Automated tracking of dolphin whistles using Gaussian Mixture Probability Hypothesis Density (GM-PHD) filters

Pina Gruden[a] and Paul R. White

Institute of Sound and Vibration Research

University of Southampton

Highfield, Hants, SO17 1BJ, UK

[a]e-mail: pg3g12@soton.ac.uk

Abstract

This work considers automated Multi Target Tracking (MTT) of odontocete whistle contours. An adaptation of Gaussian Mixture Probability Hypothesis Density (GM-PHD) filter is described and applied to the acoustic recordings from six odontocete species. From the raw data, spectral peaks are first identified and then GM-PHD filter is used to simultaneously track the whistles' frequency contours. Overall over 9000 whistles are tracked with a precision of 85% and recall of 71.8%. The proposed filter is shown to track whistles precisely (with mean deviation of 104 Hz, about one frequency bin, from the annotated whistle path) and 80% coverage. The filter is computationally efficient, suitable for real-time implementation, and is widely applicable to different odontocete species.

I. INTRODUCTION

The detection of marine mammal vocalizations plays an important role in passive acoustic monitoring. The objectives of such studies include species recognition[24,29,9], species presence and abundance estimation[21], studying species behaviour[26], mitigation during industrial activities[35]. Odontocetes (toothed whales) produce a rich variety of high-frequency vocalizations, which can be grouped into three broad categories: whistles, echolocation clicks and burst pulses[2], all of which have most of their energy above 2 kHz[31]. This work focuses on whistles, which are highly variable, narrowband, frequency modulated, tonal sounds with fundamental frequencies generally between 2 and 30 kHz and are typically used in a social context[14]. Not all odontocete species whistle, but majority of delphinid species do.

Methods used for detection and frequency estimation of odontocete whistles vary from semi-automated methods *e.g.,*[14,24] to fully automated methods *e.g.,*[8,36,11,28,12,9,13]. Most methods are based on spectrogram techniques, although alternative approaches also exist *e.g.,*[12,11].

Prior to applying a detection algorithm to the signal, some pre-processing of data is typically carried out in order to reduce background noise and interfering signals *e.g.,*[8,36,12,22,28,9]. After the noise removal, spectrogram-based algorithms usually identify strongest spectral peaks *e.g.,*[10,15,28,22,9] or apply image-processing techniques to define the pixels[20] or ridges that represent whistles[13]. The identified peaks are then connected into a continuous whistle contour using different approaches, such as particle filtering[36,28], Kalman filtering[20], combination of polynomial fitting and Kalman filtering[13]; hypothesis tracking with some gating rules[10,22]; phase tracking[11,12].

The automated methods for whistle contour detection are commonly based on the algorithms that allow for single target tracking. In this work, an alternative approach is taken in which the detection and tracking of frequency content of delphinid whistles is considered as a multi-target tracking (MTT) problem, where whistles are targets that overlap, their numbers are unknown and vary with time and there are interfering signals present. An MTT algorithm called Gaussian Mixture Probability Hypothesis Density (GM-PHD) filter[16,32], which has been previously used in sonar applications[6], was adapted here for application of dolphin whistle contour tracking. The paper is organized as follows. In Section II some background is given on target tracking and PHD filters. Section III introduces formulation of the GM-PHD filter for dolphin whistle tracking and derivation of models and parameters for this particular problem. The performance of the proposed GM-PHD filter is tested on the acoustic recordings of dolphin whistles, which have been hand-annotated and results are given in Section IV. Discussion and conclusions may be found in Sections V and VI respectively. Appendix summarizes the most frequently used symbols and their meanings.

II. BACKGROUND

A. Target Tracking

Target tracking is a process of estimating a target's state as it evolves in time, from a sequence of noisy measurements. A target is broadly defined as the entity to be tracked and the state vector, $\boldsymbol{x}_k$, contains the information about the properties of the target at time k. The only available information about the targets is given by the measurement vector, $\boldsymbol{z}_k$, which also typically contains noise. In the case of whistle frequency contour tracking, each whistle represents a target. The target state vector consists of frequency and chirp (rate of change of frequency) information and the measurement vector consists of frequency peaks. Measurements may also be contaminated by the detection of false targets (clutter) and points where there has been a failure to detect a target.

In order to perform target tracking at least two models are required; first a model describing the evolution of the state with time, called the system (or dynamic) model

$$\boldsymbol{x}_k = \Phi_k(\boldsymbol{x}_{k-1}, \boldsymbol{n}_{k-1}) \tag{1}$$

where Φ_k is a system function that describes the evolution of the state vector and $\boldsymbol{n}_{k-1}$ is a system noise process and is a vector of random variables specifying the random component of the parameter evolution[1,36]. From the system model one can define a state transition density $f_{k|k-1}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$, which characterizes the transition of the state from time $k-1$ to time k. The second model required is a model relating the noisy measurement to the state, called the measurement (or observation) model

$$\boldsymbol{z}_k = \psi_k(\boldsymbol{x}_k, \boldsymbol{\eta}_k) \tag{2}$$

where ψ_k is a function that defines the measurement process and $\boldsymbol{\eta}_k$ is the measurement noise process[1,36]. From the measurement model one can obtain a likelihood function $g_k(\boldsymbol{z}_k|\boldsymbol{x}_k)$, that describes the likelihood that a measurement $\boldsymbol{z}_k$ was generated by the target $\boldsymbol{x}_k$. These models are collectively known as a state-space model.

Target tracking is typically achieved with the use of a recursive Bayesian filter where one attempts to construct the posterior probability density function (pdf) of the state, $p_k(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$, based on the set of measurements $\boldsymbol{z}_{1:k}$ up to time k[1]. Such a filter involves a two stage process; prediction and update, where the system model is used to predict the state pdf and the measurements are used to refine that prediction[1]. This is implemented in a recursive manner and at each time step an estimate of the state is obtained from the posterior pdf.

In the case of single target tracking, it is assumed that only one target is present and that all the observations are generated by that target. If the system and measurement models are linear and the noise processes are Gaussian, then optimal target tracking is achieved with the Kalman filter[5], which in this case represents the optimal solution to Bayesian recursion. If the models are non-linear and/or the noise is non-Gaussian, particle filters can be used to perform single target tracking[36].

In the majority of real-world applications there are multiple targets present at any given time, the number of which will change through time as targets appear (*i.e.* target birth) and disappear (*i.e.* target death). At each time k there are n_k target states $\boldsymbol{x}_{k,1}, ..., \boldsymbol{x}_{k,n_k}$ and m_k measurements $\boldsymbol{z}_{k,1}, ..., \boldsymbol{z}_{k,m_k}$. The states of the targets and the observations can be modelled using the concept of a random finite set. A random finite set is an object in which the elements have random values, as in any multivariate random process, but in addition to which the number of elements in the set is also random[27]. The random set of states (multi-target state), $\boldsymbol{X}_k$, and the random set of measurements (multi-target measurement), $\boldsymbol{Z}_k$, are represented as follows:

$$\boldsymbol{X}_k = \{\boldsymbol{x}_{k,1}, ..., \boldsymbol{x}_{k,n_k}\} \in \mathcal{F}(\mathcal{X}) \tag{3}$$

$$\boldsymbol{Z}_k = \{\boldsymbol{z}_{k,1}, ..., \boldsymbol{z}_{k,m_k}\} \in \mathcal{F}(\mathcal{Z}) \tag{4}$$

where $\mathcal{F}(\mathcal{X})$ and $\mathcal{F}(\mathcal{Z})$ are the finite subsets of the state and observation spaces $\mathcal{X}$ and $\mathcal{Z}$, respectively.

In this case the use of multi target tracking (MTT) techniques is required and the objective is to jointly estimate the number of targets and their states from the noisy measurements[32]. Traditional approaches to MTT are based on data association techniques and involve explicit associations between measurements and targets that are achieved with the use of single target tracking techniques. Examples of traditional MTT include nearest neighbor (NN), joint probabilistic data association (JPDA) and multiple hypothesis tracking (MHT)[4].

