Automated Acoustic Tracking of a Sperm Whale (*Physeter macrocephalus*) using a Wide Baseline Array of Sensors

Pina Gruden
Cooperative Institute for Marine and Atmospheric Research
and
Ocean and Resources Engineering
University of Hawai'i at Manoa
Honolulu, HI 96822, USA
Email: pgruden@hawaii.edu

Eva-Marie Nosal
Ocean and Resources Engineering
University of Hawai'i at Manoa
Honolulu, HI 96822, USA
Email: nosal@hawaii.edu

E. Elizabeth Henderson
Whale Acoustic Reconnaissance Project
NIWC Pacific
San Diego, CA 92152-5001, USA
Email: erin.e.henderson.civ@us.navy.mil

***Abstract*—Underwater acoustics is a key tool for monitoring marine environments and understanding marine mammal populations. However, extracting meaningful information from passive acoustic recordings poses challenges due to overlapping signals, species-specific vocalization behavior, and missed and false detections. Many methods for marine mammal tracking and localization rely on human operators for signal detection and measurement association, which is a subjective and laborious process. In this paper we demonstrate a fully automated framework for marine mammal tracking and localization using wide-baseline arrays based on a multi-target Bayesian approach. Leveraging a "track-before-localize" strategy and fusing information from multiple sensors and virtual sensors, the framework eliminates the need for detection, classification, or association steps, thereby improving efficiency and objectivity. The feasibility and performance of the proposed framework are demonstrated using real-world data of a clicking sperm whale from the US Navy's AUTEC test range.**

I. Introduction

Underwater acoustics plays a key role in monitoring the marine environment, providing essential data to various governmental agencies and stakeholders about endangered marine mammal populations [1]. Marine mammals produce a variety of sounds, which can be used to identify and monitor the individual species [2]. Understanding species occurrence, abundance and distribution is a key element in assessing ecosystems health, implementing and assessing mitigation measures.

In order to extract meaningful information from passive acoustics recordings, information from multiple sensors must be reconciled, and overlapping signals need to be extracted, tracked and localized. This problem incorporates classical multi-sensor, multi-target tracking challenges such as signal detection, measurement origin uncertainty, and targets appearance and disappearance, which are common challenges across a wide range of engineering disciplines [3]–[6]. In addition, typical animal vocalization behaviour can lead to significant variations in signal availability and characteristics across species, which can hinder the monitoring output [7].

A. Previous work- marine mammal tracking and localization

Various types of hydrophone arrays, including fixed, towed, and drifting arrays, are used to localize and track marine mammals [8]–[11]. Most marine mammal localization methods rely on the Time Difference Of Arrival (TDOA) of the signal between pairs of hydrophones (or "virtual" hydrophones to incorporate reflections). Among these, model-based localization methods can be applied when iso-speed assumptions are violated [9], [12]–[14]. Model-based methods also provide a framework that allows environmental uncertainty (e.g. in sound speed and phone position) to be propagated forward to give corresponding error bounds and uncertainties in estimated locations.

Typically, localization methods rely on (a) detecting the signals of interest (e.g. echolocation clicks), (b) classifying, associating, and pruning direct and reflected arrivals [14]–[16], and (c) extracting TDOAs and/or Direct Reflected Time Differences (DRTDs) to locate the sources. In multiple animal situations, various approaches that use single animal localization methods, after associating viable TDOAs, have been used. These usually rely on either a source separation/association step [8], [17], or on a spurious TDOA (from incorrect associations) pruning step [16], [18]–[20]. In most methods human operators are needed to manually make decisions on the presence of animals and perform measurement-to-track associations to form TDOA tracks, or associate tracked TDOAs between different hydrophone pairs to obtain localization, which is a time consuming and subjective process. Moving toward more general methods, Ref. [9] introduced a multiple-animal model-based localization method, relying on clustering and fixed rules to associate (and reject) detections during the tracking phase. Ref. [21] introduced a multi-target tracking method based on multi hypothesis tracking to track clicking animals.

Situations involving multiple sources, false alarms, and

missed detections are ideally suited for multi-target tracking (MTT) approaches that jointly estimate the number, states and trajectories of targets. Automated approaches based on multi-target Bayesian methods have recently been proposed for marine mammal tracking and localization [6], [11], [22]. Ref. [6] used an MTT approach based on Gaussian Mixture Probability Hypothesis Density (GM-PHD) [23] to extract TDOAs of false killer whales from towed array data using joint whistle and echolocation click information. Ref. [11] used a graph-based MTT method to track clicking beaked whales with compact volumetric arrays. Ref. [22] used both GM-PHD and graph-based MTT methods to track singing humpback whales with vector sensors.

