THE REAL-TIME DETECTION OF THE CALLS OF CETACEAN SPECIES

E J Harland[1] , M S Armstrong[2]
[1]QinetiQ Ltd, Winfrith Technology Centre, Dorchester Dorset, UK. DT2 8XJ.
[2]Kaon Ltd, 5 Wey Court, Mary Road, Guildford, Surrey, UK. GU1 4QU

ABSTRACT

The UK is currently implementing the next generation of military surface ships active sonars. As part of the environmental protection work associated with the new sonars, and also in support of existing sonars, the UK Ministry of Defence has funded a programme of work to improve the capability to detect, classify and localise marine mammals. QinetiQ Ltd have been contracted to provide a software package, which can process the raw acoustic data from a number of sonar systems, including towed arrays, hull-mounted arrays and sonobuoys. The software needs to be able to adjust to the local environment and provide a cetacean presence/non-presence decision in real-time and with a very low false alarm rate. The first version, running on a standard PC, has now been completed and tested at sea during the NATO SIRENA 03 cruise. This paper describes the processing method employed and the results achieved during testing using a number of datasets.

RÉSUMÉ

Le Royaume-Uni implante présentement la prochaine génération de navires de surface avec sonar actif. En ce qui a trait au travail de protection de l'environnement associé avec les nouveaux sonars, et aussi en support aux sonars existants, le MoD a fondé un programme de travail visant l'amélioration des capacités de monitorage des vaisseaux de la marine royale afin de détecter, classifier et localiser les mammifères marins. Le QinetiQ Ltd a été mandaté afin de fournir un système pouvant traiter les données acoustiques brutes à partir d'un certain nombre de systèmes sonar, incluant les réseaux remorqués, les réseaux montés sur coque, et les bouées acoustiques. Un tel système doit pouvoir s'ajuster à l'environnement local et doit fournir une décision concernant la présence/non-présence de cétacés en temps réel avec un très faible taux de fausse alarme. La première version peut être utilisée sur un PC standard, et a été complétée et testée en mer durant la croisière expérimentale SIRENA 03. Cet article décrit la méthode d'évaluation employée et les résultats des tests en laboratoire en utilisant un certain nombre d'ensembles de données.

1. INTRODUCTION

The UK is currently implementing the next generation of military surface ship active sonars. The Ministry of Defence (MoD) has recognised the potential problems associated with the use of active sonar systems and has in place a research programme looking at minimising the possible risk to the marine environment. As part of this work QinetiQ Ltd has been tasked to provide a software package, which aims to detect, classify and localise the calls of all marine mammals. The first version of this package is known as the Marine Mammal Automated Detection System (MMADS) and is reported in this paper. Kaon Ltd were contracted by QinetiQ to write the real-time implementation. This initial version does not fully implement echolocation pulse processing and only implements the detection and classification aspects.

It is recognised that animals may not always vocalise so this package will be part of a broader system, which will integrate marine mammal detections from visual, infra-red and radar sensors and combine these with the passive acoustic detection data to provide a twenty four hour detection capability for marine mammals.

The research described here aims to detect the calls from all marine mammals. The method used extracts parameters from the detected sound that allow marine mammal calls to be distinguished from other natural and anthropogenic sounds. While the software aims to recognise calls from cetaceans, pinnipeds and sirenia, it also needs to be able to identify sounds from sonars and other anthropogenic sources in order to eliminate these sounds during the classification process.

2. BACKGROUND

Previous attempts to produce automated detection software have been reported by a number of groups. This earlier work (e.g. Potter et al. 1994; Sturtivant and Datta 1997a; Sturtivant and Datta 1997b; Mellinger and Clark 2000) looked at recognising individual species or closely-related groups of species. This work has normally been associated with behavioural research or animal censusing where it is important to be sure that the calls are from just the required species. Other workers in the field have used marine mammal calls to demonstrate the efficacy of novel signal processing algorithms (e.g.Helweg and Moore 1997; Tiemmann et al. 2001) or to demonstrate better/faster signal processing hardware (e.g.Jones et al. 1997). A number of workers have investigated the possibilities of using a number of call parameters to identify species (e.g. Wang et al. 1995; Oswald et al. 2003).