However, the uncertainty in the evolution of the multi-target state and the origin of the multi-target measurement is naturally modelled by random finite sets[27] and therefore data association-free techniques, based on Mahler's finite set statistics (FISST) framework (an overview is provided in[19]), have been increasingly used in the last decade for the Bayesian multi-target filtering problems. A multi-target Bayesian filter determines at each time step k the posterior probability density of multitarget-state $p_k(\boldsymbol{X}_k|\boldsymbol{Z}_{1:k})$[27]. The high dimensionality of the Bayes multi-target filter makes the recursion intractable in practice, a problem which is overcome using the Probability Hypothesis Density (PHD) filter[16,17].

B. Probability Hypothesis Density (PHD) filter

The PHD filter approximates the multi-target Bayes recursion by propagating the first-order statistical moment $v_k(\boldsymbol{x}|\boldsymbol{Z}_{1:k})$ of the multi-target posterior $p_k(\boldsymbol{X}_k|\boldsymbol{Z}_{1:k})$, known as the intensity function or the PHD[17,32,25,27]. The PHD is a function whose peaks identify the likely positions of the targets. By integrating the PHD on any region of the state space one obtains the expected number of targets in that region. It should be noted that PHD is a density function but is not a pdf, since its integral over the space of its variable is not unity[19]. A target with state $\boldsymbol{x}$ is more likely to be present in the region when the PHD (intensity function) is large than when it is small, which allows one to obtain state estimates of the targets based on peaks in the PHD.

The PHD filter comprises both prediction and update steps. In the prediction step, the PHD filter incorporates the motion of individual targets and accounts for disappearance of existing targets (by incorporating the probability of target's survival). In addition it incorporates the appearance of completely new targets. Hence, the predicted intensity function, $v_{k|k-1}(\cdot)$, consists of the newborn targets (introduced by the birth intensity function) and the existing targets (targets surviving from the previous time step that are represented by the posterior intensity function from the previous time step $v_{k-1}(\cdot)$). The abbreviation $v_k(\boldsymbol{x}|\boldsymbol{Z}_{1:k}) \overset{abbr}{=} v_k(\boldsymbol{x}_k)$ is used and the prediction step can be expressed as[32,25,27]

$$v_{k|k-1}(\boldsymbol{x}_k) = \gamma_k(\boldsymbol{x}_k) + \langle p_{S,k}(\boldsymbol{x}_{k-1})v_{k-1}(\boldsymbol{x}_{k-1}), f_{k|k-1}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})\rangle \tag{5}$$

where $\gamma_k(\boldsymbol{x}_k)$ denotes the PHD of target births between time $k-1$ and k; $p_{S,k}(\boldsymbol{x}_{k-1})$ denotes the probability of survival, that is probability that a target with state $\boldsymbol{x}$ at time $k-1$ will survive until time k; $f_{k|k-1}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ denotes single-target state transition density from time $k-1$ to k and $\langle g, f\rangle = \int f(x)g(x)dx$. Note that spawning terms, that define how one target can become resolved into more than one target, have been omitted from the above equation. This is because rarely, if ever, does one observe a dolphin whistle contour which splits into two distinct contours.

In the update step, the PHD filter incorporates the probability that any given target was not detected (by incorporating the probability of target detection) and updates the predicted intensity with a set of measurements by also taking into the account the measurement likelihood function and false alarms (clutter). The posterior intensity function $v_k(\cdot)$ at time step k is given by

$$v_k(\boldsymbol{x}_k) = [1 - p_{D,k}(\boldsymbol{x}_k)]v_{k|k-1}(\boldsymbol{x}_k) + \sum_{\boldsymbol{z} \in Z_k} \frac{p_{D,k}(\boldsymbol{x}_k) g_k(\boldsymbol{z}|\boldsymbol{x}_k) v_{k|k-1}(\boldsymbol{x}_k)}{\kappa_k(\boldsymbol{z}) + \langle p_{D,k}(\boldsymbol{x}_k) g_k(\boldsymbol{z}|\boldsymbol{x}_k), v_{k|k-1}(\boldsymbol{x}_k)\rangle} \tag{6}$$

where $p_{D,k}(\boldsymbol{x}_k)$ denotes the probability of detection, that is the probability that observation will be collected at time k from a target with state $\boldsymbol{x}_k$, $\boldsymbol{Z}_k$ denotes the multi-target measurement at time k, $\kappa_k(\boldsymbol{z})$ denotes denotes the PHD of clutter at time k and $g_k(\boldsymbol{z}|\boldsymbol{x}_k)$ denotes the single-target measurement likelihood function at time k.

The computational load of the PHD filter can grow significantly if target births can occur uniformly in the state space. One approach to mitigate this is to adapt the birth intensity according to the measurements[27], which results in the prediction and update steps being preformed separately for newborn and existing targets. A label β is introduced to distinguish between the two types of targets; $\beta = 0$ refers to existing targets, $\beta = 1$ refers to newborn targets. The prediction stage becomes[27]

$$
\begin{aligned}
v_{k|k-1}(\boldsymbol{x}_k, \beta) &= \gamma_k(\boldsymbol{x}_k) && \beta = 1 \\
&= \langle p_{S,k}(\boldsymbol{x}_{k-1}) v_{k-1}(\boldsymbol{x}_{k-1}), f_{k|k-1}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})\rangle && \beta = 0
\end{aligned} \tag{7}
$$

where $v_{k-1}(\cdot)$ represents posterior intensity function from the previous time step and consists of posterior intensity functions of existing and newborn targets from the previous time step $(v_{k-1}(\cdot, 0) + v_{k-1}(\cdot, 1))$.

The update stage of the filter for existing targets ($\beta = 0$) can be expressed as[27]

$$
\begin{aligned}
v_k(\boldsymbol{x}_k, 0) = {} & [1 - p_{D,k}(\boldsymbol{x}_k)] v_{k|k-1}(\boldsymbol{x}_k, 0) \\
& + \sum_{\boldsymbol{z} \in Z_k} \frac{p_{D,k}(\boldsymbol{x}_k) g_k(\boldsymbol{z}|\boldsymbol{x}_k) v_{k|k-1}(\boldsymbol{x}_k, 0)}{\mathcal{L}(\boldsymbol{z})}
\end{aligned} \tag{8}
$$

and for newborn targets ($\beta = 1$)

$$
v_k(\boldsymbol{x}_k, 1) = \sum_{\boldsymbol{z} \in Z_k} \frac{g_k(\boldsymbol{z}|\boldsymbol{x}_k)\gamma_k(\boldsymbol{x}_k)}{\mathcal{L}(\boldsymbol{z})} \tag{9}
$$

where

$$
\mathcal{L}(\boldsymbol{z}) = \kappa_k(\boldsymbol{z}) + \langle g_k(\boldsymbol{z}|\boldsymbol{x}_k), \gamma_k\rangle + \langle p_{D,k}(\boldsymbol{x}_k) g_k(\boldsymbol{z}|\boldsymbol{x}_k), v_{k|k-1}(\boldsymbol{x}_k, 0)\rangle \tag{10}
$$

Note that since newborn targets are created from the measurements, the newborn targets are always detected, *i.e.* $p_D(\boldsymbol{x}, 1) = 1$[27].

It can be seen from the above equations that in addition to the system and measurement models (from which the $f_{k|k-1}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ and $g_k(\boldsymbol{z}|\boldsymbol{x}_k)$ are obtained respectively), the PHD filter requires definition of additional models and parameters. Specifically, the target's survival ($p_{S,k}(\boldsymbol{x}_{k-1})$) and detection ($p_{D,k}(\boldsymbol{x}_k)$) probabilities and clutter ($\kappa_k(\boldsymbol{z})$) and target birth ($\gamma_k(\boldsymbol{x}_k)$) models. The formulation of these is described in the Section III.B.2.

The above equations still involve integrals that typically have no closed form solution and therefore the PHD filter needs to be approximated[32,25]. Practical implementations of PHD filters include Gaussian Mixture PHD (GM-PHD)[32] and Sequential Monte Carlo PHD (SMC-PHD)[33] filters. In this work the GM-PHD approach was chosen since it tends to be faster and more straightforward than the SMC-PHD approach[18]. The GM-PHD filter and its application to a specific problem of dolphin whistle tracking is presented in the next section.

