PROCESSING THE WORKSHOP DATASETS USING THE TRUD ALGORITHM

Edward Harland
Chickerell BioAcoustics, Chickerell, Dorset, UK

ABSTRACT

The Transient Research Underwater Detector (TRUD) is designed to search for echolocation clicks from marine mammals. It uses a spectrogram correlation method with a set of reference matrices to search for clicks from multiple species. This paper describes the algorithm and presents the results of processing the workshop datasets from the Third International Workshop on the Detection and Classification of Marine Mammals using Passive Acoustics held in Boston in July, 2007. The work shows that TRUD can detect and classify the target species. Recommendations are made for further improvements to the algorithm.

SOMMAIRE

Le Transient Research Underwater Detector (TRUD) est conçu pour rechercher les clics d'écholocation émis par certains mammifères marins. Il utilise une méthode basée sur la correlation de spectrogrammes avec un ensemble de matrices de référence pour rechercher les clics de plusieurs espèces. Cet article décrit l'algorithme et présente les résultats de la basée de donnes proposée par le 3rd International Workshop on the Detection and Classification of marine mammals using Passive Acoustics qui s'est tenu à Boston en juillet 2007. Ce travail montre que TRUD permet de détecter et de classer les espèces de cétacés que l'on recherche. Des recommandations sont formulées pour des améliorations futures de cette approche.

1. INTRODUCTION

There is an increasing need to monitor for the presence of acoustically-sensitive species such as marine mammals. Examples of such a need include as a precursor to the operation of high power sound sources (Tasker, 1998) or as part of a site survey leading to offshore installations such as wind farms or wave/tidal generators (Madsen, 2006).

One option for detecting the presence of marine mammals is to detect, classify, and, if possible, localise their calls. Ideally this process should be completely automatic. This process should have a very low false alarm rate combined with an acceptable probability of detection.

Marine mammals generally make three classes of calls, narrow bandwidth, medium bandwidth, and high bandwidth calls. Narrow bandwidth calls are tonal signals that can be processed using high-resolution FFT techniques to produce a spectrogram. Image processing techniques can then be applied to detect tonals and measure parameters about the signal to allow classification. Medium bandwidth signals are roars or grunts that have a significant instantaneous bandwidth and are prolonged in time. High bandwidth signals are typically the echolocation clicks used by odontocetes. These are very short in duration and occupy bandwidths of several octaves.

The Transient Research Underwater Detector (TRUD) was designed to detect high bandwidth signals only. It is intended to be used in conjunction with other medium and narrow bandwidth processing such as the MMADS system (Harland and Armstrong, 2004).

When designing such a classification system it is not necessary to classify every single click. Some clicks get distorted by acoustic propagation that results in miss-classificaton. It is necessary to look at the ensemble of clicks and make a majority decision on the single click classifications. It can be expected that any individual click may pass multiple species classification tests, but with varying degrees of confidence and any classifier must look across all of these outputs to arrive at the final classification decision.

This paper presents the background to the TRUD system, sets out how it operates and then presents the results of processing the workshop dataset for the Third International Workshop on the Detection and Classification of Marine Mammals using Passive Acoustics held in Boston, USA, in July 2007.

2. THE TRUD ALGORITHM

The TRUD algorithm is part of a suite of processing packages designed for the site characterisation role. Other packages process the medium and narrow bandwidth signals and also characterise the ambient noise levels. The whole package is aimed at stand-alone applications such as pre-installation monitoring for wind farms or tidal generators. However, the suite can also be used for real-time monitoring applications and the results presented in near real-time to an operator.

Previous work by the author led to the prototype version of the Porpoise Detector (POD) currently available from Chelonia in the UK (Tregenza, 1998). This prototype system used analogue processing for detection and classification of the harbour porpoise (*Phocoena phocoena*) clicks and suffered from a number of problems inherent in analogue systems including low dynamic range and filter mismatch causing false detections. More recently an attempt was made to produce an improved system using digital processing and this resulted in the Simple Porpoise Underwater Detector (SPUD) (Harland, 2007). SPUD is based on the spectrogram correlation system proposed by Mellinger and Clark (Mellinger and Clark, 2000). The SPUD system was then made more general by using multiple reference matrices and evolved into the TRUD algorithm described here.

