DETECTION AND CHARACTERIZATION OF MARINE MAMMAL CALLS BY PARAMETRIC MODELLING

A. T. Johansson[1] and P. R. White[2]
Signal Processing and Control Group, Institute of Sound and Vibration Research (ISVR),
University of Southampton, Highfield, Southampton SO17 1BJ, UK.
Email: tj@isvr.soton.ac.uk[1], prw@isvr.soton.ac.uk[2].

ABSTRACT

In this paper, we describe a parametric modeling method for detection and characterization of tonal signals and its application to marine mammal calls. The method tracks dominant frequencies with an adaptive notch filter (ANF), and couples this to a novel, simultaneous detection step. The detection statistic is derived from a measure of tracking reliability, obtained as a by-product of the tracking algorithm. Detection therefore comes at little extra computational cost from an algorithm that is fast, simple, and capable of dealing with multiple signals in low signal-to-noise ratios. Frequency estimates are derived directly from the time domain waveform, avoiding the resolution trade-off and other short-comings of the commonly used spectrogram. The performance of the algorithm is demonstrated on both simulated signals and recordings of right whale calls. The method is found to be noise robust and capable of extracting right whale and other calls with a low false alarm rate.

SOMMAIRE

Le présent article décrit une méthode de modélisation paramétrique pour la détection et la caractérisation des vocalisations tonales de mammifères marins. La méthode consiste à poursuivre les fréquences dominantes à l'aide d'un filtre à encoche adaptatif (ANF), couplé à une étape de détection simultanée innovatrice. La statistique de détection est dérivée d'une mesure de la fiabilité de poursuite, un sous-produit de l'algorithme de poursuite. La détection entraîne un coût computationnel additionnel minime, et est à la fois rapide, simple, et capable de traiter des signaux multiples et à de bas rapports signal sur bruit. Les estimations de fréquence sont dérivées directement du domaine temporel et fréquentiel, évitant ainsi les compromis de résolution de la technique spectrogramme utilisée fréquemment. Dans une application sur un fichier d'une durée de 18-min de l'ensemble de donnée de l'atelier, il est possible de détecter quatre vocalisations probables de baleines franches.

1. INTRODUCTION

Passive acoustic detection of cetaceans is an important and growing research field. The mitigation of several different threats to cetaceans, such as collisions with ships and ensonification by high-power active sonar, require the detection, localization and, ideally, identification of cetaceans in the vicinity. Visual observation requires daylight, a reasonably calm sea, and that the cetaceans are at the surface. These drawbacks do not apply to a system based on processing the cetaceans' vocalizations. Here, to be detectable an animal must vocalize, but this is not a major drawback as most cetaceans are highly vocal. One must also be able to process the calls in strong background noise and other difficult environmental conditions and handle the great variability of most cetacean calls.

The Workshop on Detection and Localization of Marine Mammals Using Passive Acoustics, held in Dartmouth, NS, 19-21 November 2003, answered to the growing need of collecting and comparing different algorithms for passive acoustic processing of marine mammal sounds. The conference focused on processing the sounds of the North Atlantic right whale, a critically endangered species of which there less than 300 individuals left [1]. Strong protection measures are already in place for this species, for example the US Marine Mammal Protection Act states that no vessel is allowed within 500 meters of a North Atlantic right whale [2]. To implement such protection measures it is crucial to be able to detect right whales in the vicinity.

Accompanying the workshop was a dataset of several underwater recordings containing right whale calls. These recordings were all made on moored single hydrophones in the Bay of Fundy, a key North Atlantic right whale habitat. One file, called L-138, was specifically meant for testing detection algorithms, and will be used for this purpose here. The file is 18 min long and sampled at 1250 Hz.

In this paper, we use model-based signal processing to detect and characterize right whale sounds. Model-based signal processing is a popular research field that has found applications in many fields, such as biomedical signal processing, speech recognition and economic forecasting. The idea here is to apply a model, controlled by a small number of parameters, to the signal. Given old samples, the model produces a prediction of the next signal sample. By minimizing the difference between the model-predicted and the actual signal, we fit the model to the signal and force information about it into the model parameters. These can subsequently be used to characterize the signal, and, as we show here, also to detect occurrences of the signal immersed in broadband noise.

We use a specific type of model known as an adaptive notch filter (ANF), which expresses the prediction error by a filtering operation on the input signal. The transfer function magnitude response of the notch filter is that of a deep notch at one or more frequencies, and a relatively flat level away from notches. On fitting the notch filter to a recording, we minimize the filter output, which forces the notches to attempt to cancel the signal frequencies at each time. After model fitting one can the use the notch frequencies as estimates of the dominant frequencies of tonal components of the signal. The frequencies can then be fed to a classifier, but this step is not reported here. The ANF model is tailored to fit narrowband signals, and is capable of modeling simultaneous signals of time-varying frequencies and amplitudes. The authors have recently reported on this for cetacean whistles [3][4]. Because it works directly on the signal waveform, the ANF model avoids the resolution problems of characterization methods that are applied to the spectrogram or another time-frequency distribution. Also, it is simple to implement and use, requires little user tuning, and can be run in real-time.

This paper exploits a novel architecture where the ANF is used both for detection and characterization. The algorithm is an adaptive scheme with a fading memory. This allows it to estimate parameters based on a finite observation interval, so giving it the ability to track time-varying parameters. We propose to run a parametric model along the whole signal and detect signals from a measure of the reliability of the parameter estimates. This measure is an internal variable in the adaptation scheme, so detection comes at very little extra computational cost. The detection decision could possibly also be made on the basis of an analysis of the frequency estimates themselves. But for time-varying signals this is difficult - for instance we cannot just use the parameters' variances because a time-varying signal will naturally impose its own variation in the parameters.

The layout of this paper is as follows: In Sections 2 and 3, we describe the theory of ANF modeling and establish theoretical grounds for a suitable detection statistic. Then, in Section 4 we describe how to apply our method to detect and characterize tonal sounds in oceanic background noise. Section 5 reports on the results of application to a simulated signal and the workshop dataset L-138. Finally, in Section 6 we draw conclusions from the findings.