III. METHODOLOGY

A. Data, pre-processing steps and obtaining the measurements

The data set used in this study was obtained from the 5th Workshop of Detection, Classification, Localization and Density Estimation (DCLDE) conference 2011 (available at MobySound archive, http://www.mobysound.org). This dataset contained raw data and analyst-annotated files for six species: long-beaked common dolphin (*Delphinus capensis*), short-beaked common dolphin (*Delphinus delphis*), melon-headed whales (*Peponocephala electra*), spinner dolphin (*Stenella longirostris*), Atlantic spotted dolphin (*Stenella frontalis*) and bottlenose dolphin (*Tursiops truncatus*). The recordings contained in this dataset were single-species recordings that were confirmed by trained visual observers. Study areas, data collection protocols and procedure for hand-annotation of the data are summarized in Roch *et al.*[28], Baumann-Pickering *et al.*[3] and Soldevilla *et al.*[30]. The raw data was used for the GM-PHD filter to track the whistles from and hand-annotations were used to evaluate the performance of the filter. In addition, a small part of raw data was set aside to be used as training data for certain parameters of the GM-PHD filter. For this purpose three files were randomly selected from the annotated dataset and a 1 minute section of each of those files was taken as the training data. These training files corresponded to three species, *D. capensis*, *D.delphis* and *S.frontalis*, and were obtained using different recording equipment. This training data was subsequently not used in the performance evaluation.

For ease of implementation, where necessary, the data was re-sampled to 192 kHz (before re-sampling 2.5% of the files were sampled at 300 kHz, 12.5% at 480 kHz and 85% at 192 kHz). After re-sampling, pre-processing was applied to the data in order to reduce the background noise and interfering signals. A pre-processing scheme was adapted from Gillespie *et al.*[9] and was applied with a sliding window that was 2048 points long and had 50% overlap, resulting in 93.8 Hz spacing between frequency bins. Within each window the following steps were performed as described in Gillespie *et al.*[9]: first echolocation clicks were removed by applying a weighting function; then spectrogram was computed on a decibel scale, using 2048 point Hanning window, and spectral peaks were enhanced by applying normalization across frequency based on a 61 point median filter; after that the normalization across time using exponential moving average (with the weighting constant of 0.02) was performed in order to remove persistent tones from the spectrogram.

In each window, after the noise was removed, spectral peaks were determined by identifying all frequencies whose normalized magnitude exceeded 8 dB. Only frequency bins between 2 and 50 kHz were searched for peaks, since most dolphin whistles will lie within this range and to be consistent with the hand annotations which were also applied to whistle harmonics. The identified spectral peaks represent the measurement set from which the whistle contours were tracked using the Gaussian Mixture PHD (GM-PHD) filter.

Measurement sets containing spectral peak measurements and a list of all files used in this study, as well as Matlab implementation of the method for obtaining spectral peak measurements was released to the MobySound archive.

B. Whistle contour tracking with Gaussian Mixture PHD (GM-PHD) filters

The GM-PHD filter algorithm[32] was implemented and used to track frequency contours of whistles from the identified spectral peaks. In this approximation to the PHD filter, the posterior intensity function $v_k(\boldsymbol{x}_k)$ is represented by a sum of weighted Gaussian components whose weights, means and covariances are propagated in time[25]. This strategy is analogous to Kalman filter[5] for single target tracking, which propagates the first moment (the mean) of the single-target state[32]. So that each whistle at time k is represented by a Gaussian component and is therefore characterized by a mean (consisting of frequency and chirp), a weight and a covariance. The means and covariances of the existing and newborn whistles are predicted using the Kalman filter prediction equations and updated with the received measurements (spectral peaks) also using the Kalman equations. The weights of the whistles are predicted and updated using the PHD equations and they can be thought of as a measure of the likelihood of presence of a component. Detailed description of the GM-PHD filter is given next.

The whistle estimates generated by the GM-PHD filter do not inherently contain identity. In order to assign a particular state to a specific whistle, tracking of Gaussian components needs to be carried out. Tracking is achieved by labelling each individual Gaussian component with a unique tag and the likelihood of each track is then given by the weight of each component *e.g.*,[7,25,34].

This section is organized as follows. First the GM-PHD algorithm is outlined, then filter's models and parameters are defined, followed by a description of the performance evaluation.

1. The GM-PHD algorithm

The GM-PHD filter approximates the intensity functions (PHDs) with Gaussian mixtures. It should be noted that these do not share the properties of GM approximations to pdfs in terms of weights summing to 1. Here the sum of weights reflects the number of whistles present at each time step. The GM-PHD filter makes the following assumptions. It is assumed that each whistle follows a linear Gaussian dynamical model and that measurements follow a linear model[32]. That is, (1) and (2) can be written as

$$\boldsymbol{x}_k = F_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{n}_{k-1} \tag{11}$$

$$\boldsymbol{z}_k = H_k\boldsymbol{x}_k + \boldsymbol{\eta}_k \tag{12}$$

where $\boldsymbol{x}_k$ and $\boldsymbol{z}_k$ denote the sate and measurement vectors respectively, F_{k-1} and H_k denote state transition and measurement matrices respectively, $\boldsymbol{n}_{k-1}$ denotes system noise with covariance matrix Q_{k-1} and $\boldsymbol{\eta}_k$ denotes measurement noise with covariance matrix R_k. So the state transition density function and measurement likelihood function are Gaussian:

$$f_{k|k-1}(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = \mathcal{N}(\boldsymbol{x}; F_{k-1}\boldsymbol{x}_{k-1}, Q_{k-1}) \tag{13}$$

$$g_k(\boldsymbol{z}_k|\boldsymbol{x}_k) = \mathcal{N}(\boldsymbol{z}; H_k\boldsymbol{x}_k, R_k) \tag{14}$$

where $\mathcal{N}(\cdot; m, P)$ denotes a Gaussian density with mean m and covariance P.

It is also assumed that the probability of survival and detection are state independent and constant between time steps

$$p_{S,k}(\boldsymbol{x}) = p_S \tag{15}$$

$$p_{D,k}(\boldsymbol{x}) = p_D \tag{16}$$

The intensity function of target birth is also assumed to be a Gaussian mixture[32]

$$\gamma_k(\boldsymbol{x}_k) = \sum_{i=1}^{J_{\gamma,k}} w_{\gamma,k}^{(i)} \mathcal{N}(\boldsymbol{x}; m_{\gamma,k}^{(i)}, P_{\gamma,k}^{(i)}) \tag{17}$$

where $J_{\gamma,k}$, $w_{\gamma,k}^{(i)}$, $m_{\gamma,k}^{(i)}$, $P_{\gamma,k}^{(i)}$, $i = 1, \cdots, J_{\gamma,k}$ are given model parameters that determine the shape of the birth intensity function, which is derived in Section III.B.2.

The algorithm then consists of the following steps:

Step 0: Initialization. At the initialization (time k=*0*) the intensity function v_0 is a mixture of J_0 Gaussian components

$$v_0(\boldsymbol{x}) = \sum_{i=1}^{J_0} w_0^{(i)} \mathcal{N}(\boldsymbol{x}; m_0^{(i)}, P_0^{(i)}) \tag{18}$$

In this study J_0 is initialized randomly to be between 1 and 10 components, means m_0 of those components are drawn randomly from a uniform distribution between 2 and 30 kHz and the initial covariance P_0 is set to be the same as the system noise covariance, Q_{k-1}. The initial weights of all components are the same and are set to $w_0 = 1/J_0$.

Each component is assigned a unique tag (identifier), $L_0^{(i)}$, to form a set $L_0 = \{L_0^{(i)}\}_{i=1}^{J_0}$[7,25].

Step 1: Prediction. In this step the Kalman filter prediction equations are used to predict means (m) and covariances (P) of the Gaussian components representing existing whistles. The weights (w) for existing whistles depend on the probability of survival, p_S.