Some of the currently available software packages for cetacean acoustic research include an automated detection capability. ISHMAEL, produced by NOAA/PMEL (Mellinger 2002; Mellinger 2004), includes three forms of automated detection software. Energy summation sums the energy across all frequencies over a limited range in time and is useful for detecting echolocation pulses. The matched filter method correlates the incoming signal with a user generated reference waveform and is useful for searching for signals with little variation from pulse to pulse. The spectrogram correlation method (Mellinger and Clark 2000) cross correlates the spectrogram with a time-frequency kernel. This method is more tolerant of variability in the call than the matched filter method, but still requires that the expected signal be constrained within fairly tight limits.

RAINBOW CLICK is available from IFAW (IFAW 2004) and is used to detect and track sperm whales (Gillespie and Chappell 1998). The software processes the incoming datastreams from two hydrophones to identify sperm whale clicks and then uses the time delay between the hydrophones to estimate the bearing of each pulse. The results are then presented to an operator for manual interpretation.

WHISTLE, also available from IFAW, searches the incoming acoustic datastream for whistle calls. The contours of the whistles are then displayed to an operator and stored in a file for further analysis.

All of the above packages were designed to assist cetacean research and this constrains their potential use in a military environment where operator involvement has to be kept to the absolute minimum and the processed data cannot be further interpreted. The MMADS package is designed from the outset to require the minimum possible operator intervention and to have a simple presence/non-presence indication as its output. The first version described here still has some research facilities i.e. the spectrogram displays and the file capture capability, in order to be able to assess how well the software is performing.

3. REQUIREMENTS AND ASSUMPTIONS

The automated detection system must be able to detect all marine mammal calls while achieving low false alarm rates and a high probability of classification. If the system is to be useful in the mitigation role it must give detections and classifications that the sonar operators and command team can trust. It must therefore detect at adequate range, classify in a timely and correct manner and with a very low false alarm rate. The software must also be able to operate with a variety of sonar sensors e.g. towed arrays, hull-mounted arrays and sonobuoys. It should also be capable of multi-sensor operation with only the minimum of re-configuration

The MMADS software package implements a number of algorithms which are based on assumptions about the calls of marine mammals. These are:-

- That animals rarely vocalise just once
- That there are sufficient vocalisations to discard crossing and corrupt calls and still be able to make a correct and timely decision
- That it is possible to identify features in the calls that allow the calls to be discriminated from all other sounds in the sea.

In preparation for this work a reference set of acoustic data has been assembled covering examples of the cetacean species groups defined below. This data has been gathered from a number of military sonar sensors, from research hydrophones and from pop-up and acoustic tag deployments. Additional data was also gathered in areas where it was known that no cetaceans were present and in areas of high levels of anthropogenic noise from a variety of sources. This data was used to visually check that the above assumptions were valid, to define the algorithms described below and then to set the limits for the parametric tests in the decision-making software. Further data became available when the implementation work was well advanced and this was used to test the implemented algorithms.

4. IMPLEMENTATION

MMADS processing splits the incoming acoustic signals into five processing channels appropriate for five groups of calls. These are:-

Odontocete echolocation	15-150 kHz
Odontocete tonal	1-22 kHz
Low-frequency echolocation	1-22 kHz
HF mysticete	150-1000 Hz
LF mysticete	10-150 Hz

For the first implementation of MMADS it was decided to use readily available commercial technology resulting in the choice of standard PC hardware with the Linux operating system and the C++ programming language. The audio input card was chosen to be a SoundBlaster-compatible audio input card.

Figure 1 MMADS Hardware

The audio card limited the sample rate to 48 kHz and this meant that the odontocete echolocation pulse processing channel was severely limited in capability as most of the available energy is above the 22 kHz bandwidth. For many of the species there is still sufficient energy remaining to make a classification decision but animals like the harbour porpoise (*Phocoena phocoena*), which transmit limited-bandwidth pulses centred on 140 kHz cannot be detected by this method. A fourth processing channel optimised for echolocation calls is being investigated and will be added in a later version of the software. The 22 kHz bandwidth limit will not accept the highest frequencies in whistles from some of the small odontocetes, but this represents a very small percentage of total energy available and is not considered a significant limitation on the performance of the detector.

