DETECTION OF FREQUENCY-MODULATED CALLS USING A CHIRP MODEL

Justin Matthews
International Fund for Animal Welfare,
87-90 Albert Embankment, London, SE1 7UD, UK
whalesong@ifaw.org

ABSTRACT

Many cetacean vocalisations are tonal and most are frequency-modulated. The detection algorithm presented here breaks the frequency contour into a sequence of elements. Each element is sufficiently short that a linear approximation to the frequency contour can be made. In this way the problem is simplified from that of detection of an unknown signal, to the detection of a known signal (a linear chirp) with unknown parameters. The method of estimation is based on maximum likelihood, and the start frequency, chirp rate and amplitude of each element are estimated. Further analysis is then carried out on groups of concatenated chirps (i.e. calls) to classify them.

Results are given on performance for the supplied test recording and for synthetic signals in white noise. The pros of the algorithm are: good detection performance, at least in white noise; high resolution; ease of interpretation; flexibility; data compression. The cons are: computational cost; deterioration of performance in non-white noise or with amplitude-modulated signals. Further development is needed to reduce errors with overlapping tonal or non-tonal signals. The algorithm is currently being applied to the problem of detecting right whale vocalisations and distinguishing them from those of humpback whales.

RÉSUMÉ

Plusieurs vocalisations de cétacés sont de type tonal et la plupart sont modulées en fréquence. L'algorithme de détection présenté ici sépare le contour de fréquence en une séquence d'éléments. Chacun des éléments est suffisamment petit pour qu'une approximation linéaire du contour de fréquence puisse être effectuée. Le problème est donc en ce sens simplifié de façon à ce que la détection d'un signal inconnu passe à celle d'un signal connu (une modulation linéaire de fréquence) avec des paramètres inconnus. La méthode d'estimation est basée sur le maximum de vraisemblance, et la fréquence de départ, le taux de modulation et l'amplitude de chacun des élements sont estimés. Des analyses plus poussées sont alors effectuées sur des groupes de modulations enchaînées (i.e. vocalisations) afin de classifier les sons comme étant du bruit ou comme faisant partis d'une espèce spécifique.

Les résultats sont tirés de la performance des données de test et de signaux synthétiques en présence de bruit blanc. Les avantages de cet algorithme sont: une bonne performance de détection, du moins à l'intérieur d'un bruit blanc; une haute résolution; la facilité d'interprétation; la flexibilité; la compression de données. Les désavantages sont: les coûts computationnels; la détérioration de la performance à l'extérieur d'un bruit blanc ou avec un signal modulé en amplitude. Des développements plus poussés sont requis afin de réduire les erreurs provenant de la superposition d'un son tonal sur un son non tonal. L'algorithme a été appliqué aux problèmes de détection des vocalisations des baleines franches ainsi qu'à celui de la distinction de leurs vocalisations avec celles des rorquals à bosses.

1. INTRODUCTION

The calls of marine mammals are highly variable, and even those species with prima facie stereotyped call repertoires, such as blue whales, have proved to be more varied under wider scrutiny (e.g. Rivers 1997, Stafford *et al.* 1999). Consequently, even a detection system directed at a single species must be capable of handling biological variation (intra- and inter-individual, geographic, seasonal etc.). Furthermore, the frequency contours of individual vocalizations can be complex and nonlinear. Both these points are illustrated in Fig. 1, which shows a sample of calls from a group of singing humpback whales (*Megaptera novaeangliae*).

In this intricate signal environment the signals of interest cannot usually be fully specified. Parametric models of sufficient flexibility and complexity to approximate real varying signals are rarely used in marine mammal bioacoustics. (The usual approach is to manually measure nonparametric features of the signal from a spectrogram, e.g. minimum and maximum frequency. These are useful quantities but don't represent the signals well.) What is more, even under a suitable model, the statistical distributions of the parameters are not generally available, because sampling the calls on a sufficiently large scale over time, space, behaviour etc. is difficult.

This paper describes a method for detecting marine mammals calls in which the calls being detected, or parts thereof, are approximated by constant-amplitude linear chirps. The received signals are in effect simplified by analyzing them in short sections; in this way a paramateric model *can* be specified. A detector can be devised based on the estimate of the signal amplitude, and the frequency-related estimates hold information useful for classification.

The performance of the system is examined by (a) simulations using synthetic signals and noise, with known properties (b) trials using real recordings of whale calls.

The investigations in this study were motivated by a project to examine the potential use of acoustic detections for detecting North Atlantic right whales (*Eubalaena glacialis*) (Gillespie and Leaper, 2001). Improved understanding of the whereabouts of these animals could reduce the high mortality rate due to ships and fishing gear. Computer assistance in acoustic detection and classification is very desirable with large volumes of data, and could also be used for remote sensing.

2. METHODS

2.1 General description of algorithm

The frequency contour of marine mammal tonal calls is usually nonlinear. However, subsequences of the data (frames) can be taken, and if they are sufficiently short, then the call contour in that frame will be approximated well by a linear chirp. When a signal is present, each data frame can be treated as possibly containing a 'partially known' signal (i.e., a linear chirp with unknown parameter values to be estimated). This framing approach is like that used in the short-time Fourier transform, but the underlying signal model is different.

A spectrogram of a synthetic tonal sound with a nonlinear frequency contour is shown in Fig. 2. The sound was divided into four parts, and each part was treated as a linear chirp. The parameters were estimated with the algorithm described in this paper; the chirps are overlaid in the figure.