The predicted intensity of existing whistles, $v_{k|k-1}(\boldsymbol{x}, 0)$, at time k is a Gaussian mixture of the form[32,7]:

$$v_{k|k-1}(\boldsymbol{x}, 0) = p_S \sum_{j=1}^{J_{k-1}} w_{k-1}^{(j)} \mathcal{N}(\boldsymbol{x}; m_{k|k-1}^{(j)}, P_{k|k-1}^{(j)}) \quad (19)$$

$$m_{k|k-1}^{(j)} = F_{k-1} m_{k-1}^{(j)} \quad (20)$$

$$P_{k|k-1}^{(j)} = F_{k-1} P_{k-1}^{(j)} F_{k-1}^t + Q_{k-1} \quad (21)$$

where J_{k-1} denotes the number of existing whistles derived from the previous time step (combination of existing and newborn whistles) and w_{k-1} denotes the weights from the previous time step.

In this step $J_{\gamma,k}$ new Gaussian components, representing newborn whistles, are also created according to the birth model (defined in Section III.B.2, Eqs. (38 and 39)).

The tags of the Gaussian components in this step are maintained separately; existing whistles keep their tags, $L_{k|k-1}$, from the previous time step and new tags, $L_{\gamma,k}^{(i)}, i =$

$1, \cdots, J_{\gamma,k}$, are assigned to Gaussians introduced by the birth model so that

$$L_{k|k-1} = L_{k-1} \tag{22}$$

$$L_{\gamma,k} = \{L_{\gamma,k}^{(1)}, \cdots, L_{\gamma,k}^{(J_{\gamma,k})}\} \tag{23}$$

Step 2: Update. In this step the predicted means and covariances of existing and newborn whistles are updated using the Kalman filter update equations. The predicted weights are updated with the PHD equation. The update is performed separately for existing and newborn whistles, Eqs. (8) and (9) respectively.

For the existing whistles the posterior intensity function at time k is given by a Gaussian mixture[32,7]:

$$v_k(\boldsymbol{x}, 0) = (1 - p_D)v_{k|k-1}(\boldsymbol{x}, 0) + \sum_{\boldsymbol{z} \in Z_k} \sum_{j=1}^{J_{k|k-1}} w_k^{(j)}(\boldsymbol{z})\mathcal{N}(\boldsymbol{x}, m_k^{(j)}(\boldsymbol{z}), P_k^{(j)}) \tag{24}$$

where $(1 - p_D)$ denotes the probability of missed detection at current time k; $\boldsymbol{z}$ denotes an individual measurement in the measurement set Z_k at time k and

$$w_k^{(j)}(\boldsymbol{z}) = \frac{p_D w_{k|k-1}^{(j)} g_k^{(j)}(\boldsymbol{z})}{\mathcal{L}(\boldsymbol{z})} \tag{25}$$

$$g_k^{(j)}(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z}; H_k m_{k|k-1}^{(j)}, R_k + H_k P_k^{(j)} H_k^t) \tag{26}$$

$$m_k^{(j)}(\boldsymbol{z}) = m_{k|k-1}^{(j)} + K_k^{(j)}(\boldsymbol{z} - H_k m_{k|k-1}^{(j)}) \tag{27}$$

$$P_k^{(j)} = [I - K_k^{(j)} H_k] P_{k|k-1}^{(j)} \tag{28}$$

$$K_k^{(j)} = P_{k|k-1}^{(j)} H_k^t (H_k P_{k|k-1}^{(j)} H_k^t + R_k)^{-1} \tag{29}$$

where K_k denotes the Kalman gain and I denotes the identity matrix.

For the newborn whistles the posterior intensity function at time k is also a Gaussian mixture

$$v_k(\boldsymbol{x}, 1) = \sum_{\boldsymbol{z} \in Z_k} \sum_{j=1}^{J_{\gamma,k}} w_{\gamma,k}^{(j)}(\boldsymbol{z}) \mathcal{N}(\boldsymbol{x}, m_{\gamma,k}^{(j)}(\boldsymbol{z}), P_{\gamma,k}^{(j)}) \tag{30}$$

where $m_{\gamma,k}^{(j)}(\boldsymbol{z})$ and $P_{\gamma,k}^{(j)}$ are calculated with Kalman update equations, in the same way as in the equations above and the weights are updated according to Eq. (9)

$$w_{\gamma,k}^{(j)}(\boldsymbol{z}) = \frac{w_{\gamma,k}^{(j)} g_{\gamma,k}^{(j)}(\boldsymbol{z})}{\mathcal{L}(\boldsymbol{z})} \tag{31}$$

where

$$\mathcal{L}(\boldsymbol{z}) = \kappa_k(\boldsymbol{z}) + \sum_{l=1}^{J_{\gamma,k}} w_{\gamma,k}^{(l)} g_{\gamma,k}^{(l)}(\boldsymbol{z}) + p_D \sum_{l=1}^{J_{k|k-1}} w_{k|k-1}^{(l)} g_k^{(l)}(\boldsymbol{z}) \tag{32}$$

$$g_{\gamma,k}^{(l)}(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z}; H_k m_{\gamma,k}^{(l)}, R_k + H_k P_{\gamma,k}^{(l)} H_k^t) \tag{33}$$

At the end of the update step, there are $(1+|Z_k|)J_{k|k-1}$ Gaussian components, $(1+|Z_k|)$ for each predicted Gaussian[32] for existing whistles and $|Z_k|J_{\gamma,k}$ Gaussian components for newborn whistles. The same tag is assigned to each of the associated predicted and updated Gaussian components to form the set[7,25]

$$L_k = L_{k|k-1}^{v_{k|k-1}} \cup L_{k|k-1}^{z_1} \cup \cdots \cup L_{k|k-1}^{z_{|Z_k|}} \tag{34}$$

for existing whistles and for newborn

$$L_{\gamma,k} = L_{\gamma,k}^{z_1} \cup \cdots \cup L_{\gamma,k}^{z_{|Z_k|}} \tag{35}$$

The intensities and tags of existing and newborn whistles are then joined and predicted jointly in the next time step.

With every iteration the number of Gaussian terms will increase, increasing the computational cost of the algorithm. To control this, pruning and merging schemes are applied to the mixture at the end of the update step.

Step 3: Pruning and Merging. Pruning is achieved by truncating all components with small weights by applying a pruning threshold, T_r. In the merging stage, the components that are close together are merged into a single Gaussian component based on a merging threshold U. The distance is computed with a Mahalanobis distance measure[32].

Additionally, to further reduce the computational load, if the number of Gaussian components exceeds the desired maximum number of components (J_{max}), only the J_{max} Gaussian components with the largest weights are kept in the recursion.

The values for T_r, U, J_{max} are discussed in Section III.B.2 and listed in Table I.

Step 4: State estimation and tracking. At the end of each recursion the pruned Gaussian mixture represents the posterior intensity function $v_k(\cdot)$ and the means of the Gaussian components therefore represent local maxima of $v_k(\cdot)$. By taking the Gaussians that have weights greater than some threshold w_{th} (derived in Section III.B.2 and listed in Table I), the multi-target states are estimated[32,7]. This step does not affect the main GM-PHD recursion. The individual whistles are then tracked from the estimated states based on their tags. When a track of a whistle exceeds 150 ms then it is labelled as a detection. The 150 ms length threshold was selected based on the study by Roch *et al.*[28] and serves to reduce the false detections.

2. Definition of the models and parameter selection

State space models for dolphin whistles

The whistle state vectors in this study consist of frequency f and chirp rate α (rate of change of frequency)[36]:

$$\boldsymbol{x}_k = [f, \alpha]^t \tag{36}$$

where $[\cdot]^t$ denotes the transpose.