2. REVIEW OF ADAPTIVE NOTCH FILTER THEORY

For a signal composed of one or more slowly evolving narrowband signals, an AR model is appropriate. However, when such a signal is observed in moderate or strong background noise, the AR model does not perform well. Better performance can be obtained by including the noise in the model. To this end, we first pre-whiten the noise by estimating and equalizing its spectrum. This will be further discussed in Section 4. Including white background noise in the model results in an ARMA model with equal AR and MA parts. However, such a model is not identifiable as its transfer function is undefined at the signal frequencies. The standard way of dealing with this is to contract the poles slightly towards the origin using the pole contraction factor ρ, $0<\rho\leq 1$. The pole contraction factor controls the notch bandwidth, and therefore implicitly the trade-off between tracking ability and noise robustness. This is because the algorithm is only able to track a signal if its instantaneous frequency falls within the current position of a notch, but the wider the notch width the more noise energy slips into the notch and influences the estimation.

The adaptive notch filter model expresses the prediction error $\varepsilon(n)$ as

$$\varepsilon(n)=\frac{1+\sum_{i=1}^{P}a_i(n)q^{-i}}{1+\sum_{i=1}^{P}\rho^i a_i(n)q^{-i}}\,y(n)=H(q^{-1},n)y(n) \qquad (1)$$

where $y(n)$ is the recording, $H(q^{-1},n)$ is the transfer function of the notch filter, and $a_i(n)$ is the i^{th} AR coefficient at time n. The model order P is twice the number of components M tracked by the model, $P=2M$. The model order is a user-defined parameter, however one for which we argue that the exact value is not critical. Since we shall couple the estimation to detection, wherein we shall determine when a component is locked on to a signal, it suffices to choose the model order as the maximum number of simultaneous tonals that one wishes to track. If more tonals are present, the model will track the strongest ones at each time.

Several different filter parameterizations can be used.

Estimating the AR coefficients $a_i(n)$ directly is the most intuitive approach, but it is better to use parameters that relate to only one tonal each. This can be accomplished by either using a direct frequency parameterization based on the fact that the transfer function pole angles are equal to the normalized angular notch center frequencies [5], or by writing the notch filter in cascaded form [6][7]. Here, the direct frequency parameterization is chosen. Classification of a tonal marine mammal sound is usually based on its frequency contour, that is the evolution of its dominant frequency with time [8]-[10], so this is suitable for the application. Moreover, by defining the model through the poles one obtains a representation in which the parameters are nearly independent. This permits the development of the detection methodology described in Section 3. However, cascaded filter forms promise better convergence properties [7] and will be studied in the future.

For slowly evolving tonals, the AR polynomial $A(q^{-1},n)$

$$A(q^{-1},n) = 1 + \sum_{i=1}^{P} a_i(n) q^{-i} \tag{2}$$

is necessarily monic symmetric. This implies that the transfer function of each cascaded filter stage only has M free filter parameters (ρ is usually taken as a user parameter).

We use the popular Gauss-Newton type recursive prediction error (RPE) algorithm [11] to estimate the model parameters of a direct frequency parameterized adaptive notch filter. This algorithm has several attractive properties, including a fast operation (it has been implemented in real-time), good convergence properties, and a minimal parameter variance when applied to stationary signals [11]. The properties of ANFs estimated with the RPE algorithm and applied to both stationary and non-stationary signals have also been much studied [6], [12] -[14]. With the RPE algorithm, estimation works by stepwise minimization of a cost function $\beta(n)$, which is a weighted sum of squared prediction errors,

$$\beta(n) = 2\iota(n) \sum_{m=1}^{n} \Gamma(n,m) \varepsilon^2(m) \tag{3}$$

Here, $\Gamma(n,m)$ is a weighting function that defines the

1. Initialize: $\boldsymbol{\chi}(0) = \mathbf{0}, \boldsymbol{\psi}^c(0) = \mathbf{0}, \mathbf{S}(0) = \left[100/\overline{y^2}\right]\mathbf{I}, \omega_i(0) = \pi i/(M+1)$. Design parameters: $\rho(n), \lambda(n), P(=2M)$.

2. For n=1,2,…,N do:

$$\chi_i(n) = -y(n-i) - y(n-P+i) + \rho^i(n)\varepsilon(n-i) + \rho^{P-i}(n)\varepsilon(n-P+i) \qquad ,1 \le i < M$$

$$\chi_M(n) = -y(n-M) + \rho^M(n)\varepsilon(n-M)$$

$$\varepsilon(n) = y(n) + y(n-P) - \rho^P(n)\varepsilon(n-P) - \boldsymbol{\chi}^T(n)\mathbf{a}(n-1)$$

$$\psi_i^c(n) = -y_F(n-i) - y_F(n-P+i) + \rho^i(n)\overline{\varepsilon}_F(n-i) + \rho^{P-i}(n)\overline{\varepsilon}_F(n-P+i) \qquad ,1 \le i < M$$

$$\psi_M^c(n) = -y_F(n-M) + \rho^M(n)\overline{\varepsilon}_F(n-M)$$

$$\mathbf{J}(n) = \mathrm{J}\{\boldsymbol{\omega}(n-1), \mathbf{a}(n-1)\}$$

$$\boldsymbol{\psi}(n) = \mathbf{J}(n)\boldsymbol{\psi}^c(n)$$

$$\mathbf{S}(n) = \frac{1}{\lambda(n)}\left[\mathbf{S}(n-1) - \frac{\mathbf{S}(n-1)\boldsymbol{\psi}(n)\boldsymbol{\psi}^T(n)\mathbf{S}(n-1)}{\lambda(n) + \boldsymbol{\psi}^T(n)\mathbf{S}(n-1)\boldsymbol{\psi}(n)}\right]$$

$$\boldsymbol{\omega}(n) = \boldsymbol{\omega}(n-1) + \mathbf{S}(n)\boldsymbol{\psi}(n)\varepsilon(n)$$