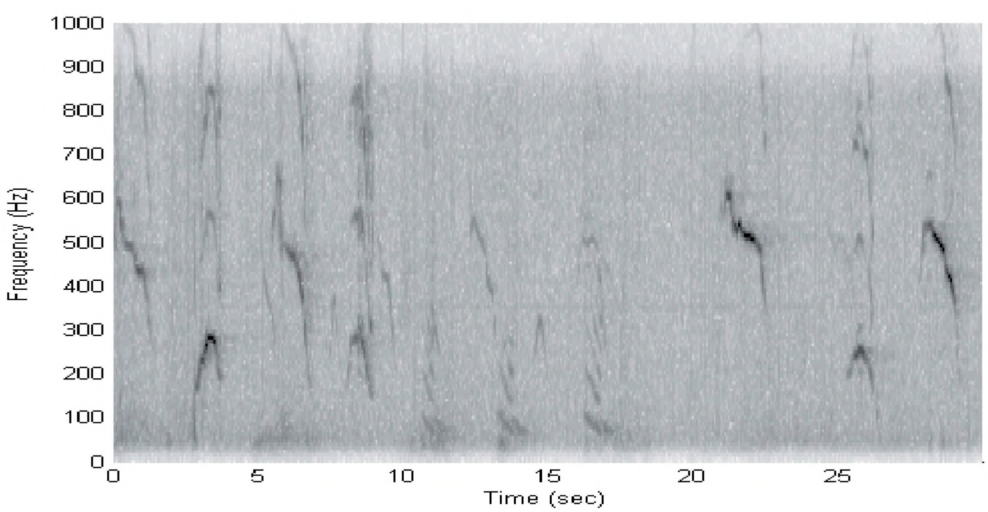

Figure 1. Spectrogram of a sequence of calls from a group of singing humpback whales (*M. novaeangliae*).

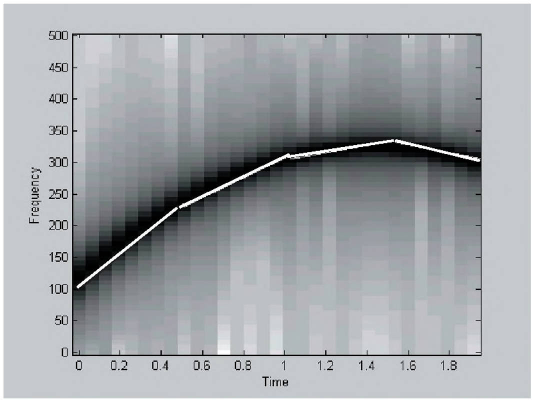

Figure 2. Spectrogram of a synthetic tonal signal with a nonlinear frequency contour, characterised by a sequence of four linear chirps (white lines).

If a chirp is present, its parameters may be estimated (amplitude, start frequency and chirp rate). The problem before us then is to find a good detector *for short signals and low SNR*. Together, these are particularly challenging conditions for detection. Boashash (1992) reviewed several methods under these circumstances, and with one exception, a maximum likelihood estimation (MLE) method worked best. This is the method used here.

Detection and estimation in white noise

In the case of white Gaussian noise (WGN), the combined (complex) signal and noise model is **y**=**z** + **w**, where **y** is the observed signal and noise, **z** represents the underlying signal, and **w** the noise. The signal is modeled as a linear chirp so that:

$$z[n] = A\exp(j(a_0 + a_1\Delta n + a_2\Delta^2 n^2))$$

where Δ is the sampling interval; a_0, a_1 and a_2 are phase parameters; and A is the amplitude (assumed constant). The phase parameters are related to the start frequency f and chirp rate c by $f=a_1/2\pi$ and $c=a_2/\pi$.

The MLE estimates of a linear chirp in white noise are shown by Boashash (1992) to be obtainable using a 'dechirping' operation:

$$\max L(a_1, a_2) =$$

$$\max\left|\frac{1}{N}\sum_{n=0}^{N-1} y[n]\exp(-j[a_1\Delta n + a_2\Delta^2 n^2])\right|^2$$

(For those readers familiar with time-frequency analysis, there is a connection here with the Wigner distribution. Kay and Boudreaux-Bartels (1985) describe a detector in which the Wigner distribution of the observed data is integrated along all lines in the time-frequency plane. They show that this is the optimal detector for sufficiently long chirp signals in white noise. Li (1987) proves that the dechirping approach used here is equivalent.)

An example likelihood surface from a 1024-point synthetic chirp signal is shown in Fig. 3.

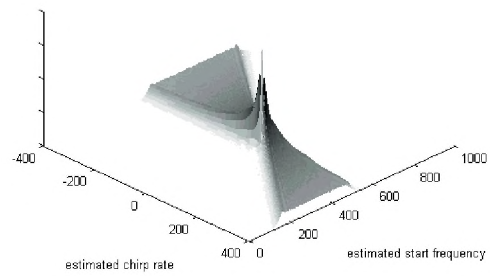

Figure 3. The likelihood surface from a 1024 point frame, for a synthetic, constant-amplitude linear chirp with a start frequency of 500 Hz and chirp rate of 100 Hz/second.

The likelihood can be maximised numerically over a grid of frequency and chirp rate values. The allowable combinations form a parallelogram: values chosen outside of this region would lead to negative frequency values. This region is given by:

$$-\frac{f}{T} \le c \le \left(\frac{f_s}{2T} - \frac{f}{T}\right)$$

where f_s is the sample rate, T is the duration of frame (seconds).

