Automatic Discrimination Of Beaked Whale Clicks In Noisy Acoustic Time Series

L.J. Hamilton and J. Cleary
Defence Science & Technology Organisation (DSTO)
13 Garden Street
Eveleigh NSW 2015, Australia

***Abstract*-A matched filter bank methodology is investigated for the automatic detection and classification of stereotypical beaked whale chirps in long acoustic time series. Energy normalization and non-dimensional thresholding techniques are employed. A primary confidence level is provided for each detection based on an indirect measure of distortion or malformation of the detection relative to its template waveform. The matched filter correlator peak output value is used to derive a secondary confidence level. Ambiguity experienced by matched filtering is recognized, and removed for some marine mammal clicks by simple time domain or frequency rules. Matched filters are very good at signal detection, but very poor at signal discrimination in some circumstances.**

I. Introduction

Passive acoustic means are presently regarded as the best method of obtaining information on the distribution of some of the more elusive marine mammals, especially the beaked whales. Automatic detection and classification of marine mammal vocalizations is therefore a topic receiving much attention. Beaked whales belong to the odontocetes (toothed whales), a suborder of the Cetacea which include whales, dolphins, and porpoises.

Marine mammals produce a wide variety of vocalizations, e.g. clicks, chirps, and whistles, and different cetaceans utilise frequencies from tens of Hz to hundreds of kHz. This implies that a range of methods may be required to detect and classify all whale sounds in a particular area. A further difficulty in automation is that different species in the same family, and even different families, may have quite similar calls, or may share the same frequency range. The present paper investigates an automatic matched filtering method to detect the stereotypical clicks of beaked whales (Fig. 1). Whale songs (a repeated and organized or structured set of sounds) and whistles are not considered. Signals accepted as possible detections are described by two measures of confidence relating the matched filter cross-correlation function to the autocorrelation function of the targeted waveform. If only a knowledge of presence or absence is required, this allows the user to first investigate detections with higher confidence levels, potentially saving much time. The scheme is suitable for real time applications, provided that suitable template signals are available beforehand. The use of proxy templates may be feasible in some circumstances.

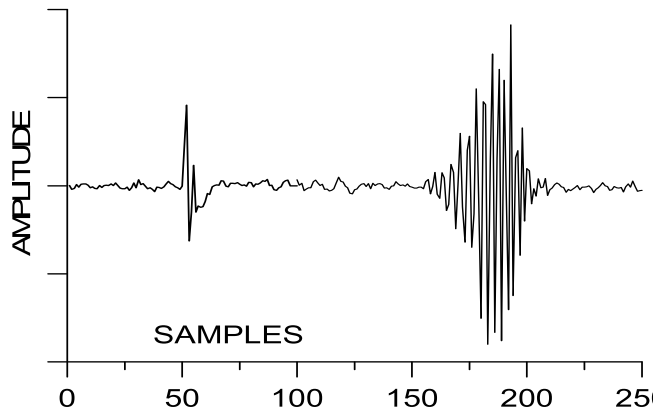

Figure 1. Examples of stereotypical whale vocalizations from the digital library of the Marine Biological Library, Woods Hole Oceanographic Institute (www.mblwhoilibrary.org). The right hand waveform is a probable Blainville's beaked whale chirp. Note the relative differences in duration of the two waveforms.
(Filename Set3_A1_042705_CH11_H11_A0300_0330.WAV, Time 11:14.0500 to 11:14.0534).

II. Methods for Automatic Detection of Marine Mammal Clicks and Chirps

[1] consider that automatic classification of marine mammal calls involves at least three steps: signal detection, feature extraction, and classification. [2] list the following methods as having been used for automated detection of sperm whale clicks: Fourier Transform, wavelet transform, step change method, high order moments, Hilbert-Huang Transform, Teager-Kaiser energy operator, Stochastic Matched Filter, spectrogram, neural network, and a spectral null method. [1] state that odontocete call identification is difficult owing to the complexity of their calls. They list other automatic methods as linear discriminant analysis, hyperplanes with classification and regression trees, dynamic time warping, unsupervised classifiers, adaptive resonance theory, and Gaussian mixture models. Many of these methods couple spectrograms (see Fig. 2) and feature extraction. None of the above methods has as yet received general acceptance, although some have achieved

978-1-4244-5222-4/10/$26.00 ©2010 CROWN

extremely high detection rates against sets of hand picked signals.