$$\mathbf{a}(n) = G\{\boldsymbol{\omega}(n)\}$$

$$\overline{\varepsilon}(n) = y(n) + y(n-P) - \rho^P(n)\overline{\varepsilon}(n-P) - \boldsymbol{\chi}^T(n)\mathbf{a}(n)$$

$$\begin{aligned}\overline{\varepsilon}_F(n) = {} & \overline{\varepsilon}(n) - \rho^P(n)\overline{\varepsilon}_F(n-P) - \rho^M(n)\overline{\varepsilon}_F(n-M)a_M(n) - \\ & - \sum_{i=1}^{M-1}\left[\rho^i(n)\overline{\varepsilon}_F(n-i) + \rho^{P-i}(n)\overline{\varepsilon}_F(n-P+i)\right]a_i(n)\end{aligned}$$

$$\begin{aligned}y_F(n) = {} & y(n) - \rho^P(n)y_F(n-P) - \rho^M(n)y_F(n-M)a_M(n) - \\ & - \sum_{i=1}^{M-1}\left[\rho^i(n)y_F(n-i) + \rho^{P-i}(n)y_F(n-P+i)\right]a_i(n)\end{aligned}$$

Table 1. Gauss-Newton RPLR estimation algorithm for adaptive notch filter with direct frequency parameterization [5].

analysis window and *ι(n)* is a normalization factor equal to the inverse of the window sum. The RPE algorithm is in essence an algorithm for stationary signals, so we use the window function *Γ(n,m)* to define a short analysis window within which the signal is nearly stationary. The width of the window is controlled by a design parameter known as the forgetting factor, λ. The name arises because λ controls the duration of the algorithm's memory. For a constant forgetting factor, the window is exponential and given by

$$\Gamma(n,m)=\lambda^{n-m} \tag{4}$$

The Gauss-Newton RPE-type algorithm for estimation of the pole angles of an adaptive notch filter is given by Table 1. The algorithm was first reported by Chen *et al* [5] and is based on the ANF of Nehorai [5] and the findings of Nehorai and Starer on a pole-parameterized AR model [13]. In Table 1, $\psi^{c}(n)$ denotes the negative gradient of the prediction error with respect to the vector of transfer function coefficients ***a****(n)*, $\psi(n)$ the gradient with respect to the angular notch center frequencies ***ω****(n)*,

$$\psi_i(n)=-\frac{\partial\varepsilon}{\partial\omega_i}(n) \tag{5}$$

and ***S****(n)* the inverse of the pseudo-Hessian matrix (see [11], [5] and Section 3). **G***(n)* is the mapping from ***ω****(n)* to ***a****(n)*. One can implement it by denoting $a_i=a_i^{(M)}$ and for *m*=0,1,..,*M* calculating [13]

$$\mathrm{a}_{\mathrm{i}}^{(\mathrm{m})}=\mathrm{a}_{i}^{(\mathrm{m\text{-}1})}-2\mathrm{a}_{\mathrm{i\text{-}l}}^{(\mathrm{m\text{-}1})}\cos 2\pi\omega_{\mathrm{m}}+\mathrm{a}_{i-2}^{(m-1)} \tag{6}$$

for 1≤*i*<*P*, given that $a_0^{(0)}=1$ and $a_i^{(m)}=0$ for all other *i* and *m*. **J***(n)* is the Jacobian of this mapping, which can be estimated iteratively from [13]

$$\begin{cases}\dfrac{\partial a_0}{\partial\omega_p}(n)=0\\[2mm] \dfrac{\partial a_1}{\partial\omega_p}(n)=2\sin\omega_p(n)\\[2mm] \dfrac{\partial a_i}{\partial\omega_p}(n)=2\cos\omega_p(n)\dfrac{\partial a_{i-1}}{\partial\omega_p}(n)-\dfrac{\partial a_{i-2}}{\partial\omega_p}(n)+\\ \qquad\qquad +2a_{i-1}(n)\sin\omega_p(n),\qquad 2\le i<P\end{cases} \tag{7}$$

for 1≤*p*<*M*. Further, *y*$_F$*(n)* and *ε*$_F$*(n)* are the recording and prediction error, respectively, filtered by the AR part of the notch filter, and $\bar{\varepsilon}(n)$ is the a posteriori prediction error. Using $\bar{\varepsilon}(n)$ instead of $\varepsilon(n)$ where possible improves the convergence properties of adaptive estimation algorithms [11]. See [5] for a more thorough discussion. The tonal frequencies *f*$_i$*(n)* can be estimated from

$$f_i(n)=f_s\frac{\omega_i(n)}{2\pi} \tag{8}$$

where *fs* is the sampling frequency.

The selection of the design parameters λ and ρ is important. A higher forgetting factor gives a better noise robustness, but a reduced tracking ability, and vice versa. This trade-off is similar to that which controls the choice of the pole contraction factor, wherefore a relationship between them can be determined. Previous authors have studied how to choose this relationship according to different criteria [15]–[18]. For simplicity, we choose to adopt the result of Dragosevic and Stankovic [15], which is that the optimal pole contraction and forgetting factors are equated. This result was derived assuming that components are strictly narrowband and that their frequencies evolve according to a random walk model, which describes frequency increments as small and normally distributed with zero mean. Frequency increments on tonals are not well described by a random walk model, although by the central limit theorem if we average the increments over a great many signals we might expect this to result in a Gaussian pdf. Whilst this forms a partial justification for our choice we make no claim about absolute optimality. Our limited information about the signals of interest precludes determining a relationship that is certain to give better performance.

In this study, we keep the forgetting factor λ (and, consequently, the pole contraction factor ρ) constant during the course of the whole recording. This is not the common approach on previously detected signals [12], [16]-[18]. There, one usually increases λ and ρ exponentially from a low starting value. But we do not beforehand know where we will find detections, and have no reason to change our trade-off between tracking ability and noise robustness during the recording.