However, for reasons of efficiency the algorithm searches the entire rectangular region using FFTs:

$$\hat{a}_1, \hat{a}_2 =$$

$$\arg\max_{a_1,a_2}\left(\left|\frac{1}{N}\sum_{n=0}^{N-1} y[n]\exp(-j[a_1\Delta n + a_2\Delta^2 n^2])\right|\right)$$

$$= \arg\max_{a_1,a_2}\left(\left|DFT(y'(a_2))\right|^2\right)$$

which uses the maximum of the discrete Fourier transform (DFT) of the dechirped signal $y'[n;a_2]=y[n]\exp(-ja_2\Delta^2n^2)$. Although the search is over the entire rectangular region, the likelihood values in the out-of-bound (negative frequency) regions will be small.

The MLE estimate of A is given by (Boashash 1992):

$$\hat{A}_1 = \left|\frac{1}{N}\sum_{n=0}^{N-1} y[n]\exp(-j[\hat{a}_1\Delta n + \hat{a}_2\Delta^2 n^2])\right|$$

but a better estimate of amplitude (see results, especially fig. 7) is:

$$\hat{A}_2 = \max_{a_1,a_2} \left| \frac{1}{N} \sum_{n=0}^{N-1} y[n] \exp(-j[a_1 \Delta n + a_2 \Delta^2 n^2]) \right|$$

Detection and estimation in coloured noise

In the more general and probably more realistic case of coloured noise, we have the signal and noise model $\mathbf{y}=\mathbf{z}+\mathbf{w_c}$, where $\mathbf{w_c}$ is now distributed as $N(\mathbf{0},\mathbf{C})$, where $\mathbf{C}$ is the sample autocorrelation matrix of the background noise. The spectrum of coloured noise is not flat.

In this case, a whitening approach can be used as described by Kay (1993). The matrix $\mathbf{C}$ is factored to give a 'prewhitening matrix' $\mathbf{D}$ such that $\mathbf{C}^{-1}=\mathbf{D}^T\mathbf{D}$. Applying $\mathbf{D}$ to the combined signal and noise:

$$\mathbf{Dy} = \mathbf{Dz} + \mathbf{Dw}_c = \mathbf{z'} + \mathbf{w'}$$

where $\mathbf{w'}$ is distributed as $N(\mathbf{0},\mathbf{I})$, i.e. the noise has been whitened. The resulting signal $\mathbf{z'}$, the parameters of which are now estimated, is a distorted version of the original. The potential advantage of prewhitening is that the flat noise spectrum means that consecutive estimates when the signal is absent or weak will not be correlated due to correlation in successive noise spectra.

Efficiency of estimators

MLEs are usually statistically efficient, i.e. as N tends to infinity and at a sufficiently high SNR, the estimates are unbiased and have variances that attain the Cramer-Rao Lower Bound (CRLB) (a lower bound on the variance of all unbiased estimates). In other words, under these ideal conditions, the resolution of the MLE is as high as can be. Peleg and Porat (1991) give the CRLB variance bounds for constant-amplitude linear chirps in white noise.

DFT-based estimates, on the other hand, do not attain this resolution even under high SNR, high N conditions (see e.g. Boashash 1992, fig 3). Of course, whether high resolution is required for a classification problem needs to be assessed on a case-by-case basis.

However, in the present study the estimates do not necessarily have these desirable MLE properties, because of the following conditions used:

(i) N is not large because short frames are used to obtain a linear approximation to the frequency contour.
(ii) The SNR can be low for long-range signals.
(iii) The use of the DFT for computationally efficient estimation of frequency introduces a limit to the resolution.
(iv) With the prewhitening transformation of a signal in coloured noise, the underlying signal is distorted.
(v) The assumption of constant amplitude is almost certainly violated.

The performance of MLE estimators at small N cannot usually be found analytically (Kay 1993). Simulations were therefore carried out using synthetic signals and noise to provide an understanding of the bias of the estimators as a function of SNR and N.

After detection of individual chirps, sequences of chirps can be concatenated to represent the entire signals or calls. The information can then be used jointly for call detection (as opposed to chirp detection) or call classification.

2.2 Testing of performance

The following simulations used 1000 runs each and a sample rate of 2kHz. Results were obtained as a function of signal-to-noise ratio (SNR) and signal length (N). The SNR is defined as A^2/σ^2, where σ^2 is the noise variance.

Bias of estimators

Synthetic linear chirps (start frequency 100 Hz, chirp rate 200 Hz/second) were created and embedded in WGN. The biases in estimates of start frequency, chirp rate and amplitude were examined.

Detection using synthetic signals and noise

Synthetic linear chirps were created and embedded in noise. The parameter values were randomly chosen in such a way that there were no negative-going frequencies. Some examples are shown in Fig. 4. The algorithm was used to estimate the signal amplitude, and this was used as a test statistic for detection. The probability of false and true detections was determined by simulation.

Two types of noise were used: WGN and a coloured spectrum based on a sample from the ocean. Synthetic noise, with a similar magnitude spectrum to the sample of ocean noise, was generated by passing white noise through a specially designed filter. An example power spectrum is shown in Fig. 5. The filter was designed using the Yule-Walker method (MATLAB signal processing toolbox).