Figure 2 shows a typical screen display produced by MMADS. The three spectrograms correspond with the three frequency ranges used by the processing software. Alongside each spectrogram window is a vertical bar, which displays the confidence of there being marine mammal calls in that frequency range. A background scrolling text window also displays the confidence level of the five processing chains. The top window is the 1-22 kHz window and the side bar is a combination of the confidence levels from the odontocete tonal, odontocete echolocation and sperm whale processing channels. The lower left window is the LF mysticete channel and the lower right window is the HF mysticete channel.

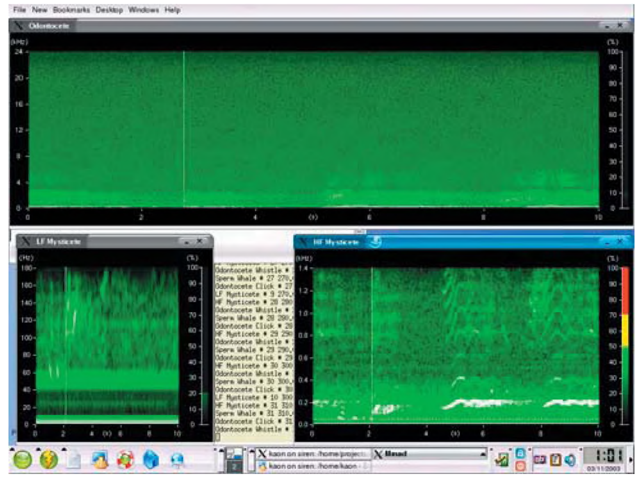

Figure 2 MMADS display

5. ALGORITHMS

5.1 Odontocete echolocation pulses

This class of signal includes the echolocation pulses emitted by animals ranging from common dolphins to killer whales. Incoming signals are digitised at a 48 kHz sample rate. If amplitude clipping is detected, the time sequence is rejected and a message generated in the text box. Good data is then transformed into the frequency domain using a 256 point FFT with Hamming window and 50% overlap. Frequency domain data is accumulated into 10 second patches. The patch is then normalised using the background mean. The resulting normalised data is converted to a binary spectrum using a fixed threshold 8 dB above the mean. The initial detection criteria is that 40% of points within the required bandwidth must exceed the threshold. The centre of the pulse is located in time using a split window test and the duration of the pulse has to be appropriate for echolocation pulses. The spectral slope is then tested to confirm the signals are echolocation pulses. A confidence level is calculated based on the number of echolocation pulses per patch.

5.2 Low-frequency echolocation pulses

For this class of signal, which is primarily clicks from sperm whales, all processing is as for the odontocete echolocation pulses except that the spectral test looks for the lower frequencies used by sperm whales.

5.3 Odontocete tonal

This class of signal includes the tonal signals emitted by animals ranging from common dolphins to killer whales. The incoming data stream, sampled at 48 kHz, is transformed to the frequency domain using a 1024 point FFT with Hamming window and 75% overlap. This spectral data is accumulated into patches ten seconds long and then normalised using a mean value offset in time. The data is converted to a binary spectrogram using a threshold 8 dB above the mean. The binary data is then searched for connected components using an 8-connectivity neighbourhood. The first test is that each connected component must have at least twenty time-frequency bins.

Each connected component is then tested to extract the following parameters:

i. Minimum frequency
ii. Maximum frequency
iii. Start frequency
iv. Stop frequency
v. Duration
vi. Bandwidth
vii. Instantaneous bandwidth
viii. Area ratio
ix. Porosity
x. Mean centre

The area ratio is the ratio of the number of time-frequency bins in the signal to total number of bins within the rectangle formed by start and stop times and minimum and maximum frequencies.

$$AreaRatio = \frac{SignalArea}{TotalArea} \quad 1$$

The porosity is the ratio of enclosed non-signal time-frequency bins to signal bins within the rectangle defined by min and max frequency and start and stop times.

$$Porosity = \frac{ZeroArea}{SignalArea} \quad 2$$

The porosity value is used to reject complex signals such as crossing tones or noisy signals.

The measured parameters are then tested against expected values for odontocete tonals. Successful classifications are accumulated across three successive ten second patches and used to calculate the confidence levels.