3. USING THE MODEL OUTPUT FOR DETECTION

Near an extremum point of a one-dimensional function, the second derivative provides a measure of how "sharp" the stationary point is. Equivalently, a measure of the width of the extremum peak or trough is given by the inverse of the second derivative. In multiple dimensions, the analog to the second derivative is the Hessian matrix ***P****$_{ij}$* of second order partial derivatives. The Hessian of the cost function of equation (3) is

$$\mathbf{P}_{ij}(n)=\frac{\partial^2\beta(n)}{\partial\omega_i\partial\omega_j}= \\ =\iota(n)\sum_{m=1}^{n}\Gamma(n,m)\left\{\psi_i(m)\psi_j(m)+\varepsilon(m)\frac{\partial^2\varepsilon}{\partial\omega_i\partial\omega_j}(m)\right\} \tag{9}$$

The diagonal elements of the inverse pseudo-Hessian ***S****(n)* provide a measure of the peak or trough width. In the pseudo-Hessian, the second term of equation (9) has been discarded, but this is a good approximation close to an extremum point [11]. It also ensures that the diagonal elements of ***S****(n)* are positive. ***S****(n)* can then be used to measure the reliability of the current parameter estimate. This has also been formalized in the Cramer Rao theorem on a lower bound to the variance of any unbiased estimate of a stationary parameter. The Cramer-Rao lower bound is estimated from the diagonal elements of the inverse of the Fisher information matrix, which in white background noise and for stationary parameters is closely related to the Hessian. Although the formalities are not detailed here, we note that for stationary signals the diagonal elements of the inverse Hessian are related to the variance of the parameter estimates. This nice property does not strictly hold for the non-stationary signals of interest here, but if they are indeed nearly stationary within the analysis window there should still be a strong relation between the diagonal of ***S(n)*** and the estimate variances.

To use the above reliability estimation method in practice, we need to know that the estimate is close to the extremum point. There is no guarantee for this, but in the RPE algorithm updating step it is assumed that the previous parameters actually minimized the cost function [11]. If this approximation were not a good one at each time, the algorithm would lose tracking and drift off into noise space. The fact that it is much used and recognized for its good tracking performance [11], as is also found here, constitutes a heuristic validation for the approximation.

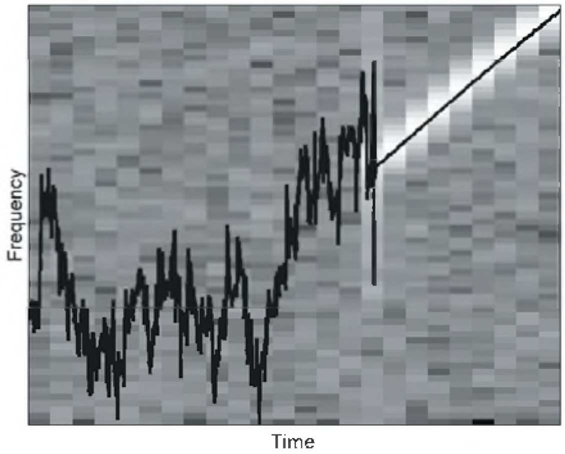


Figure 1. Gray-scale spectrogram of linear chirp in white noise, with frequency estimates overlaid (thick solid line).

For detection from a reliability measure of the parameter estimates, we need to convince ourselves that a reliable estimate only occurs when the adaptive notch filter is tracking a signal. The model is constantly looking for locally dominant frequencies, and even in white noise there are local time-frequency regions where the noise power is stronger. When signals are absent, the noise therefore produces a fluctuating parameter estimate. Figure 1 shows the evolution of a single component frequency upon moving from a noise-only section to tracking a signal (a linear chirp). It is apparent that the estimate fluctuates until the signal starts, but that as the algorithm starts to track the signal it becomes stable.

The background noise adds a stochastic component to the reliability estimate. In zero-mean noise, we can theoretically remove the effect of this by calculating the expectation value of the Hessian before inverting it. However, this appears very difficult in any practical application. It is commonplace in adaptive estimation schemes to simply ignore the expectation operation and use the "raw" quantity instead, accepting that the estimate becomes more variable. This approach is also taken here.

4. APPLICATION

To apply our detection and characterization method to a signal, we first need to pre-whiten it. Here the background noise spectrum is equalized and normalized, and constant frequency tonals are also attenuated. This is necessary because the model describes the background noise as white. It also helps to remove unwanted components such as ship noise.

In this study, pre-whitening is implemented by estimating the noise magnitude spectrum and then dividing the total magnitude spectrogram by it, thus preserving the phase information. The noise power spectrum is estimated from spectrograms of long data blocks using order statistics such as the median and trimmed mean [19]. This approach allows us to estimate the spectrum from the noise-dominated smaller values of the spectrogram only.

An alternative approach more suitable for streaming data is to estimate the noise power spectrum from a moving average on the recording spectrogram. If the window is exponential no memory of previous data is required to update the noise spectrum estimate with current data. This method is fast and simple but its spectrum estimates are easily influenced by the presence of signals. It is therefore not used here.

The pre-whitening is the only processing step that operates on the spectrogram. We therefore subsequently inverse transform to obtain the pre-whitened time waveform. Then, the RPE parameter estimation algorithm of Table 1 is run on the whole signal. It is then interrogated for the diagonal elements of the inverse pseudo-Hessian matrix, ***S(n)***, at each time instant. The detection statistic used is developed from each of these elements on a logarithmic scale,

$$\alpha_{k,raw}(n) = \log_{10} S_{kk}(n) \qquad (3)$$

This is referred to as the *raw* detection statistic. The logarithm makes its range more manageable. Note that the inverse Hessian diagonal elements decrease when the model starts to track a signal, so detections are made from small values of the detection statistic.

In white Gaussian noise (WGN), the raw detection statistics on different components are independent. Figure 2 shows histograms of the detection statistics on components 1, 2, and 3 for 100000 samples of WGN. The line shows the mean of the data fitted to a normal probability density function (pdf). As the figure shows, raw detection statistics on different components all have approximately the same distribution, which is well approximated by a normal pdf.

As predicted in Section 3, the raw detection statistic fluctuates rapidly and may be difficult to threshold. Therefore, a smoothing filter is applied to it. The smoothed detection statistic is a weighted linear combination of raw detection statistics, so is also approximately normally distributed on WGN input.