The present authors have trialled only a few of the listed methods. Several marine mammals have rather similar waveforms and similar peak frequency, and simple Fourier Transform methods do not always provide sufficient discrimination for these cases, especially in the presence of noise and signal distortion. Our investigations have shown that low degrees of signal distortion can considerably change the spectrogram of a call, rendering it unreliable for automatic classification. Spectrograms (Fig. 2) are a function of several user choices, e.g. window size and degree of window overlap. Visually they are very helpful, but may not be so for automatic processing. Neural networks have had good results for some cases [3]. Although neural nets do not necessarily require the input of user rules, they require intensive and specialized training, are susceptible to noise, and may possibly work only for a particular data set [4][5]. These points detract from their usefulness. Additionally their complexity can mask their actual behaviour, which is sometimes very simple.

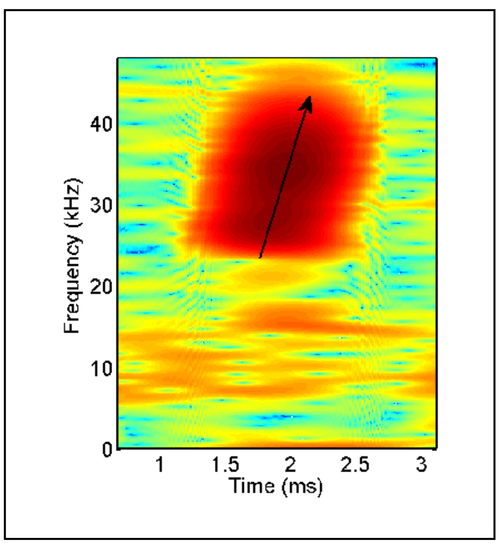

Figure 2. Spectrogram of a chirp with frequency upsweep.

Our requirement is for a simple and reliable automatic detection and classification method requiring a minimum of user interaction, and ideally suitable for real-time applications. The detection method must be chosen to suit the characteristics of the particular vocalizations of interest, which for beaked whales are stereotypical clicks and chirps. Matched filtering is an efficient method to identify stereotyped chirp signals, and is also efficient at identifying signals in noise.

III. Matched Filtering

A matched filter is constructed from an acoustic signature or template waveform known to be for a particular whale species, and the template is cross-correlated against the acoustic record to act as a detector for that particular signature. Conceptually this is an uncomplicated form of signal processing which does not begin with the overheads of spectrogram correlation or neural nets.

The output from the matched filter is a cross-correlation function (Fig. 3). Typically the only part of this function that is used is the peak value. A detection is reported if the peak value is above some user decided threshold. Matched filtering with a single template is available in the publicly available ISHMAEL software program [6], a well presented package specifically designed for marine mammal detection. ISHMAEL also includes spectrogram correlation and energy summation over user selected frequency bands. For the ISHMAEL matched filter implementation the user inputs a threshold, then manually verifies the detections. There are some difficulties with ISHMAEL usage. Firstly, what is the threshold? Since matched filters are efficient at detecting signals in noise, the usual signal to noise thresholds do not apply. Trial and error must be used to find a suitable threshold for a particular task. For a noisy 30-minute test file, the number of ISHMAEL matched filter detections rose without apparent limit to millions as the threshold was decreased. The number of detections did not stabilize, and no stopping rule could be formulated. Our observation was that the ISHMAEL matched filter found every wiggle with a hint of periodicity, without any suggestion of discrimination. Secondly, ISHMAEL does not use energy normalization, so that high/low amplitude signals produce a high/low cross-correlation output. This further precludes discrimination between calls from different species. It also means that no allowance is made for the attenuation of calls received from different ranges.

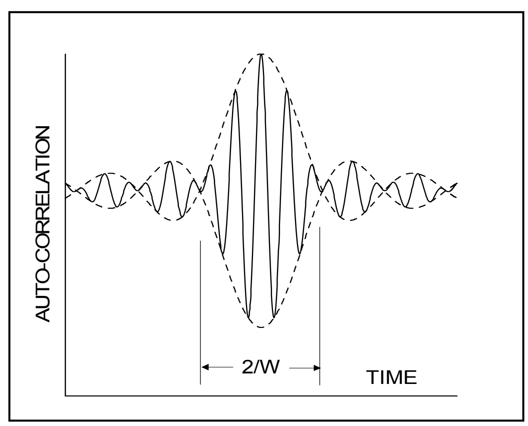

Figure 3. The autocorrelation function for a chirp signal, approximated as a sinc function. W = bandwidth (Hz) of the chirp.

