Applied Acoustics 106 (2016) 36–41

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Improved passive acoustic band-limited energy detection for cetaceans

Joey D. Hood [a,*], David G. Flogeras [a], James A. Theriault [b]

[a] Akoostix Incorporated, Dartmouth, NS B3B 1J4, Canada
[b] Defence Research and Development Canada (DRDC) – Atlantic, Dartmouth, NS B2Y 3Z7, Canada

ARTICLE INFO

Article history:
Received 4 December 2014
Received in revised form 27 November 2015
Accepted 2 December 2015
Available online 7 January 2016

Keywords:
Marine animals
Signal detection
Monitoring
False alarm reduction
PAMGUARD

ABSTRACT

Acoustic detection of cetaceans is challenged by the variability of calls and the presence of variable background noise. One detection method is to start with frequency band-limited signal and noise estimates, and apply a likelihood ratio test (LRT). These detectors suffer from false alarms when broadband signals overlap the band of interest, triggering detection. Some detectors only consider previous samples, causing further false alarms. The authors propose a method of reducing false alarms by defining a guard band that is not expected to contain energy from the species of interest. A second LRT is performed by testing the ratio of the signal estimate in the signal band with the signal estimate of the guard band. This method is shown to reduce false alarms with a small reduction in detection performance. A detection method is also presented that can be optimized for high processing efficiency, while improving false-alarm rejection from signals that are longer in duration than the signal of interest. Performance is demonstrated on real cetacean recordings and ocean noise. The detection algorithm is implemented in PAMGUARD, an open source Java application designed for passive acoustic monitoring (PAM) of cetaceans.

1. Introduction

Acoustic detection, localization, and classification of cetaceans are required for a number of applications, including anthropogenic-impact mitigation and habitat surveys. Both detection performance and false-alarm rejection are important for these two applications. Impact mitigation further requires that operators quickly make decisions with respect to contact. Population-survey data analysis takes place over a longer timeframe, but can include vast amounts of data with an increased emphasis on correct classification. A staged processing approach can be used with detection being the first stage. The goal is to have a high detection rate and moderate or low false-alarm rate to significantly reduce the volume of data. Subsequent stages can be used to further refine classification and filter false alarms.

The challenge is that marine-mammal vocalizations are highly variable with frequency content ranging from 10 Hz to, at least 150 kHz [1]. Durations may be as little as a few milliseconds for an echo location click, but may extend to tens of minutes for a complex humpback song. Vocalizations may be described by adjectives such as tonal, pulsive, song, shriek, FM sweep, and many others. Robust cetacean detection is a difficult problem and a number of techniques have been developed, each with their own merits. Generic algorithms that use matched filters to correlate time-series data or spectrogram correlation to detect patterns in spectrograms have been developed [2,3]. These techniques are useful when the cetacean vocalization is predictable but are limited by requiring detailed priori knowledge and stability of the vocalization [4,5]. Another technique based on band-limited energy detection is also common [6,7]. This technique is well suited to signals that are characteristic in duration and bandwidth, but too variable for correlation-based processing. One example of this class of call is produced by the North Atlantic right whale (*Eubalaena glacialis*), which has been shown to adapt its vocalization to the environment. One of its vocalization types is nominally represented by a frequency-modulated (FM) upsweep that is found in the 50–400 Hz band with duration of approximately 1 s. However, initial frequency, FM sweep rate, duration, and bandwidth will vary [5].

For energy detection, a detection function can be formed by estimating the signal over a short-term average and dividing it by a longer average background noise estimate [6,7]. This function takes the form of a likelihood ratio test (LRT) common to many detection strategies [8]. The primary difference between LRT based detectors is the statistic used to generate the LRT [5,8]. In the LRT case, the detection function is sensitive to abrupt changes in band-limited energy level, characteristic of a cetacean vocalization. In similar detectors, such as the sperm-whale click detector [6], the estimates have been computed using a single-pole IIR (infinite

* Corresponding author. Tel.: +1 902 404 7464; fax: +1 902 405 3855.
E-mail address: jhood@akoostix.com (J.D. Hood).

http://dx.doi.org/10.1016/j.apacoust.2015.12.011
0003-682X/

impulse response) filter, or exponential average. This approach prevents the detector from discriminating between a short and long duration signal, causing one form of false alarm. The detection function is also limited to band-energy data from the band of interest, preventing it from discriminating between band-limited and broadband signals. False alarms can be a frequent problem in this type of detector as non-stationary, non-Gaussian background noise from machinery, other animals, or the environment can cause detection [10].

The objective is to develop a flexible, generic detector that will provide adequate first-stage detection performance over a wide variety of species. False alarms from longer duration signals and wide-band noise could be filtered by subsequent processing stages as done in [9], but many can also be efficiently eliminated in the first detection stage, reducing complex processing on high volumes of data in noisy environments. This article proposes improvements to band-limited energy detection that significantly reduce false alarms. Detection performance and processing load are very similar to other energy detectors, whereas processing load is less than that of correlation-based detectors, especially as the number of species is increased. This allows the algorithm to be implemented in low-power processors, which are commonly used for embedded processing.