The system model (11) in current application uses the state transition matrix $F_{k-1} = \begin{bmatrix} 1 & \triangle \\ 0 & 1 \end{bmatrix}$, where $\triangle$ denotes the time interval between overlapping spectral windows and is related to the sampling frequency (f_s), $\triangle = (w_w/2)/f_s$, where w_w denotes the length of the window. The system noise, $\boldsymbol{n}_{k-1}$, in this model is independent Gaussian white noise with a covariance matrix Q_{k-1}. Initially, Q_{k-1} was defined as $Q_{k-1} = diag[\sigma_f^2, \sigma_\alpha^2]$, where σ_f and σ_α denote the standard deviations of the frequency and chirp respectively, here $\sigma_f = 70.7$ and $\sigma_\alpha = 3.2 \times 10^3$.

This noise covariance matrix was then refined by running the GM-PHD filter (described in previous Section III.B.1) on the training data and calculating the mean noise covariance, resulting in

$$Q_{k-1} = \begin{bmatrix} \sigma_f^2 & \sigma_{f,\alpha} \\ \sigma_{f,\alpha} & \sigma_\alpha^2 \end{bmatrix} \tag{37}$$

where the refined standard deviations of frequency and chirp are $\sigma_f = 70.8$ and $\sigma_\alpha =$

7.35×10^3 and the off-diagonal element is $\sigma_{f,\alpha} = 408.4^2$.

The measurement model (12) uses the measurement matrix $H_k = [1, 0]$, indicating that only the frequency information is measured. The measurement noise, $\boldsymbol{\eta}_k$, is independent Gaussian white noise with covariance matrix R_k. R_k in this study is defined as a variance of a uniform random variable and is therefore ${b_w}^2/12$ where b_w denotes bin width and is equal to $b_w = f_s/w_w$.

Other models and parameters

In addition to the system (11) and measurement (12) models required by standard tracking methods, the PHD filter requires definition of additional models and parameters that govern the GM-PHD recursion. All of these are application dependent. Some of the parameters can be determined analytically, but some parameters need to be estimated from training data.

The additional models needed for the GM-PHD filter, model the birth and the clutter intensities. The birth model defines where in the state space new whistles are likely to appear. If a whistle appears in a region that is not covered by the predefined birth intensity then the PHD filter will not detect it[27]. Since dolphin whistles typically occur in a frequency band between 2 and 30 kHz[14], making the birth intensity diffuse over such a large region would increase the computational load. Therefore, the birth intensity in this study is based on the available measurements[27] and the new whistles are created as follows. In each time step k,

$J_{\gamma,k}$ newborn whistles are created, where $J_{\gamma,k}$ corresponds to the number of measurements in the measurement set $\boldsymbol{Z}_k$ at time k. Each newborn whistle is a Gaussian component and is therefore characterized by a mean ($m^{(i)}_{\gamma,k}$), a weight ($w^{(i)}_{\gamma,k}$) and a covariance ($P^{(i)}_{\gamma,k}$), where $i = 1, \cdots, J_{\gamma,k}$. The covariance of the *i-th* newborn whistle is set to be Q_{k-1} (Eq. 37). The frequency component of the mean of the *i-th* newborn whistle ($\{m^{(i)}_{\gamma,k}\}_f$) is obtained by drawing from a Gaussian mixture centred on the measurements and the chirp component of the mean ($\{m^{(i)}_{\gamma,k}\}_\alpha$) is set to zero:

$$\begin{aligned}\{m^{(i)}_{\gamma,k}\}_f &\sim \frac{1}{J_{\gamma,k}} \sum_{j=1}^{J_{\gamma,k}} \mathcal{N}(x; z^{(j)}_{f,k}, 0.01 z^{(j)}_{f,k}) \\ \{m^{(i)}_{\gamma,k}\}_\alpha &= 0 \end{aligned} \tag{38}$$

where $z_{f,k}$ denotes frequency measurements at time k. The weight of the *i-th* newborn whistle is computed as

$$w^{(i)}_{\gamma,k} = \frac{p_{start}(z^{(i)}_{f,k})}{J_{\gamma,k}} \tag{39}$$

where $p_{start}(z^{(i)}_{f,k})$ is a value of the log-normal pdf of starting frequencies of whistles (that was obtained from the training data) at a particular frequency $z^{(i)}_{f,k}$.

The clutter (false detections) intensity used in the present study was computed as follows. It is assumed that clutter is uniformly distributed over the frequency range (2 to 50 kHz) and is constant with respect to time. The average number of clutter points (r) per time step was estimated based on the training data. The training data were pre-processed using the technique described in Section III.A. The number of identified spectral peaks per time step was compared to the number of annotated whistle peaks from the analyst-annotated data. From this the average number of clutter points can be computed. It was determined that our pre-processing technique results in $r = 10$ clutter points per time step, giving the clutter intensity of $\kappa_k = r/A$, where A denotes the bandwidth over which clutter can occur, which is 48 kHz for this study.

In addition to the models for birth and clutter intensities, the GM-PHD filter requires the selection of five other parameters; p_S, p_D, U, T_r, w_{th}. Parameters determined analytically in this study were probability of survival (p_S) and merging threshold (U). Probability of survival, p_S, determines how likely the whistle is to survive from one time step to another. As such it will depend on the average length of the whistles, specifically one can show that $p_S = 1 - (1/\bar{k})$, where $\bar{k}$ is the average length of the whistles expressed in time steps.

The average length of whistles was calculated from the study by Oswald *et al.*[23], where four species were the same as in the present study. The average length was 0.875 s, which equates to 165 time steps (since the time step used in this study is 5.3 ms), giving a p_S of 0.994.

The merging threshold, U, determines which components are merged and is based on the Mahalanobis distance between two Gaussians. Mahalanobis distances are characterized by the Chi-squared distribution with d-degrees of freedom (where d equals the number of variables; in our case, where the state vector consists of frequency and chirp rate, d is equal to 2). For a Chi-squared distribution with 2-degrees of freedom, 99% of all the values coming from this distribution will lie within 9.2. Therefore merging threshold U was set to 10.

Parameters determined experimentally from the data were probability of detection (p_D), pruning threshold (T_r) and weight threshold (w_{th}). All three parameters were determined experimentally by running the GM-PHD filter on the training data and by selecting the values that resulted in the best performance. The parameters used in the GM-PHD for dolphin whistle tracking are summarized in Table I.

3. GM-PHD performance evaluation

After applying the GM-PHD filter described above to the acoustic recordings of dolphin whistles, the detected list of time against frequency peaks for each whistle was compared to the ground truth hand-annotated data in order to evaluate the filter's performance. First the whistles in the hand-annotated data were evaluated in terms of whistles' duration and SNR. The ground truth whistle was only expected to be detected if its duration exceeded 150 ms and if its SNR exceeded 10 dB for at least one third of its duration (following Roch *et al.*[28]). Ground truth whistles meeting these selection criteria were termed valid.

Next the output of the GM-PHD filter was compared to the ground truth whistles. The detected whistle was considered a match (true positive) to a ground truth whistle if its timing overlapped with the ground truth whistle and if the mean difference between the detected whistle path and ground truth whistle path did not exceed 3 frequency bins (281 Hz). If the detected whistle exceeded that criteria, it was considered as false positive. It should be noted that detected whistles were matched to ground truth whistles regardless of whether the ground truth whistles met the selection criteria (*i.e.* if they were valid). However only the whistles that matched valid ground truth whistles were considered in the evaluation metrics that describe the quality and quantity of matches[28]. Also, since the hand-annotations were only applied to the frequencies between 4.5 kHz and 50 kHz, all the detected whistles that had over 40% of the contour below the 4.5 kHz were not taken into account in the evaluation.

The performance of the GM-PHD filter was measured in terms of recall, precision, fragmentation, deviation and coverage. For detailed description see Roch *et al.*[28]. Recall measures the percentage of the expected detections that are retrieved, precision measures the percentage of the detections that are correct. For the detected whistles that matched valid ground truth whistles (true positives), three additional performance metrics are computed; fragmentation, mean deviation and coverage. Fragmentation measures the average number of detections per ground truth whistle, deviation measures the average frequency deviation between the path of ground truth whistle and its corresponding detection and coverage measures the average percentage of a ground truth whistle that is matched.