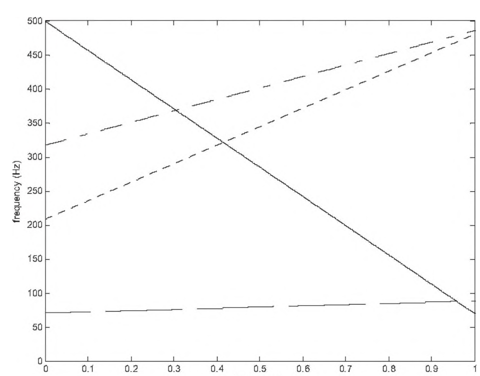

Figure 4. Examples of four linear chirp signals with randomly chosen phase parameter values. The parameter values are chosen so that there are no negative-going frequencies.

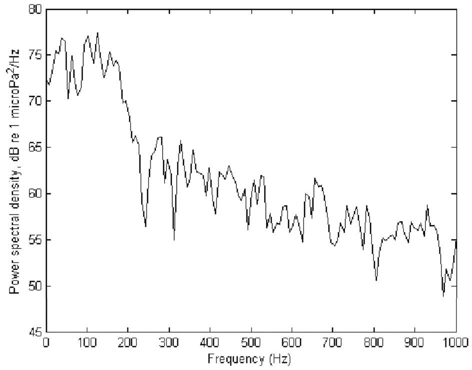

Figure 5. Power spectrum of a sample of synthetic ocean noise.

Detection using real recordings

In order to allow a dynamic response to changing noise levels and spectra in real noise conditions, the noise variance (σ^2) and the prewhitening matrix were re-estimated periodically, based on a block of sound. The signal samples used to calculate these quantities are based on order statistics.

Let

$\{\underline{s}_i\}$, i=1,…,m be a set of sound vectors each of length N, selected randomly from the current sound block.

v_i=var($\underline{s}_i$), the variance of vector i

e_j and e_k = the jth and kth percentiles of the distribution of $\{v_i\}$.

e_j provides an estimate of σ^2. The sample vector corresponding to e_k is used to estimate the prewhitening matrix **D**.

Table 1 shows the parameter settings used for the analysis presented in this paper. In addition, files were high-pass filtered at 50 Hz. The allowable chirp rate was set between +/- 250 Hz per second and searched over 100 points. Chirps were detected and estimated with start frequency up to 800 Hz. These settings were chosen after some initial experimentation showed they produced reasonable fits to call contours (see Fig. 11) and few false detections on the spectrogram, but they are in no sense optimal.

Table 1. Parameter settings for analysis of real recordings used in this study.

Parameter	Description	Value
f_n	Sampling rate (per second) for noise blocks	1/60
m	Size of set of sample noise vectors	400
j	Percentile for noise estimation	0.9
k	Percentile for prewhitening matrix	0.5
N	Frame length	512
H	Hop length	256

After detection and estimation of chirps, sequences of contiguous chirps that did not differ in frequency by more than 30 Hz were combined into detected 'calls'. The characteristics of calls from recordings of right whales, humpback whales, and a recording thought to be free of both, were then compared. Information about these recordings is shown in Table 2.

The comparison of humpback and right whale call characteristics showed that right whales commonly produced upsweeping calls. An upsweep detector was then made; this simply selected the subset of upsweeps from the detected calls.

The upsweep detector was optimised as a function of the SNR threshold in the following sense. The number of upsweep detections was maximized in recordings where right whales were known to be present, and simultaneously minimized in the recording where right whales were thought to be absent.

3. RESULTS

3.1 Simulations using synthetic signals and noise

Bias of estimators

Fig. 6 shows the bias in frequency and chirp rate as a function of N and SNR, from simulations using a synthetic chirp in WGN. Below a threshold SNR for all N, bias increases rapidly. Above the threshold, the bias is generally small.

Table 2. Information about recordings used for analysis.

Location	Date	# hours recorded	No. of channels	Human listening	Visual survey
Cape Cod Bay	16/3/01	4	3	Many right whales heard.	Right whales seen on 17/3/01.
Great South Channel	26/5/00	4	6	Right whales and humpbacks heard.	Right whales seen on 26/5/00.
Great South Channel	01/05/01	12	1	Few right whale calls heard by human listeners; of these, none definite.	None
Great South Channel	16/5/00	4	6	Many humpbacks heard. No right whales heard.	None

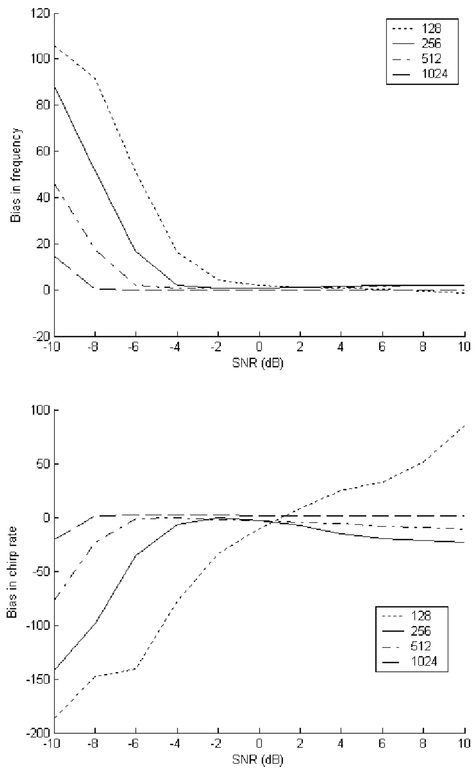

Figure 6. Simulation results showing biases of start frequency (top) and chirp rate (bottom) for linear chirps in WGN, as a function of SNR and N. When the number of samples and SNR are sufficiently high, the bias approaches (but does not necessarily reach) zero. The threshold SNR decreases as N increases. N is shown in the legend.