Figure 1 shows the block diagram of the TRUD processing chain. The incoming datastream is processed in 128 kilosample blocks using a 64 point FFT with Blackmann-Harris weighting and 75% overlap. These settings were chosen as a good compromise between time and frequency resolution for use with the range of pulse lengths of echolocation clicks. The resulting spectrogram is then searched for clicks using a sparse reference matrix. There is one matrix for each species. The reference matrix of weighting coefficients is cross-multiplied on a cell by cell basis with the spectrogram to form the classification factor at each time increment.

$$clf = \sum_{f_r, t_r} s_{f,t} * w_{f,t}$$

where s is the spectrogram value at f,t and w is the weighting coefficient at f,t within the reference matrix.

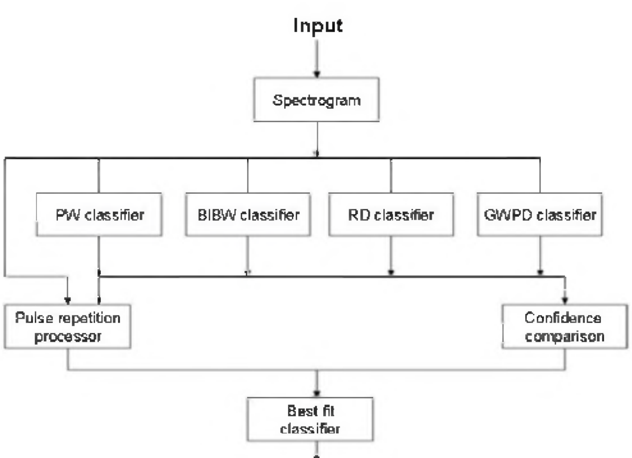

Figure 1 TRUD processing chain

The reference matrix can be up to 16 time samples by 31 frequency bins in size. The reference matrix is then moved along the time axis of the spectrogram, repeating the cross-multiplications at each sample interval to form the classification factor. By using negative and positive weightings classification occurs whenever the classification factor is positive.

In addition to the individual species reference matrices a General Wideband Pulse Detector (GWPD) is also implemented. This has two functions. It acts as a catch-all detector so clicks from species whose calls are not documented are not missed, and it can also be used as a pre-processor to minimise power consumption. The aim is to process the data initially using only the GWPD and to only search with the reference matrices when cetacean-like clicks are encountered. The GWPD uses energy summation over the frequency range 15-45 kHz and compares this with similar sums 5 samples ahead and behind the summation point. The detection threshold is chosen to be 10dB.

The outputs of each of the individual species classifiers are compared to decide which species any one click originated from. In order to aid this comparison it is better to use a confidence factor rather than the classification factor. The confidence factor is normalised and independent of the amplitude of the input click and reflects the degree of confidence in the single click classification. For the work described here the confidence factor was calculated as the ratio of the classification factor to the same summation carried out across only those cells with positive weighting. The confidence factor varies from 0 to 1 depending on the signal/noise ratio and the degree of match to the reference matrix.

The single click classification is then combined with the pulse train processing output and compared with references sets of expected parameters for individual species to form the classification decision. In a fully implemented system this would then be combined with the outputs from narrow and medium bandwidth classifiers and used with geographic information to arrive at the final classification decision.

TRUD is currently implemented in MATLAB (Mathworks, Release 14) running on a PC.

3. TRAINING TRUD

The initial training of the TRUD algorithm was carried out using a dataset from Blainville's beaked whale (*Mesoplodon densirostris*) recorded on dTAGs from animals off El Hierro in the Canary Islands (Johnson *et al*, 2005). This dataset is available from the MOBYSOUND website (Mellinger and Clark, 2006). This dataset has the advantage of a high signal/noise ratio and a sampling frequency of 192 kHz. The high sample rate is important as it means the data contains all parts of the click signal. Past experience working with harbour porpoise clicks suggest that the algorithm works best when a sample of the spectrum above the call is available as part of the classification test.

