Passive acoustic detection and estimation of the number of sources using compact arrays

Ildar R. Urazghildiiev[1,a)] and David E. Hannay[2]
[1]*JASCO Applied Sciences (USA) Incorporated, Silver Spring, Maryland 20910, USA*
[2]*JASCO Applied Sciences (Canada) Limited, Victoria, Canada*

(Received 25 October 2017; revised 19 April 2018; accepted 20 April 2018; published online 11 May 2018)

The problem of estimating the number of sound-producing sources detected using a compact array of hydrophones is addressed. Closed form expressions representing the techniques of automatic detection and estimation of the number of callers are given. Their performance is evaluated on a year-long dataset (1 October 2015–6 October 2016) containing humpback whale and killer whale calls collected in the Strait of Georgia, near Vancouver, British Columbia. Manual verification of the automatic detections produced by the approach required ∼40 h.
 https://doi.org/10.1121/1.5037361

[WWA] Pages: 2825–2833

I. INTRODUCTION

Detecting vocalizing marine mammals and estimating their locations and numbers is a requirement of new noise mitigation strategies aimed at reducing interactions of animals with human activities such as shipping. Visual surveys, traditionally used to detect and count animals,[1] require extensive manual effort, have limited effectiveness in poor visibility conditions or when animals do not surface frequently, and are prone to human error.[2] Passive acoustic methods can monitor continuously and are not limited by visibility conditions, but they are only able to detect sound-producing animals. Methods for counting the numbers of detected vocalizing animals inherently lend themselves to improved acoustic animal density estimation, although here we limit our focus to the counting technique itself.

Most passive acoustic monitoring (PAM) studies use isolated omni-directional sensors to detect vocalizing animals (see Refs. 3–9, and references therein). The sound detection algorithms applicable for isolated omni-directional sensors require high signal-to-noise ratio (SNR) to provide a high probability of true positives and a low probability of false alarms. The SNR, in turn, depends on the distance from the recorder to a vocalizing animal and on the local ambient noise characteristics. Increasing the detection range by decreasing the detection threshold leads to increased probability of false alarms. Rejecting false alarms and verifying the automatic detection results is labor-intensive and significantly increases data processing time and cost.

Isolated omni-directional sensors provide information about the arrival times of the detected sounds, but they produce little information about the relative directions of the callers. This information is insufficient for directly counting the number of detected animals. The existing "cue-based",[5–8] and other indirect methods of estimating the number of sources are based on the ratio of all detected sounds to the average number of sounds (acoustic cues) produced per unit time by an individual of the species of interest. The average number can differ significantly from the transmitted number of sounds, which changes with time, season, habitat, and other factors. The number of detectable sounds also changes as a source travels within the detection area because the probability of detecting a signal depends on the sensor-to-source distance and the ambient noise conditions. For animals that vocalize frequently (such as those producing echolocation clicks), the manual evaluation[6] of counting detections is labor-intensive.[5]

Omni-directional sensors can be used more effectively if they are synchronized and integrated into arrays because this enables the estimation of a source's position. For example, Ref. 10 considers counting vocalizing North Atlantic right whales using tracks of a whale's positions. Despite the efficiency of this technique, it is difficult to use large arrays of omni-directional sensors in applications that require near real-time performance. The other important disadvantage of large arrays is the practical difficulty of synchronizing omni-directional sensors over long deployment periods, although highly precise clocks are increasingly being used to overcome that problem. In practice, a source with a known position transmitting sounds over the entire array deployment period is also often used to synchronize sensors.[11]

Fixed compact sensor arrays with all hydrophones sampled by a synchronized multi-channel recorder can provide information about the bearings of detected sounds.[12–19] They can be cabled to a data processing station to provide near real-time capability or they can work autonomously.[18] These important advantages of compact arrays are being used in a new generation of accurate and cost-effective systems for various long-term PAM projects.

This study focuses on the problem of using fixed compact arrays to jointly detect and estimate the number of vocalizing animals within a continuously monitored area. The goal of this work is to develop a new technique based on using the bearings of the detected sounds to count the number of animals that produced them. The proposed technique provides short-time estimates of the number of

[a)]Electronic mail: ildar.urazghildiiev@jasco.com

vocalizing animals but it does not directly estimate abundance or density. The ability to count vocalizing animals near an array should be highly applicable to density estimation problems, but we do not develop that specific use here. We demonstrate that bearing estimates can be used to increase the efficiency-to-cost ratio of a PAM system by decreasing the human hours required to analyze large amounts of data and increasing the information about the number of sources that can be extracted from the data. Section II formulates the problem of joint detection and estimation of the number of vocalizing animals in a standard framework of hypotheses testing. Section III presents the closed-form representations of the hypotheses testing algorithms. A practical technique suitable for long-term PAM applications is described in Sec. IV. The results of processing a year-long data set collected in the Strait of Georgia, near Vancouver, British Columbia, are used to demonstrate the performance of the proposed technique in Sec. V. The discussion is presented in Sec. VI. Section VII provides conclusions.

II. DATA MODEL AND PROBLEM FORMULATION

Let us assume that a fixed compact array continuously monitors some area over a long observation interval, T_{obs}. The value of T_{obs} considered in this work is from weeks to a year or longer. It is much longer than the time period any animal would remain in the area. On a certain short time interval, $U(t_0) = [t_0,\, t_0 + T]$ starting at time t_0 and having the duration $T \ll T_{\text{obs}}$ (an interval of $T = 60\,\text{s}$ is used in this work), the automatic detector extracts $N \geq 0$ sounds arriving at the recorder input at times $t_1, t_2, \ldots, t_N \in U(t_0)$. The extracted sounds can include noise transients (the detections of noise transients are referred to as false alarms) and signals produced by $M > 0$ vocalizing animals (vocalizing animals are referred to as point sources). The main problem considered in this work is to find an estimator of the number of sources, M. As known from the statistical detection and estimation theory,[20] an optimal estimator is reduced to maximization of a likelihood function over the unknown parameter, M. Unfortunately, designing a statistically optimal estimator is difficult because a suitable likelihood function does not exist. Apart from that, implementing any kind of estimator makes sense only for the time intervals containing sounds from the sources. Therefore, we propose a solution based on hypotheses testing and introduce the following hypotheses:

$$H_0 : N = N_0; \quad H_1 : N = N_0 + P, \tag{1}$$

where $N_0 \geq 0$ is the number of noise transients, $P = \sum_{m=1}^{M} N_m$ is the total number of all detected signals transmitted by M sources, and $N_m > 0$ is the number of detected signals transmitted by the mth source. The hypothesis H_0 assumes that all detected sounds are noise transients and the hypothesis H_1 states that at least one sound is from a source of interest.