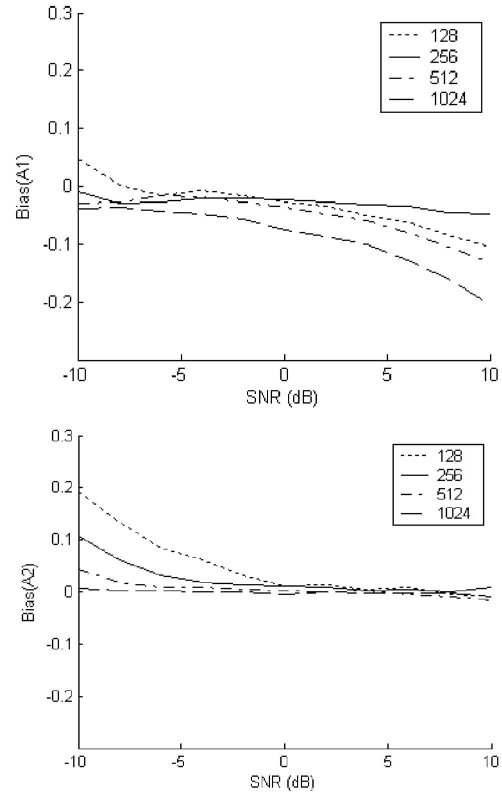

Figure 7. Simulation results showing biases of two estimates of signal amplitude: $\hat{A}_1$ (top) and $\hat{A}_2$ (bottom), for linear chirps in WGN, as a function of SNR and N. See text for details of these estimators. $\hat{A}_2$ has less bias and appears to stabilise as SNR increases. N is shown in the legend.

The bias of the amplitude estimator $\hat{A}_1$ is shown in Fig. 7 (top), and does not approach zero for these values of N and SNR. The estimator is sensitive to the biases in the estimates $\hat{a}_1$ and $\hat{a}_2$. The bias of estimator $\hat{A}_2$ is shown in Fig. 7 (bottom). The bias stabilises above a threshold SNR and is relatively small. Therefore, $\hat{A}_2$ is used for detection purposes in this study.

Detection performance

The performance of the detector on random linear chirps is shown in Fig. 8. The results show that, as anticipated, the increase in N leads to improved performance at low SNR. However, the real cost of increased N is hidden in these simulations. Real signals usually have a nonlinear contour, and increasing N in that case would make the linear approximation poorer, leading to a trade-off in performance.

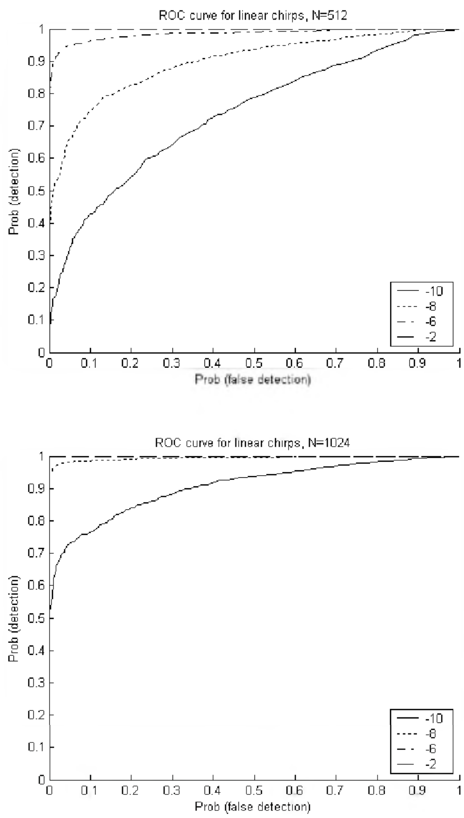

Figure 8. Performance of detector on synthetic linear chirps in WGN, for N=512 (top) and 1024 (bottom). Detection performance improves as N increases. SNR (dB) is shown in the legend.

Fig. 9 shows the results for a detection of linear chirps in WGN by peak-picking the DFT. Performance is poorer than the chirp detector. However, these results may exaggerate what would happen using real signals, because modulation rates are likely to be more limited.

The performance of the detector in coloured noise with and without prewhitening is shown in fig. 10. Prewhitening in this case improves the detector.

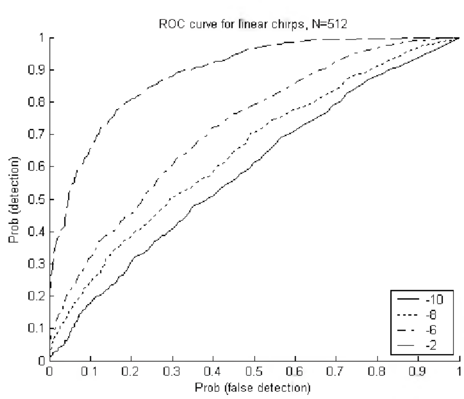

Fig. 9 Detection performance by peak-picking the DFT using synthetic linear chirps in WGN (N=512). SNR (dB) is shown in the legend.