2. Signal processing

Data is processed using a consistent processing stream with the processing parameters tuned for each species. First, a fast Fourier transform (FFT) is performed with overlapping segments and the power spectral density (PSD) is estimated. A Hann window is used with overlap and optional spectral averaging. Fifty to seventy-five percent overlap with no averaging are typically used for processing. Specific examples of parameters are presented in Table 3. Next an arbitrary number of band-limited energy time series are computed by averaging PSD estimates between a minimum and maximum frequency chosen to match the vocalization of the species of interest. Guard bands are also defined where no target (signal) energy is expected for use later in the detector.

The band-limited energy data are used to compute a signal and noise estimate. Two estimation methods are provided. The exponential average has typically been used in [6], but an alternative and often more optimal, in terms of detection performance, split-window average is suggested here.

The exponential average estimate can be computed using

$$y[n] = \alpha y[n-1] + (1-\alpha)x[n] \tag{1}$$

where α is a weighting constant that must be less than 1.0. The variable x represents an input band-limited energy sample for sample n; y represents the estimate at samples n and $n-1$ respectively. An α value of 0.0 produces output samples y identical to the input samples x (i.e. no averaging). The closer α gets to the value 1.0, the more weight the previous estimate output samples are given and thus the longer the averaging. For the noise estimate, Eq. (1) can be further enhanced by allowing α to be changed based on the noise estimate's level relative to the signal estimate. This change is done to allow the noise estimate to quickly recover after a loud transient as further explained in Section 3. This decision takes the form

$$\alpha = \begin{cases} \alpha_S, & y_N[n-1] \leqslant \lambda \cdot y_S[n-1] \\ \alpha_N & \text{otherwise} \end{cases} \tag{2}$$

where subscripts S and N denote respective signal and noise estimates.

The split-window estimate uses two rectangular windows of different widths to estimate the signal and noise levels. The windows have an odd number of samples to allow precise definition of the window center. The noise window, W_N must be longer than the signal window, W_S which is chosen based on priori knowledge of the vocalization characteristics of interest as will be shown in Section 4.2. The signal estimate for each sample n is estimated using

$$y_S[n] = \frac{1}{W_s} \sum_{i=-(W_S-1)/2}^{(W_S-1)/2} x[n+i]. \tag{3}$$

The noise is estimated using

$$y_N[n] = \frac{1}{W_N - W_S} \left[\sum_{i=-(W_N-1)/2}^{(W_N-1)/2} x[n+i] - y_S[n] W_S \right]. \tag{4}$$

The split-window estimator is non-causal, that is, to calculate output sample $y[n]$ it must have access to input sample $x[n + (W_N - 1)/2]$. In a real-time processing system, this means that the output will lag the input by approximately half of the length of the noise window. Depending on the time resolution and the length of the estimator's windows, the lag could be several seconds long, resulting in a delay before detections are declared. The PAMGUARD [9] implementation of the detector provides both estimation options.

Implementation of the split-window estimator occurs in the time domain and uses a processing optimization. The optimization exploits the sliding window by simply adding the newest sample and removing the oldest, requiring a processor load that is very similar to that of the exponential average. This method requires management of a circular buffer containing all active samples, but saves considerable processor load over a brute-force implementation.

It is assumed that the signal is restricted to a given bandwidth. A two-stage LRT compares the signal detection using the restricted bandwidth with that for a broader bandwidth. By eliminating the cases where there is no difference in the performance, broadband transient signals do not trigger false detections. First the restricted band likelihood is computed using

$$L_t = \frac{y_S}{y_N} > \tau_t \tag{5}$$

where L_t represents the estimated likelihood ratio test for the restricted signal band and τ_t is the in-band threshold. With wider guard bands defined above and below the signal band, the first stage LRT test passed, the secondary LRT is also performed using

$$L_b = \frac{y_S}{\overline{y_{SG}}} > \tau_b \tag{6}$$

where $\overline{y_{SG}}$ is the average signal estimate of all associated guard bands and τ_b is the guard-band threshold. Any number of guard bands may be defined, though one are two are typically sufficient. The upper and lower guard bands are chosen based on *a priori* knowledge of the expected signal. The lower guard band, signal band, and upper guard band need not be contiguous. By not requiring continuous spectrum coverage, spectral gaps may be introduced to ensure optimal performance of the first stage while minimizing signal leakage into the guard bands. The second test ensures that the detected signal is limited to the frequency band of interest. τ_t and τ_b are constant signal to noise ratio thresholds.

3. Synthetic examples

The differences between the two estimation options are best demonstrated on synthetic data representing ideal band-limited energy time series. Two synthetic representations of $x[n]$ with a sample rate of 1 Hz are processed through each detector illustrating the output of Eq. (5). Both pulses have equal SNR (20 dB), but