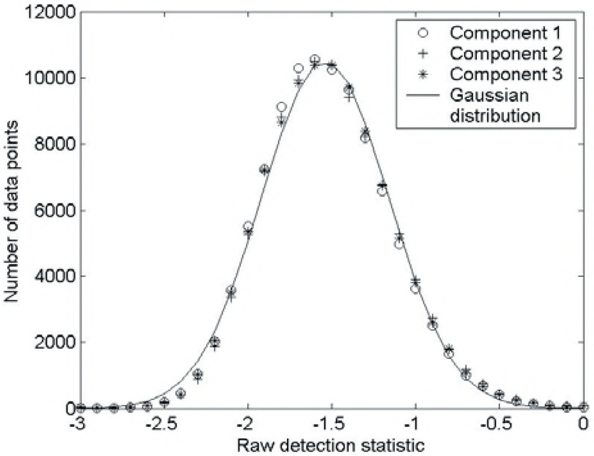


Figure 2. Histograms of raw detection statistics on 100,000 samples of WGN, fitted to a Gaussian distribution.

The means and standard deviations of the smoothed detection statistics are used to define a simple detection threshold; for detection, we require that the smoothed detection statistic is lower than a specified number of standard deviations below the mean. Here, the Page test can be employed to improve the performance [20], and this will be investigated in the future.

To as far as possible prevent signals from affecting the threshold levels, we define an equal detection threshold for all components from the highest mean estimate and the smallest estimated standard deviation. A signal lowers the detection statistic, so we expect these estimates to be the least influenced by signal presence. We estimate the mean and the standard deviation of the detection statistic by averaging over the whole recording. Since the distribution of each detection statistic is approximately normal on WGN input, we use the median as a (nearly) unbiased estimator of the mean.

On streaming data one could instead estimate the mean and standard deviation of the detection statistic via sliding window averaging. It would then be possible to prevent signals from affecting the detection thresholds by only updating the threshold estimates from non-tracking components. This approach was not taken here.

Despite the smoothing of the detection statistic, the algorithm also picks up short duration transients such as clicks. Cetacean clicks can be so much stronger than tonals that in the short processing window applied to the recording, they can contain more energy even in narrow frequency bands. Therefore, even if the model is tracking one or more tonals there is a high risk of it switching to tracking the dominant frequencies of the click. (Note that it is probably not possible to describe the click as a sum of constant frequency tonals even within our short analysis window, so the term “dominant frequencies” should be interpreted as the peak frequencies in the spectrum of the current analysis window.)

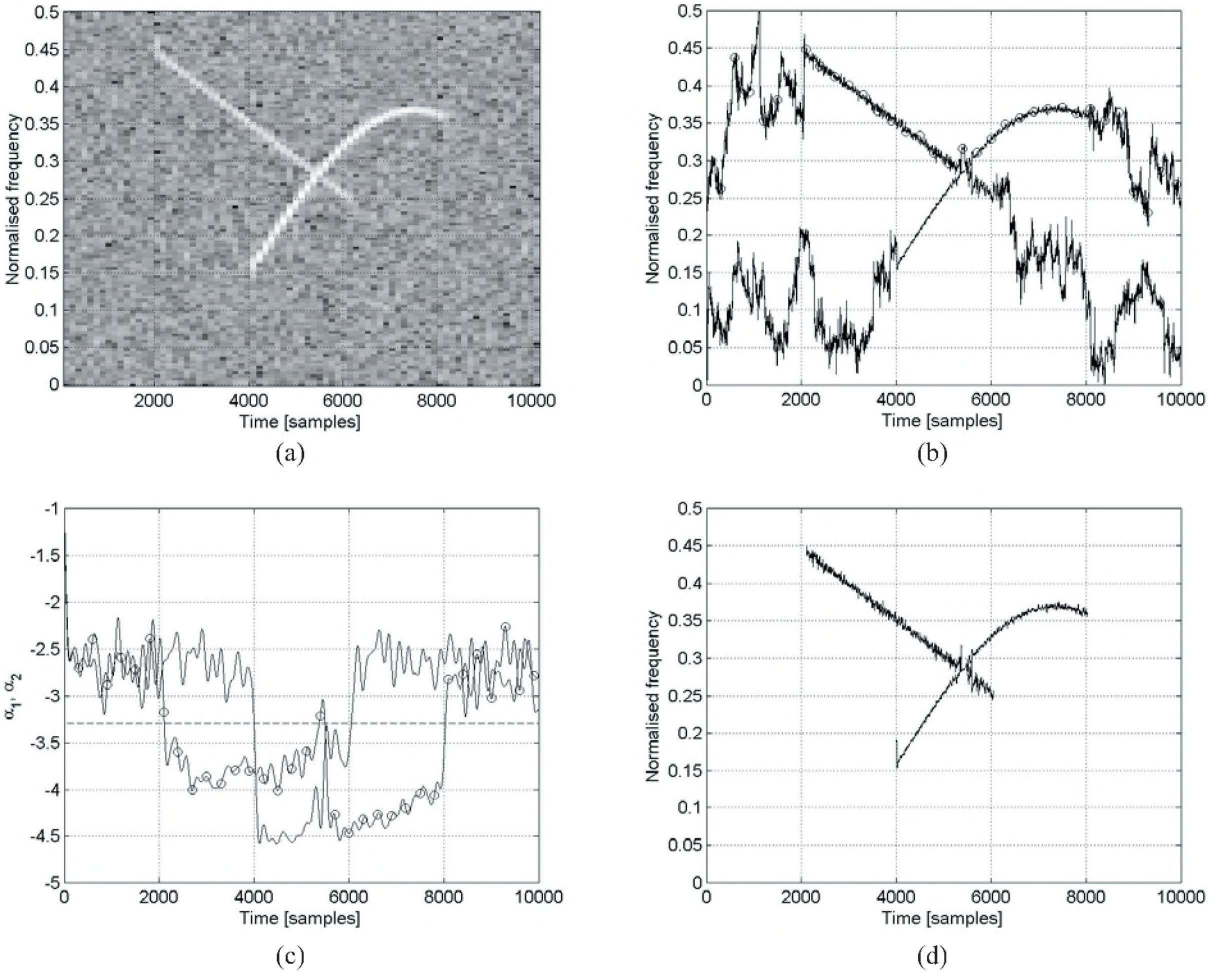
Disturbance to the tracking by clicks is of course undesirable. It could be avoided by lengthening the analysis window to reduce the ratio of click power to tonal power, but that would also reduce the tracking ability. An alternative is to introduce a pre-processing step, which detects clicks and reduces their influence. However, none of these measures have yet been implemented.