IV. RESULTS

Across all six species in the selected database, 9192 ground truth whistles met the selection criteria. The performance of the GM-PHD detector for each species is summarized in Table II. The GM-PHD detector tracked whistles successfully with overall precision of 85% and overall recall of 71.8%. Across all species, the whistles were tracked precisely with average deviation from the whistle path of 104 Hz and with coverage of 80.3%. An example of GM-PHD tracking is shown in Figure 1. The detector tracked the paths of individual whistles when overlapping whistles were present, although occasional "breaking" of the whistle contours still occurred (on average there were 1.2 fragments per whistle across all species). An example is shown in Figure 2, where both successful tracking through a crossing and some breaking of the whistle track can be observed.

V. DISCUSSION

This study demonstrated the use of a MTT technique for tracking odontocete whistle contours. The proposed adaptation of the GM-PHD filter successfully simultaneously tracked whistles in complex environments (overlapping whistles, missed detections, clutter present) for all species investigated, despite the parameter optimization being performed on only three of the species in the overall dataset. This suggests that the GM-PHD detector formulation in this study is widely applicable to whistle tracking problems across a wide range of species.

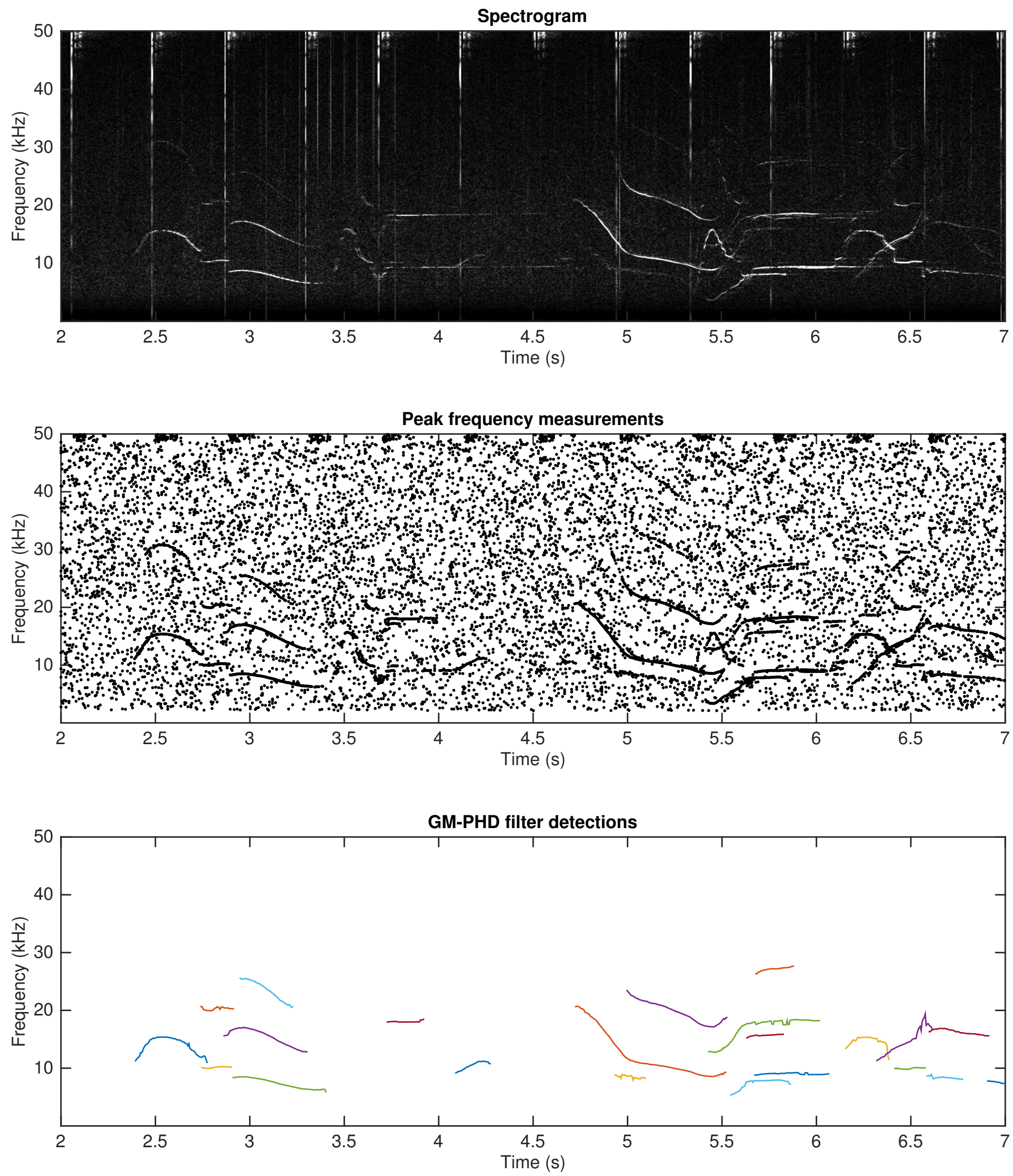

Figure 1: (Color online) Detected whistles with the GM-PHD filter. Spectrogram of raw data is shown (top), peak frequencies measurements (peaks 8 dB above background noise) (middle) and tracked whistles (bottom) where GM-PHD filter detections are shown.

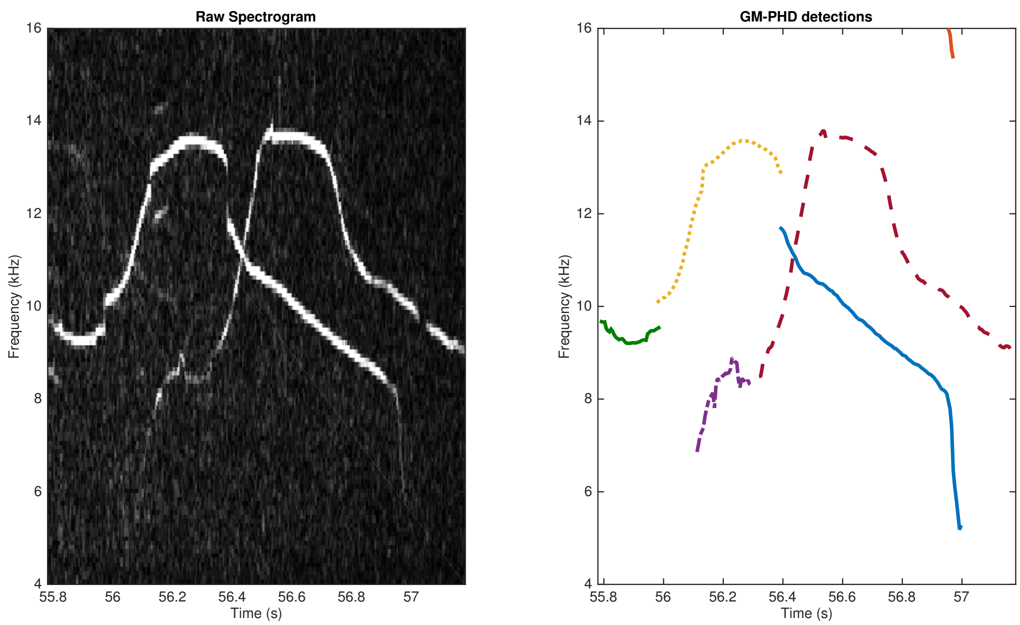

Figure 2: (Color online) Detection of crossing whistles with the GM-PHD filter. Spectrogram of raw data (left) and tracked whistles with GM-PHD filter (right) are shown.

The precision for all species was generally higher than the recall. It should be noted that the precision for *T.truncatus* is slightly lower than the precision for other species (Table II). When *T.truncatus* files were investigated, it was observed that one file in particular contained many burst pulses, which the GM-PHD filter detected as whistles, resulting in a high number of false positives. In general, there is a trade-off between the precision and recall, and the recall could be increased by allowing shorter fragments to be detected (currently a 150 ms threshold is used). However this would, in turn, lower the precision since it would likely increase the number of false positive detections.