Using a compact array, the bearings of all detected sounds, $\alpha_{mn} = \alpha(t_{mn}),\ n = 1, \ldots, N_m,\ m = 0, \ldots, M$, are estimated.[19] Here $t_{mn} \in U(t_0)$ is the time of arrival of the nth sound produced by the mth source and the index $m = 0$ corresponds to noise.

We assume that bearings of noise transients, α_{0n}, are independent identically distributed (iid) random variables having uniform probability distribution within the interval $[0,\, 2\pi]$. The probability density function (PDF) of bearings of noise transients is a product of one-dimensional PDFs

$$w(\alpha_{01}, \ldots, \alpha_{0N_0}) = \prod_{n=1}^{N_0} w(\alpha_{0n}) = w_0, \tag{2}$$

where

$$w(\alpha_{0n}) = \frac{1}{2\pi}. \tag{3}$$

Substituting Eq. (3) in Eq. (2) gives

$$w_0 = \frac{1}{(2\pi)^{N_0}}. \tag{4}$$

The bearing estimates of the signals can be represented as

$$\alpha_{mn} = \mu_{mn} + \varepsilon_{mn}, \quad n = 1, \ldots, N_m, \quad m = 1, \ldots, M, \tag{5}$$

where $\mu_{mn} = \mu(t_{mn})$ is the true bearing of a signal produced by the mth point source at time instance t_{mn}; and $\varepsilon_{mn} = \varepsilon(t_{mn})$ is the bearing estimation error. We assume that for all sources and for all time instances, bearing estimation errors are iid Gaussian variables with zero mean and variance σ_1^2. Then the PDF of signal bearings (5) is also is a product of one-dimensional PDFs,

$$w(\alpha_{m1}, \ldots, \alpha_{mN_m}) = \prod_{n=1}^{N_m} w(\alpha_{mn}) = w_m, \tag{6}$$

where

$$w(\alpha_{mn}) = \frac{1}{(2\pi)^{1/2}\sigma_1} \exp\left\{-\frac{1}{2\sigma_1^2}(\alpha_{mn} - \mu_{mn})^2\right\}. \tag{7}$$

Substituting Eq. (7) in Eq. (6) gives

$$w_m = \frac{1}{(2\pi)^{N_m/2}\sigma_1^{N_m}} \exp\left\{-\frac{1}{2\sigma_1^2}\sum_{n=1}^{N_m}(\alpha_{mn} - \mu_{mn})^2\right\}. \tag{8}$$

A complete statistical representation of the hypotheses (1) can be given using PDFs (4) and (8),

$$\begin{aligned} H_0 &: w(\boldsymbol{\alpha}|H_0) = w_0, \\ H_1 &: w(\boldsymbol{\alpha}|H_1) = w_0 \prod_{m=1}^{M} w_m, \end{aligned} \tag{9}$$

where $\boldsymbol{\alpha} \in R^N$ is the N-dimensional vector of bearing estimates of all detected sounds.

The problem considered in this work can be formulated as follows: (i) to develop an algorithm for testing the

hypotheses (9) and for estimating the number of sources, M, if the hypothesis H_1 is accepted; (ii) to design a practical technique for jointly detecting and estimating the number of sources; and (iii) to test the performance of the technique using a long-term data set.

III. TESTING THE HYPOTHESES

The hypotheses (9) is tested with a commonly used statistical approach based on the likelihood ratio test.[20–22] The generalized likelihood ratio[22] is

$$\zeta(t_0) = \frac{\max w(\boldsymbol{\alpha}|H_1)}{w_0}. \tag{10}$$

Taking into account Eq. (8), the PDF $w(\boldsymbol{\alpha}|H_1)$ can be represented as

$$w(\boldsymbol{\alpha}|H_1) = \frac{1}{(2\pi)^{P/2}\sigma_1^P}\exp\left\{-\frac{1}{2\sigma_1^2}\sum_{m=1}^{M}\sum_{n=1}^{N_m}(\alpha_{mn}-\mu_{mn})^2\right\}. \tag{11}$$

The PDF (11) contains unknown parameters M, N_m, and μ_{mn}. Maximizing Eq. (10) over all unknown parameters is computationally expensive, such that it would be difficult to test the hypotheses (9) by implementing a likelihood ratio-based approach. Therefore, we test the hypothesis H_0 only. If the hypothesis H_0 is accepted, then the signal number estimate is $\hat{M} = 0$ and no other action is needed. Rejecting the hypothesis H_0 leads to accepting H_1 and estimating the number of sources.

To test the hypothesis H_0, we use a statistically optimal approach based on the Kolmogorov-Smirnov test.[23,24] The decision to accept the hypothesis H_0 is made by comparing the statistic

$$\hat{W} = \max_{\alpha} \hat{w}(\alpha, t_0), \tag{12}$$

with a certain threshold, c. Here $\hat{w}(\alpha, t_0)$ is a one-dimensional empirical PDF of all N bearings, α_{mn}, $n = 1, \ldots, N_m$, $m = 0, \ldots, M$, measured over the time interval $U(t_0)$. The threshold was selected using the Neyman-Pearson approach minimizing a false alarm probability for a given detection probability, such that the source detector can be represented as

$$\begin{cases} \hat{W} < c \Rightarrow H_0 \text{ is accepted}, \\ \hat{W} \geq c \Rightarrow H_1 \text{ is accepted}. \end{cases} \tag{13}$$

For the empirical PDF $\hat{w}(\alpha, t_0)$, we use a histogram representing one-dimensional empirical bearing distribution such that $\hat{w}(\alpha, t_0)$ is a function defined as the number of all bearing estimates observed over the interval $U(t_0)$ that fall into the bearing bin $[\alpha_k - \delta/2, \alpha_k + \delta/2]$, $k = 1, \ldots, K$, where δ is the bin width and K is the number of bins. (This work used non-overlapped bearing bins with $\delta = 2^\circ$ wide.)