B. Contributions

The fundamental question addressed in this paper is the feasibility of developing a completely automated framework, and thus eliminating human operator steps, for marine mammal tracking from passive acoustic data from wide-baseline arrays. The contributions are:

- Development of a fully automated framework for tracking biological sources underwater by leveraging a "track-before-localize" strategy that does not require detection, classification, or association steps.
- Fusing information from multiple sensors, as well as virtual sensors, to better inform 3D spatial estimates (virtual sensors are constructed by exploiting underwater multipath arrivals in the recordings, and are the mirror image of the real sensor positions with respect to the sea surface/seabed).
- Demonstration of the tracking and localization performance on the real-world data from the US Navy's AUTEC test range.

II. Background

A. Random Finite Set tracking: the PHD filter

Target tracking is often achieved using Bayesian methods [24], [25]. Among these is the random finite sets (RFS) framework [25]. RFS prompted development of non-traditional MTT approaches that are data-association free and can sometimes outperform traditional MTT methods [25]. RFS provides a Bayesian framework for recursive update of the multi-target posterior density based on the noisy measurements, and it incorporates the missed detections and false alarms in the problem formulation. Detailed information can be found in Refs [25] and [3]. The filters developed within this framework have been successfully applied to MTT applications across a broad range of disciplines including tracking targets in sonar [4], tracking multiple speakers in reverberant environments [5], and tracking multiple biological sources in spectrograms and correlograms [6], [26], [27].

The Probability Hypothesis Density (PHD) filter is one of the filters formulated within the Random Finite Sets (RFS) framework [28]. It is an approximation to the multi-target Bayes filter, and propagates the first-order statistical moment (termed the PHD) of the multi-target posterior distribution at discrete time intervals [28]. An analytical solution to the PHD filter, termed the Gaussian Mixture Probability Hypothesis Density (GM-PHD) [23], is obtained by assuming linear Gaussian models for the underlying dynamics and noise processes. The GM-PHD filter approximates the PHD function by a mixture of weighted Gaussian components, and propagates it recursively via a two-stage prediction and update procedure. At each time step, new targets are introduced via the birth model, target states are estimated based on the posterior PHD, and computational efficiency is maintained through pruning and merging techniques [23]. Extensions to the measurement model that incorporate additional features have been proposed to better discriminate between targets and clutter, and strategies to reduce the bias in the number of estimated targets when initiating new targets based on measurements have also been proposed [6], [29], [30].

TABLE I
HYDROPHONE RELATIVE LOCATIONS AND DEPTHS (REF. [31])

Hydrophone	x (m)	y (m)	z (m)
G	10658.04	-14953.63	-1530.55
H	12788.99	-11897.12	-1556.14
I	14318.86	-16189.18	-1553.58
J	8672.59	-18064.35	-1361.93
K	12007.50	-19238.87	-1522.54

III. Tracking approach

A. Data

Development and testing were accomplished on the sperm whale (*Physeter macrocephalus*) "bench-mark" data from the 2nd International Workshop on Detection Classification Localization and Density Estimation (DCLDE). The data come from the US Navy's Atlantic Undersea Test and Evaluation Center (AUTEC) test range and were prepared by the Naval Undersea Warfare Center (NUWC) [31]. The AUTEC range consists of wideband, bottom-mounted hydrophones. The DCLDE dataset consists of data from five sensors located approximately 2 - 4 nmi apart, at a depth of roughly 1500 m (Table I). One dataset consists of 25 min of recordings of a single vocalizing sperm whale. The dataset was corrected for a time offset of 2.3359 seconds to properly time-align the recordings [31]. While no real ground truth data exists for this dataset, prior analyses [14]–[16] provide a comparison point for this study.

B. Signal processing and tracking workflow

Our proposed framework consists of two main parts (Fig. 1). In the first part, a "track-before-localize" strategy is employed to track TDOA information from multiple pairs of sensors. In the second part, a "localize-then-track" strategy is used to track sources in the 3-D spatial domain.

The data is first preprocessed by applying a fourth-order bandpass Butterworth filter with cutoff frequencies of 2 and 20 kHz. We use both cross- and auto-correlation techniques [32], [33] to obtain correlograms (cross- and auto-correlograms) for each sensor pair. Cross-correlograms and auto-correlograms

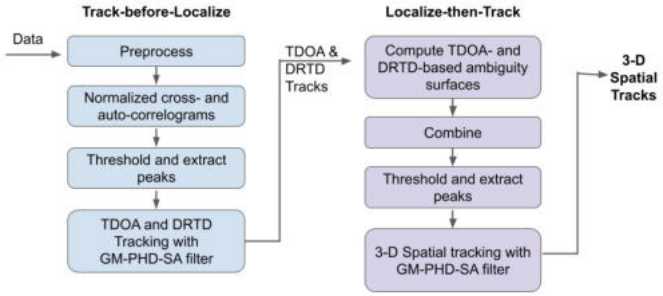

Fig. 1. Overall workflow of our framework. The framework consists of two parts: Part 1- "Track-before-Localize" and Part 2- "Localize-then-Track".

