DETECTION AND CLASSIFICATION OF RIGHT WHALE CALLS USING AN 'EDGE' DETECTOR OPERATING ON A SMOOTHED SPECTROGRAM

Douglas Gillespie

Song of the Whale Research Team,
International Fund for Animal Welfare
87-90 Albert Embankment
London, SE1 7UD
UK.

ABSTRACT

A detector has been developed which can reliably detect right whale calls and distinguish them from those of other marine mammals and industrial noise. Detection is a two stage process. In the first, the spectrogram is smoothed by convolving it with a Gaussian kernel and the 'outlines' of sounds are extracted using an edge detection algorithm. This allows a number of parameters to be measured for each sound, including duration, bandwidth and details of the frequency contour such as the positions of maximum and minimum frequency. In the second stage, these parameters are used in a classification function in order to determine which sounds are from right whales. The classifier has been tuned by comparing data from a period when large numbers of right whales were known to be in the vicinity of bottom mounted recorders with data collected on days when it was believed, based on ship and aerial surveys, that no right whales were present. Overall, the detection system is capable of picking out a high proportion of right whale calls logged by a human operator, while at the same time working at a false alarm rate of only one or two calls per day, even in the presence of background noise from humpback whales and seismic exploration. Although it is impossible to reduce the false alarm rate for individual calls to zero whilst still maintaining adequate efficiency, by requiring the detection of several calls within a set waiting time, it is possible to reduce false alarm rate to a negligible level.

SOMMAIRE

Un détecteur capable de déceler avec efficacité les vocalisations de baleines franches et de distinguer ces dernières des autres mammifères marins et du bruit industriel, a été développé. La détection se fait en deux étapes. En premier lieu, le spectrogramme est lissé par convolution avec un noyau de distribution gaussienne et le « contour » du son est extrait en utilisant un algorithme de détection d'arête. Pour chacun des sons, ceci permet de mesurer un certain nombre de paramètres incluant la durée, la largeur de bande et les détails sur les contours de fréquence, telle la position des fréquences maximale et minimale. Dans un second temps, ces paramètres sont utilisés lors de la fonction de classification, dans le but de déterminer quels sons proviennent des baleines franches. Le classificateur est optimisé en comparant les données découlant d'une période où il était établi qu'un large nombre de baleines franches était à proximité des appareils d'enregistrement ancrés sur le fond marin, avec des données recueillies les jours où il était plausible, basé sur des sondages maritimes et aériens, qu'aucune baleine franche n'était présente. Ce système de détection est capable de choisir une large proportion de vocalisations de baleines franches consignées par un opérateur, tout en opérant avec un taux de fausse alarme d'une ou deux vocalisations par jour et cela même en présence de bruit de fond provenant des baleines à bosses et de l'exploration sismique. Il est impossible de réduire à zéro le taux de fausse alarme pour chacune des vocalisations tout en maintenant une efficacité adéquate. Cependant, en imposant la détection de plusieurs vocalisations à l'intérieur d'un temps d'attente pré déterminé, il est possible de réduire le taux de fausse alarme à un niveau négligeable.

1 INTRODUCTION

The problems facing the North Atlantic right whale (*Eubalaena glacialis*) are well documented (IWC 2001). Although once widely distributed throughout the North Atlantic, only a remnant population of approximately 300 individuals survives. The known habitats of the North Atlantic right whale are all along the Eastern seaboard of the United States and Canada, the feeding and breeding grounds and the migratory routes between them coincide with major ship routes and important fishing grounds. It is believed that the North Atlantic right whale will become extinct within approximately 200 years (Caswell *et. al.*, 1999; Fujiwara and Caswell 2001) unless steps can be taken to reduce anthropogenic mortality due to collisions with ships and entanglement in fishing gear. Despite considerable efforts to better manage these activities in order to protect right whales, efforts are hampered by the lack of reliable, up to date, surveillance data in the areas where right whales are most at risk. Current survey methods rely primarily on the use of light aircraft which are expensive to run, require large numbers of personnel and cannot operate effectively during inclement weather. Passive acoustic monitoring has been proposed as a tool that could provide the information required for effective management action (IFAW, 2001).

Right whales make a variety of vocalizations (Clark, 1983). The work described in this paper is concerned solely with detection of one of the most commonly heard sounds from right whales, the frequency modulated (fm) up-sweep or 'contact call'. These are typically about a second long and sweep upwards in frequency between approximately 100 and 200 Hz. There is however considerable variability between individual upsweep calls, examples of which are shown in **Figure 1**. Vocalization rates of North Atlantic right whales are highly variable and individuals have been known to remain silent for several hours (Matthews *et. al.*, 2001).