If H_1 is accepted, then the short-time estimate of the number of sources, $\hat{M}$, are found. The proposed estimator is based on the assumption that over the time interval $U(t_0)$, sources travel with relatively small angular velocities such that the bearings of signals produced by the mth source are clustered around some area, $\alpha_{mn} \in \Delta_m = [\alpha_{m1}, \alpha_{mN_m}]$, centered near the average bearing $\bar{\alpha}_m(t_0) = (N_m)^{-1}\sum_{n=1}^{N_m}\alpha_{mn}$. If the bin width, δ, is comparable with Δ_m, the one-dimensional empirical PDF $\hat{w}(\alpha, t_0)$ can have up to M peaks that exceed the threshold c.

After applying a threshold, an empirical bearing distribution is represented as

$$\tilde{w}(\alpha, t) = \begin{cases} \hat{w}(\alpha, t), & \text{if } \hat{w}(\alpha, t) \geq c, \\ 0, & \text{otherwise}. \end{cases} \tag{14}$$

Let $u(\alpha)$ be an indicator function that equals unity if $\tilde{w}(\alpha, t_0)$ has a local peak at α and equals zero otherwise. Then, the proposed short-time estimator of the number of sources produces the following estimates:

$$\hat{M}(t_0) = \begin{cases} 0, & \text{if } H_0 \text{ is accepted}, \\ \sum_{\alpha} u(\alpha), & \text{if } H_1 \text{ is accepted}. \end{cases} \tag{15}$$

The position of the mth peak specifies the estimate of the average bearing, $\bar{\alpha}_m(t_0)$, of the mth source

$$\hat{\bar{\alpha}}_m(t_0) = \arg\max_{\alpha} \tilde{w}_m(\alpha, t_0). \tag{16}$$

The algorithm for combining the short-time estimates (16) when processing long-term data recordings is considered in Sec. IV.

IV. PROCESSING LONG-TERM DATA RECORDINGS

A practical technique suitable for processing long-term data recordings should include the steps for

(1) detecting impulsive signals;
(2) estimating the bearings of detected signals;
(3) detecting sources [Eq. (13)];
(4) estimating the number of sources [Eq. (15)];
(5) combining the bearing estimates [Eq. (16)] obtained for short-time intervals $U(t_j)$ for which the hypothesis H_1 is accepted.

Various algorithms can be used at different steps. This section presents an example of the algorithms used by the authors to demonstrate the method. The parameters of the algorithms have not been optimized using statistical procedures; they were selected based on prior information about the signals of interest observed manually in the data.

A. Detecting signals

Signal detection implies using a standard procedure to extract impulsive sounds of interest from continuous background noise. Marine mammals produce a variety of sounds, but arguably the most distinctive sounds are clicks and tonals, where tonals refer to one or more narrowband sounds with short durations of say 0.1 s to a few seconds. There are many detection techniques in the literature that would

potentially detect clicks and tonals.[3,4,9,25–27] Most known detectors are species-unique in the sense that signal parameters typical for a certain species are used when designing the corresponding detection and classification algorithms. Species-unique detectors are generally required when a PAM system must detect sounds from a variety of species. We applied a generic detection algorithm, described below, that was applicable to sounds produced by killer whales and humpback whales. We assumed that all sounds to be detected would have sufficient SNR to allow their visual detection in either their time-series pressure waveforms or spectrograms.

Clicks can be detected in the time domain using signal waveforms. The proposed click detector runs independently, using acoustic pressure data recordings, $x(t)$, collected over short time intervals $U(t_j) = [t_j,\ t_j + T]$, $j = 0, 1, \ldots$ (a duration $T = 60$ s was used in this work). The click detector takes the following steps:

(1) Compute bandpass filtered data, $x_{BF}(t)$, from the waveforms $x(t)$.
(2) Compute normalized filtered data, $\tilde{x}_{BF}(t) = x_{BF}(t)/\sigma_{\text{noise}}$, where σ^2_{noise} is ambient noise variance [the following robust estimate of variance[29] was used in this work: $\sigma^2_{\text{noise}} = \text{median}\{|x_{BF}(t)|^2\}$].
(3) Select a false alarm threshold, c_1, based on the specific needs (the value of $c_1 = 10$ was used in this work).
(4) Pick the samples of $|\tilde{x}_{BF}(t)|$ that exceed the false alarm threshold. Samples with time differences less than Δ_C ($\Delta_C = 20/F_s$ in this work, where F_s is the sampling frequency), are combined in a cluster. All clusters created in this way are defined as click candidates.
(5) Test the duration, peak frequency, and inter-click intervals (more parameters can be tested if needed) of click candidates and reject those that do not satisfy the predefined conditions.
(6) Evaluate the number of click candidates per unit of time and reject all candidates if their number is less than $N_{\min}$ per the interval $U(t_j)$ ($N_{\min} = 10$ in this work).

The click candidates passing these tests are considered as the detected clicks for which bearings were evaluated in step (2).

To detect tonals, a spectrogram-based approach is proposed. The tonal detector runs over the same short time intervals $U(t_j)$. The tonal detector takes the following steps:[9]

(1) Compute the data spectrogram, $X(f,t)$, using short-time Fourier transform. (The spectrogram is scaled in units of power spectrum density in this work.)
(2) Compute normalized spectrogram $X'(f,t) = X(f,t)/\sigma^2_{\text{noise}}(f)$, where $\sigma^2_{\text{noise}}(f) = \text{median}_t\{X(f,t)\}$ is a robust estimate of ambient noise power spectrum density at frequency f.
(3) Select a false alarm threshold, c_2, based on the specific needs (the value of $c_2 = 4$ was used in this work).
(4) Create a binary image, $B(f,t)$, from the normalized spectrogram $X'(f,t)$. The pixels of $B(f,t)$ equals unity if $X'(f,t) > c_2$ and equals zero otherwise.
(5) Create objects as sets of connected pixels with unity value, and apply morphological operations to remove isolated and spurious pixels and fill interior pixels [function *bwmorph* implemented in the Image Processing Toolbox of MATLAB R2017b (The MathWorks Inc., Natick, MA) were used in this work].
(6) Reject objects having the number of pixels, duration, frequency band, and other parameters[9] that do not satisfy predefined conditions.

The objects passing these tests are considered detected tonals for which bearings are evaluated in step (2). The start time of the kth object, t_k, is used as the time of arrival of the kth detected signal.