5.4 HF mysticete

This class of signals includes tonal sounds emitted by humpback whales and some pinnipeds. The data, sampled at 48 kHz, is low-pass filtered at 1.3 kHz and then decimated by a factor of 16 to a sample rate of 3 kHz. This data is then transformed into the frequency domain using a 1024 point FFT and 75% overlap. In each FFT, 256 points are real, the rest are zeroes. The resulting spectra are accumulated into patches ten seconds long and then normalised using a 3 second window median normaliser. The output is then converted to binary using a 10 dB threshold. The rest of the processing is then as described for odontocete tonals in 4.3 above, except that the parameter testing is optimised for this class of sounds. Some calls from the mysticetes are much broader in bandwidth than pure tonals. Provided the porosity value stays below the rejection threshold, the instantaneous bandwidth parameter allows this type of call to be classified.

5.5 LF mysticete

This class of signals includes the sounds made by fin, blue and right whales. The incoming data, sampled at 48 kHz is low-pass filtered at 140 Hz and then decimated by a factor of 128 to give a sample rate of 375 Hz. This data is then transformed to the frequency domain using a 512 point FFT with Hamming window. 64 points of real data are overlapped by 75%, the other 448 points in each FFT are set to zero. The resulting spectrogram is normalised using the median of a 5 second window with no offset and converted to a binary spectrogram using a threshold of 8 dB. The remainder of the processing is as described for the odontocete tonals in paragraph 4.2 above.

5.6 Operator display/control

The package also produces a display so that the operator can monitor the output from the MMADS software. Figure 2 shows a typical display. In addition, it is possible to automatically capture the raw audio into files of type WAV whenever a preset confidence level is exceeded. To allow evaluation of the algorithm there is a facility to capture the partially processed data in a form suitable for importing into MATLAB. This was used to produce the displays in Figures 3-8.

6. KNOWN DEFICIENCIES

Testing of the software against the reference dataset and using the NATO SIRENA 03 and DRDC/Dalhousie datasets has suggested that this initial version of the MMADS software has a number of deficiencies.

The algorithm to detect sperm whales was designed around a limited dataset obtained by using sonobuoys. Data subsequently obtained from wideband hydrophones suggests that the implemented algorithm, optimised for use with sonobuoys suffers from an unacceptably high false alarm rate on higher bandwidth data. This will be addressed in the next version of the software.

The odontocete echolocation pulse detector is missing the majority of available energy in the pulse. This means that the detection range for these pulses is significantly reduced compared with that achievable using the full bandwidth. The limitation is due to the choice of 48 kHz A/D converters. Use of the next generation of A/D cards operating at 192 kHz would alleviate this problem, although it would still not work with harbour porpoise pulses.

The odontocete tonal detector works well with animal calls and has successfully detected calls from a range of animals. However, when used in a sonar environment it can be partially activated by complex transmit sequences. A processing channel which identifies sonar waveforms and removes them from the input to the odontocete tonal classifier is currently being investigated.

The testing of the HF mysticete channel has not been as extensive as the authors would like due to the lack of good quality data. Alternative sources of data are currently being sought to allow this testing to continue.

The LF mysticete channel was tested extensively against fin whales (*Balaenoptera physalus*) during the NATO SIRENA 03 cruise. This revealed an unexpected problem. The algorithm worked well with individual calls but when a group of animals started calling in the confines of the Ligurian Sea the reverberation levels built up to the point where the normaliser was suppressing all of the calls. The algorithm will need to be modified to ensure the normaliser does not cancel out continuous calling.

7. MMADS TEST RESULTS

DRDC and Cornell University issued two acoustic datasets gathered using autonomous seabed recorders for use by participants in *the Workshop on Detection and Localisation of Marine Mammals using Passive Acoustics* held in Halifax, Nova Scotia, in November 2003. These datasets featured right whale and fin whale calls.

The data was read into the ISHMAEL software to extract single channel data from the multi-channel data supplied. The single files were then processed using MMADS in file-input mode to process the data. The processed data is then extracted at three stages through the processing chain and displayed using MATLAB. The figures that follow illustrate this processing of the data. In each group of three figures the first figure is the raw spectrogram of the data. The second figure is the spectrogram overlaid with the detected connected-components. The third figure is the spectrogram overlaid with the connected components that the decision-making process has chosen as originating from marine mammals.