are constructed by computing an envelope of the generalized cross- or auto-correlation function across a 10 s window, then advancing the window by 2.5 s step (i.e., 75% overlap) across the full recording length. Correlograms are normalized so that the background has a Rayleigh parameter of 1 [6]. Cross-correlograms are used to identify TDOAs of the signal's direct path between a pair of sensors. They also contain "ghost tracks" that originate from the close agreement between a signal's direct and surface-reflected paths. Auto-correlograms are used to identify DRTDs of the signal's direct and surface-reflected paths on the same sensor. Auto-correlograms also contain "ghost tracks" from agreement between direct arrivals of different clicks and direct and surface-reflected arrivals of different (successive) clicks.

In contrast to most marine mammal tracking approaches, we do not try to classify and/or eliminate "ghost tracks". Rather, we track all targets with a version of the GM-PHD filter: GM-PHD-SA filter ("S" stands for the separate prediction and update for the newborn and persistent targets, and "A" stands for the amplitude) [6]. This filter uses amplitude information as an additional feature (to time) and updates newborn and persistent targets separately [6], [29], [30]. It outputs automatically extracted TDOA and DRTD tracks, and can be thought of as a "decluttering" step for the next part ("Localize-then-Track") of the framework (Fig. 1). The GM-PHD-SA tracking process is performed on each sensor pair (for TDOAs) and each sensor with its virtual counterpart (for DRTDs). This results in a set of TDOA tracks for each sensor pair and a set of DRTDs for each sensor.

In the second ("Localize-then-Track") part of our framework, the extracted TDOA and DRTD tracks are used to compute ambiguity surfaces. Ambiguity surfaces, which are probabilistic indicators of source positions [9], [12], [14], are created for each time step k by comparing modeled TDOAs/DRTDs to the measurements (extracted tracks). The ambiguity values for a candidate source position $\boldsymbol{w}$ between phones i and j are calculated as:

$$\mathrm{AS}_{\mathrm{DRTD}}(\boldsymbol{w},k) \propto \prod_{ii} \exp\left[\frac{-1}{2\sigma_{ii}^2}\left(\tau_{ii}-\hat{\tau}_{ii}(\boldsymbol{w},k)\right)^2\right], \quad (1)$$

$$\mathrm{AS}_{\mathrm{TDOA}}(\boldsymbol{w},k) \propto \prod_{ij} \exp\left[\frac{-1}{2\sigma_{ij}^2}\left(\tau_{ij}-\hat{\tau}_{ij}(\boldsymbol{w},k)\right)^2\right], \quad (2)$$

where τ_{ii} and τ_{ij} are measured DRTDs and TDOAs, respectively; $\hat{\tau}_{ii}(\boldsymbol{w})$ and $\hat{\tau}_{ij}(\boldsymbol{w})$ are modelled DRTDs and TDOAs, respectively, at position $\boldsymbol{w}=[p_x,p_y,p_z]$; and σ_{ii} and σ_{ij} are standard deviations in DRTD and TDOA errors, respectively, and account for errors in arrival-time measurements, sensor positions, and propagation model. The product in Eqs. (1) and (2) is taken over minimum number of phones $n_{\min}=4$, and the combination of $n_{\min}$ that results in the highest ambiguity value at given $\boldsymbol{w}$ is retained.

In results presented here, the modelled $\hat{\tau}_{ii}(\boldsymbol{w})$ and $\hat{\tau}_{ij}(\boldsymbol{w})$ are obtained using a Bellhop ray tracing propagation model [34]. Bellhop was used to create a lookup table of direct and surface-reflected arrivals for a list of candidate source ranges and depths for each receiver. We used the same historic sound speed profile as in [14] (taken from the Generalized Digital Environmental Model at 24° 45' N, 77° 45' W for March). The depth list varied from 0 to 1700 m with 25 m increments, and the range list varied from 0 m to 10 km with 25 m increments. Since arrival times varied smoothly for the depths and ranges of interest, required TDOAs and DRTD are interpolated from the values in the lookup table.

For each sensor pair, the ambiguity surfaces have the highest values (close to 1) along the hyperboloids of possible source locations. The hyperboloids intersect with high combined ambiguity value, Eqs. (1)-(2), at a source location (Fig. 2a). Low ambiguity values when a constant sound speed is assumed (Fig. 2b) illustrate the importance of using a propagation model to account for depth-dependent sound speed profiles in the dataset.