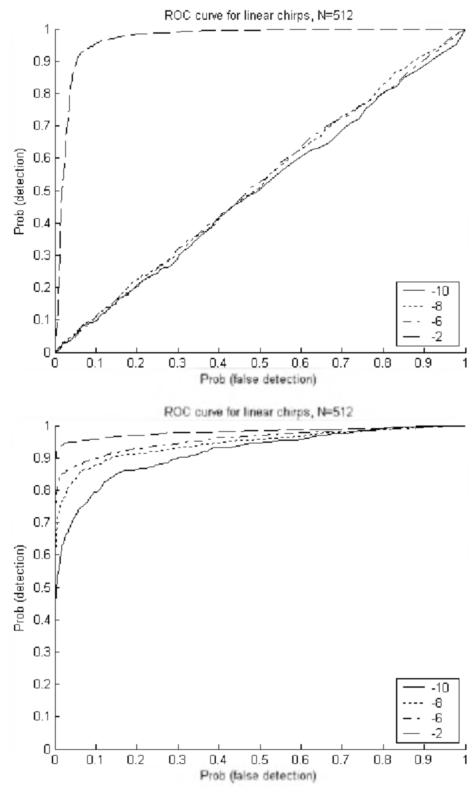

Figure 10. Detection performance in synthetic ocean noise with (bottom) and without (top) prewhitening. SNR (dB) is shown in the legend.

3.2 Tests using real recordings

The spectrogram of an underwater recording is shown in Fig. 11, with the detector results overlaid in white. In the figure are the sounds of a disk drive, two right whale upsweeps and another, more complex whale call. The detector has found and estimated short chirps and these are shown in white. After detection and estimation of chirps, nearby chirps are joined together into calls or sounds as described in the methods section.

The start frequency and sweep of calls from the recordings of right whales and humpbacks is shown in Fig. 12. A cloud of points representing upsweeps is evident from the right whales at about 50-150 Hz start frequency and sweeping up by about 30 Hz or more. This type of call is in fact well known from previous studies as a low-frequency 'contact' (upsweep) call (Clark 1982, McDonald and Moore 2002). As the scatter of points in Fig. 12 shows, they vary considerably in start frequency and sweep rate. Some example spectrograms are shown in Gillespie (2004). Calls with these characteristics were not found in the humpback recording.

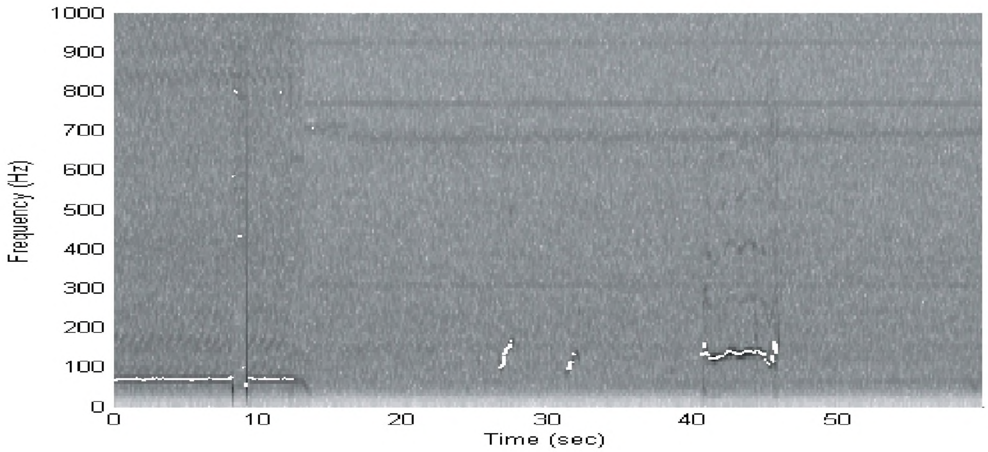

Figure 11. Illustration of the detection and estimation algorithm. The spectrogram of the received sound is shown with the detected and estimated chirps overlaid in white. The algorithm has detected the sound of a disk drive over the first 12 seconds; right whale upsweeps at about 26 and 32 seconds; and a more complex call at about 40 seconds.

There is considerable overlap elsewhere in the parameter space, though humpback signals with start frequencies >200 Hz tend to have greater sweep magnitudes. A large proportion of the signals with small sweep are false detections of ambient noise. In fact, the magnitude of the frequency modulation is a good indication of a biological signal.

The results show that, at least in the recordings tested here, upsweeps are a common and fairly distinctive right whale call. An upsweep detector was created to find calls with start frequency between 50 and 200 Hz, and with an end frequency at least 30 Hz higher than the start frequency (i.e. >30 Hz upsweeps). The duration of the signals was between 0.5 and 3 seconds. It is possible, however, that calls with inflexions in the contour may also be selected i.e., which are not strictly upsweeps.

The results of applying the upsweep detector to real recordings are shown in Fig 13. The figure shows the number of upsweeps detected per hour per channel as a function of threshold SNR. The y-axis in each plot shows the rate from a recording when right whales were present, and the x-axis shows the rate from a recording where right whales were thought to be absent. As the SNR threshold is decreased, more right whale calls are detected (y-axis) but there are also more false detections (x-axis). The appropriate threshold setting depends on the false detection rate required by the application. In the case of the northern right whale, a very low false alarm rate is likely to be required.

The effect of the prewhitening process is also shown in Fig. 13. Prewhitening appears to reduce the performance of the detector, except at high SNR thresholds in the Cape Cod Bay recording. No general conclusions should be drawn about using prewhitening though: its usefulness or otherwise will depend on the nature of the noise; and furthermore only one particular setting was tested here (Table 1).