B. Estimating the bearings of detected signals

Azimuths of the detected clicks and tonals can be measured using an array consisting of L hydrophones and the maximum likelihood (ML) time-difference-of-arrival (TDOA)-based algorithm[19]

$$\hat{\alpha}(t_k) = \arg\min_{\alpha} \sum_i \sum_j \left(\hat{\tau}_{i,j}(t_k) - \tau_{i,j}(\alpha, \beta_0, d_0)\right)^2. \quad (17)$$

Here, $\hat{\tau}_{i,j}(t_k) = \text{argmax}_\tau \rho_{i,j}(\tau)$ is the TDOA estimate between the ith and jth recorder of the array ($i = 1, \ldots, L-1$, $j = 2, \ldots, L$) computed for the time t_k; $\rho_{i,j}(\tau)$ is the correlation function of signals in the ith and jth recorder; $\tau_{i,j}(\alpha, \beta_0, d_0)$ is the expected TDOA computed for the azimuths $\alpha = 0, \ldots, 360°$, elevation angle β_0, and range d_0. The correlation function can be computed in frequency domain using inverse Fourier transform of signal spectrums, $\rho_{i,j}(\tau) = \text{IFFT}\{S_i(f,t_k)S_j^*(f,t_k)\}$, where $S_i(f,t_k) = \text{FFT}\{x_i(t_k)\} = \sum_t^{t+T} x_i(t_k + t)e^{-j2\pi ft}$ is the spectrum of the signal in the ith recorder; and symbol "*" denotes complex conjugation.

C. Detecting and estimating the number of sources

The straightforward extension of the short-time source detector (13) and the estimator (15) to process long-term data recordings is similar to that used to compute a short-time Fourier transform (also known as a spectrogram). It is based on using empirical short-time bearing distribution (STBD). We define STBD as a two-dimensional (2D) matrix of bearing estimates and detection time intervals: $\hat{w}(\alpha,t)$, $\alpha \in [\alpha_1, \ldots, \alpha_K]$, $t \in [t_0, t_1, \ldots]$. The jth column of $\hat{w}(\alpha,t)$ is composed of the bearing histograms computed for time interval $U(t_j) = [t_j,\ t_j + T]$ starting at time t_j and having the duration T. The time intervals can overlap.

The proposed source detection algorithm consists of three steps. In the first stage, the empirical STBD is computed for the entire observation interval T_{obs} and is compared with a false alarm threshold, c [Eq. (13)]. After applying the threshold, the resulting STBD is $\tilde{w}(\alpha,t)$ [Eq. (14)].

For each time interval $U(t_j)$ [or for each column of $\tilde{w}(\alpha,t_j)$], the detector [Eq. (13)] is applied in the second stage. If H_1 is accepted, the average bearing estimates, $\hat{\bar{\alpha}}_m(t_j)$ [Eq. (16)], are found for this interval.

In the third stage, clusters of the average bearing estimates [Eq. (16)] are created. The clustering algorithm

combines two short-time average bearings $\hat{\bar{\alpha}}_k(t_i) \neq \hat{\bar{\alpha}}_m(t_j)$ into the same cluster if the distance between them satisfies the condition

$$\frac{\left(\hat{\bar{\alpha}}_k - \hat{\bar{\alpha}}_m\right)^2}{2\sigma_\alpha^2} + \frac{(t_i - t_j)^2}{2\sigma_t^2} \leq 1. \tag{18}$$

Note that Eq. (18) represents an ellipse in the azimuth and time plane with the center at $(\hat{\bar{\alpha}}_m, t_j)$, and the parameters σ_α and σ_t defining the lengths of the axes. In this work, these parameters were taken empirically based on the bearing estimation accuracy provided by the array and the maximum time between signal arrivals (we used $\sigma_\alpha = 5^\circ$ and $\sigma_t = 75$ s for clicks and high-frequency tonals, and $\sigma_\alpha = 15^\circ$ and $\sigma_t = 600$ s for low-frequency tonals). The pth cluster is defined as the set of bearings that satisfy the ellipse equation (18),

$$U_p = \{\hat{\bar{\alpha}}_k, k = k_1, k_1 + 1, \ldots, k_1 + K_p - 1\}, \quad p = 1, \ldots, P, \tag{19}$$

where K_p is the number of average bearings composing the pth cluster and P is the total number of clusters created over the entire observation interval T_{obs}. Here, we number all short-time average bearing estimates $\hat{\bar{\alpha}}_m(t_j)$ [Eq. (16)] in ascending order.

In a final stage, the duration of all clusters [Eq. (19)] are tested. The ones with a duration of T_C ($T_C = 90$ s in this work) or longer are accepted as the clusters associated with the sources. Let the indicator function v_p specify the accepted cluster, i.e., $v_p = 1$ if the cluster U_p is accepted and $v_p = 0$ otherwise. Then, the estimate of the number of sources detected over the entire observation interval T_{obs} is

$$\hat{M} = \sum_{p=1}^{P} v_p. \tag{20}$$

Note that a source can be undetectable for a certain time (either because it stops vocalizing or because ambient or anthropogenic noise masks its sounds), such that more than one cluster [Eq. (19)] can be associated with the source. Therefore, the estimate [Eq. (20)] can be higher than the actual number of vocalizing animals, so it should not be interpreted as an abundance estimate. Apart from that, point sources producing sounds mistakenly recognized as signals (such as ships) may contribute to the estimate [Eq. (20)].

V. TEST RESULTS

In this study, acoustic recordings were collected on a tetrahedral compact array of four hydrophones spaced at 1.85 m and sampled by an Autonomous Multichannel Acoustic Recorder[18,19] (AMAR, JASCO Applied Sciences). The array and AMAR were deployed on the VENUS cabled ocean observatory operated by Ocean Networks Canada in the Strait of Georgia, near Vancouver, BC, Canada. The AMAR had four GeoSpectrum M-36 hydrophones (GeoSpectrum Technologies Inc., Dartmouth, NS, Canada) sampled at 64 kHz using the AMAR's four channel analog-to-digital converter to ensure simultaneous sampling. Combined hydrophone sensitivity and current-to-voltage converter board sensitivity were −165 dB re 1 V/μPa at 250 Hz. The array was deployed on the ocean bottom at a depth of 168 m. The array position and orientation were measured during the deployment and calibrated using surface vessels with known GPS coordinates. Data collected from 1 October 2015 to 6 October 2016 were used for *in situ* testing.

Tonals and clicks were detected using the algorithms presented in Sec. IV A. The ML TDOA-based algorithm (17) was used to estimate bearings of the detected sounds. The algorithm described in Sec. IV C was used to detect and estimate the number of sources.