In general at least two matched filters would be needed to provide discrimination between a wanted signal and other unwanted signals, together with some form of normalization.

Our approach is to trial a matched filter bank with energy and threshold normalizations.

IV. Matched Filter Bank With Normalisation

In the present post-processing scheme ISHMAEL matched filtering is used to make the initial detections (it is very efficient at this), then energy normalization is introduced, and the detections are input to a matched filter bank. For each matched filter, the shape of the cross-correlation function produced by the template and the detected waveform is described by its peak height and kurtosis. These two shape parameters are normalized by the height and kurtosis of the corresponding auto-correlation function. Signals with less than 50% correlation amplitude matching are discarded as having little chance of matching the targeted waveform. The normalized kurtosis is used as the primary measure of the reliability of the detection, and is observed to be related to the degree of distortion or malformation of the waveform compared to its template.

Three normalized cross-correlation amplitude thresholds were used, these being 50, 75, and 85% (low, medium, and high amplitude matching). However, the height ratio only becomes useful as a secondary indicator of detection confidence at a later stage of the processing. Because of the energy normalizations, detections with only a few cycles and very short duration may take high signal amplitudes. These can act like delta functions or spikes, returning high cross-correlation values across the full width of longer duration signals. These cases must be recognized and removed before using the amplitude ratios to form detection classes. An unrelated factor is that noise may occasionally cause cross-correlation amplitudes to exceed 100%.

Normalised kurtosis was used to assign a confidence to each detection as follows: (1) > 90%, taken as positive identification of a whale species, (2) > 80%, probable identification, (3) > 70%, possible identification, (4) > 65%, marginal identification, and (5) < 65%, not the target whale, or too distorted a signal for a reliable automatic result. A detection was assigned to the matched filter returning the highest kurtosis detection class, or to the highest kurtosis when kurtosis classes from different matched filters were the same level. The latter case is flagged for possible ambiguity.

For each file three subdirectories are formed corresponding to the three amplitude ratios, and four subdirectories are used within each of these for the four kurtosis classes. Each accepted detection is stored to allow manual verification. By drilling down into the confidence levels the present scheme allows the user to initially examine perhaps 50 to 100 detections instead of thousands to hundreds of thousands per 30-minute file.

Through the normalisations a low amplitude well formed signal scores higher than a distorted high amplitude signal. The degree of distortion of the waveform is important to the classification, not the signal to noise ratio.

Table I shows sample results for four test files. The test data were provided by [7] for a Coral Sea survey. The first file (row 1 of the table) has only 2 detections, each at the lowest possible confidence level for both amplitude and kurtosis. File 2 has no detections. Files 3 and 4 clearly indicate the presence of the targeted whale species.

V. Matched Filter Ambiguity

The output of a matched filter is a cross-correlation function (e.g. Fig. 3). When the detection has the same properties as a template the output becomes the autocorrelation function of that template Several marine mammals emit chirps, a waveform in which the frequency increases or decreases linearly with time. A property of chirp signals can result in detection ambiguity, even though chirp waveforms may have obvious time domain differences.

The autocorrelation function of a chirp is approximately a cosine modulated by a sinc function. The width between the primary nulls of the sinc function is 2/W, where W is the chirp bandwidth (Hz). The frequency of the cosine is the median frequency of the chirp [8][9]. Chirps with similar bandwidth will produce sinc envelopes with similar widths between the primary nulls, regardless of chirp duration or centre frequency. Chirps with similar median frequencies and similar bandwidth have similar autocorrelation functions, including phase, regardless of differences in waveforms of duration or number of cycles.

TABLE I. Sample detection classes derived for four 30-minute test files.
Asterisks (*) denote no misclassifications, and bracketed values are the number of erroneous detections in the total shown, for file 4 only.

File	Definite BW				Probable BW				Possible BW				Marginal BW			
	Low	Med	High	> 100	Low	Med	High	> 100	Low	Med	High	> 100	Low	Med	High	> 100
1	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	15	31	0	14	26	9	0	16	14	0	0	91	10	0	0
4	0	20*	52*	0	6*	39*	31*	0	(2)17	24*	6*	0	(7)20	16*	0	0

It is also possible for amplitude modulated clicks with constant frequency to have similar autocorrelations to chirps. A third case of possible ambiguity caused by short duration high frequency waveforms consisting of one or two cycles was mentioned in the previous section.