5. RESULTS

To illustrate the use of the proposed detection and characterization method, we commence by applying it to a simulated signal consisting of one linear and one non-linear chirp immersed in white noise. This signal is not intended to directly simulate a marine mammal call, although dolphin whistles usually have a narrow bandwidth and a smooth frequency evolution.

The amplitude of the linear chirp is constant at 7.9, whereas the non-linear one varies quadratically from 19.1 to 9.6 (RMS value 16.2). White noise of variance 7.9^2 is added so that the effective average SNRs for the linear and non-linear components are –3.0 and 3.2 dB, respectively.

The detection threshold is determined from the noise-only sections at the start and end of the signal. It is set at 2.5 times the estimated standard deviation below the estimated mean. This gives a low false alarm rate. The normalized cut-off frequency of the smoothing filter is 0.01, corresponding to an averaging length on the order of 100 samples.



(a)

(b)

(c)

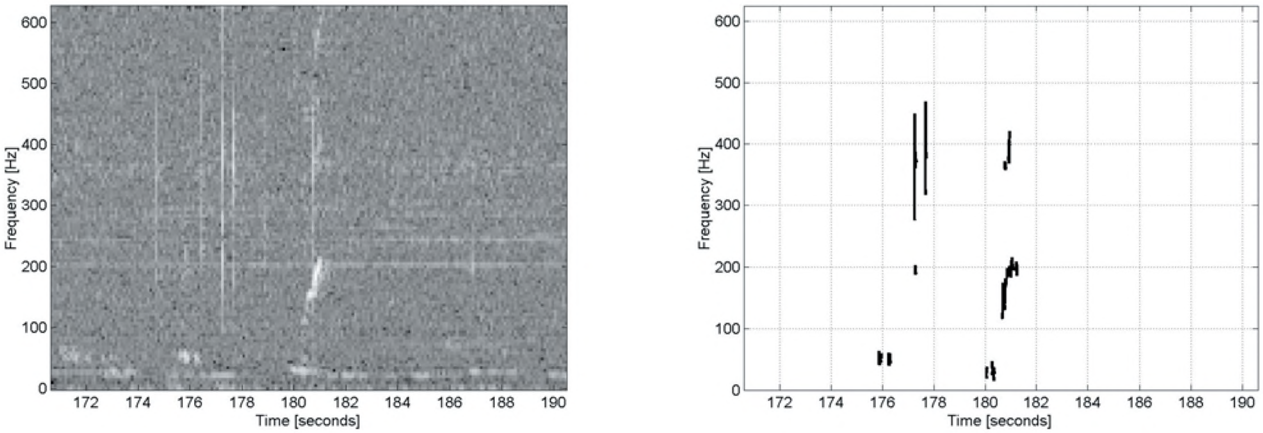
(d)

Figure 3. The proposed algorithm applied to the simulated two-component signal. (a) Spectrogram of the signal. (b) Estimated instantaneous frequencies. (c) Detection statistics and threshold level (dashed). (d) Detected components.

The results are shown in Figure 3. Subfigure (a) shows a spectrogram of the simulated signal. Estimated instantaneous frequencies can be found in subfigure (b). The evolution of the detection statistics on components 1 and 2 are shown in subfigure (c). Finally, subfigure (d) shows the detected components. It is evident from Figure 3 that the frequency estimates are highly variable in noise-only sections, but quickly lock on to signals when they appear. The proposed detection statistic provides a good measure of the estimate reliability. Note that the estimate is less variable on the stronger component. This is to be expected, and is also reflected in lower detection statistics on this component. Concluding, as subfigure (d) shows, the method is capable of detecting and characterizing simultaneous components in strong background noise.

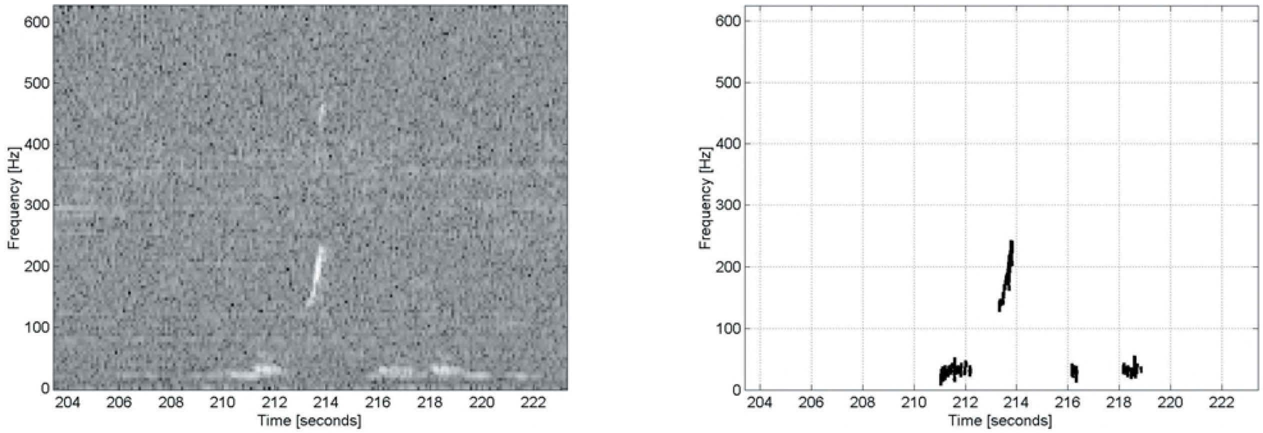
We now turn the attention to the problem of detecting marine mammal vocalizations in general, and North Atlantic right whales in particular. The "Report of the Workshop on Right Whale Acoustics: Practical Applications in Conservation" [21] classifies right whale calls as "gunshot", "low frequency", or "high frequency". A gunshot call is what is usually referred to as a "click". It is an impulsive, broadband sound of duration less than 0.5 s. The low frequency sounds are narrowband, with duration of 0.2-5.0 s and frequencies around 70 Hz. Finally, the high frequency calls have durations of 0.5-3.0 s and fundamentals at 100-600 Hz. A specific common type of high frequency call is the "FM upsweep". The duration of such a call is 0.5-1.5 s and its frequency rises monotonically in the band 100-400 Hz. The FM upsweep call is thought to be used as a contact call. It is the most well known call of the North Atlantic right whale and to the best of the authors' knowledge the only one species-specific enough for detection and discrimination from other whales with a reasonable degree of certainty.