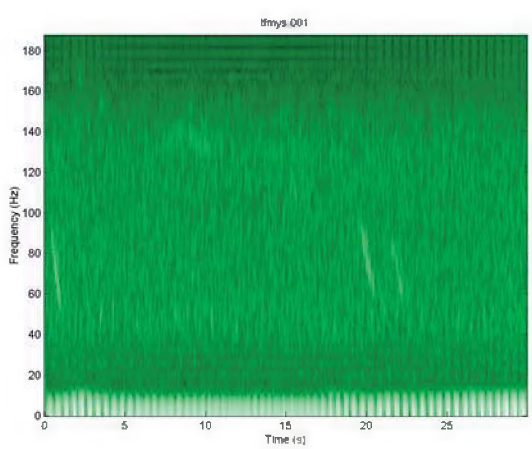

Figure 3 Raw spectrogram

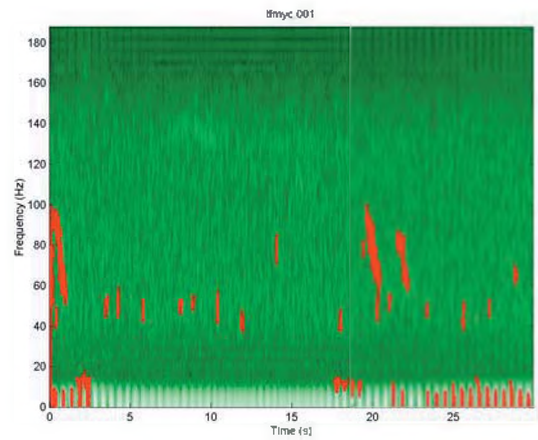

Figure 4 Connected components overlay

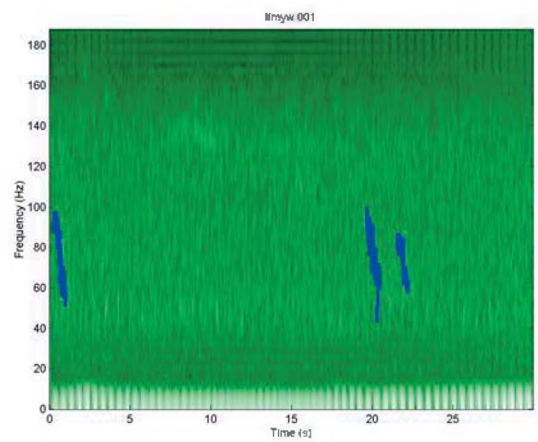

Figure 5 processed data

The white band at low frequencies appears to be cable strumming. The three pictures are visually inspected to determine a) how many animal calls there are in the raw spectrogram, b) how many animal calls are identified in the connected component (C-C) scan, c) how many non-animal sounds are identified in the C-C scan, d) how many animal call C-C are classified as marine mammals and e) how many non-animal C-C are classified as marine mammals. The data was processed using the LF Mysticete channel and both right whale and fin whale calls were processed. The table below includes the calls from both species.

Referring to the categories of detection (a-e) above, the results for a number of files from the Cornell data set are shown in Table 1 below.

The majority of the DRDC dataset was not suitable for processing through MMADS because the files were too short. Only one file is long enough to process and this is the L138b file. Again using the same detection categories (a-e) defined above the results for this file are shown in Table 2 below.

	Detection categories				
Data file	a	b	c	d	e
GSC0420	33	33	250	19	4
CCB0800	46	46	196	12	0
CCB0810L	37	36	182	16	1
CCB0810H	65	65	351	23	2
CCB0825	44	44	178	25	1
GSC0650	20	20	183	14	0
Totals	245	244	1340	109	8

Table 1. Analysis of Cornell data.

	Detection categories				
Datafile	a	b	c	d	e
L138b	168	128	276	68	22

Table 2. Analysis of DRDC data.

From these results it can be seen that just over 40% of calls are classified correctly while 3% of classified calls are false for the Cornell dataset and 10% for the DRDC data set.