Because of the intense ship traffic in the sensor deployment area, vessel bearings were also measured. TDOAs of ship noise were computed within the frequency range from 50 to 25 000 Hz for consecutive data segments, each 1 s long and overlapped by 0.5 s. Vessel bearings were estimated using the ML algorithm [Eq. (17)].[19]

Sounds produced by killer whales and humpback whales were detected in the data set. The killer whale sounds included clicks and pulsed calls, call types that are well described in the literature.[26,27] The humpback whale sounds were low-frequency, single song units that are typically produced by singing males during winter.[28] We refer to pulsed calls produced by killer whales as high-frequency tonals and song units produced by humpback whales as low-frequency tonals.

Figure 1(a) shows a spectrogram of automatically detected clicks and high-frequency killer whale tonals in the presence of ship noise. Figure 1(b) presents the corresponding bearing estimates. Figure 1 illustrates how compact arrays make it possible to separate ship noise from impulsive sounds produced by animals in the bearing domain.

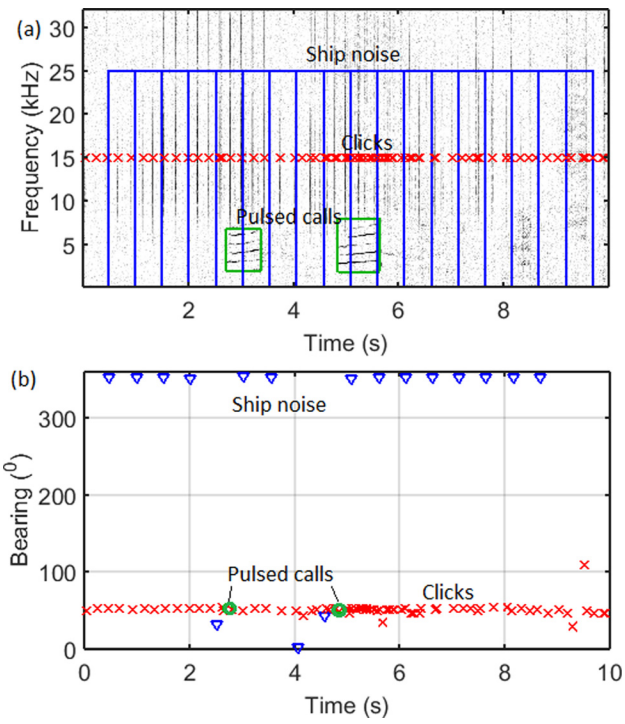

FIG. 1. (Color online) (a) Data spectrogram displaying detected killer whale clicks (red ×'s), pulsed calls (green boxes), and ship noise segments (blue boxes) used to compute bearings. (b) Bearing estimates of the killer whale pulsed calls (green circles), clicks (red ×'s), and ship noise (blue triangles).

Figures 2(a) and 3(a) display bearing estimates of ship noise and detected impulsive sounds (clicks, plus high- and low-frequency tonals). The detected impulsive sounds include an unknown number of true positives and false alarms. Figure 2(b) shows the STBD [Eq. (14)] of clicks (computed using the threshold $c = 6$) and Fig. 3(b) shows the STBD of low-frequency tonals ($c = 3$). To compute STBD, sliding time frames $U(t_j) = [t_j, t_j + T]$, with the duration $T = 60$ s overlapped by $t_{j+1} - t_j = 30$ s, and nonoverlapped bearing bins $\delta = 2^\circ$ wide were used.

Figures 2(c) and 3(c) show the automatically created bearing clusters [Eq. (19)]. According to Eq. (20), the number of clusters is accepted as the number of detected sources. The automatic clustering algorithm (18) created $\hat{M} = 24$ clusters for the killer whale sounds [Fig. 2(c)] and $\hat{M} = 1$ cluster for humpback whale low-frequency sounds [Fig. 3(c)]. The actual number of vocalizing whales were evaluated by a human operator as the number of bearing tracks visually detected from Figs. 2(a) and 2(b) for killer whales and 3(a) and 3(b) for humpback whales. A total of $M = 13$ killer whales and $M = 1$ humpback whale was detected by the human operator.

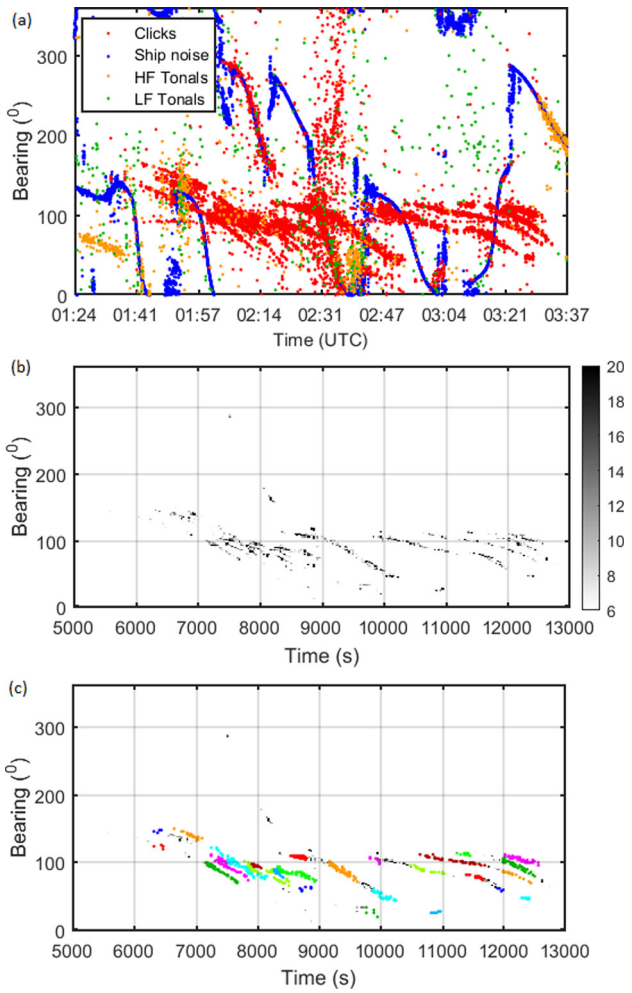

FIG. 2. (Color online) (a) Bearing estimates of ship noise and detected impulsive sounds. (b) The empirical STBD of killer whale sounds (clicks and high-frequency tonals). Gray color represents the number of bearing estimates in the time and bearing bin. (c) Automatically created clusters of killer whale bearings. Colors represent different clusters. Data collected on 4 October 2015.