Detection and accurate classification of right whale sounds was complicated by the wide variety of different sources of background noise present in the study area. As well as ships, either on passage or engaged in fishing activities, sounds from distant seismic exploration and other species of marine mammal were regularly heard on recordings made in areas frequented by right whales. The frequency range of many of these background noises overlaps that of right whale sounds. The most similar sounds to those of right whales encountered in this study, and the ones causing the greatest problem in classification, were found to be those of humpback whales (*Megaptera novaeangliae*). The problem was exacerbated by the fact that humpbacks are more numerous than right whales and also appear to vocalise more often and to be louder than right whales, so for every right whale vocalisation detected, it was necessary to avoid potential false detections from many thousands of humpback calls.

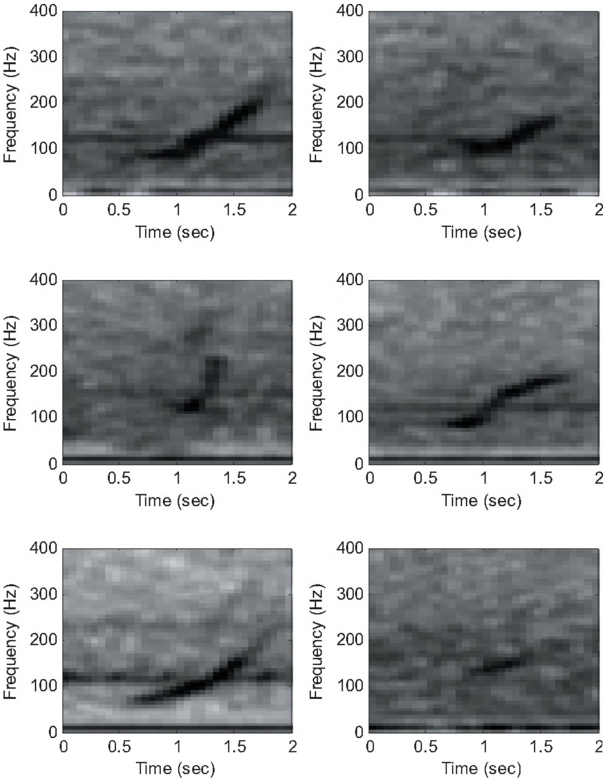

Figure 1. Example spectrograms of right whale upsweep calls selected by a human operator.

A good detection algorithm should be sensitive to up-sweeping sounds of varying sweep rate, but at the same time detect differences between the time-frequency contours of sounds from right whales and those from other sources. For the algorithm to be useful, it must work at a known efficiency and false alarm rate. Clearly it is desirable to maximise the former and minimise the latter. The coincidence in frequency of background noise sources means that simple energy detectors would have a high false alarm rate. Conversely, the considerable variation in sweep rate of the right whale calls is likely to cause detectors based on correlation techniques using a fixed template in either time or frequency to have low efficiency.

The algorithm described here used a detector which was sensitive to any type of sound rising above a predetermined threshold. The output of the detector was *edges* of the sound in time and frequency. From those edges, and the contour of maximum amplitude between them, a number of parameters were measured which were used in a statistical classifier to correctly identify right whale sounds.

2 METHODS

2.1 Data collection

Data were collected using bottom mounted recorders ('Pop-Ups') developed by Cornell University. They include a hydrophone, a microprocessor, a computer hard drive and a