The total ambiguity surface at a given position $\boldsymbol{w}$ for a given time k is then:

$$\mathrm{AS}(\boldsymbol{w},k) \propto \mathrm{AS}_{\mathrm{DRTD}}(\boldsymbol{w},k) \times \mathrm{AS}_{\mathrm{TDOA}}(\boldsymbol{w},k) \quad (3)$$

Total ambiguity surfaces are thresholded, then peaks (that correspond to potential source location) are extracted and clustered using the k-means clustering algorithm. The number of clusters is selected automatically through unsupervised "gap" evaluation statistics [35]. The extracted peaks are connected/tracked using the GM-PHD-SA filter and thus spatial 3-D tracks of animal movement are obtained.

C. Models

We use the GM-PHD-SA filter [6] to track targets in both TDOA and DRTD and spatial domains. Some of the parameters in "Track-before-Localize" part of the framework were estimated based on a small set of hand-annotated data. For this purpose cross- and auto-correlograms were annotated, then the parameters were statistically estimated from manually annotated tracks and from comparison between extracted measurements and hand-annotations. The following assumptions and models are used.

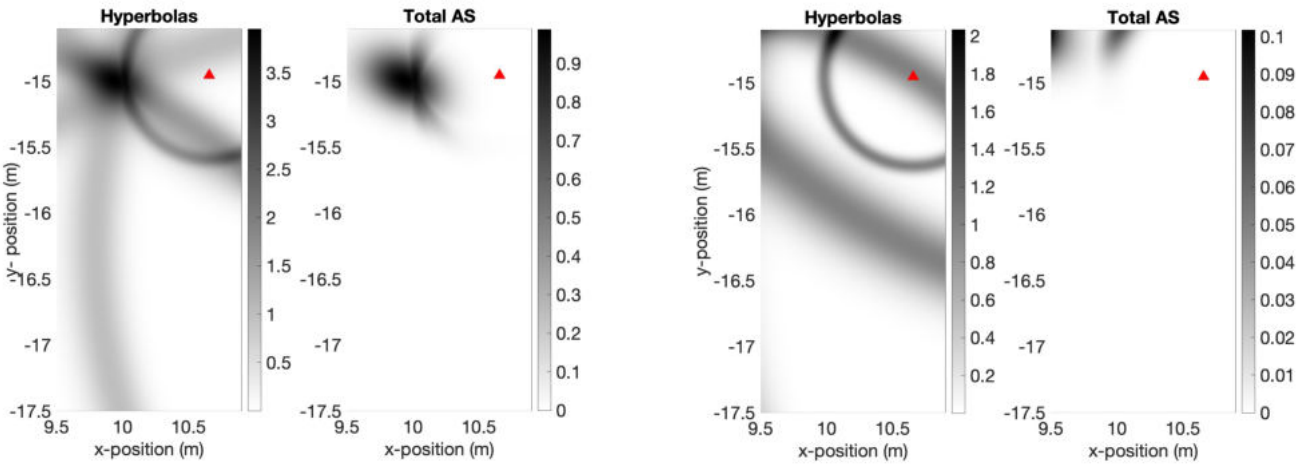

Fig. 2. DRTD-based ambiguity surfaces (AS) for the first 10 s of data at 690 m depth (depth at which source is likely located based on this and previous studies [14], [16]), when using (a) depth-dependent sound speed profile and (b) isospeed profile with c = 1510 m/s. Triangles indicate receiver position. The colorbars indicate the value of the surface: in case of hyperbolas they reflect the total summed value, and in case of AS they reflect the total product.

The target state $\boldsymbol{x}_k$ develops according to a nearly constant velocity (NCV) model [36] and has survival probability $p_S = 0.99$. At time step k, the target state is defined as:

$$\boldsymbol{x}_k = \boldsymbol{F}\boldsymbol{x}_{k-1} + \boldsymbol{n}_{k-1} = \begin{bmatrix} \boldsymbol{I}_n & \triangle \boldsymbol{I}_n \\ \boldsymbol{0}_n & \boldsymbol{I}_n \end{bmatrix} \boldsymbol{x}_{k-1} + \boldsymbol{n}_{k-1}\,, \quad (4)$$

where $\boldsymbol{F}$ is the state transition matrix, $\boldsymbol{I}_n$ and $\boldsymbol{0}_n$ denote $n \times n$ identity and zero matrices respectively, $\triangle$ denotes the time step between windows, $\boldsymbol{n}_{k-1}$ is the zero-mean white noise process with a system noise covariance matrix $\boldsymbol{Q}$ defined as [36]:

$$\boldsymbol{Q} = \sigma_\nu^2 \begin{bmatrix} \frac{\triangle^4}{4}\boldsymbol{I}_n & \frac{\triangle^3}{2}\boldsymbol{I}_n \\ \frac{\triangle^3}{2}\boldsymbol{I}_n & \triangle^2 \boldsymbol{I}_n \end{bmatrix} \quad (5)$$

where σ_ν is the standard deviation of the system noise.

