Proceedings of Meetings on Acoustics

Volume 17, 2012 http://acousticalsociety.org/

ECUA 2012 11th European Conference on Underwater Acoustics
Edinburgh, Scotland
2 - 6 July 2012
Session UW: Underwater Acoustics

UW115. Generalized marine mammal detection based on improved band-limited processing

Benjamin B. Bougher*, Joey Hood, James Theriault and Hilary Moors

***Corresponding author's address: Akoostix Inc, Dartmouth, B3B 1J4, nova scotia, Canada, bbougher@akoostix.com**

Akoostix continues to experiment with flexible, low-processing-load marine mammal detection options suitable for implementation in both workstations and low-power embedded systems. Building on previous work, additional processing stages have been added to normalize and de-noise spectrograms using a wide variety of user-configurable options. These pre-processing options can be optimized for the signals of interest, which enhances targets while significantly reducing the impact of structured noise that cause false alarms with other methods. Band-limited signal excess is then computed using one of several weighting functions, after which detection is performed on the signal excess time series. Out-of-band tests are also performed to ensure band-limited signals. The detector is tested on the workshop Minke whale localization dataset with configurations of varying complexity and performance (detection and processing load) are compared to other detection algorithms.

Published by the Acoustical Society of America through the American Institute of Physics

Received 7 Sep 2012; published 14 Dec 2012
Proceedings of Meetings on Acoustics, Vol. 17, 070067 (2012)

1 INTRODUCTION

Passive acoustic monitoring (PAM) can be used as a non-invasive technique for determining the presence of marine mammals based on detection of their vocalizations. PAM provides researchers the ability to monitor marine mammals in situations where traditional visual surveys are difficult to conduct or when the results obtained from visual surveys are not adequate [1]. However, PAM often results in large volumes of data that can be inefficient to analyze manually, and therefore the development of efficient and effective methods of vocalization detection is important for successful implementation of marine mammal mitigation measures involving PAM. Methods for automated detection of marine mammal vocalizations are actively being developed to improve the ability to accurately process large quantities of acoustic data in a timely manner.

Developing a general marine mammal vocalization detector presents a difficult problem as vocalizations vary substantially between species, which is further complicated by dynamic, non-Gaussian noise levels in the ocean [2]. These dynamic characteristics of background noise as well as interference from anthropogenic signals, such as shipping noise, may cause false-positive detections during real-time PAM, such as for mitigation during military and seismic activities. False alarms may result in unnecessary stoppages of these activities, which is often expensive and time consuming. False detection rates can be decreased by increasing detection thresholds, which may also result in missed detections, or by applying higher-order filtering, which reduces the generality, increases the cost, and increases complexity in configuring and operating the system.

There has been significant advancement in the optimization of automated marine mammal detection algorithms in recent years. The optimal method for known signal detection in white noise is a matched filter. This method has been adapted for marine mammal detection by applying synthetic matched filtering, where a synthetic vocalization is developed and correlated against an acoustic dataset [3]. Spectrogram correlation, which follows a similar approach using spectrogram signal kernels, has been

successful in right whale detection [4]. A spectrogram approach with additional filtering using impulsive noise kernels and feature extraction has also been applied to right whale data [5]. These approaches require significant *a priori* information and predictable signal characteristics and thus may not be suitable for dynamic situations involving multiple species, such as during real-time PAM. When general detection is required and discrimination between similar signals is not a priority, band-limited summed energy methods may be appropriate.

For maximum efficiency, human operators should be able to adjust a PAM detector with basic training. An effective PAM detector should also have a high probability of signal detection and low false-alarm rate, be robust across many signal types, and require a low-processing load allowing scalability to multiple platforms. Considering these constraints, this study examines several detection processing streams based on band-limited spectral energy. The objective was to minimize false alarms without significantly increasing the complexity of the system. A generalized detection-processing stream developed by Akoostix Inc. and Defense Research and Development Canada (DRDC) is presented. Processing streams of varying complexity designed to detect minke whale (*Balaenoptera acutorostrata*) vocalizations were applied to a dataset with manually identified minke whale vocalizations, a dataset consisting of only shipping noise, and a dataset with vocalizations of other species present.

2 GENERALIZED BAND-LIMITED DETECTION

To aid in analysis of large PAM data sets, as well as development of PAM processing for Autonomous Underwater Vehicles (AUV) [6], a generalized low-power band-limited detector has been developed (Figure 1). Each stage in the processing stream is described in the sections below.