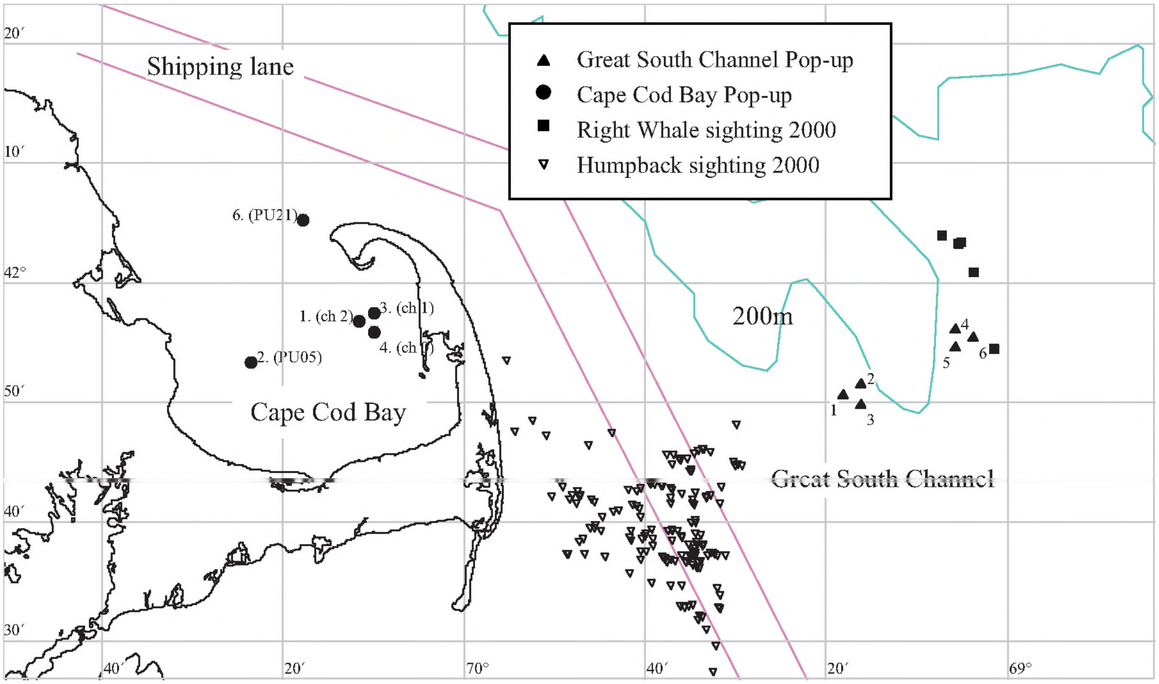

Figure 2. The locations of pop-up recorders in the Great South Channel 2000 and Cape Cod Bay 2001. Right Whale and humpback sightings form aerial surveys in 2000 are also shown. For clarity, sightings from 2001 in Cape Cod Bay have not been shown.

release mechanism. The units used in the study were capable of recording continuously to the hard disk for approximately 30 days. Each pop-up was moored two metres from the sea bottom to which it was attached using a disposable anchor (biodegradable sacks of gravel and sand). At the end of the recording period, an acoustically transmitted command from the surface caused the units to separate from their anchors and return to the surface. Data used in this study were collected using six pop-ups deployed in two groups of three at approximately 200m depth in the Great South Channel between 13 May and 12 June 2000 and three units deployed in a triangular configuration at the eastern side of Cape Cod Bay at 30m between 8 March and 2 April 2001 (**Figure 2**). Once on shore, data from the pop-ups were combined into six and three channel sound files for 2000 and 2001 data respectively. Synchronisation between channels was achieved by dropping light bulbs in 2000 (Marshall, 1993) and by playing FM sweeps from an underwater speaker in the vicinity of the 2001 pop-ups. All recordings were made at a sampling rate of 2000 Hz.

All recordings were browsed by a human operator viewing a spectrogram of the multi channel sound files and also listening on headphones whenever necessary. In 2000, the operator logged right whale calls on all six channels, but in 2001 only the loudest or the first occurrence of each sound was logged when it was observed on more than one channel.

Aerial surveys were flown over the Great South Channel in 2000 by NOAA Fisheries and over Cape Cod Bay in 2001 by the Center for Coastal Studies. The primary motivation behind these surveys was to provide information for ship and fisheries management and to collect identification photos of individuals for long term population monitoring. The surveys were not concentrated on the precise locations of the pop-ups and could only take place during relatively calm conditions. The surveys did not therefore provide constant or even regular coverage over the pop-ups but could provide a general overview of the presence / absence of right whales during a pop-up deployment.

The 2000 aerial surveys found few right whales in the vicinity of the pop-ups. The only occasion on which right whales were spotted within ~10km of the pop-ups occurred on the morning of 26 May, 2000. The surveys did however spot a considerable number of right whales between 38 and 140 km NW of the pop-ups and also found large numbers of humpback whales ~20km to the SW, close to the shipping lanes to the East of Cape Cod. The Cape Cod Bay surveys indicted that large numbers of right whales were present in the Bay throughout the 2001 deployment period.

The detector and classifier were tuned using two days of data from each location, having no right whale calls and several thousand right whale calls respectively. Once tuning had been completed, the detector / classifier was used to analyse the entire data set.