Applying the present algorithm to the workshop dataset file L-138 results in a total of 486 detections, using the same threshold and cutoff frequency as for the simulated signal. Among these 486 detections, many are gunshot or click sounds, and many others are low-SNR calls split up into



(a) (b)

Figure 4. (a) Spectrogram of 20 s of data from dataset file L-138, centered at the detected right whale call starting at 180.6 s. (b) Detections extracted from this 20 s data batch.



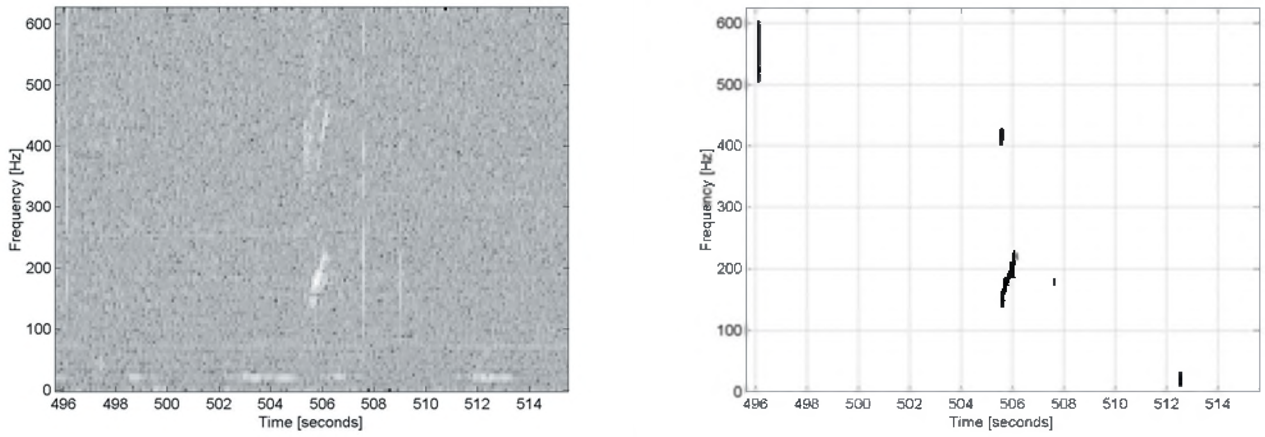
(a) (b)

Figure 5. (a) Spectrogram of 20 s from dataset file L-138, centered at the detected right whale call starting at 213.4 s. (b) Detections extracted from this 20 s data batch.

many detections. These are not of interest here as they cannot directly be used to identify HF upsweep calls. Only detections lasting more than 0.2 seconds – 155 in total – are included in the search for right whales. Among these, 86 have frequencies below 50 Hz and are likely to be fin whale calls. Out of the remaining 69 calls, 8 calls are identified as candidate right whale calls. These are between 0.3 and 1.5 seconds long, start and end within 50 to 450 Hz, and sweep up at least 50 Hz.

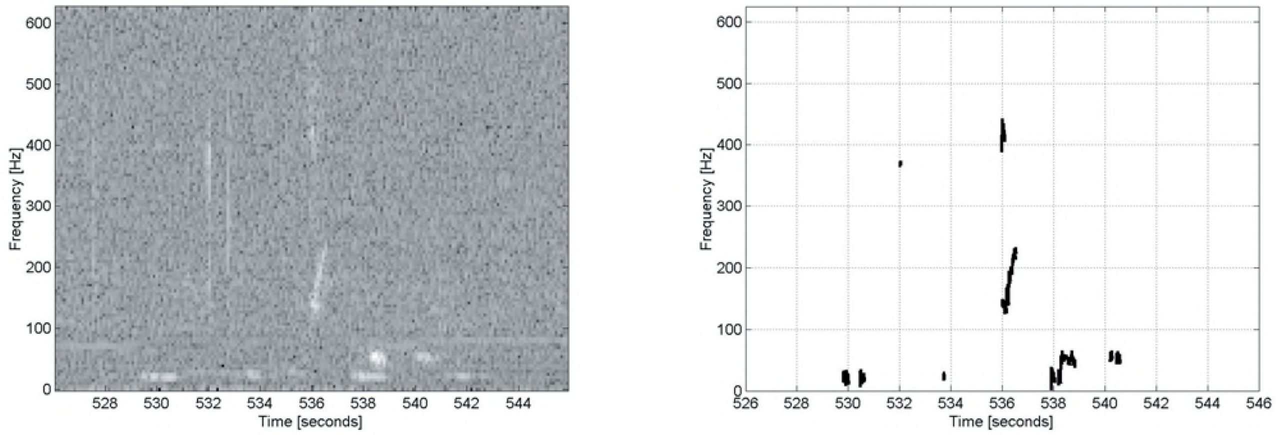
These selection criteria are based on the reported characteristics of HF upsweep calls [21], "loosened up" to allow for algorithm imperfections. Three of these 8 candidate right whale calls could directly be discarded because their frequency evolutions started low and almost immediately jumped to a nearly constant higher frequency. This is probably caused by the detection firing too early on a strong and suddenly onset call. Future fine-tuning of the algorithm should alleviate this problem.