The end result is that it is possible for completely different waveforms, whether chirps or not, to have near the same autocorrelation functions. This matched filtering ambiguity must be resolved by other information on the signal. Simple time or frequency domain comparisons suffice for some marine mammals. For example, beaked whales have much longer signal durations than click sounds of other odontocetes. This makes the end problem for beaked whales rather simple. However, use of the present methodology should allow several types of whales to be reliably identified, not only beaked whales.

VI. Results

The scheme was run on detections identified for a range of ISHMAEL matched filter thresholds from 70 down to 2 mU (milliUnits of unspecified dimensions). Over this threshold range the number of ISHMAEL detections increased from 4 to more than 150,000 for one noisy 30-minute file. For the present scheme the number of identifications for a particular targeted test vocalization stabilized within all detection classes for ISHMAEL thresholds below 8 mU. Essentially this means that the scheme is largely self verifying. The combined number for all 12 possible classes (3 amplitude times 4 kurtosis) was 240, somewhat less than the initial 150,000. Fifty detections were classed as having high amplitude and high kurtosis. These results mean that ISHMAEL (or any other detector) need be run once only with a single low threshold to obtain the detections to be supplied to the present scheme. Because of the normalizations, and because signal to noise thresholds are not used, the scheme is also impervious to gain changes.

VII. Verification Of Results

Manual verification must be performed at least once for each whale type, although there are many direct and indirect ways to do it. Several authors test their detection and classification schemes against hand picked well formed signals with high signal to noise characteristics. Reference [10] varied this procedure by forming Receiver Operating Characteristic (ROC) curves for signals with a range of signal to noise ratio. As noted earlier signal distortion may be more important than signal to noise ratio in deciding the performance of an algorithm. [10] noted cases where performance for calls with high signal to noise ratio was occasionally worse than for calls with lower signal to noise ratio. The degree of distortion could be a possible cause.

Signal to noise ratio could be used as a basis to choose test signals for the present scheme, but there seems little point. The actual acoustic record may be used, without excluding lower quality signals from the testing. Verification then begins with manual scanning of the accepted detections having highest confidence levels. This examination is not difficult if sets of overplots of the signals in each class are viewed. For one 30-minute test file (see file 4 of Table I), 187 detections were made for the high and medium cross-correlation amplitude ratio classes. All were manually verified as being from the target whale. False detections were made only for the low amplitude cross-correlation ratio range for the two lowest kurtosis ratio classes of possible and marginal. However, some were missed for the medium amplitude ratio. Reference [10] considered a single-point score of 26% false detections as quite tolerable for many applications. Combined false detections for all twelve detection classes of the present scheme were better than 4% (9 of 231) for this one noisy test file, with no misclassifications for detections with high and medium cross-correlation amplitude ratio. The detection and classification scheme is obviously reliable and useful. The nine false detections were generated by a signal from a whale other than the target whale which were closely followed by structured noise or other signals.

An indirect verification of the detections and their classes was also obtained by examining overplots of Fourier Transforms of the detections for each class. Peak power tended to decrease with kurtosis ratio for each amplitude ratio class. The variation in power spectral density for frequency bins was observed to increase with decreasing kurtosis ratio only for the low amplitude ratio class. Obvious anomalies in transforms occurred only in the low amplitude ratio class.

Missed Detections

The ISHMAEL matched filter implementation makes detections highly efficiently albeit highly indiscriminately. For this reason missed detections in the initial phase are not likely to be a problem. After ambiguity in classification is resolved, the main problem for the present scheme is likely to be valid detections erroneously excluded by the lowest thresholds used for kurtosis and amplitude. The threshold levels were selected to correspond to visually assessed changes in signal quality, but additional detection classes with lower thresholds than those used here can be added if necessary.

Visual examinations indicated that for medium amplitude ratios, 16 signals discarded by the lowest cross-correlation kurtosis ratio threshold of 65% were all target whale clicks. Kurtosis ratio ranged from 51 to 64.3%, with a median of about 60%. This indicates the lowest kurtosis threshold of 65% is too high (50% may be more suitable). Adding another kurtosis class would allow for these cases. However, it is not necessary to have 100% detection rate when the target whale makes numerous vocalizations in a short interval (e.g. 30 minutes) in order to be sure that the target is present. Others have previously exploited this point, e.g. [10].