The optimum reference matrix was determined by a manual iterative trial and error method. As may be expected with such a high signal/noise ratio, 100% detection was achieved with no false alarms.

The reference matrix was then truncated at the lower Nyquist frequency used by the workshop dataset and used to process a selection of the training files from the workshop dataset known to contain calls from Blainville's beaked whale. It soon became clear that this reference matrix was sub-optimal because of the characteristics of the workshop datasets (see below). The reference matrix was then re-optimised using the same manual iterative trial and error method. This resulted in a significant improvement in the detection rate and a lowering of the false positive rate.

An example of a reference matrix is shown in figure 2 for Blainville's beaked whale. The areas shown in black have a positive weighting, while those shown in grey have a negative weighting. Blank cells are ignored and not used in the calculation.

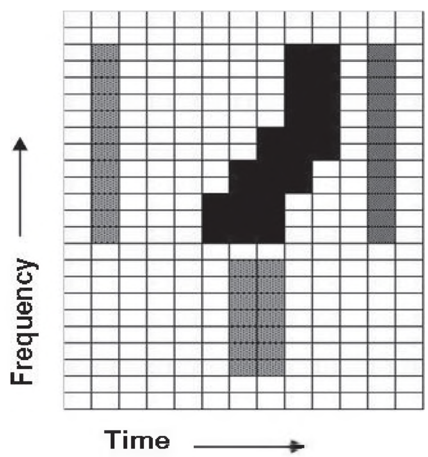

Figure 2 Reference matrix for Blainville's beaked whale

Different regions of the matrix are weighted depending on the frequency response of the system and the characteristics of the target species signal. There is also an additional differential weighting between the black and grey areas to implement the detection threshold. For these tests it was chosen to be 10 dB. The cell by cell weighted cross multiplies are summed together and a detection occurs when the sum is positive.

The same methodology was then applied to generating reference matrices for the other target species: pilot whale (*Globicephalus macrorhynchus*) and Risso's dolphin (*Grampus griseus*). No published information on the clicks of the pilot whale could be found so the reference matrix was initially chosen as a wideband pulse with spectral content from 15-48 kHz. The matrix was then optimised from the training dataset.

The Risso's dolphin click classifier was trained to look for the off-axis clicks. Madsen *et al* (Madsen *et al*, 2004) described the clicks of this species and showed that the off-axis spectral content was significantly different from the on-axis click. The training dataset contains examples of both types of click, but the off-axis type predominates.

From the testing it became clear that rejection of a non-target species click was as important as accepting the target species click. As an example, the pilot whale reference was adjusted to minimise false classification of some sperm whale clicks. A further round of optimisation was then carried out to improve the false positive rate at the expense of a small reduction in the true positive rate.

The GWPD reference matrix was also tested and partially optimised during this process to improve rejection of non-cetacean clicks.

4. THE DATASET

The dataset consisted of two groups of files. The training files were fully annotated with the sounds in the files, while the test files had no annotation. The aim was to train the classifiers using the training set and then use the trained classifiers to search and classify the test files. All files were sampled at 96 kHz and saved as WAV format single channel files.

During the training process described above it became clear that there were characteristics of the workshop dataset that were impacting operation of the TRUD algorithm and these are described below:

a) The low sample rate results in loss of the high frequency components of the call, reducing classification performance
b) The limited available bandwidth does not allow testing of the spectrum above the call which can result in a higher rate of incorrect classifications
c) System non-linearity introduces artefacts when signal levels are sufficiently high to drive the data collection system into non-linearity. This could be caused by clipping or slew-rate limitations (see figure 4). This can lead to incorrect classification
d) Different hydrophone channels have different characteristics, requiring a modified reference matrix to maintain optimum performance
e) Some clicks suffer significant dispersion due to acoustic propagation (see figure 3) which can lead to incorrect classification

The effect of the limited bandwidth and reduced spectral test is to reduce the effectiveness of the TRUD algorithm. Regrettably time did not permit an evaluation of the degree of degradation. The non-linearity was only a problem at high signal levels and resulted in a number of incorrect classifications when the animal was close to a hydrophone.