In the "Track-before-Localize" part of the framework, the state $\boldsymbol{x}_k = [\tau, \dot{\tau}]^T$ consists of TDOA/DRTD (τ) and rate of change of change of TDOA/DRTD ($\dot{\tau}$), where $[\cdot]^T$ denotes transpose. In Eqs. (4)-(5) $n = 1$, and the time step $\triangle = 2.5$ s. The standard deviation of the system noise σ_ν in Eq. (5) is $\sigma_\nu = 2 \times 10^{-4}$ (1/s) and $\sigma_\nu = 1.2 \times 10^{-4}$ (1/s) for TDOAs and DRTDs respectively. The standard deviation values were estimated from a set of hand-annotated data.

In the "Localize-then-Track" part of the framework, the state $\boldsymbol{x}_k = [p_x, p_y, p_z, \dot{p_x}, \dot{p_y}, \dot{p_z}]^T$ consist of position (p_x, p_y, p_z) and velocity $(\dot{p_x}, \dot{p_y}, \dot{p_z})$. In Eqs. (4)-(5) $n = 3$, and the time step $\triangle = 10$ s. The standard deviation of the system noise σ_ν in Eq. (5) is $\sigma_\nu = 0.1$ (m/s^2).

A target is detected with a probability of detection p_D and generates a measurement. The measured TDOA/DRTD positions $\boldsymbol{z}_k$ at time step k are related to the target states through the following measurement model:

$$\boldsymbol{z}_k = \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{\eta}_k = \begin{bmatrix} I_n & 0_n \end{bmatrix} \boldsymbol{x}_k + \boldsymbol{\eta}_k\,, \quad (6)$$

where $\boldsymbol{H}$ denotes the measurement matrix, and $\boldsymbol{\eta}_k$ is the zero-mean white noise process with a measurement noise matrix $\boldsymbol{R}$:

$$\boldsymbol{R} = \sigma_r^2 \boldsymbol{I}_n\,, \quad (7)$$

where σ_r denotes the standard deviation of the measurement noise.

In the first part of the framework $n = 1$ in Eqs. (6)-(7). The standard deviation of the measurement noise σ_r in Eq. (7) is $\sigma_r = 0.025$ s and $\sigma_r = 0.001$ s for TDOAs and DRTDs respectively. Targets are detected with $p_D = 0.5$ and $p_D = 0.9$ for TDOAs and DRTDs, respectively. σ_r and p_D values were estimated from a set of hand-annotated data.

In the second part of the framework, targets are detected with $p_D = 0.8$ and $n = 3$ in Eqs. (6)-(7). The standard deviation of the measurement noise is $\sigma_r = 10$ m in Eq. (7).

In addition to time/position, the amplitude (of the envelope of the cross-/auto-correlation and of the ambiguity surfaces) is also measured but is not directly propagated through the filter. Rather, it is used to inform the weights of the newborn targets and the weights of targets and clutter in the update step of the filter [6]. New targets are initialized based on measurements [6], [29], [30] with a birth rate ν_b. In the first part of the framework, the birth rate is estimated from hand-annotated data to be $\nu_b = 5 \times 10^{-3}$ and $\nu_b = 2 \times 10^{-3}$ for TDOAs and DRTDs, respectively. In the second part of the framework $\nu_b = 5 \times 10^{-4}$.

The clutter measurements are assumed to follow a Poisson model. They are uniformly distributed over the observation space: between τ_{max} and τ_{min} with clutter rate $r_c = 26$ in the "Track-before-Localize" part of the framework, and between $[8673, 14319] \times [-19239, -11897] \times [0, -1362]$ with clutter rate $r_c = 1$ for the "Localize-then-Track" part.

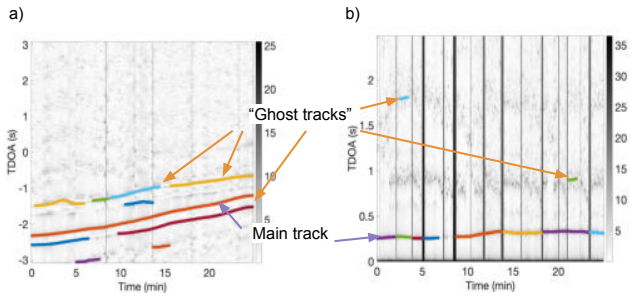

Fig. 3. Examples of extracted TDOA tracks from phones G and I (a) and DRTD tracks from phone I (b) with the GM-PHD-SA filter are shown. The extracted tracks contain both the true target track (which can be detected in fragments) and false positive tracks (which result from "ghost tracks"). In a) "ghost tracks" originate from the close agreement between a signal's direct and surface-reflected paths (reflected phone G - reflected phone I, direct phone G - reflected phone I, reflected phone G - direct phone I). In b) "ghost tracks" originate from agreement between direct arrivals of different clicks and direct and surface-reflected arrivals of different (successive) clicks. The dark vertical lines occurring in both a) and b) are instances between click trains (when animal stops vocalizing and no animal generated signals are produced).