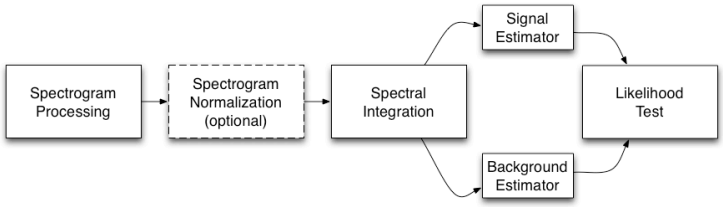

Figure 1: Block diagram of a general band-limited energy detector.

2.1 Spectrogram Processing

The first stage of the band-limited detection processing stream generates a spectrogram from acoustic time-series data using a fast Fourier transform (FFT) to form the series of spectra, $P(f,t)$. Each spectrogram is generated using overlapping Hann-shaded windows, with a power-of-two window size that provides a spectral resolution that best matches the short-time bandwidth of the signal of interest. Windows should typically be overlapped by 50% - 75% of the window size, which helps to reduce signal splitting between time frames.

2.2 Spectrogram Normalization

The spectral data may then be passed to an optional spectral processing stage, which acts to normalize the background level and enhance the signal of interest. The normalization method depends on the signal characteristics and available processing power. For band-limited signals, such as moans, squeals, or songs, a background estimate taken across each spectrum can be used to filter broadband signals by dividing each spectral bin by the background estimate. The detector streams presented make use of a block average background estimate and a sliding median background estimate. The median is a more expensive operation, but can be configured so that large signal energy does not bias

the background estimate. The window used for the median calculation should be at least twice the short-time bandwidth of the signal to avoid a self-polluted background estimate.

2.3 Spectral Integration

Time-series data, $E(t)$, are generated by integrating spectra between user-defined band limits. Overlapping spectral sub-bands are used to limit signal energy splitting between adjacent bands. The integration is performed using one of three algorithms, which are described below.

2.3.1 Energy Summation

Energy summation is the simplest integration method, which generates a time-series for each band by summing the energy contained in each spectral bin:

$$E(t) = \sum_{f} P(f,t) \tag{1}$$

2.3.2 Entropy

The entropy algorithm, first applied to marine-mammal detection by Erbe [7], uses band-limited Shannon entropy to form the time-series data. First, each spectral band is normalized to unit energy:

$$P'(f,t) = \frac{P(f,t)}{\sum_{f} P(f,t)} \tag{2}$$

This normalization stage is equivalent to a block average spectral normalization with window equal to the number of spectral bins in the band of interest.
The information entropy is then calculated:

$$E(t) = -\sum_{f} P'(f,t) \cdot \log P'(f,t) \tag{3}$$

The entropy calculation acts to increase the contribution of higher energy spectral bins when forming the output time-series.

2.3.3 Power law

Following a similar approach to entropy summation, a power law detector raises each spectral value to a user-defined power before summing the spectrum [8]. Spectral bins with higher energy are given more weight in the summation by applying an *n-th* order power law:

$$E(t) = \sum_{f} P(f,t)^{n} \tag{4}$$

The power law approach provides a similar effect as an entropy sum while using less processing power, as logarithms are a computationally expensive operation.

2.4 Likelihood Test

The final stage in the detection stream tests the likelihood that the target signal is present. The simple test uses an estimated background level from $E(t)$ to assess the detection likelihood. A gapped-split window then generates estimates in three different sized windows:
A total window of length w_t:

$$W(t) = \sum_{t-\left(\frac{w_t}{2}\right)}^{t+\left(\frac{w_t}{2}\right)} E(t) \tag{5}$$

A signal window of length s_t:

$$S(t) = \sum_{t-\left(\frac{s_t}{2}\right)}^{t+\left(\frac{s_t}{2}\right)} E(t) \tag{6}$$

And a gap window of length g_t:

$$G(t) = \sum_{t-\left(\frac{g_t}{2}\right)}^{t+\left(\frac{g_t}{2}\right)} E(t) \tag{7}$$

To assess the likelihood that the target signal is contained in $S(t)$, a detection function, $D(t)$, is formed using the mean signal-to-noise ratio, where the noise term is determined by subtracting the gap window level from the total window:

$$D(t) = \frac{S(t) \Big/ s_t}{\left[\frac{W(t) - G(t)}{w_t - g_t}\right]} \tag{8}$$

Or a mean signal-noise difference for time-series generated from power law or entropy algorithms:

$$D(t) = \frac{S(t)}{s_t} - \left[\frac{W(t) - G(t)}{w_t - g_t}\right] \tag{9}$$

The signal window size should be set to approximately the same length as the signal duration that contains the majority of the energy. The gap-window size should be at least the duration of the target signal, ensuring that the noise estimate is not biased if there is signal present. The total window size should be set so that it provides a statistically relevant measurement for the background, but is short enough that neighbouring signals will not interfere. Figure 2 illustrates the three windows. A threshold, δ is used to determine detection whenever $D(t) \geq \delta$. The threshold can be set based on sensitivity requirements of the particular situation.