The number of vocalizing animals detected visually using plots of bearing estimates [Figs. 2(a) and 3(a)] by the human operator was used as ground truth data to evaluate the performance of the proposed techniques.

Like any detector, the source detector considered in this work can potentially produce two types of errors: false alarms and misses. We refer to a false alarm as an automatically created cluster of bearings [Eq. (19)] that does not correspond to a biological source. In our tests, the vast majority of false alarms were produced by ship noise. A miss occurs when no clusters [Eq. (19)] correspond to a bearing track visually detected by the human operator.

To manually detect and classify animals, software for processing and visualizing bioacoustical information[30] was used. The human operator checked automatically detected bearing clusters [such as those shown in Figs. 2(c) and 3(c)], marked false and true clusters, respectively, counted the number of animals as the number of bearing tracks visible on the STBD image [Figs. 2(b) and 3(b)], and classified the detected animals using the data spectrogram [Fig. 1(a)].

Figure 4 shows an example of 230 279 bearing estimates (17) of ship noise, clicks, high-, and low-frequency sounds

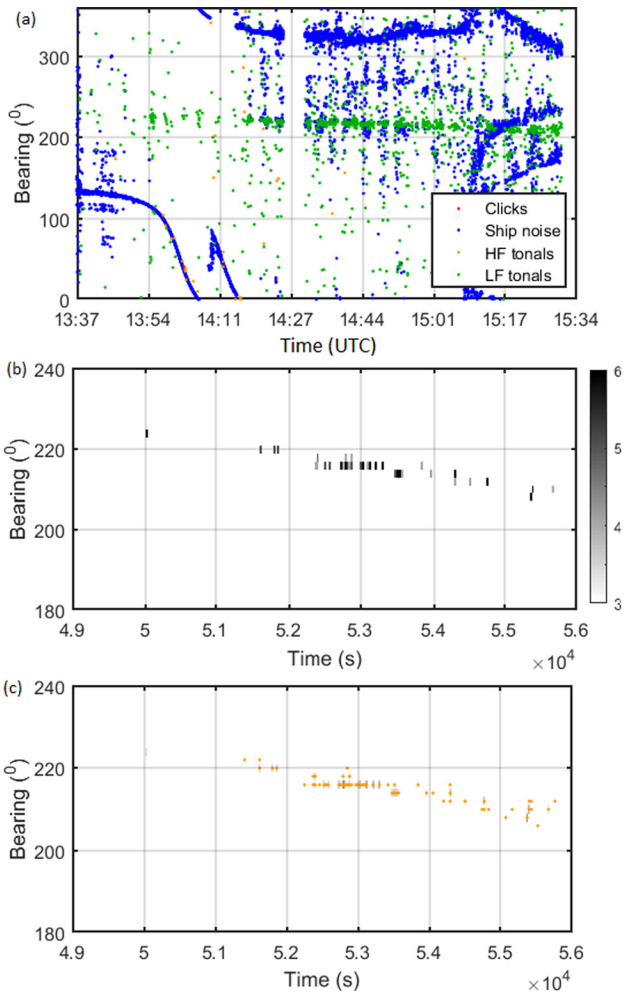

FIG. 3. (Color online) (a) Bearing estimates of ship noise and detected impulsive sounds. (b) The empirical STBD of humpback whale low-frequency sounds. Gray color represents the number of bearing estimates in the time and bearing bin. (c) Automatically created cluster of humpback whale bearings. Data collected on 25 December 2015.

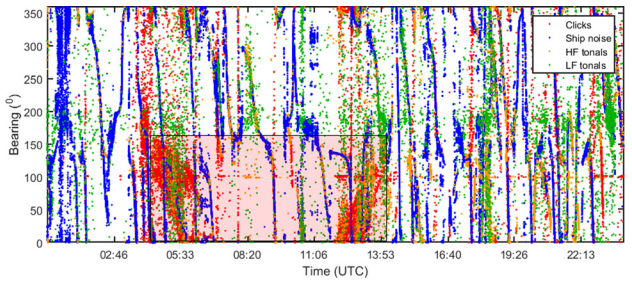

FIG. 4. (Color online) (a) Bearing estimates of ship noise and detected impulsive sounds. The interval of times and bearings corresponding to a group of 19 killer whales manually detected by the human operator is shown by the semitransparent red box. Data collected on 22 September 2016.

computed from 24 h of data collected on 22 September 2016. From these data, the human operator manually detected that a group of 19 killer whales stayed within the detection range of the compact array from 4:26 AM to 6:18 AM and from 12:05 PM to 2:07 PM. In Fig. 4, 23 816 bearings of the detected killer whale clicks are marked by a semitransparent red box.

To find missed sources, the human operator visually inspected 10 min in each 1 h of data recordings using the plots of data spectrograms [Fig. 1(a)] and bearing estimates (17) (Fig. 4). No missed sources were found.

Figure 5(a) shows the number of false alarms [i.e., the number of false clusters of bearings, Eq. (19)], M_{fa}, per one month of observation. The number of manually detected sources, M_{man}, are shown in Fig. 5(b).

The human hours needed to manually process the data using the processing and visualizing software[30] were also calculated. The average time needed to check clusters created over 24 h of data recordings was ~5 min, and the total time needed to process 1 yr of data recordings was ~40 h because some days (like that shown in Fig. 4) took much longer than the average processing time.

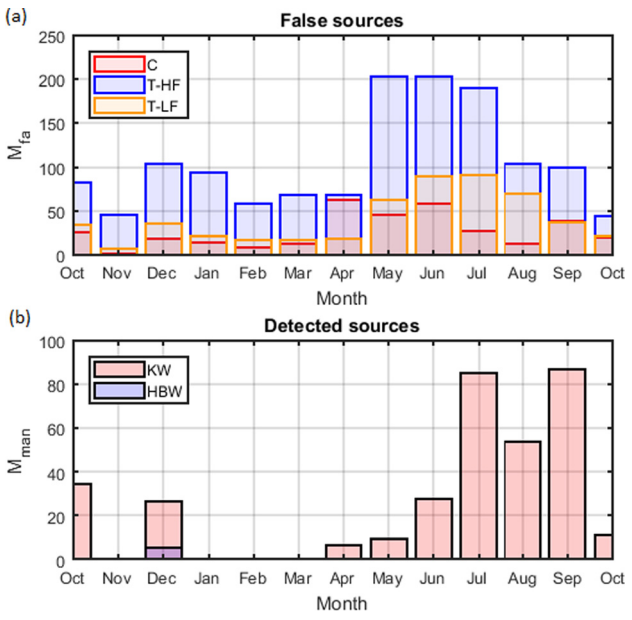

FIG. 5. (Color online) (a) The number of false alarms. (b) The number of manually detected sources.