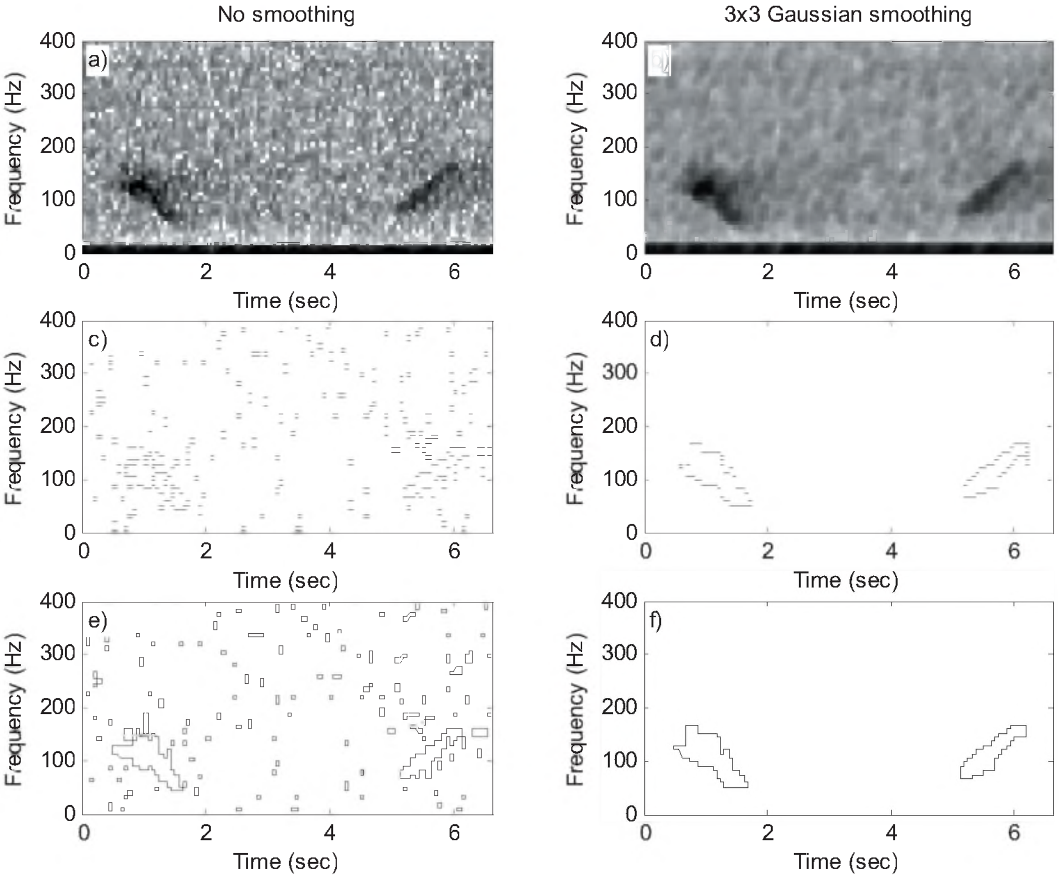

Figure 3. Sound Detection: a) Spectrogram containing two calls, one of which is a right whale upsweep; b) Spectrogram after Gaussian smoothing; c) Edge detection without Gaussian smoothing; d) Edge detection with Gaussian smoothing; e) sound outlines without Gaussian smoothing; f) sound outlines with Gaussian smoothing.

2.2 Sound Detection

The detection stage of the algorithm is not optimised to be any more sensitive to right whale sounds than any other type of sound. This lack of optimisation is important since, if the detector were optimised to only detect up-sweeping signals, it is possible that it may select up-sweeping parts of sounds having more complicated time-frequency contours and thereby create false detections.

Sounds were detected by searching for 'edges' in a spectrogram matrix and linking edges together to form the outlines of sounds. A number of edge detection algorithms of the type used in image analysis were tested, a simple threshold detector, which was found to have the best overall performance, in terms of efficiency verses false alarm rate, is described here.

Spectrogram smoothing

The power spectrogram S of the data was first calculated with a frame length of 256 samples (128ms) giving a frequency resolution of 7.8Hz. A Hanning window function was used and successive frames overlapped by 131 samples to give a frame separation in time of 1/16 second. These values were chosen so that the spectrogram had approximately equal resolution in time and frequency, i.e. a typical right whale upsweep spanned 13 bins in frequency and 16 in time.

A common technique used in image edge detection is smoothing of the image matrix by convolving it with a Gaussian kernel (Embree and Kimble, 1991, Sonka *et. al.*, 1999). This has the beneficial effect of preventing edges breaking up into many parts, but also has the detrimental effect of reducing the resolution of the image if the smoothing kernel is too large. In this study a 3×3 smoothing kernel was used to compute the smoothed spectrogram $S' = S * G$, where

$$G = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}.$$

Edge detection

The edges of sounds were detected using a simple threshold detector where the signal at any point in the spectrogram $S'_{(t,f)}$ is compared with a background measurement $B_{(t,f)}$.

$B_{(t,f)}$ was continuously updated and computed independently for each frequency using

$$B_{(t,f)} = B_{(t-1,f)} + \left(\frac{S'_{(t,f)} - B_{(t-1,f)}}{\alpha} \right),$$

where α is the time constant for the background update, thereby allowing the detector to respond to changes in noise level such as would be caused by a passing ship.

Regions of the spectrogram were over threshold if

$$\frac{S'_{(t,f)}}{B_{(t-1,f)}} > Th,$$

where the threshold *Th* was set to 4 (6 dB).

Since the background measurement $B_{(t,f)}$ would tend to increase in the presence of any sound rising above the mean noise level, two different values of α were used in the background calculation – a high value (160), giving a long time constant (~10s) when the signal at any given frequency was above threshold and a lower value (16, giving a time constant of 1s) when the signal was below threshold.

Table 1. Data used in classifier training.

	Right Whale Sample	Non-Right Whale Sample
Location and Date	Cape Cod Bay, 16-17 March 2001	Great South Channel, 16-17 May 2000
Total Number of Detections	44,672	177,080
Total Number of Candidate upsweeps	6,294	19,098
Human Operator detected upsweeps	2077	0
Human upsweeps also found by detector	1879	0

Regions of the spectrogram in successive time slices which are adjacent or overlapping in frequency were then joined together to form a 'sound'. The edges of the sound were then the frequency at which the smoothed spectrogram rose above threshold and the frequency at which it fell back below threshold in each time frame. Rules were built into the joining process which allowed gaps of a single time frame containing no data above threshold within a sound. This helped prevent sounds which rose and fell above threshold from breaking into several parts.