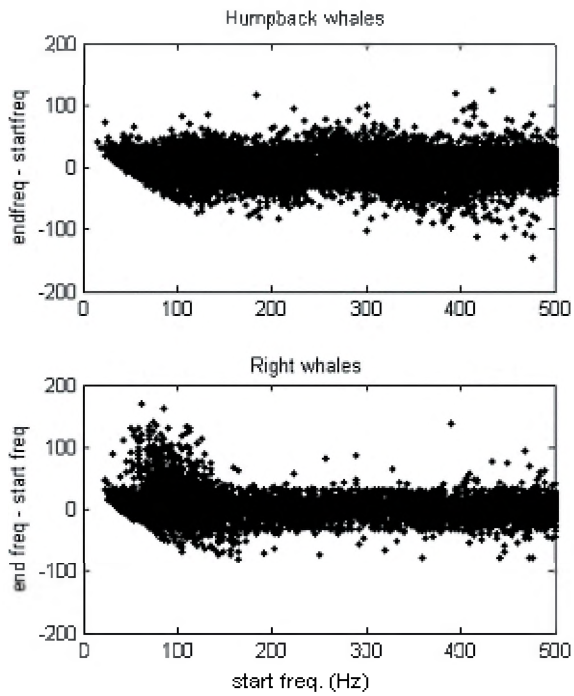

Figure 12. Scatterplot of start frequency versus sweep for calls of duration between 0.5 and 3 secs with an SNR threshold of –6 dB. The top plot is from the recording of humpback whales (Great South Channel 16/5/00) and the bottom from right whales (Cape Cod Bay 16/3/01). There are many upsweeping calls from right whales.

4. DISCUSSION

In an ideal detection environment of known, fixed signals, template-based detection methods such as matched filters perform well. In biologically realistic environments, however, signals are highly diverse and variable, and often only partially known. Methods are required which handle signals of this nature. The problem of detection and estimation of real animal signals will often be intractable under a fully parametric approach, because this requires (i) a sophisticated and flexible parametric signal model, and (ii) large, high-quality datasets for determining the statistical distributions of the parameters. The system described here attempts to simplify the problem somewhat by 'forcing' FM tonal signals to be approximately chirp-like using a short data frame. The result is a highly flexible system, which allows any slow-varying frequency-modulated signal to be detected and characterised as a sequence of linear chirps.

4.1 Detection and estimation of linear chirps

A problem introduced by the deliberately short frames is that detection performance generally falls with signal length. Short signal length and low SNR are particularly challenging detection conditions. Nevertheless, performance may still be superior to techniques based on the DFT because the underlying model in that case is that of stationary (sinusoidal) signals, while the signals of interest are known to be frequency-modulated. The simulation results appear to confirm this.

It is likely that the method will usually give a reasonable fit to the actual phase of the original signal (with sufficiently high SNR) because of the linear approximation. There are biases because the data frames are short (i.e. N is not large), but these biases are small.

However, the model assumes a constant amplitude signal, which is certainly incorrect. Real signals will have at the very least a finite attack and decay. No results have been obtained in this paper on the effects of amplitude modulation.

A potential disadvantage of the method described here is that it only searches for the global maximum of the likelihood surface, and therefore can only detect a single signal in any frame. Generalising the technique for use with multicomponent signals should be possible but would require more processing. The method of simulated annealing may be a useful way of finding multiple local maxima (Press *et al.*, 1992). The current inability of the algorithm to detect multicomponent signals means that it is not practical for analysing the whistles of wild groups of odontocetes, in which simultaneous calls are common. But the call rate of baleen whales, with the exception of singing humpbacks but including right whales, is generally low and simultaneous calls may be unusual.

Further signal processing may be required to eliminate detections of broadband sounds. The algorithm will trigger on these but they may be eliminated using a further measure of spectral concentration.

Extending the present estimation method for use with a higher order polynomial is fairly straightforward but necessitates further computational costs. Some initial simulations have also indicated that the detection performance deteriorates with this type of model, although further investigation is needed. In comparisons of methods for estimating nonstationary signals, Boashash (1992) found that for estimation with short signals and low SNR, the 'cross-Wigner-Ville' method outperformed the ML method, but this method requires more intensive processing.

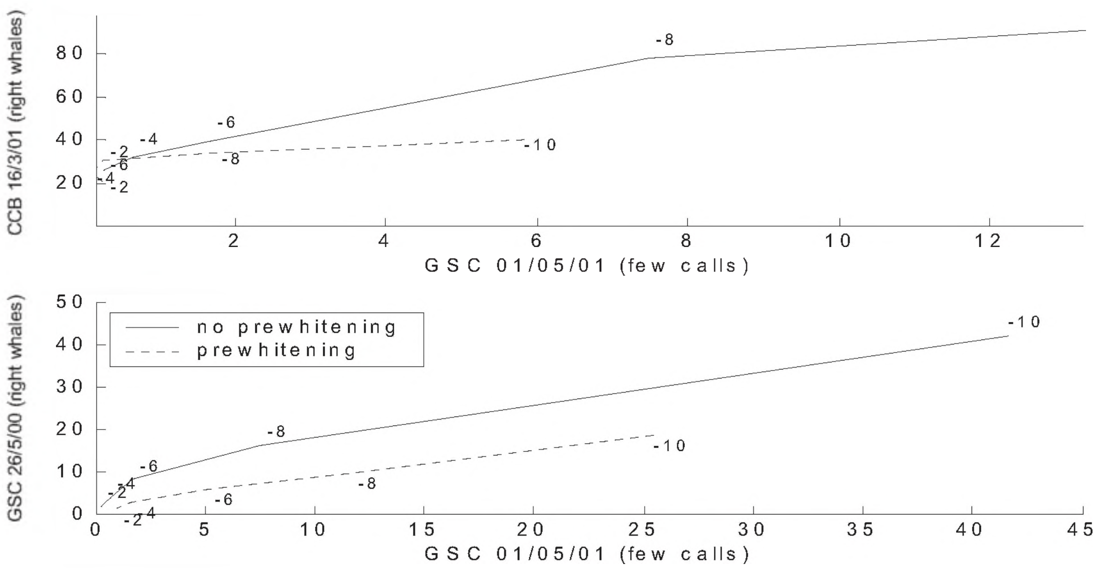

Figure 13. Performance of the upsweep detector using real recordings. Numbers of calls per channel per hour are shown for the two workshop test datasets (y-axes), in which right whale calls are present, against a dataset thought to be free of right whales (x-axis). The curves show the effect of varying the SNR threshold from –2 to -10 dB, and applying or not applying the prewhitening procedure. The characteristics of the calls are described in the text.