The reference matrix is optimised for a particular acoustic environment. This can be achieved by either pre-whitening the background or by optimising for a particular hydrophone system. The latter is the simpler method and was chosen for this test. Unfortunately the acoustic background was not consistent across the different hydrophones. This resulted in a number of incorrect classifications.

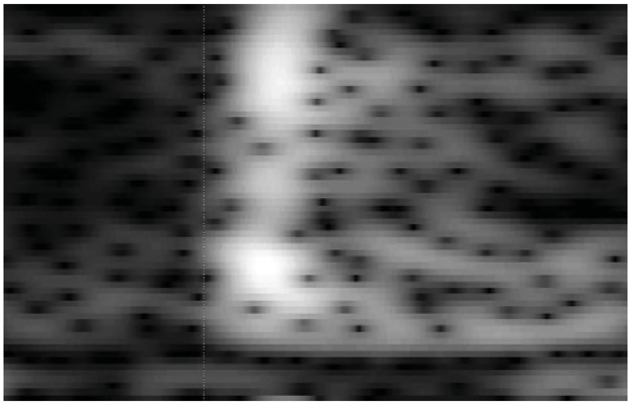

Figure 3 Spectrogram of pilot whale click with dispersion

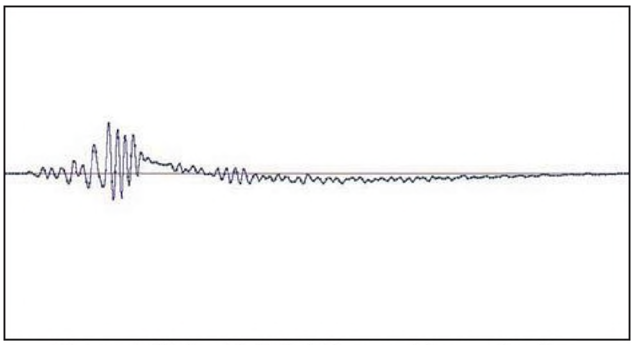

Figure 4 Click with DC shift caused by overloading

Some of the clicks suffered significant dispersion (see figure 3). It was not clear where this was occurring. It was probably predominantly acoustic dispersion but analogue transmission over long cables to the shore may have contributed to the effect. The effect was to turn the wideband clicks of species such as pilot whales into up-sweeps. This effect resulted in false negatives and/or false positives under some conditions.

5. PROCESSING THE DATASET

On completion of the training sequence the TRUD system was then used to process the test dataset. The initial processing suggested that there were other signals in the test files than those declared in the training dataset. This was seen as high GWPD counts but low counts from the target classifiers. Visual/aural checking suggested that the majority of these were sperm whale clicks and this was further confirmed by testing with a previously developed variant of the SPUD system optimised to detect regular sperm whale clicks.

Because of memory size limitations in the computer used to process the data, only five minutes of data from each file were processed. The cumbersome file names in the dataset were discarded and replaced with the simple terminology used in the following tables. The first five files for each of the species as listed in the dataset were used. The nomenclature used is that the column headed GWPD are the results for the general wideband pulse detector, BBW are the results for the beaked whale detector and PW are the results for the pilot whale detector. In later tables, RD are the results for the Risso's dolphin detector.