The pruning parameters of the GM-PHD filter are: maximum allowed number of Gaussian components $J_{max} = 100$, pruning threshold $T_r = 1 \times 10^{-3}$, merging threshold for Mahalanobis distance $U = 4$, and weight threshold $w_{th} = 0.5$.

IV. Results

In the first part of the framework, the TDOA and DRTD tracks are successfully extracted from the normalized cross- and auto-correlograms. An example based on phone pair G and I is shown in Fig. 3. The extracted tracks are often fragmented (as seen in the main DRTD track in Fig.3.b). This happens when there are no measurements available to the filter for several consecutive steps, either due to animal stopping vocalizing, change in the signal-to-noise-ratio or when no reflections are present in the recordings. In addition to the main TDOA/DRTD tracks "ghost tracks" are also extracted.

Total ambiguity surfaces, Eq. (3), are formed every 10 s based on both TDOA and DRTD extracted tracks, using $\sigma_{ii} = \sigma_{ij} = 0.009$. Peaks in the surface that exceed a threshold of 0.6 are extracted and clustered to form the 3-D spatial measurements. Based on these measurements, the whale is tracked using the GM-PHD-SA filter (Figs. 4-6). The false positive tracks ("ghost tracks") from the first part of the framework do not result in consistent 3-D measurements, and the filter tracks the whale successfully through gaps in measurements (Figs. 4-6). The results compare well with previously reported trajectory in Refs. [15], [16]. In contrast to other methods applied to this dataset, no explicit detection of the echolocation clicks, or pruning of the reflections or false targets was performed: these get resolved automatically within our proposed signal processing workflow.

V. Conclusion

This paper presents a fully automated framework for tracking marine mammals using passive acoustic data from wide-baseline arrays. The approach consists of "Track-before-Localize" and "Localize-then-Track" steps, fusing information from multiple sensors, including virtual sensors. The GM-PHD-SA filter is used to track targets in TDOA and 3-D spatial domains, and the results compare well with previous studies on the same real-world dataset.

This work is an initial proof of concept showing that the proposed framework can track a biological target from raw passive acoustic data without the input from a human operator. This is a significant advancement compared to conventional bio-acoustics methods, as it drastically reduces processing time

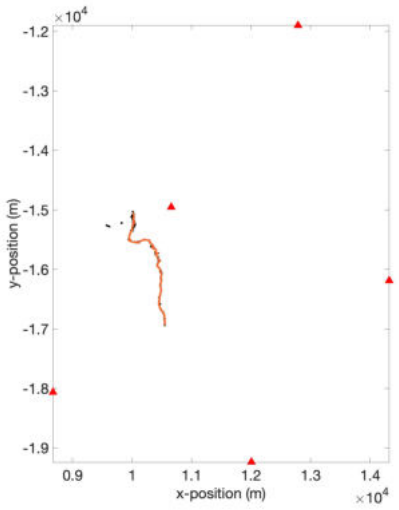

Fig. 4. 2-D view of the automatically extracted spatial track of a sperm whale with the GMPHD-SA filter. Black dots denote extracted measurements from the combined ambiguity surfaces; colored line denotes extracted track; triangles denote receiver positions. The results compare well to Fig. 6 in Ref [16] and Fig. 3 in Ref [15].

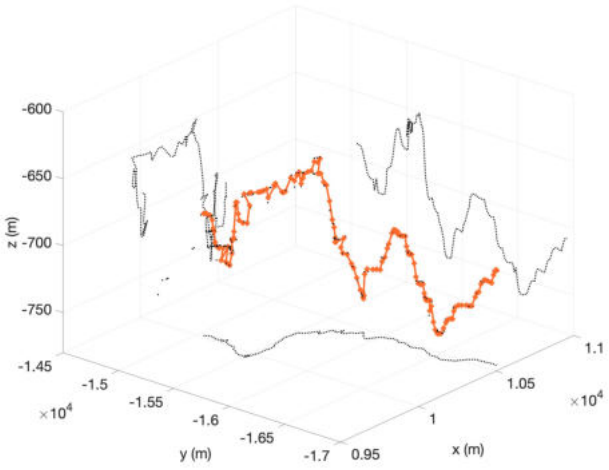

Fig. 5. 3-D view of the automatically extracted spatial track of a sperm whale with the GMPHD-SA filter. Black dots denote extracted measurements from the combined ambiguity surfaces; colored line denotes extracted tracks. Projections onto the three planes are shown with dotted lines.