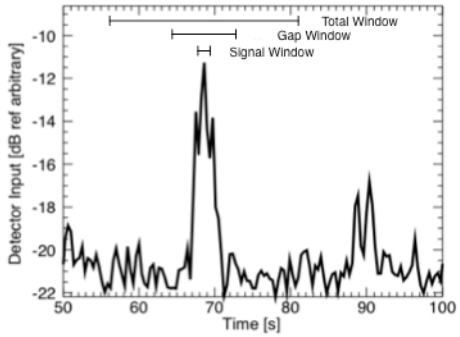

Figure 2: Gapped split windows. The signal window is used for the signal estimate, and the noise estimate is calculated by subtracting the gap window from the total window.

3 METHODOLOGY

3.1 Data Sets

A dataset that contained central minke whale vocalizations collected with seven hydrophones off the island of Kaua'i, Hawaii was provided as part of a 2011 workshop [9]. This dataset was manually analyzed and identified minke whale boings were annotated. The boings were annotated and reviewed

by two analysts to increase confidence in the identified vocalizations. The data set also contained similar, interfering vocalizations from other species, such as humpback whales [10]. In total, 210 minutes of acoustic data were analyzed and 268 boings were annotated.

In order to test detector robustness against false detections caused by ship noise, a dataset collected during a Defence Research and Development Canada (DRDC) sea trial in Exuma Sound, Bahamas was selected. The dataset contained 250 minutes of loud intermittent shipping noise recorded on sonobuoys. No marine mammal vocalizations were present on these recordings.

Robustness against detections caused by interfering vocalizations was tested using a dataset collected in the Gully submarine canyon south of Nova Scotia in the summer of 2011. This dataset was collected over the course of four days (27 May – 30 May 2011) using a combination of sonobuoy, towed array, and passive acoustic reusable buoys (PARB). This dataset was dense with delphinid clicks and squeals, as well as sensor related noises, but did not contain any minke whale vocalizations. In total, the Gully data contained approximately 90 hours of acoustic recording.

3.2 Detector Configuration

Detector streams were configured for detection of central-Pacific minke-whale boing vocalizations. These minke whale boing vocalizations range from 1.0 seconds to 4.0 seconds in duration and include frequencies from approximately 1 kHz to 10 kHz, although the dominate energy is normally between 1.3 kHz and 1.5 kHz [9],[11]. This vocalization consists of an initial burst-pulse, followed by frequency modulation that transitions into a constant frequency tail. The parameters for each detection stream are defined in Table 1.

Table 1: Detection stream configurations for central Pacific Minke boing vocalization

Detector Stream	FFT duration [s]*	Spectral normalization	Integration algorithm	Integration band [Hz]**	Detector window sizes [s]
Energy	1.5	None	Energy summation	1200-1600	$w_t = 10.0$ $s_t = 1.0$ $g_t = 5.0$
Normalized energy	1.5	100 Hz sliding median window	Energy summation	1200-1600	$w_t = 10.0$ $s_t = 1.0$ $g_t = 5.0$
Entropy	1.5	Block average	Entropy	1200-1600	$w_t = 10.0$ $s_t = 1.0$ $g_t = 5.0$
Power law	1.5	Block average***	Power law, power of 2	1200-1600	$w_t = 10.0$ $s_t = 1.0$ $g_t = 5.0$
*An overlap of 75% was used for each FFT frame **The integration band was divided into 100 Hz sub-bands with 20 Hz overlap ***Also tested with median normalization, but results did not differ significantly. Chosen to match entropy normalization.					

3.3 Detection Performance

The automated detection streams described in Section 3.2 were used to detect whale vocalizations on the minke whale dataset. The automated detections were compared to manually identified vocalizations using receiver-operator-characteristic (ROC) analysis [12], which compares automated detections generated by varying threshold sensitivities to manually annotated vocalizations (*e.g.* Figure 3).