The results of processing the training set are shown in the following tables:

Table 1. Processing the pilot whale files

File	GWPD	BBW	PW
Pilot1	424	*11*	**205**
Pilot2	645	*14*	**399**
Pilot3	3313	*12*	**1495**
Pilot4	2023	*5*	**880**
Pilot5	1780	*508*	**1140**

The counts shown in bold are true-positives, the counts in italics are false-positives. The numbers in the GWB column are a guide to the possible number of clicks in the file. The level of false-positives are well within acceptable levels except for Pilot5. The pulses in Pilot5 are distorted and also suffer dispersion, resulting in the high number of false positive detections of Blainville's beaked whale and an increased number of missed pilot whale detections.

Table 2. Processing the Blainville's beaked whale files

File	GWPD	BBW	PW
BlBW1	91	**128**	*1*
BlBW2	244	**330**	*19*
BlBW3	603	**917**	*1*
BlBW4	1082	**1432**	*2*
BlBW5	1003	**1429**	*0*

It should be noted that for this species the number of detections by the GWPD is fewer than for the Blainville's beaked whale detector. This is due to the characteristics of the clicks from this species which are not a good match to the GWPD. The level of false positive detections of pilot whales is well within acceptable limits.

Table 3 Processing the Risso's dolphin files

File	GWPD	BBW	PW	RD
Risso1	6340	*37*	*78*	**86**
Risso2	2816	*868*	*255*	**685**
Risso3	5330	*501*	*1480*	**6004**
Risso4	14497	*2339*	*7522*	**10814**
Risso5	297	*5*	*3*	**69**

The Risso's dolphin reference matrix was not available until a few days before the workshop due to time constraints so it was not possible to process all the training set with this classifier. Neither was it possible to complete the optimisation procedure. Nevertheless, the performance is generally satisfactory.

The test files were then processed to give the following results. NP denotes that the data was not processed. The Test1 and Test2 files were processed in two blocks; a is 0-5 minutes and b is 5-10 minutes.

Table 4 Processing the test files

File	GWPD	BBW	PW	RD
Test1a	2902	2466	93	397
Test1b	1037	612	91	436
Test2a	5276	1328	376	100
Test2b	7870	986	200	88
Test3	21007	182	7989	NP
Test4	15213	110	3445	NP
Test5	5576	2663	665	6188
Test6	960	1237	110	NP
Test7	2884	1773	252	2965
Test8	4184	1915	1620	NP
Test9	1673	0	361	5

Test8 has distorted clicks due to both electrical and propagation effects. Test9 appears to be all sperm whale clicks. The PW classifications in Test9 occur on very strong sperm whale clicks with energy to 40 kHz.

A visual inspection of the test files reveals that a number of species are present. If the TRUD classification criteria is such that the species chosen is the classifier with the most outputs then these classifications are compared with the manual results in table 5. UNK are pulses from an unknown species. Note that the sperm whale classifier was not running for these tests so TRUD should not have found these clicks.

Table 5 Comparison of TRUD and visual results

File	TRUD	Visual
Test1a	BlBW	BlBW
Test1b	BlBW	BlBW, Unk
Test2a	BlBW	BlBW, Unk
Test2b	BlBW	BlBW, Unk
Test3	PW	PW
Test4	PW	PW
Test5	RD	RD
Test6	BlBW	SW, BlBW
Test7	RD	RD
Test8	BlBW	PW
Test9	PW	SW

The only incorrect classification is in Test8 where pilot whale calls are classified as Blainville's beaked whale. Table 4 shows that this was a marginal decision and inspection of the file suggests that it is caused by the propagation distortion of the pulses in this file.

A detailed inspection of TRUD operation for the Test1 and Test2 files shows that the unknown pulses were rejected by all three classifiers. The high false-positive count for Risso's dolphin is caused by propagation distortion of the Blainville's beaked whale clicks.

6. PULSE TRAIN TESTING

In addition to the single pulse classification testing, the pulses were associated into trains and the statistical properties measured to aid classification. A simple pulse train follower was written using a two parameter pulse association test (time and amplitude). The pulse interval was then measured between each successive pulse in the pulse train and the histogram plotted. The training dataset was used to gather the statistics for each of the target species. However, some of the files could not be used because the simple train follower could not cope with the multiply-interleaved pulse trains.