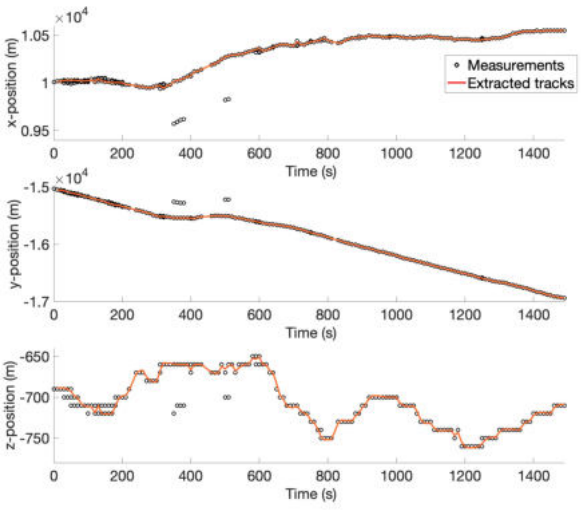

Fig. 6. Automatically extracted spatial track of a sperm whale with the GMPHD-SA filter. x (top), y (middle) and z (bottom) coordinate through time. Black *o* denote extracted measurements from the combined ambiguity surfaces; and colored line denotes extracted track. The results compare well to Fig. 2 in Ref. [15].

and removes the subjective aspect of human analyst decision-making.

While in theory the two steps of MAMBAT could be combined into a single step by using a non-linear measurement model and a corresponding non-linear filter, we have opted to reduce the computational complexity and cost by using two linear models instead. In addition, by using the two step approach we can examine the intermediate results and fine tune the two parts separately.

To asses the generalization capabilities of the proposed framework, future studies will employ it to different multiple animal scenarios, different species (that produce diverse narrowband and broadband signals), and a wider spatial range (and bigger number of sensors).

Acknowledgment

The authors would like to thank Office of Naval Research (ONR) Marine Mammals and Biology (MMB) program [Award No. N00014-22-1-2772] for funding this research. Data were provided by the Naval Undersea Warfare Center; special thanks to D. Moretti, R. Morrissey, and N. DiMarzio.

References

[1] S. M. Van Parijs, C. W. Clark, R. S. Sousa-Lima, S. E. Parks, S. Rankin, D. Risch, and I. van Opzeeland, "Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales," *Marine Ecology Progress Series*, vol. 395, pp. 21–36, 2009.

[2] J. N. Oswald, J. Barlow, and T. F. Norris, "Acoustic identification of nine delphinid species in the eastern tropical Pacific Ocean," *Marine Mammal Science*, vol. 19, no. 1, pp. 20–37, 2003.

[3] B.-N. Vo, M. Mallick, Y. Bar-Shalom, S. Coraluppi, R. Osborne III, R. Mahler, and B.-T. Vo, "Multitarget tracking," in *Wiley Encyclopedia of Electrical and Electronics Engineering*. John Wiley and Sons, Inc., 2015, pp. 1–23.

[4] A.-A. Saucan, T. Chonavel, C. Sintes, and J.-M. Le Caillec, "CPHD-DOA tracking of multiple extended sonar targets in impulsive environments," *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1147–1160, 2015.

[5] C. Evers, A. H. Moore, P. A. Naylor, J. Sheaffer, and B. Rafaely, "Bearing-only acoustic tracking of moving speakers for robot audition," in *2015 IEEE International Conference on Digital Signal Processing (DSP)*. IEEE, 2015, pp. 1206–1210.

[6] 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," *The Journal of the Acoustical Society of America*, vol. 150, no. 5, pp. 3399–3416, 2021.

[7] P. Gruden, Y. M. Barkley, and J. L. McCullough, "Vocal behavior of false killer whale (pseudorca crassidens) acoustic subgroups," *Frontiers in Marine Science*, vol. 10, p. 1147670, 2023.

[8] A. Thode, "Tracking sperm whale (*Physeter macrocephalus*) dive profiles using a towed passive acoustic array," *The Journal of the Acoustical Society of America*, vol. 116, no. 1, pp. 245–253, 2004.

[9] E.-M. Nosal, "Methods for tracking multiple marine mammals with wide-baseline passive acoustic arrays," *The Journal of the Acoustical Society of America*, vol. 134, no. 3, pp. 2383–2392, 2013.

[10] J. Barlow and E. T. Griffiths, "Precision and bias in estimating detection distances for beaked whale echolocation clicks using a two-element vertical hydrophone array," *The Journal of the Acoustical Society of America*, vol. 141, no. 6, pp. 4388–4397, 2017.

[11] J. Jang, F. Meyer, E. R. Snyder, S. M. Wiggins, S. Baumann-Pickering, and J. A. Hildebrand, "Bayesian detection and tracking of odontocetes in 3-d from their echolocation clicks," *arXiv preprint arXiv:2210.12318*, 2022.