VI. DISCUSSION

Test results demonstrated that the automated algorithms proposed here for detecting clicks and tonals were effective. Bearing estimates (17) produced by the algorithms made it possible to separate detected sounds in the bearing domain, making it possible to automatically estimate the number of detected animals.

The algorithm was more effective for localizing killer whale clicks than tonal calls. Clicks are produced at a much higher rate than pulsed calls and whistles, and their wider bandwidth leads to better TDOA-based bearing estimation accuracy. Humpback whale tonal calls have a smaller frequency bandwidth, thereby resulting in less accurate bearing estimates. However, this was still sufficient for clustering calls and estimating the number of detected whales (Fig. 3). Thus, test results demonstrated that the proposed bearing-based technique can be used to detect and localize the calls from a variety of marine animals, including both baleen and toothed whales, which produce different types of sounds.

The performance of the proposed technique was evaluated in terms of probabilities of misses and false alarms. An important result of this study is that the human operator did not find any sources that were missed by the automatic detector. All marine animals that produced sounds visible on a spectrogram were detected automatically.

The main cause of false alarms was ship noise generated by nearby vessels. It can be explained by the fact that an assumption [Eq. (2)] that bearings of noise transients have uniform distribution does not hold if false alarms are generated by a point source such as a ship or an individual of a different species than the one we are looking for. Designing and using more sophisticated data processing algorithms could potentially decrease the false alarm rate. However, an efficient way of rejecting false alarms is to have a human operator check the bearing clusters [Eq. (19)] created automatically and manually count the number of sources using images of bearing estimates [Figs. 2(a), 3(a), and 4] or STBD [Figs. 2(b) and 3(b)]. Figure 5(a) shows that even in an area of intense ship traffic, the number of false alarms was less than three hundred per month of data, which is a reasonably small number to check manually. Using the proposed techniques in combination with appropriate software (such as the one used in this study[30]) makes it possible to dramatically decrease human hours needed to process long-term data recordings. The human operator took ~40 h to process one year-long data set used in this study.

The performance of the proposed technique depends on the angular distance between the sources. If the angular distance is smaller than the angular resolution provided by the compact array, the signals produced by the sources become undistinguishable in the bearing domain. However, this situation occurs for relatively short time intervals. As Figs. 2 and 4 show, the angular distance between killer whales traveling in a group changed over time, such that automatically detecting and visual counting killer whales was possible over the time intervals when their bearings were distinguishable.

Another factor affecting the accuracy of automatic estimation of the number of sources is the presence of time

intervals when no signals from an animal are detected. This occurs when a source either does not produce sounds or the sounds are masked by noise. Large time gaps between detected sounds result in multiple detections of the same source and in an overestimation of the number of vocalizing animals. In other words, the number of detected clusters [Eq. (19)] specifying the source number estimate [Eq. (20)] becomes higher than the actual number of vocalizing animals, $\hat{M} > M$. Figure 2(c) shows a situation where the bearing tracks of killer whales were automatically detected more than once. In this situation, the time gaps between bearing clusters were relatively short, and the human operator could extrapolate visible bearing tracks to combine them and count the number of individual whales producing sounds with sufficiently high confidence. A more complicated situation is shown in Fig. 4. Specific positions of bearing tracks before and after the 5.5 h time gap indicated that the same group of 19 killer whales probably entered the detection area at 4:26 AM and 12:05 PM. However, the confidence in that decision is lower than that in a situation shown in Fig. 2. In general, as the time gap between visible bearing tracks increases, the confidence of associating the tracks to the same individual decreases.

Thus, test results demonstrated that the proposed technique is very efficient for automatically detecting the presence of animals, but the more complicated problem of estimating abundance and density would require a larger number of compact arrays covering a larger area. The proposed bearing-based approach implemented in a network of compact arrays providing required coverage can be used to design efficient abundance and density estimation techniques satisfying many practical needs.

VII. CONCLUSIONS

A compact array of hydrophones can provide important information about the bearings of the detected sounds. This capability makes it possible to design a new generation of accurate and cost-effective systems applicable to various long-term PAM projects.

This study provides closed-form representations of the statistically optimal hypothesis testing algorithms for bearing measurements. Using these algorithms, a practical source detection technique suitable for processing long-term data recordings was developed. The proposed technique includes the algorithms for detecting clicks and tonal sounds, estimating the bearings of detected sounds, computing the STBD, and clustering bearings that exceed a false alarm threshold.

In situ tests demonstrated that the proposed technique provides very low probabilities of false alarms and missed sources relative to those detectable by a manual review of spectrograms. The main factors affecting the performance of the proposed technique are the angular distance between the sources and the presence of time intervals when sounds from the source are not produced or cannot be detected. A human operator using appropriate software can, to some extent, cope with these situations. Test results demonstrated that 40 human hours were sufficient to process one year-long dataset used in this study. However, applying the proposed technique in different conditions may require different processing times.

The results of this work demonstrated that the proposed technique can potentially increase the efficiency-to-cost ratio of a PAM system by decreasing the time needed for a person to analyze a large amount of acoustic data and increasing the amount and accuracy of the information extracted from acoustic recordings. It can be applied to solve more complicated problems of abundance and density estimation using a network of compact arrays covering large marine areas. Possible uses of the proposed technique for these and other applications should be a topic of future research.

ACKNOWLEDGMENTS

The authors thank the Port of Vancouver for its support of the data acquisition project through the Enhancing Cetacean Habitat and Observation program, Ocean Networks Canada for maintaining the VENUS undersea data network. We also thank K. Hiltz for editing a draft version of the manuscript and two anonymous reviewers for their comments that helped to improve the quality of this work.