8. SIRENA 03 DATA

During the NATO SIRENA 03 cruise in August/September 2003, the MMADS system was deployed and used to process hydrophone data from the SACLANTCEN towed array and the University of Pavia two-hydrophone array. This provided an opportunity to test the system against fin whales, sperm whales, striped dolphins and common dolphins. This data has yet to be fully analysed but Figures 6 - 8 show the sequence of data for a striped dolphins call. The vertical bars are echolocation pulses but in this example only the output from the odontocete tonal channel is shown.

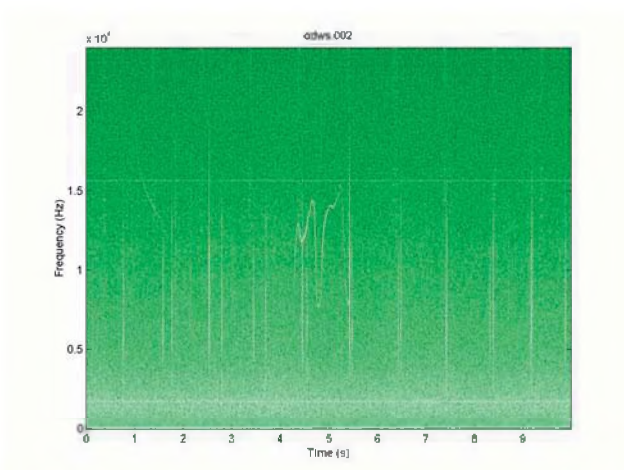

Figure 6 SIRENA 03 data spectrogram

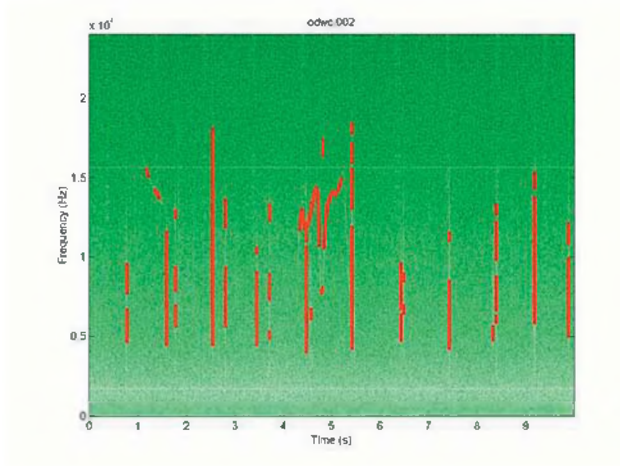

Figure 7 SIRENA 03 Connected components

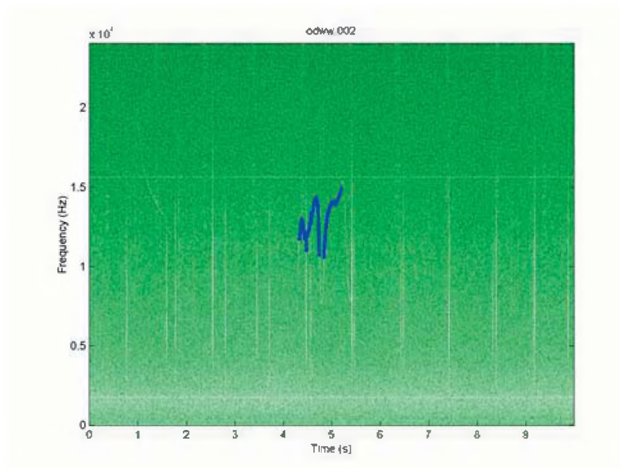

Figure 8 SIRENA 03 classified components

9. DISCUSSION

This work has demonstrated that algorithms based on classic spectral mapping techniques can provide a useful automated system for the detection and classification of marine mammal calls. Although the present system is constrained in performance by the limited reference data set used to choose the parameter tests, the system has still performed well in laboratory and at-sea testing. The 40% success at classifying calls is well within acceptable limits but the false alarm rate at up to 10% needs to be further reduced. The processing within the MMADS package to combine the individual call detections to form confidence levels provides this false alarm reduction but this has not been tested in detail yet and will be reported separately. It is still desirable to improve the individual call classification processing to reduce the load on the confidence calculation. This area will be addressed during the next phase of the work.