For the detected whistles that matched the ground truth data (true positive detections), the performance was quite good. The detected whistles followed the path of the annotated ground truth data closely (within about 1 frequency bin width) and covered the majority of the individual contours. The whistles were mainly detected as a single contour, but were occasionally "broken" into more fragments. The breaking of contours mainly occurred where the amplitude of the whistle dropped below the SNR used to detect spectral peaks and therefore there were no measurements passed to the GM-PHD detector. While the GM-PHD filter allows for missed detections, it cannot continue to track a target if the measurements are absent for several continuous time steps (an example is shown in Figure 2). Also, while the analyst constructing the ground truth data attempted not to trace whistles where the whistle path was not obvious, it was observed from manual inspection of some of the annotated files and corresponding spectrograms, that this was not universally applied. This leads to an increase in the measured fragmentation rate.

Comparing the performance of the GM-PHD filter to other filters is difficult, mainly due to different sound files being used, different pre-processing techniques and different methods to estimate the SNR. However, the results in Table II demonstrate that performance results are comparable to those of the graph filter and better than particle filter detailed in Roch *et al.*[28]. In order to facilitate the comparison between different detectors (that operate on identified spectral peaks), datasets containing detected spectral peaks used in this study have been released to MobySound archive.

Further improving the GM-PHD filter performance is not a trivial task. In general, the performance of the filter will greatly depend on the parameter selection and therefore needs further discussion. In the present study some of the parameters; p_D, κ_k, T_r, w_{th}, were estimated from the training data set. In particular the probability of detection (p_D) was selected by running the GM-PHD filter on all training data and choosing the value that on average resulted in the best performance. Since p_D depends on the SNR, which will change depending on the environment, the animal's location relative to the sensor and the recording equipment, significantly different values might have been obtained if different training files were used. During the GM-PHD recursion the p_D is assumed to be constant, but between different recordings the performance could potentially be improved if the p_D was adjusted to that particular situation. We are currently exploring the methods that would facilitate this. Another parameter estimated from the training data was the clutter intensity, κ_k. While the average number of clutter points per time step appeared to be consistent between species and files in the training set, the value will mainly depend on the threshold used to generate measurements (detected spectral peaks). The value of κ_k would need to be adjusted if a different threshold was used or if a different pre-processing or spectral peak detection strategy was adopted. The selection of pruning (T_r) and weight (w_{th}) thresholds mainly affects the computational speed of the algorithm. By selecting higher values for the two thresholds, fewer Gaussian components remain in the recursion and the speed of the recursion increases. However, if the selected values are too high, the components representing whistles start to be excluded from the recursion, which results in a decrease in performance since fewer whistles are tracked.

In addition, the performance of the GM-PHD filter will also crucially depend on the state-space and birth models used. The birth model in this study was developed from the proposition by Ristic *et al.*[27], where the birth model is based on the measurements. The weights of the newborn whistles were determined based on the probability distribution of the whistles' start frequencies, which were obtained from the training data. Since training data encompassed only three species, future work will investigate whether a model based on more species enhances the performance. The state models, used in this study, describing the evolution of the whistles are based on a simple linear model. Refining this model and developing a more rigorous method to fit its parameters to the training data should also be considered.

One attraction of the GM-PHD filter is that the formulation of the filter is based on the mathematical principles and is not *ad-hoc* as some of the other tracking algorithms. Since the filter is data-association free, it is more computationally efficient than the traditional MTT methods and can be implemented in real-time. It should be noted that the computational speed of the algorithm will not only depend on the parameter selection, but also on the amount of clutter in the measurements. If lower SNR thresholds are used in the measurement generation (spectral peak detection), more clutter is present in the measurements and the computational load increases, which results in slowing the algorithm. Using higher thresholds in the spectral peak detection increases the speed of the algorithm, but some spectral peaks associated with whistles are then missing from the measurements, which affects the tracking performance. So there is an inherent trade-off between the performance and computational speed. To illustrate, for the parameters used in this study, the GM-PHD algorithm implemented in MATLAB (version 8.5 (R2015a)) on a Mac (Os X, processor 2.7 GHz and 8 GB RAM), took 1 min and 48 s to process a file of 1 min duration at 192 kHz sample rate, that contained 103 hand-annotated whistles.

VI. CONCLUSIONS

The proposed formulation of the GM-PHD filter provides a general and powerful tool for simultaneous tracking of odontocete whistle contours. Its performance is comparable with the best existing methods, it is computationally efficient and well suited for real-time implementation.

Acknowledgements

We would like to thank MobySound archive, DCLDE committee and associated analysts for providing the datasets and hand annotations used to test detector's performance in this study. We would like to thank Marie Roch and anonymous reviewer for their helpful comments on an earlier version of this manuscript. We would also like to thank Slovene human resources development and scholarship fund (Ad futura) for funding this research.

APPENDIX

A list of the most frequently used symbols and their meanings.

η_k and R_k	Measurement noise process and its covariance matrix
F_{k-1}	State transition (system) matrix
$f_{k\|k-1}(\boldsymbol{x}_k\|\boldsymbol{x}_{k-1})$	State transition density
$g_k(\boldsymbol{z}\|\boldsymbol{x}_k)$	Likelihood function
$\gamma_k(\boldsymbol{x}_k)$	Intensity function (or PHD) of target births at time k

H_k	Measurement matrix
J_{k-1}	Number of existing targets deriving from previous time step $k-1$
$J_{\gamma,k}$	Number of newborn targets at time k
$\kappa_k(\boldsymbol{z})$	Intensity function (or PHD) of clutter at time k
$\boldsymbol{n}_{k-1}$ and Q_{k-1}	System noise process and its covariance matrix
$p_{D,k}(\boldsymbol{x}_k) \overset{abbr}{=} p_D$	Probability of detection
$p_{S,k}(\boldsymbol{x}_{k-1}) \overset{abbr}{=} p_S$	Probability of target's survival from time $k-1$ to time k
$p_k(\boldsymbol{X}_k \| \boldsymbol{Z}_{1:k})$	Posterior pdf of the multi-target state
β	Label β denotes newborn targets ($\beta = 1$) or existing targets ($\beta = 0$)
$v_{k\|k-1}(\boldsymbol{x}, \beta)$	Predicted intensity function (or PHD)
$v_k(\boldsymbol{x}, \beta)$	Posterior intensity function (or PHD)
w and w_γ	Weights for existing and newborn Gaussian components (whistles) respectively.
$\boldsymbol{x}_k$ and $\boldsymbol{z}_k$	State and measurement vectors at time k
$\boldsymbol{Z}_k$	Multi-target measurement at time k

REFERENCES

1. 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 **50**, 174–188 (2002).

2. W. W. Au, "Hearing in whales and dolphins: An overview", in *Hearing by whales and dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay, (Springer-Verlag, New York, 2000), pp. 1–42.

3. S. Baumann-Pickering, S. M. Wiggins, J. A. Hildebrand, M. A. Roch, and H.-U. Schnitz, "Discriminating features of echolocation clicks of melon-headed whales (*Peponocephala electra*), bottlenose dolphins (*Tursiops truncatus*), and gray's spinner dolphins (*Stenella longirostris longirostris*)", Journal of the Acoustical Society of America **128**, 2212–2224 (2010).

4. S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems* (Artech House, Norwood, MA, 1999), p. 1230.

5. S. Bozic, *Digital and Kalman filtering* (Edward Arnold Ltd, London, UK, 1979), p. 157.

6. D. Clark, B.-N. Vo, and J. Bell, "GM-PHD filter multitarget tracking in sonar images", in *Proceedings of SPIE 6235 - Signal Processing, Sensor Fusion and Target Recognition*

XV, edited by I. Kadar, (International Society for Optics and Photonics, Orlando, Florida, US), pp. 1–8 (2006).

7. D. E. Clark, K. Panta, and B.-N. Vo, "The GM-PHD filter multiple target tracker", in *Proceedings of the 9th International Conference on Information Fusion*, (IEEE, Florence, Italy), pp. 1–8 (2006).

8. S. Datta and C. Sturtivant, "Dolphin whistle classification for determining group identities", Signal processing **82**, 251–258 (2002).