Detections that occurred at times that overlap manual annotations were considered true positives, while detections falling outside of manual annotations were considered false alarms. As the detection sensitivity was decreased, the false alarm rate decreased at the cost of missed positive detections. The percentage of manual annotations that contain auto-detections was plotted against a false alarm rate, and the best performance was considered the curve nearest to a unit step function (eg. 100% probability of correct detection and zero false alarms). The detector streams were tested on the two datasets without minke whale vocalizations, and false alarm rates corresponding to 50%, 75%, and 90% probability of correct detection were compared.

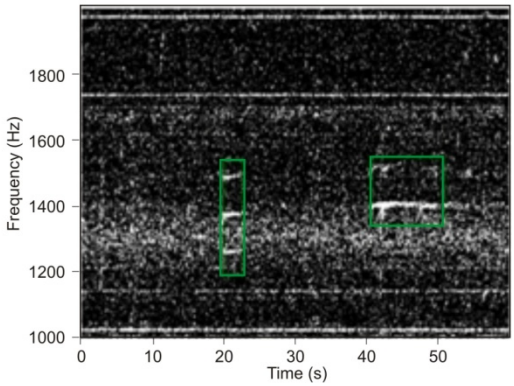

Figure 3: Example of two annotated minke whale boings from the minke whale dataset.

4 RESULTS

The ROC curves (Figure 4) show that the entropy and power law streams performed best on the minke whale dataset. The energy bands detection stream performed the poorest as it yielded the highest false alarm rate for a given probability of correct detection. The false alarm rate for each data set (Table 2-Table 4) shows that detection streams with spectral normalization (entropy, power law, normalized energy bands) were very robust against shipping noise, generating zero false detections with a threshold corresponding to 75% probability of correct detection. At high sensitivity settings, the performance of the normalized energy band detector surpassed the power law in some instances. This is demonstrated by the curves crossing at approximately 95% probability of correct detection, and the lower false alarm rate at 90% probability of correct detection on the Gully data (Table 4).

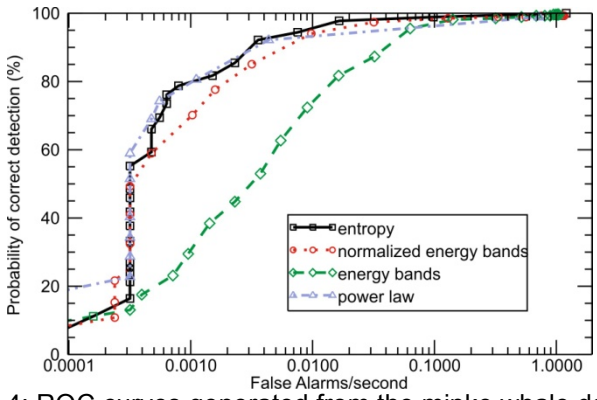

Figure 4: ROC curves generated from the minke whale dataset.

Table 2: False alarm rates [count/hour] for probability of correct detection of 50%.

Detection Stream	Minke whale	Shipping noise	Gully
Energy bands	10.85	0	70.74
Normalized energy bands	1.14	0	0.38
Power law	1.14	0	0.07
Entropy	1.14	0	0.11

Table 3: False alarm rates [count/hour] for probability of correct detection of 75%.

Detection Stream	Minke whale	Shipping noise	Gully
Energy bands	36.00	21.83	293.99
Normalized energy bands	4.57	0	3.70
Power law	2.00	0	1.87
Entropy	2.28	0	0.9

Table 4: False alarm rates [count/hour] for probability of correct detection of 90%.

Detection Stream	Minke whale	Shipping noise	Gully
Energy bands	137.45	235.93	684.67
Normalized energy bands	21.14	0	13.24
Power law	10.85	0	26.11
Entropy	11.14	3.52	7.93

5 DISCUSSION

The results can be compared to [10], which used a similar minke whale dataset to test an optimized spectrogram contour detector. At 75% probability of correct detection, the contour detector showed a false detection rate of approximately one false alarm for every 9 minutes of data, or 6.67 false detections per hour. This is comparable to the results presented in Table 3. The contour detector is more complex and requires additional user input, but also provides spectral classification information in the output.