An unknown number of rejections for the low amplitude ratio class were also observed to be possible target whale clicks. Many of these were not detected because of high instrumentation noise, which acted to add a noisy semi-random sinusoidal up-ramp or down-ramp on the signals. Following a suggestion by John Wendoloski of DSTO, removal of this noise may be investigated by Savitzky-Golay filtering. However, it is accepted that not every detection can be made automatically, or by matched filtering. This may particularly apply to signals received away from the main lobe of the hydrophone detector ("off-axis" signals). If this is a problem additional templates can be used for off-axis signals.

Many of the rejections were single, multiple, and joined or near joined echoes associated with the targeted whales. The primary signal was accepted but the echo or echoes associated with the primary signal was not. This is regarded as a good feature of the present scheme, but depends on user requirements.

VIII. Discussion

A matched filter bank and normalization techniques were used for discrimination of stereotyped whale clicks and chirps, with the expectation that this would provide better discrimination performance than the output of a single matched filter. This is certainly true, however, ambiguity in matched filter classification can arise for several different reasons. Particular examples are for chirps of similar bandwidth, and for high frequency signals of a few cycles which are much shorter than the target signal. These and other cases may return cross-correlation functions with the target template similar to the target autocorrelation. Matched filters are efficient at signal detection, including detection of signals buried in noise, making them a good choice for the first stage of detection and classification of stereotypical clicks and chirps. However, to achieve reliable discrimination between signals, user rules must be introduced, or other processing methods must be used in tandem with matched filters.

Acknowledgements

The test data used to verify the matched filter bank scheme were obtained by a team managed by Dr Doug Cato, with survey leader Mr Mark Savage, both of DSTO. The Royal Australian Navy provided the vessel HMAS Labuan and crew for an extended period to obtain the data.

References

[1] M.A. Roch, Soldevilla, M.S., Burtenshaw, J.C., Henderson, E.E. and Hildebrand, J.A. (2007). Gaussian mixture model classification of odontocetes in the Southern California Bight and the Gulf of California. *Journal of the Acoustical Society of America* 121(3), 1737-1748.

[2] A. Sánchez-García, A. Bueno-Crespo, J.L. Sancho-Gómez. (2009). An efficient statistics-based method for the automated detection of sperm whale clicks. *Applied Acoustics* Vol. 71, Issue 5, 451-459. [doi:10.1016/j.apacoust.2009.11.005Pages 451-459]

[3] S.O. Murray, Mercado, E. and Roitblat, H.L (1998). The neural network classification of false killer whale (*Pseudorca crassidens*) vocalizations. *Journal of the Acoustical Society of America* 104, 3626-3633.

[4] D.K. Mellinger and C.W. Clark. (2000). Recognizing transient low-frequency whale sounds by spectrogram correlation. *Journal of the Acoustic Society of America* 107(6): 3518–3529.]

[5] D.K. Mellinger (2008). A neural network for classifying clicks of Blainville's beaked whales (Mesoplodon densirostris). *Canadian Acoustics* 36(1):55-59.

[6] D.K. Mellinger (2001). ISHMAEL 1.0 User's Guide. NOAA Technical Report OAR-PMEL-120, 30pp. [http://www.pmel.noaa.gov/vents/acoustics/whales/ishmael/]

[7] D.H. Cato, Blewitt, M., Cleary, J., Savage, M., Parnum, I., McCauley, R.D., Dunlop, R.A., Gibbs, S. and Donnelly, D. (2009). Preliminary results of acoustic surveying for beaked whales in the Coral Sea near Australia. 3rd International Conference & Exhibition on "Underwater Acoustic Measurements: Technologies & Results", Nafplion, Greece, 21-26 June 2009; pp1399-1402.

[8] C. de Moustier. Fourth Asia-Pacific Coastal Multibeam Sonar Training Course. Cairns, Australia. 14-19 August, 2000.

[9] A. Hein. (2004). Processing of SAR data: fundamentals, signal processing, interferometry. Springer. ISBN 3-540-05043-4. 291pp.

[10] D.K. Mellinger. (2004). A comparison of methods for detecting Right Whale calls. *Canadian Acoustics* Vol. 32, No. 2, 55-65.