Four of the remaining sounds sweep up from 120-140 Hz to 200-220 Hz in 0.4-0.6 seconds. The authors believe that these are right whale calls. Their start times are approximately 180.6, 213.4, 505.6, and 536.0 s. The last remaining candidate call, starting at 63.3 s, sweeps up from approximately 80 to 150 Hz in 0.4 s. This is probably too low frequency for a right whale HF upsweep call.



(a) (b)

Figure 6. (a) Spectrogram of 20 s from dataset file L-138, centered at the detected right whale call starting at 505.6 s. (b) Detections extracted from this 20 s data batch.



(a) (b)

Figure 7. (a) Spectrogram of 20 s from dataset file L-138, centered at the detected right whale call starting at 536.0 s. (b) Detections extracted from this 20 s data batch.

Subfigures (a) of Figures 4-7 show spectrograms of 20 seconds of data centered on each right whale call; the data has been pre-whitened. Estimated frequency evolutions of all detections, that is also those shorter than 0.2 s, within these 20 seconds of data are shown in subfigures (b). These figures show that the proposed algorithm is able to track the frequency contour of the right whale calls. There are also several fin whale detections and some brief click detections. In Figures 4,6, and 7, note that the algorithm has also picked up on what are probably harmonics of the right whale call.

6. CONCLUSIONS

In this paper, we have described a new detection and characterization method for tonal marine mammal vocalizations, and have shown that the method works well with simultaneous sounds, in low signal-to-noise ratios, and with sounds, such as right whale calls, that do not appear to be strictly narrowband.

The method is simple to use and controlled by only a small number of user parameters. It has not yet been implemented in hardware, but in an off-line software implementation it processes data at a rate that exceeds that necessary for real-time implementation at sample rates of 50-60 kHz.

The algorithm picks up and is disturbed by click sounds, so for a fully automatic operation it is necessary to attenuate these prior to application. Also, despite the smoothing of the detection statistic, calls are sometimes divided into several detections. To counteract this, one can apply a detection-merging algorithm, or change the detection criteria. These improvements will be studied in the future.

ACKNOWLEDGEMENTS

We wish to extend our gratitude to EPSRC and ISVR for funding this research, and to QinetiQ for their interest and financial support.

REFERENCES

[1] M. Fujiwara and H. Caswell. Demography of the endangered North Atlantic right whale. Nature, 414(6863):537–541, Nov 2001.

[2] 500-yard protection zone to prevent human activity from disturbing endangered right whales. United States National Oceanic and Atmospheric Administration (NOAA) press release 97-R107, Nov 1997.

[3] A. T. Johansson and P. R. White. Characterisation of cetacean whistles by parametric modelling. In *Proceedings of the 17th annual conference of the European Cetacean Society (ECS)*, Las Palmas de Gran Canaria, Spain, March 2003.

[4] A. T. Johansson, P. R. White, and R. E. Sel- way. Automatic cetacean sound classification –application to sonar transient false alarm reduction and marine environmental monitoring. In *Proceedings of Underwater Defense Technology (UDT) Europe*, Malmo, Sweden, June 2003.

[5] B. S. Chen, T. Y. Yang, and B. H. Lin. Adaptive notch filter by direct frequency estimation. *Signal Processing*, 27:161-76, 1992.

[6] K.-H. Rew, S. Kim, I. Lee,and Y. Park. Real- time estimations of multi-modal frequencies for smart structures. *Smart materials and structures*, 11(1):36-47, Feb 2002.

[7] G. Li. A stable and efficient adaptive notch filter for direct frequency estimation. *IEEE Transactions on Signal Processing*, 45(8):2001-9, Aug 1997.

[8] V. M. Janik. Pitfalls in the categorization of behaviour: A comparison of dolphin whistle classification methods. *Animal Behaviour*, 57(1):133-43, Jan 1999.

[9] D. M. Mellinger and C. W. Clark. Recognizing transient low-frequency whale sounds by spectrogram correlation. *Journal of the Acoustical Society of America*, 107(6):3518-29, June 2000.

[10] S. Datta and C. Sturtivant. Dolphin whistle classification. *Signal Processing*, 82(2):251-8, 2002.

[11] L. Ljung. *System identification: Theory for the user.* 2^{nd} Edition, Prentice-Hall, 1999.

[12] A. Nehorai. A minimal parameter adaptive notch filter with constrained poles and zeros. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(4):983-96, 1985.

[13] A. Nehorai and D. Starer. Adaptive pole es- timation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(5):825-38, 1990.

[14] T. S. Ng. Some aspects of an adaptive digital notch filter with constrained poles and zeros. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2):158-61, Feb 1987.

[15] M. V. Dragosevic and S. S. Stankovic. An adaptive notch filter with improved tracking properties. *IEEE Transactions on Signal Processing*, 43(9):2068-78, 1995.

[16] P. Stoica and A. Nehorai. Performance analysis of an adaptive notch filter with constrained poles and zeros. *IEEE Transactions on Acoustics, Speech and Signal Processing* 36(6): 911-9, 1988.

[17] P. Tichavsky and P. Händel. Two adaptive algorithms for adaptive retrieval of slowly time-varying multiple cisoids in noise. *IEEE Transactions on Signal Processing*, 43(5):1116-27, 1995.

[18] P. Händel and A. Nehorai. Tracking analysis of an adaptive notch filter with constrained pole and zeros. *IEEE Transactions on Signal Processing*, 42(2):281-91, 1994.

[19] T. S.-T. Leung and P. R. White. *Mathematics in Signal Processing IV*, pages 369-82. Clarendon Press, Oxford, 1998.

[20] D. A. Abraham. Analysis of a starting time estimator based on the Page test. *IEEE Transactions on Aerospace and Electronic Systems* 33(4): 1225-34, 1997.

[21] D. Gillespie and R. Leaper, eds. Report of the Workshop on Right Whale Acoustics: Practical Applications in Conservation. Woods Hole Oceanographic Institution, March 2001.