The complete sound detection process is shown in **Figure 3** where the benefit of Gaussian smoothing (subplots b,d,f) compared to the raw spectrogram (subplots a,c,e) is clear.

2.3 Sound Parameterisation

Once a sound was detected, it was described by a relatively small number of parameters as listed below and shown in **Figure 4**.

1. Duration
2. Start Frequency[1]
3. Minimum Frequency[1]
4. Sweep Frequency (Maximum Frequency minus Minimum Frequency)[1]
5. Position of minimum frequency
6. Position of maximum frequency
7. Maximum instantaneous bandwidth (between the lower and upper frequency bounds of the sound outline)

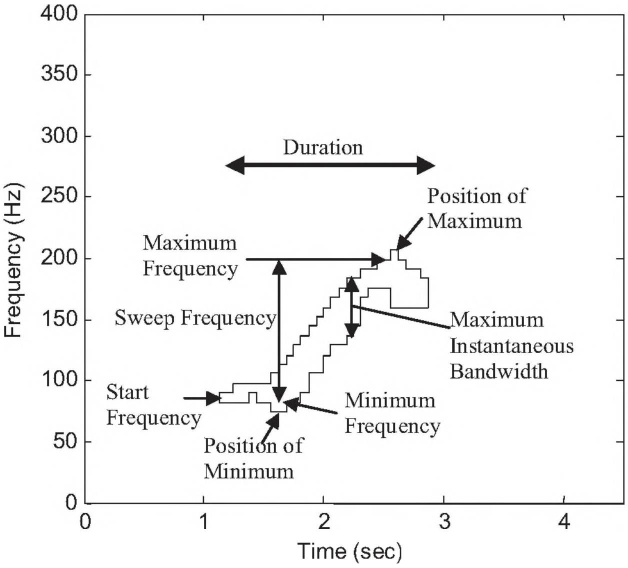

Figure 4. Parameters extracted from each sound for use in classification.

---

[1] Frequencies were taken as the frequency of maximum amplitude within each spectrogram time frame.

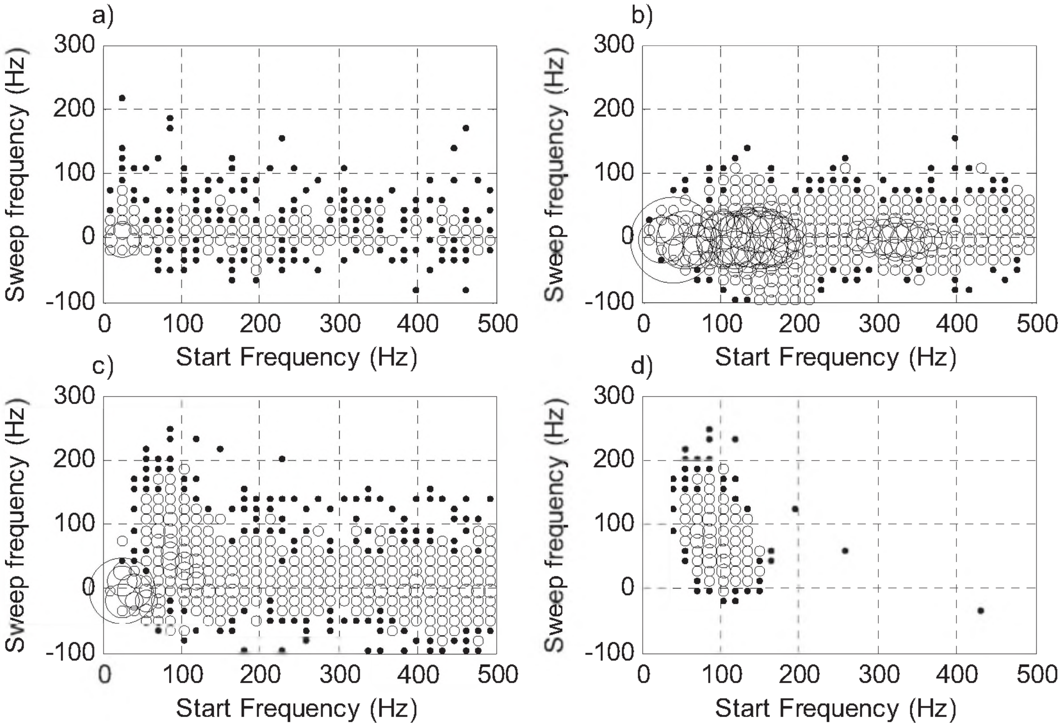

Figure 5. Sound start and sweep frequency distributions. a) A quiet day in the Great South Channel; b) a day in the Great South Channel when humpback whales are present; c) a day in Cape Cod Bay when both right whales and humpback whales are present; d) human operator selected calls from Cape Cod Bay. The area of each circle is proportional to the number of calls.

