebrand, "MPL technical memorandum 631: Beaked whale passive acoustic tracking offshore of Cape Hatteras 2017," Tech. Rep. N62470-15-D-8006, MPL, La Jolla, CA, Oct. 2018.
- [7] Y. M. Barkley, E.-M. Nosal, and E. M. Oleson, "Model-based localization of deep-diving cetaceans using towed line array acoustic data," *J. Acoust. Soc. Am.*, vol. 150, no. 2, pp. 1120–1132, 2021.
- [8] P. M. Baggenstoss, "A multi-hypothesis tracker for clicking whales," *J. Acoust. Soc. Am.*, vol. 137, no. 5, pp. 2552–2562, 2015.
- [9] P. Gruden, E.-M. Nosal, and E. Oleson, "Tracking time differences of arrivals of multiple sound sources in the presence of clutter and missed detections," *J. Acoust. Soc. Am.*, vol. 150, no. 5, pp. 3399–3416, 2021.
- [10] T. Kropfreiter, F. Meyer, S. Coraluppi, C. Carthel, R. Mendrzik, and P. Willett, "Track coalescence and repulsion: MHT, JPDA, and BP," in *Proc. Int. Conf. Inf. Fusion (FUSION)*, Nov. 2021, pp. 1–8.
- [11] B. N. Vo and W. K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4091–4104, 2006.
- [12] F. Meyer, A. Tesei, and M. Z. Win, "Localization of multiple sources using time-difference arrival measurements," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2017, pp. 3151–3155.
- [13] A. Tesei, F. Meyer, and R. Been, "Tracking of multiple surface vessels based on passive acoustic underwater arrays," *J. Acoust. Soc. Am.*, vol. 147, no. 2, pp. EL87–EL92, 2020.
- [14] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, 1976.
- [15] F. Meyer, T. Kropfreiter, J. L. Williams, R. Lau, F. Hlawatsch, P. Braca, and M. Z. Win, "Message passing algorithms for scalable multitarget tracking," *Proc. IEEE*, vol. 106, no. 2, pp. 221–259, 2018.
- [16] F. Daum and J. Huang, "Nonlinear filters with log-homotopy," in *Proc. SPIE*, Aug. 2007, pp. 423–437.
- [17] W. Zhang and F. Meyer, "Graph-based multiobject tracking with embedded particle flow," in *Proc. IEEE Int. Radar Conf.*, May 2021, pp. 1–6.
- [18] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*, Yaakov Bar-Shalom, Storrs, CT, 2011.
- [19] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Upper Saddle River, NJ, 1993.
- [20] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
- [21] F. Meyer, P. Braca, P. Willett, and F. Hlawatsch, "A scalable algorithm for tracking an unknown number of targets using multiple sensors," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3478–3493, July 2017.
- [22] Y. Li and M. Coates, "Particle filtering with invertible particle flow," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4102–4116, Aug. 2017.
- [23] A. Frantzis, J. C. Goold, E. K. Skarsoulis, M. I. Taroudakis, and V. Kandia, "Clicks from Cuvier's beaked whales, *Ziphius cavirostris* (L)," *J. Acoust. Soc. Am.*, vol. 112, no. 1, pp. 34–37, 2002.
- [24] F. Meyer and J. L. Williams, "Scalable detection and tracking of geometric extended objects," *IEEE Trans. Signal Process.*, vol. 69, pp. 6283–6298, 2021.
- [25] T. Li, M. Bolic, and P. M. Djuric, "Resampling methods for particle filtering: Classification, implementation, and strategies," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 70–86, 2015.
- [26] F. Daum and J. Huang, "Curse of dimensionality and particle filters," in *Proc. IEEE Aerosp. Conf.*, Mar. 2003, vol. 4, pp. 1979–1993.
- [27] B. Ristic, B.-T. Vo, B.-N. Vo, and A. Farina, "A tutorial on Bernoulli filters: Theory, implementation and applications," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3406–3430, 2013.
- [28] S. M. Wiggins and J. A. Hildebrand, "High-frequency acoustic recording package (HARP) for broad-band, long-term marine mammal monitoring," in *Proc. IEEE Int. Symp. Underw. Technol. (UT)*, Apr. 2007, pp. 551–557.