9. D. Gillespie, M. Caillat, J. Gordon, and P. White, "Automatic detection and classification of odontocete whistles", Journal of the Acoustical Society of America **134**, 2427–2437 (2013).

10. X. C. Halkias and D. P. Ellis, "Call detection and extraction using bayesian inference", Applied Acoustics **67**, 1164–1174 (2006).

11. C. Ioana, C. Gervaise, Y. Stphan, and J. I. Mars, "Analysis of underwater mammal vocalisations using timefrequency-phase tracker", Applied Acoustics **71**, 1070–1080 (2010).

12. A. T. Johansson and P. R. White, "An adaptive filter-based method for robust, automatic detection and frequency estimation of whistles", The Journal of the Acoustical Society of America **130**, 893–903 (2011).

13. A. Kershenbaum and M. A. Roch, "An image processing based paradigm for the extraction of tonal sounds in cetacean communications", The Journal of the Acoustical Society of America **134**, 4435–4445 (2013).

14. M. O. Lammers, W. W. L. Au, and D. L. Herzing, "The broadband social acoustic signaling behavior of spinner and spotted dolphins", Journal of the Acoustical Society of America **114**, 1629–1639 (2003).

15. S. Madhusudhana, E. Oleson, M. Soldevilla, M. Roch, and J. Hildebrand, "Frequency based algorithm for robust contour extraction of blue whale B and D calls", in *OCEANS 2008 - MTS/IEEE Kobe Techno-Ocean*, (IEEE, Kobe, Japan), pp. 1–8 (2008).

16. R. P. Mahler, "A theoretical foundation for the Stein-Winter "Probability Hypothesis Density (PHD)" multitarget tracking approach", in *Proceedings of the 2000 MSS National Symposium on Sensor and Data Fusion* , San Antonio, Texas, US, pp. 99–117 (2000).

17. R. P. Mahler, "Multitarget bayes filtering via first-order multitarget moments", IEEE Transactions on Aerospace and Electronic Systems **39**, 1152–1178 (2003).

18. R. P. Mahler, "A survey of PHD filter and CPHD filter implementations", in *Proceedings of SPIE 6567 - Signal Processing, Sensor Fusion and Target Recognition XVI*, edited by I. Kadar, (International Society for Optics and Photonics, Orlando, Florida, US), pp. 1–12 (2007).

19. R. P. Mahler, *Statistical multisource-multitarget information fusion*, (Artech House, Norwood, MA, 2007), p. 856.

20. A. Mallawaarachchi, S. Ong, M. Chitre, and E. Taylor, "Spectrogram denoising and automated extraction of the fundamental frequency variation of dolphin whistles", The Journal of the Acoustical Society of America **124**, 1159–1170 (2008).

21. T. A. Marques, L. Thomas, J. Ward, N. DiMarzio, and P. L. Tyack, "Estimating cetacean population density using fixed passive acoustic sensors: An example with blainvilles beaked whales", The Journal of the Acoustical Society of America **125**, 1982–1994 (2009).

22. D. K. Mellinger, S. W. Martin, R. P. Morrissey, L. Thomas, and J. J. Yosco, "A method for detecting whistles, moans, and other frequency contour sounds", The Journal of the Acoustical Society of America **129**, 4055–4061 (2011).

23. J. N. Oswald, J. Barlow, and T. F. Norris, "Acoustic identification of nine delphinid species in the eastern tropical pacific ocean", Marine Mammal Science **19**, 20–37 (2003).

24. J. N. Oswald, S. Rankin, J. Barlow, and M. O. Lammers, "A tool for real-time acoustic species identification of delphinid whistles", Journal of the Acoustical Society of America **122**, 587–595 (2007).

25. K. Panta, D. E. Clark, and B.-N. Vo, "Data association and track management for the Gaussian mixture probability hypothesis density filter", IEEE Transactions on Aerospace and Electronic Systems **45**, 1003–1016 (2009).

26. N. J. Quick and V. M. Janik, "Whistle rates of wild bottlenose dolphins (*Tursiops truncatus*): influences of group size and behavior", Journal of Comparative Psychology **122**, 305–311 (2008).

27. 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 **48**, 1656–1668 (2012).

28. M. A. Roch, T. S. Brandes, B. Patel, Y. Barkley, S. Baumann-Pickering, and M. S. Soldevilla, "Automated extraction of odontocete whistle contours", Journal of the Acoustical Society of America **130**, 2212–2223 (2011).

29. M. A. Roch, M. S. Soldevilla, J. C. Burtenshaw, E. E. Henderson, and J. A. Hildebrand, "Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California", Journal of the Acoustical Society of America **121**, 1737–1748 (2007).

30. M. S. Soldevilla, E. E. Henderson, G. S. Campbell, S. M. Wiggins, J. A. Hildebrand, and M. A. Roch, "Classification of risso's and pacific white-sided dolphins using spectral properties of echolocation clicks", Journal of the Acoustical Society of America **124**, 609–624 (2008).

31. P. L. Tyack and C. W. Clark, "Communication and acoustic behavior of dolphins and whales", in *Hearing by whales and dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay, (Springer-Verlag, New York, 2000), pp. 156–224.

32. B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter", IEEE Transactions on Signal Processing **54**, 4091–4104 (2006).

33. B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multitarget filtering with random finite sets", IEEE Transactions on Aerospace and Electronic Systems **41**, 1224–1245 (2005).

34. Y. Wang, H. Meng, H. Zhang, and X. Wang, "Improved GM-PHD tracker with delayed decision", in *10th International Conference on Signal Processing*, (IEEE, Beijing, China), pp. 255–258 (2010).

35. C. R. Weir and S. J. Dolman, "Comparative review of the regional marine mammal mitigation guidelines implemented during industrial seismic surveys, and guidance towards a worldwide standard", Journal of International Wildlife Law and Policy **10**, 1–27 (2007).

36. P. White and M. Hadley, "Introduction to particle filters for tracking applications in the passive acoustic monitoring of cetaceans", Canadian Acoustics **36**, 146–152 (2008).

Table I: Summary of parameters used in GM-PHD filter for odontocete whistle tracking. p_S and p_D denote probabilities of survival and detection respectively; U, T_r and w_{th} denote merging, pruning and weight thresholds respectively and J_{max} denotes maximum allowed number of Gaussian components in one iteration.

p_S	p_D	U	T_r	w_{th}	J_{max}
0.994	0.85	10	0.001	0.009	100

Table II: Performance of the GM-PHD filter for detection of odontocete whistle contours. *N files* denotes number of audio files used, *Valid whistles* denotes the number of ground truth whistles that met the selection criteria, *μDeviation* denotes average deviation, *SD* denotes standard deviation. The summary performance is computed across all ground truth whistles that met the criteria and is not the average of file performances.

Species	N files	Valid whistles	Recall	Precision	Coverage ±SD (%)	Fragments ±SD	μDeviation ±SD (Hz)
D.capensis	7	1859	72.1	91.1	80.6±22.3	1.2±0.4	94±51
D.delphis	10	1931	71.6	85.7	79.2±23.2	1.2±0.4	96±53
P.electra	3	756	66.8	91.3	79.8±21.6	1.1±0.3	92±54
S.longirostris	3	869	76.4	93.5	77.2±22.2	1.2±0.5	100±51
S.frontalis	2	242	70.7	88.6	86.1±19.4	1.1±0.3	117±63
T.truncatus	15	3535	71.7	78.3	81.2±21.2	1.2±0.5	117±53
OVERALL	**40**	**9192**	**71.8**	**85.0**	**80.3±22.0**	**1.2±0.4**	**104±54**

Figure Captions

Figure 1. Detected whistles with the GM-PHD filter. Spectrogram of raw data is shown (top), peak frequencies measurements (peaks 8 dB above background noise) (middle) and tracked whistles (bottom) where GM-PHD filter detections are shown.

Figure 2. Detection of crossing whistles with the GM-PHD filter. Spectrogram of raw data (left) and tracked whistles with GM-PHD filter (right) are shown.