[1]S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas, *Introduction to Distance Sampling—Estimating Abundance of Biological Populations* (Oxford University Press, Oxford, UK, 2001), pp. 77–78.
[2]G. A. Bortolotto, D. Danilewicz, A. Andriolo, and A. N. Zerbini, "Humpback whale *Megaptera novaeangliae* (Cetartiodactyla: Balaenopteridae) group sizes in line transect ship surveys: An evaluation of observer errors," Zoologia **33**(2), e20150133 (2016).
[3]W. M. X. Zimmer, *Passive Acoustic Monitoring of Cetaceans* (Cambridge University Press, Cambridge, UK).
[4]*Detection, Classification and Localization of Marine Mammals Using Passive Acoustics*, edited by O. Adam and F. Samaran (Dirac NGO, Paris, France, 2013).
[5]T. A. Marques, L. Thomas, S. W. Martin, D. K. Mellinger, J. A. Ward, D. J. Moretti, D. Harris, and P. L. Tyack, "Estimating animal population density using passive acoustics," Biol. Rev. **88**(2), 287–309 (2013).
[6]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 Blainville's beaked whales," J. Acoust. Soc. Am. **125**(4), 1982–1994 (2009).
[7]E. T. Küsel, D. K. Mellinger, L. Thomas, T. A. Marques, D. J. Moretti, and J. Ward, "Cetacean population density estimation from single fixed sensors using passive acoustics," J. Acoust. Soc. Am. **129**(6), 3610–3622 (2011).
[8]J. Ward, L. Thomas, S. Jarvis, N. DiMarzio, D. Moretti, T. A. Marques, C. Dunn, D. Claridge, E. Hartvig, and P. Tyack, "Passive acoustic density estimation of sperm whales in the tongue of the ocean, Bahamas," Mar. Mammal Sci. **28**(4), E444–E455 (2012).
[9]I. R. Urazghildiiev, C. W. Clark, T. P. Krein, and S. E. Parks, "Detection and recognition of North Atlantic right whale contact calls in the presence of ambient noise," IEEE J. Ocean. Eng. **34**(3), 358–368 (2009).
[10]I. R. Urazghildiiev, "Statistical analysis of North Atlantic right whale (*Eubalaena glacialis*) signal trains in Cape Cod Bay, spring 2012," J. Acoust. Soc. Am. **136**(5), 2851–2860 (2014).
[11]I. R. Urazghildiiev and C. W. Clark, "Comparative analysis of localization algorithms with application to passive acoustic monitoring," J. Acoust. Soc. Am. **134**(6), 4418–4426 (2013).
[12]R. Hirotsu, M. Yanagisawa, T. Ura, M. Sakata, H. Sugimatsu, J. Kojima, and R. Bahl, "Localization of sperm whales in a group using clicks received at two separated short baseline arrays," J. Acoust. Soc. Am. **127**(1), 133–147 (2010).
[13]S. M. Wiggins, M. A. McDonald, and J. A. Hildebrand, "Beaked whale and dolphin tracking using a multichannel autonomous acoustic recorder," J. Acoust. Soc. Am. **131**(1), 156–163 (2012).

[14]S. M. Wiggins, K. E. Frasier, E. E. Henderson, and J. A. Hildebrand, "Tracking dolphin whistles using an autonomous acoustic recorder array," J. Acoust. Soc. Am. **133**(6), 3813–3818 (2013).
[15]M. Gassmann, E. E. Henderson, S. M. Wiggins, M. Roch, and J. A. Hildebrand, "Offshore killer whale tracking using multiple hydrophone arrays," J. Acoust. Soc. Am. **134**(5), 3513–3521 (2013).
[16]M. Gassmann, S. M. Wiggins, and J. A. Hildebrand, "Three-dimensional tracking of Cuvier's beaked whales' echolocation sounds using nested hydrophone arrays," J. Acoust. Soc. Am. **138**(4), 2483–2494 (2015).
[17]D. Mathias, A. M. Thode, J. Straley, and R. D. Andrews, "Acoustic tracking of sperm whales in the Gulf of Alaska using a two-element vertical array and tags," J. Acoust. Soc. Am. **134**(3), 2446–2461 (2013).
[18]C. Hillis, X. Mouy, I. Urazghildiiev, and T. Dakin, "Autonomous multichannel acoustic recorders on the VENUS Ocean Observatory," in *Oceans'2014*, St. John's (2014).
[19]I. Urazghildiiev and D. Hannay, "Maximum likelihood estimators and Cramer–Rao bound for estimating azimuth and elevation angles using compact arrays," J. Acoust. Soc. Am. **141**(4), 2548–2555 (2017).
[20]H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part I* (Wiley, New York, 2001).
[21]E. L. Lehman, *Testing Statistical Hypotheses* (Wiley, New York, 1986).
[22]A. Hero, "Signal detection and classification," in *Digital Signal Processing Handbook*, edited by E. Madisetti and D. Williams (CRC Press, New York, 1999).
[23]A. N. Shiryayev, "On the empirical determination of a distribution law," in *Selected Works of A. N. Kolmogorov. Mathematics and Its Applications* (Soviet Series), edited by A. N. Shiryayev (Springer, Dordrecht, 1992), Vol. 26, Chap. 15.
[24]T. B. Arnold and J. W. Emerson, "Nonparametric goodness-of-fit tests for discrete null distributions," R J. **3**(2), 34–39 (2011).
[25]W. M. X. Zimmer, J. Harwood, P. L. Tyack, M. P. Johnson, and P. T. Madsen, "Passive acoustic detection of deep-diving beaked whales," J. Acoust. Soc. Am. **124**(5), 2823–2832 (2008).
[26]J. C. Brown, A. Hodgins-Davis, and P. J. O. Miller, "Classification of vocalizations of killer whales using dynamic time warping," J. Acoust. Soc. Am. **119**(3), EL34–EL40 (2006).
[27]A. D. Shapiro and C. Wang, "A versatile pitch tracking algorithm: From human speech to killer whale vocalizations," J. Acoust. Soc. Am. **126**(1), 451–459 (2009).
[28]K. Payne, P. Tyack, and R. Payne, "Progressive changes in the songs of humpback whales (*Megaptera novaeagliae*): A detailed analysis of two seasons in Hawaii," in *Communication and Behavior of Whales*, edited by R. Payne (Westview, Boulder, CO, 1983), pp. 9–57.
[29]R. A. Maronna, R. Douglas Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods* (Wiley, West Sussex, 2006).
[30]"System and methods for processing and the visualization of bioacoustical information," U.S. patent application 2016/0252635 (September 1, 2016).