2.4 Sound Classification

Since there is a possibility that the human operator had missed genuine right whale calls and since the operator had only marked calls on a single channel, data from Cape Cod Bay, when many right whales were present, were not used to measure false alarm rate. Instead, the Cape Cod Bay data was only used to measure detection and classification efficiency and false alarm rate was measured using data from the Great South Channel from a period when no right whales were seen and none detected on the spectrograms by the operator. Two days of data from each location were used for classifier tuning.

The classifier was realised firstly by selecting right whale-like sounds using the selection criteria listed in Table 2. Final separation of right whale and non right whale sounds was then carried out with a multivariate discriminant analysis function utilising four of the parameters (start frequency, sweep frequency, duration and maximum instantaneous bandwidth) measured from the sounds which had passed the initial selection. Right whale sounds were then selected by choosing an appropriate cut on the first canonical variable resulting from the discriminant analysis.

The primary motivation for developing this algorithm is for use in dynamic ship management systems to mitigate against ship strikes. For such a management system to be effective, the false alarm rate must be extremely low if it is to be accepted both by the industry and national / international regulators. Classifier tuning was carried out with this in mind by tuning it to detect the highest possible number of true right whale sounds for a maximum false alarm rate of 1 – 2 calls per pop-up per day.

3 RESULTS

Table 1 shows the total numbers of sounds detected from the Great South Channel and Cape Cod Bay data on the two days used to tune the detector and classifier. Of 2077 right whale upsweeps detected by a human operator, the detector found 1897 (90%). Of these 1897 calls, the measured parameters showed that only 1753 (84%) swept up in frequency by at least 7 Hz (more than 1 frequency bin). It is believed that this is due to errors in the detection and parameterisation process occurring at low signal to noise ratio rather than errors on the part of the operator. 19,098 upsweeps were detected on the two days of Great South Channel data.

The numbers of detected sounds in the Great South Channel was considerably higher than that in Cape Cod Bay. This is primarily due to the presence of humpback whales, but airgun arrays used in seismic surveys are also audible on the Great South Channel recordings. **Figure 5** shows distributions of two of the parameters (start frequency and

Table 2. Selection criteria applied to calls before the multivariate discriminant analysis

	Loose	Medium	Tight
Minimum Duration	>=0.5 s	>=0.5s	>= 0.5 s
Maximum Duration			< 2 s
Sweep Frequency	>= 7 Hz	>= 23 Hz	>= 54 Hz
Start Frequency		50 ñ 160 Hz	50 ñ 160 Hz

sweep frequency) describing detected sounds for the different data sets. The upsweeping right whale calls are clearly visible on the distributions in subplots c and d. However, it is also clear from **Figure 5** that there is an overlap in the distribution of right whale call parameters and those of humpback whales.

Figure 6 shows plots of combined right whale detection and classification efficiency against the number of false alarms from non-right whale sounds for varying cuts on the canonical variable from the discriminant analysis. If no pre-selection of calls was made, the classifier performance was poor, particularly at low false alarm rates. If only sounds which started at between 50 and 160Hz, and swept through at least 23 Hz were selected, detector performance improved. It was found that the best detector performance at a false alarm rate of 1-2 calls per pop-up per day could be obtained by making the 'tight' pre-selections listed in Table 2. In this case, the algorithm correctly detected and classified approximately 60% of human detected calls.

Figure 7 shows the number of calls classified as right whale every 4 hours in the Great South Channel in 2000 using the detector operating point shown in **Figure 6**. Significant numbers of right whale calls were only detected between 0400 and 0800 UTC (0000 to 0400 local time) on 26 May. Obviously, it was impossible that the aerial surveys would have spotted them at that time, but right whales were seen at the locations shown in **Figure 2** later that morning, close to the three pop-ups on which the calls were detected.

4 DISCUSSION

The critical parameters describing any detection system are its efficiency and false alarm rate. False alarm rate was measured using data from a period when no right whales appeared to be present. If false alarm rate had been measured using data when right whales were present, counting calls found by the detector and not by humans as false alarms, it is possible that an artificially high false alarm rate would have resulted from the operator having missed genuine calls. Since the number of calls produced can never be know precisely, efficiency can only be measured relative to that of a human. In spite of this, measurements of efficiency and false alarm rate made in this way can still be used to compare and optimise detectors.

The detector and classifier described here are capable of finding right whale sounds with a reasonable efficiency (~60%) while at the same time achieving a false alarm rate of 1-2 calls per pop-up per day, even in the presence of many tens of thousands of sounds from humpback whales and seismic exploration which are in the same frequency band as the right whale calls.