Once the histograms had been built for each of the species declared in the training files, a similar set of measurements were made for the test files and the results compared as shown in figure 5, 6, and 7.

This suggests that Test1 and Test6 contained Blainville's beaked whale, Test2 appears to have two species present, of which one is Blainvilles beaked whale, and Test7 is Risso's dolphin. Test8 is a good match to pilot whale. Test6 also contains another species with a low repetition rate as can be seen in figure 7. From table 5 it is likely that these are the sperm whale pulses. Test3, Test4, Test5, and Test9 were not processed.

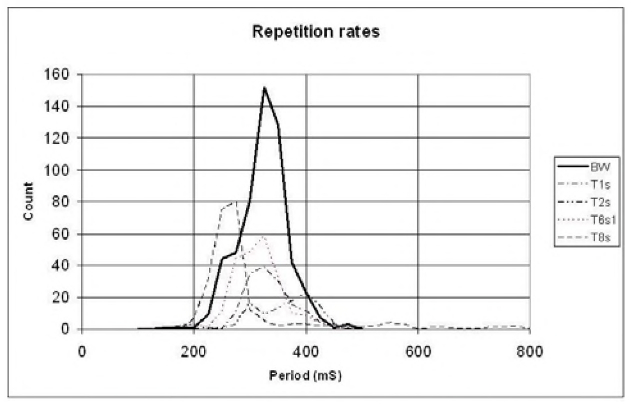

Figure 5 Beaked whale pulse train results

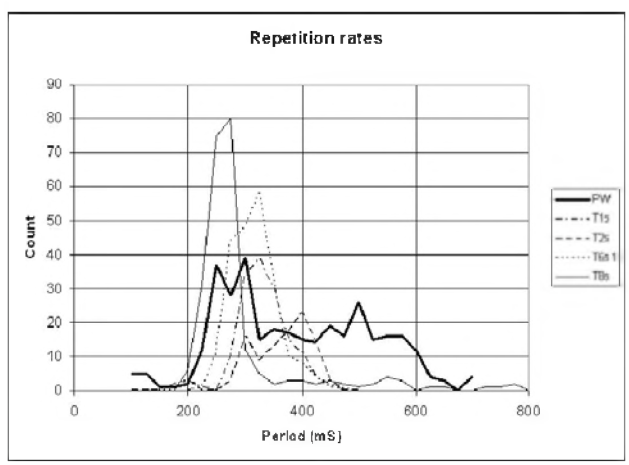

Figure 6 Pilot whale train processing results

These classifications were determined by visual comparison of the curves. Work is still in progress to carry out this comparison automatically.

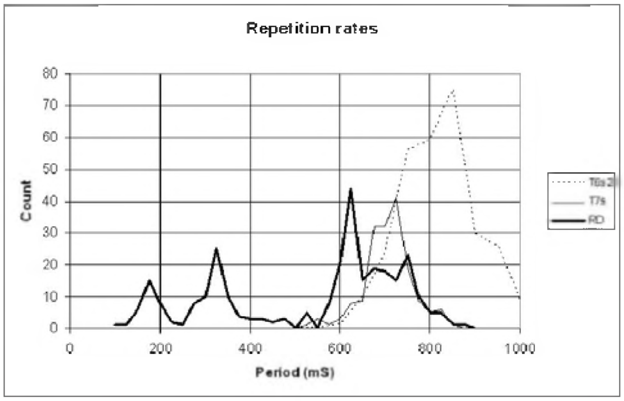

Figure 7 Risso's dolphin pulse train processing results

7. DISCUSSION

It must be emphasised that the results presented here are for the first stages in a multi-stage classification process. It is not possible, except in a very limited number of cases, to design a classifier that will uniquely identify the species that originated any one click. The complete classification process consists of building up a weight of evidence leading to a 'best guess' at the species. The results presented here are based on the classification factor output. Many of the false positives will be eliminated when the confidence factor is used and pulse association processing eliminates multiple classifications on a single pulse.