The largest performance gain came from the spectral normalization stage of the detector. This is demonstrated by the large detection performance improvement between the energy band and normalized energy band detector (Table 2-4), as spectral normalization is the only stage that differs

between the two streams. Normalizing across each individual spectrum acts to remove high-energy broadband signals, whereas time-based spectrogram normalization methods, e.g. exponential averaging, will emphasize any transient signal. When spectral normalization was configured for minke whale boings, the detector was robust against false alarms caused by both shipping noise and other vocalization types (such as clicks produced by delphinids). It should be noted that as the bandwidth of the signal increases, the effect of normalizing across each spectrum will diminish, and thus this approach should not be used for detection of broadband signals such as clicks. Time-based normalization would be more effective for clicks.

Using power law and entropy spectral integration algorithms showed additional detection improvement (Tables 2-4). The performance gain would occur for lower-SNR signals where the peak energy is contained in fewer bins, causing the signal to be obfuscated during simple energy summation. Since these algorithms increase the contribution of higher-energy bins, the bins containing signal had more influence during summation, thus allowing for detection to be triggered. The entropy and power law algorithms showed a similar performance, but the entropy calculation was computationally more intensive. Using 500 000 records of 1024-element summations to test performance loading, an x^2 calculation used 20% of the processing load required for an $x \log x$ calculation. Future work to examine higher-order weighting factors is required to determine optimal performance.

The general detector described would improve the efficiency of both real-time PAM and post-analysis of acoustic data. A single operator, who could be cued by the auto-detections, would be capable of monitoring multiple sensors in real-time. The detector can also be used to extract presence/absence information from large datasets, improving the efficiency of PAM-based research. For example, using a threshold corresponding to 75% probability of detection, the absence of minke whale boings in the 90 hour Gully data set, which was dense with cetacean vocalizations, could be determined by assessing just 81 false detections.

6 CONCLUSION

An approach to detecting marine mammals acoustically using an automated vocalization detector has been presented. User-configurable signal-processing streams were used to test the approach for detection of central minke whale boings and robustness against false detection. Normalization across each spectrum showed significant improvements in detection performance of band-limited signals, while applying functions such as power law and entropy during integration of each spectrum showed additional performance gain over a simple energy summation. The detector was shown to be robust against false detection from both shipping noise and vocalizations produced by other species.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the 5th International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics for providing the minke whale data set, and the Captain and crew of Canadian Forces Auxiliary Vessel (CFAV) Quest for the collection of the shipping noise and Gully datasets.

REFERENCES

1. D.K. Mellinger, K.M. Stafford, S.E. Moore, R.P. Diak and H. Matsumoto, An over of fixed passive acoustic observation methods for cetaceans, Oceanography, 20(4) 36-45. (December 2007).
2. R.J. Urick. Ambient Noise in the Sea, Peninsula Publishing, 7.1-7.19. (1984).

3. D.K. Mellinger and C.W. Clark, Methods for automatic detection of mysticete sounds, Mar.Fresh.Behav.Physiol. 29(1) 163-181. (June 1996).
4. D.K. Mellinger, A comparison of methods for detecting right whale calls, Can.Acoust., 32(2), 55-65. (2004).
5. I.R. Urazghildiev, Detection and recognition of North Atlantic right whale calls in the presence of ambient noise, IEEE. J. Ocean. Eng., 34(3) 358-368. (July 2009).
6. J.A. Theriault, J.D. Hood, D. Mosher and T. Murphy, Performance of automated detection algorithms for beaked whales using autonomous underwater gliders, 4th International Workshop on Detection, Classification and Localization of Marine Mammals Using Passive Acoustics, Abstract only. 28. Pavia (2009).
7. C. Erbe and A.R. King, Automatic detection of marine mammals using entropy, J.Acoust.Soc.Am., 124(5) 2833-2840 (August 2008).
8. A.H. Nuttall, Detection performance of power-law processors for random signals of unknown location, structure, extent, and strength, NUWC-NPT Technical Report 10,751, 15-31. (September 1994).
9. Workshop Data Set, 5th International Workshop on Detection, Classification, Localization, and Density Estimation of Marine Mammals using Passive Acoustics, 22-25. (August 2011) http://www.bioacoustics.us/dcl.html.
10. D.K. Mellinger, S.W. Martin, R.P. Morrissey, L. Thomas and J.J. Yasco, A method for detecting whistles, moans, and other frequency contour sounds, J.Acoust.Soc.Am., 129(6) 4055-4061. (June 2011).
11. S. Rankin and J. Barlow, Source of the North Pacific “boing” sound attributed to minke whales, J.Acoust.Soc.Am., 118(5) 3346-3351 (November 2005).
12. T. Fawcett, An introduction to ROC analysis, Pattern Recognition Letters, 27 861-874 (2006).