To obtain a reasonable detection efficiency it is not possible to reduce false alarm rate to zero. For many applications, such as management of shipping, detection of whales or

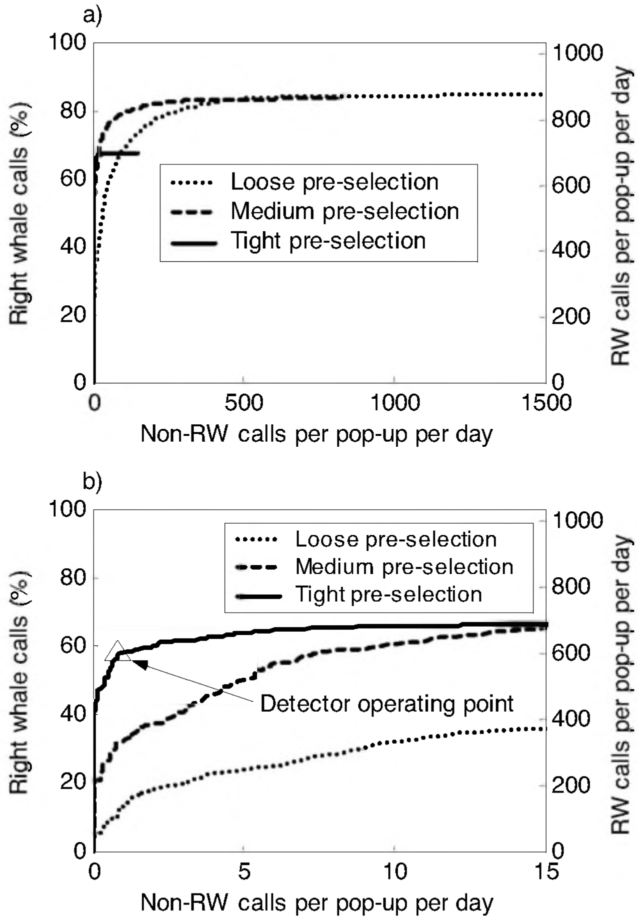

Figure 6. Efficiency / false alarm rate plots for different pre-selections of the calls (see table 2). Each plot shows the percentage and the number of detected right whale calls plotted against the number of false detections a) is scaled to show the full range of numbers of false detections, b) is scaled to show only the region of the curves in a) which are of interest, i.e. false alarm rates of only a few calls per pop-up per day.

groups of whales is of more interest than the individual calls. It is therefore possible to reduce false alarm rate to a negligible level by requiring a minimum number of sounds within a given waiting period. If false alarms are randomly distributed in time, then the number of calls which can be expected within a given waiting time is described by the Poisson distribution. As an example, if the false alarm rate were 10 calls per 24 hour period, then the probability of receiving 10 or more false calls in a one hour period is approximately 10^{-12}.

Even though human observers are more efficient than the automatic system, the automatic system has the advantage over the human that it is more objective and will not be affected by inter or intra observer variability. Although not impossible, the tasks of manually analysing many months of data from bottom mounted recorders is an onerous one and the bigger the dataset, the more likely it is to require more than one observer to analyse it. On the other hand, the adaptability of human observers may make them less likely to become confused by an unexpected sound which was not present in the data used to tune the detector and classifier.

The current classification system analyses each sound in isolation. It does not use other available information such as the rate of call production or the presence of other types of sound. If this 'contextual' information were used, it should be possible to make the classifier adapt, using stricter criteria when sources such as humpback whales are known to be present and less strict criteria on quiet days.

The current classification system relies on a multivariate discriminant analysis. Such analysis assumes that the parameter distributions are Gaussian and is only optimal if this is the case. A Neural Network using the same parameters (start frequency, sweep frequency, etc.) describing the sounds as its input may give better performance. A preliminary investigation showed this not to be the case, although further studies are planned for the