As an example, many of the incorrect Risso's dolphin classifications listed in the results tables are caused by a low confidence classification immediately before a high confidence beaked whale classification caused by the partial overlap of the reference matrix and the signal. Similarly, a number of beaked whale and Risso's dolphin false classifications are caused by the reverberation tail from a pilot whale click. Both of these would be eliminated by pulse association processing.

The pulse train classification needs a lot more data to fully define the pulse interval reference statistics for each species. The work here shows that it can be a useful classification aid for the specific circumstances of the training and test datasets, but much more work is needed to explore how useful it is across the full acoustic and geographic ranges of an individual species.

The present version of TRUD provides useful initial classification processing stages. The single pulse classification and pulse train statistics are two of the most important factors in the weight of evidence processing and this testing has shown that TRUD has the potential to fulfil this role. However, the testing has also shown that improvement is desirable in a number of areas:

a) For best performance the incoming data needs to be of a high quality with good linearity and a bandwidth sufficient to allow all of the spectrum tests. When used in its intended role as a stand-alone site monitor it will generally be possible to provide the requisite high quality input stage.
b) The Risso's dolphin reference matrix needs to be further optimised to reduce the false alarm rate from distorted Blainville's beaked whale clicks.
c) For best efficiency the general wideband pulse detector needs to perform better. The aim is to run only the GWPD while searching for clicks and to only activate the more detailed classification processes when candidate clicks are found that pass this lower threshold.
d) TRUD alone cannot provide a unique classification for many species. It will be able to classify to an acoustic clade level, but to refine the classification decision it will need to be combined with medium and narrow bandwidth classifiers.

8. CONCLUSIONS

This work has shown that spectrogram correlation is a viable classification method for echolocation pulses of the species in the workshop dataset. It has also shown that pulse train processing can aid the classification process for echolocation click sequences.

9. REFERENCES

Harland, E.J. (2007) The performance of the SPUD algorithm detecting harbour porpoise (*phocoena phocoena*) echolocation clicks. Third International conference on Bioacoustics. April 2007. Institute of Acoustics, Loughborough University, UK.

Harland, E.J., Armstrong, M.S. (2004) The real-time detection of the calls of cetacean species. Canadian Journal of Acoustics, 32: 76-82

Johnson, M., Madsen, P., Tyack, P., de Soto, N.A. (2005) A binaural acoustic recording tag reveals details of deep foraging in beaked whales. Journal of the Acoustical Society of America, 117: 2524

Madsen, P.T., Kerr, I., Payne, R. (2004) Echolocation clicks of two free-ranging, oceanic delphinids with different food preferences: false killer whales *Pseudorca crassidens* and Risso's dolphins *Grampus griseus*. J. Exp. Biol. 207: 1811-1823

Madsen, P.T., Wahlberg, M., Tougaard, J., Lucke, K., Tyack, P. (2006) Wind turbine underwater noise and marine mammals: implications of current knowledge and data needs. Marine Ecology Progress Series 309: 279-295

Mellinger, D.K., Clark, C.W. (2000) Recognizing transient low-frequency whale sounds by spectrogram correlation. Journal of the Acoustical Society of America 107: 3518-3529

Mellinger, D.K., Clark, C.W. (2006) MobySound: A reference archive for studying automatic recognition of marine mammal sounds. Applied Acoustics Special Issue: Detection and localization of marine mammals using passive acoustics 67: 1226-1242

Tasker, M. (1998) Guidelines for minimising acoustic disturbance to marine mammals from seismic surveys. JNCC, Aberdeen, UK.

Tregenza, N. (1998) Site acoustic monitoring for cetaceans - a self-contained sonar click detector. In: Tasker, M.L., Weir, C. (eds) Proceedings of the seismic and marine mammals workshop. London, 23-25th June 1998. Sea Mammal Research Unit, St Andrews, UK.