4.2 Formation of calls

Two further processes are required after the detection/estimation of chirps. Firstly, a method of forming calls from sequences of linear chirps is needed. Secondly, classification requires the allocation of subsets of calls with the required characteristics to the appropriate classes. In this study, these secondary processes were purposefully kept simple. Contiguous chirps that did not differ in frequency by more than some upper limit were combined into calls, and calls with appropriate start and end frequencies and durations were selected. Both of these processes could easily be improved to make further use of phase and amplitude information, perhaps also using continuity conditions.

4.3 Application to right whale detection

The numerical representation using the linear chirp model is very compact. In the example of Fig. 2, a ~2 second signal (2048 samples at 1000 Hz sample rate) was characterized by four 512-point linear chirps, requiring storage of 12 numbers (4 chirps x 3 estimates). This reduced the original digitized sound information by a factor of the order 100. In practical terms, satellite transmission is much more feasible after processing in this way.

In this study, right whale detection has focused on the relatively simple (yet variable; see Fig. 12) contact call or upsweep. Recent studies, including this one, have now shown this type of call to be reasonably common from northern right whales (McDonald and Moore 2002, Laurinolli *et al.* 2003).

For this particular signal type the 512-point linear chirp model, as applied in the present paper, may not be ideal. Firstly, a longer chirp model ought to provide better detection performance if the linear approximation to the contour holds fast. Secondly, the search space of the algorithm could be reduced to only include the region corresponding to upsweeps.

A more realistic model for variable upsweeps is a polynomial phase signal. Clark (1982) estimated polynomial coefficients by regression on digital images of spectrograms and used these as descriptors of call shape in southern right whales. Since then, computer functionality has developed considerably to allow this kind of parametric approach. For right whale upsweeps, an extension of the method used here to higher-order polynomials ought to be possible, though the additional computational cost could cause practical problems.

On the other hand, there are advantages to using the general, short chirp detector applied in this paper. Although not optimal for detecting longer upsweeps, information on other signals present over a larger frequency range (50-800 Hz, say) could be used to identify characteristics of humpback calls. This more general approach might, for example, provide suitable quantitative information within which to search for periodic, similar signals, and for practical purposes these may be sufficiently distinctive features of humpback song.

REFERENCES

Boashash, B. 1992. Estimating and interpreting the instantaneous frequency of a signal - part 2: algorithms and applications. *Proc. IEEE* 80: 540-568.

Clark, C. 1982. The acoustic repertoire of the southern right whale, a quantitative analysis. *Animal Behaviour* 30: 1060-1071.

Gillespie, D. 2004. Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram, Canadian Acoustics, Vol 32 #2, June 2004.

Gillespie, D. and Leaper, R. 2001. Report of the workshop on right whale acoustics: practical applications in conservation. Paper SC/53/BRG2 presented to Scientific Committee of International Whaling Commission, London, 2001.

Kay, S. 1993 *Fundamentals of Statistical Signal Processing I: Estimation Theory*. Prentice-Hall, New Jersey. 595 pp.

Kay, S., and Boudreaux-Bartels, G. 1985. On the optimality of the Wigner distribution for detection. *IEEE ICASSP Proc.* 3:1017-1020.

Laurinolli, M., Hay, A., Desharnais, F. and Taggart, C. 2003. Localization of North Atlantic right whale sounds in the Bay of Fundy using a sonobuoy array. *Mar. Mamm. Sci.* 19:708-723.

Li, W. 1987. Wigner distribution method equivalent to dechirp method for detecting a chirp signal. *IEEE Trans. Acoust. Speech and Sig.Proc.* 35: 1210-1211.

McDonald, M. A., and Moore, S. 2002. Calls recorded from North Pacific right whales (*Eubalaena japonica*) in the eastern Bering Sea. *J. Cet. Res. Manage.* 4: 261-266.

Peleg, S., and Porat, B. 1991. The Cramer-Rao lower bound for signals with constant amplitude and polynomial phase. *IEEE Trans. Sig. Proc.* 39: 749-752.

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B. 1992. *Numerical Recipes in C*. CUP.

Rivers, J.A., 1997. Blue whale, *Balaenoptera musculus*, vocalizations from the waters off central California. *Mar. Mamm. Sci.* 13: 186-195.

Stafford, K., Nieurik, S., and Fox, C. 1999. Low-frequency whale sounds recorded on hydrophones moored in the eastern tropical Pacific. *J. Acoust. Soc. Am.* 106: 3687-3698.

ACKNOWLEDGEMENTS

I wish to thank Cornell Bioacoustics Laboratory for providing the recordings used in this analysis; M. Brown and P. Gerrier for visual aerial survey data. Doug Gillespie provided useful comments on a draft; any errors are mine. Two anonymous reviewers made helpful comments.