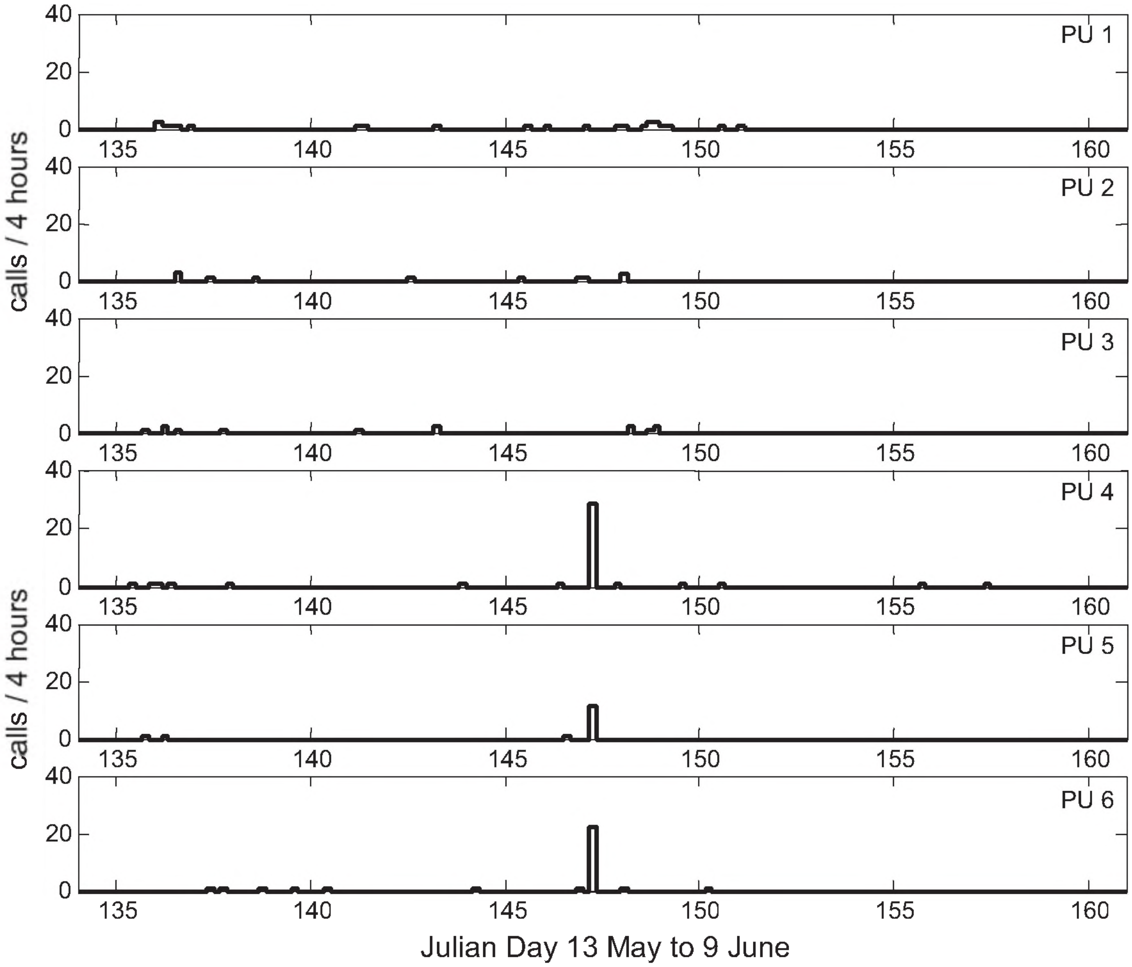

Figure 7. 4 hour call counts from pop-ups 1 to 6 deployed in the Great South Channel in 2000.

future.

Although the classifier described here has been tuned to detect only a certain type of right whale sound, the detector is designed to pick out and measure parameters of any sounds rising above threshold (section 2.2; **Figure 3**). Developing additional classifiers for other types of right whale sound, or sounds from other species, should be a relatively straight forward task if sufficient training data can be obtained. When using the detection algorithm with other sounds, careful consideration should be given to the FFT length and time frame overlap in order to optimise the quality of data provided to the classifier. Classifiers for more complex sound types, which may have a number of inflexions in their time frequency contour may also require the extraction of a different set of descriptive parameters for each sound.

5 ACKNOWLEDGEMENTS

Funding for this research was provided by the International Fund for Animal Welfare. Pop-up recorders were provided by Cornell Lab of Bioacoustics and I am most grateful to Dr. Christopher Clark the director of the lab and to Melissa Fowler who did all of the human browsing of sound data. Aerial survey data was provided by Pat Gerrior of NOAA Fisheries and by Moira Brown of the Center for Coastal Studies. I also wish to thank two anonymous reviewers for their comments on an earlier draft of this paper.

6 REFERENCES

Caswell, H., Fujiwara, M. and Brault, S. 1999. Declining survival probability threatens the North Atlantic right whale. *Proc. Natl. Acad. Sci* 96: 3308-3313.

Clark, C.W. 1983. Acoustic communication and behavior of the southern right whale. In: R.S. Payne (ed.), *Behavior and Communication of Whales*. *Westview Press*: Boulder, CO. pp. 163-198.

Embree, P., and Kimble, B., 1991. C Language Algorithms for Digital Signal Processing. *Prentice Hall*, NJ. pp386-388

Fujiwara, M. and Caswell, H. 2001. Demography of the endangered North Atlantic right whale. *Nature* 414: 537-541.

IFAW, 2001. Report of the Workshop on Right Whale Acoustics: Practical Applications in Conservation. (eds. Gillespie, D. and Leaper, R.) *International Fund for Animal Welfare*, Yarmouth Port, MA 02675, USA. 23 Pages.

IWC. 2001. Report of the workshop on status and trends of western north Atlantic right whales. *J Cet. Res. Manage Special Issue* 2: 61-87.

William J. Marshall, Jr. 1993, Acoustic signatures of imploding underwater light bulbs, JASA 94:3, p. 1845

Matthews, J., Brown, S., Gillespie, D., Johnson, M., McLanaghan, R., Moscrop, A., Nowacek, D., Leaper, R., Lewis, T., and Tyack, P.. 2001. Vocalisation rates of the North Atlantic Right Whale. . *J. Cet. Res. Manage* 3(3):271-282.

Sonka, M, Hlavac, V., and Boyle, R., 1999. Image Processing, Analysis and Machine Vision. 2nd ed. PWS Publishing, CA. p445.