The reference data set has been expanded considerably since the MMADS processing was designed and will now be reviewed to ensure the parameter tests can provide the correct classification for an increased range of species.

The current software detection and classification is based on parametric extraction from single calls, be they tonals or echolocation pulses. In the case of sperm whales this leads to high false alarm rates because of all the other sounds, both natural and anthropogenic, having very similar characteristics to sperm whale pulses. The main thrust of future work will investigate the possibilities of context processing in which the whole ensemble of calls is examined to further characterise the signals and enhance rejection of non-animal sounds. The acoustic background will also be examined to ensure there are no frequency or time effects, which could skew the classification results. To be useful an automated system needs to know how well it will perform in order to warn the user when the acoustic environment may adversely affect its ability to detect marine mammals.

If a very high quality reference dataset is available, such as those obtained by tagging studies or work with captive animals, and the acoustic environment parameters, such as sound velocity profile, surface roughness etc, are known then it becomes possible to refine the classification testing by predicting the characteristics of the calls likely to be received after propagation through the medium. This could include time dispersion due to multi-path or frequency dependent velocity of propagation, and frequency dependent absorption. This could significantly improve detection performance in some acoustic environments.

10. ACKNOWLEDGEMENTS

The authors would like to acknowledge the sources of acoustic data used in the reference dataset employed in the course of this work. These include the UK Ministry of Defence, Woods Hole Oceanographic Institute, USA, University of Pavia, Italy, DRDC, Canada, Cornell University, USA, SACLANTCEN, Italy and the Durlston Marine Mammal Research Project, UK. This work was funded by the UK Ministry of Defence as part of its Environmental Protection Programme within the Applied Research Programme (Entity 25).

We also acknowledge the helpful comments on the manuscript from Graham Jackson of Dstl, UK and two anonymous reviewers.

11. REFERENCES

Gillespie D, Chappell OP (1998) Automated cetacean detection and monitoring. In: Tasker ML, Weir C (eds) Proceedings of the seismic and marine mammals workshop, London, 23-25 June 1998. Sea Mammal Research Unit, St Andrews.

Helweg DA, Moore PWB (1997) Classification of aspect-dependent targets by a biomimetic neural network. Naval Command, Control and Ocean Surveillance Center, San Diego

IFAW (2004) Rainbow Click. IFAW, www.ifaw.org/ifaw/general/default.aspx?oid=35704.

Jones J, Browne RW, Di Meglio A, Wang LS (1997) Dolphin vocalisation analysis using an ADSP21020. Institute of Electrical Engineers, London.

Mellinger D (2004) Ishmael. PMEL/NOAA, http://cetus.pmel.noaa.gov/cgibin/getinfo.pl?dirname=ishmael

Mellinger DK (2002) Ishmael 1.0 User's Guide. NOAA/PMEL, OAR PMEL-120, Seattle, USA.

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

Oswald JN, Barlow J, Norris TF (2003). Acoustic identification of nine delphinid species in the Eastern Tropical Pacific Ocean. Marine Mammal Science 19: 20-37.

Potter JR, Mellinger DK, Clark CW (1994) Marine mammal call discrimination using artificial neural networks. Journal of the Acoustical Society of America 96: 1255-1262.

Sturtivant CR, Datta S (1997a) Automatic dolphin whistle detection, extraction, encoding and classification. In: Goodson AD, Harland EJ (eds) Underwater bio-sonar and bioacoustics symposium. Institute of Acoustics, Loughborough University, pp 259-266.

Sturtivant CR, Datta S (1997b) Dolphin whistle classification with the "dolphin" software. In: Goodson AD, Harland EJ (eds) Underwater bio-sonar and bioacoustics symposium. Institute of Acoustics, St Albans, pp 149-156.

Tiemmann CO, Porter MB, Frazer LN (2001) Automated model-based localisation of marine mammals near Hawaii. In: Holland RC (ed) OCEANS 2001. Marine Technology Society, Washington, pp 1395-1400.

Wang D, Wursig B, Evans W. (1995) Comparison of whistles among seven odontocete species. In: Kastelein RA, Thomas JA, Nachtigall PE (eds). Sensory systems of aquatic mammals. De Spil Publishers, Woerden, The Netherlands. ISBN 90-72743-05-9. pp 299-323.