[12] C. O. Tiemann, M. B. Porter, and L. N. Frazer, "Localization of marine mammals near Hawaii using an acoustic propagation model," *The Journal of the Acoustical society of America*, vol. 115, no. 6, pp. 2834–2843, 2004.

[13] A. Thode, "Three-dimensional passive acoustic tracking of sperm whales (*Physeter macrocephalus*) in ray-refracting environments," *The Journal of the Acoustical Society of America*, vol. 118, no. 6, pp. 3575–3584, 2005.

[14] E.-M. Nosal and L. N. Frazer, "Track of a sperm whale from delays between direct and surface-reflected clicks," *Applied Acoustics*, vol. 67, no. 11-12, pp. 1187–1201, 2006.

[15] ——, "Sperm whale three-dimensional track, swim orientation, beam pattern, and click levels observed on bottom-mounted hydrophones," *The Journal of the Acoustical Society of America*, vol. 122, no. 4, pp. 1969–1978, 2007.

[16] P. R. White, T. G. Leighton, D. C. Finfer, C. Powles, and O. N. Baumann, "Localisation of sperm whales using bottom-mounted sensors," *Applied Acoustics*, vol. 67, no. 11-12, pp. 1074–1090, 2006.

[17] P. M. Baggenstoss, "Separation of sperm whale click-trains for multipath rejection," *The Journal of the Acoustical Society of America*, vol. 129, no. 6, pp. 3598–3609, jun 2011.

[18] P. Giraudet and H. Glotin, "Real-time 3d tracking of whales by echo-robust precise tdoa estimates with a widely-spaced hydrophone array," *Applied Acoustics*, vol. 67, no. 11-12, pp. 1106–1117, 2006.

[19] J. L. Spiesberger, "Finding the right cross-correlation peak for locating sounds in multipath environments with a fourth-moment function," *The Journal of the Acoustical Society of America*, vol. 108, no. 3, pp. 1349–1352, 2000.

[20] R. P. Morrissey, J. Ward, N. DiMarzio, S. Jarvis, and D. J. Moretti, "Passive acoustic detection and localization of sperm whales (*Physeter macrocephalus*) in the tongue of the ocean," *Applied acoustics*, vol. 67, no. 11-12, pp. 1091–1105, 2006.

[21] P. M. Baggenstoss, "A multi-hypothesis tracker for clicking whales," *The Journal of the Acoustical Society of America*, vol. 137, no. 5, pp. 2552–2562, may 2015.

[22] P. Gruden, J. Jang, A. Kügler, T. Kropfreiter, L. Tenorio-Hallé, M. O. Lammers, A. Thode, and F. Meyer, "Automating multi-target tracking of singing humpback whales recorded with vector sensors," *The Journal of the Acoustical Society of America*, vol. 154, no. 4, pp. 2579–2593, 2023.

[23] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, 2006.

[24] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking,"

IEEE Transactions on signal processing, vol. 50, no. 2, pp. 174–188, 2002.

[25] R. P. Mahler, *Statistical multisource-multitarget information fusion*. Norwood, MA, USA: Artech House, Inc., 2007.

[26] P. Gruden and P. R. White, "Automated tracking of dolphin whistles using Gaussian mixture probability hypothesis density filters," *The Journal of the Acoustical Society of America*, vol. 140, no. 3, pp. 1981–1991, 2016.

[27] ——, "Automated extraction of dolphin whistles - A sequential Monte Carlo probability hypothesis density approach," *The Journal of the Acoustical Society of America*, vol. 148, no. 5, pp. 3014–3026, 2020.

[28] R. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1152–1178, 2003.

[29] B. Ristic, D. Clark, B.-N. Vo, and B.-T. Vo, "Adaptive target birth intensity for PHD and CPHD filters," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1656–1668, 2012.

[30] D. Clark, B. Ristic, B.-N. Vo, and B. T. Vo, "Bayesian multi-object filtering with amplitude feature likelihood for unknown object SNR," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 26–37, 2010.

[31] O. Adam, J.-F. Motsch, F. Desharnais, N. DiMarzio, D. Gillespie, and R. C. Gisiner, "Overview of the 2005 workshop on detection and localization of marine mammals using passive acoustics," *Applied Acoustics*, vol. 67, no. 11-12, pp. 1061–1070, 2006.

[32] G. C. Carter, "Coherence and time delay estimation," *Proceedings of the IEEE*, vol. 75, no. 2, pp. 236–255, 1987.

[33] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE transactions on acoustics, speech, and signal processing*, vol. 24, no. 4, pp. 320–327, 1976.

[34] M. Porter, "BELLHOP (a Gaussian beam/finite element beam code)," Available in the *Acoustics Toolbox*, http://oalib.hlsresearch.com/AcousticsToolbox/, 2005, last accessed January 2024.

[35] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411–423, 2001.

[36] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation*. John Wiley & Sons Inc., 2001.