



Detection of sperm whale clicks based on the Teager–Kaiser energy operator

V. Kandia^{a,*}, Y. Stylianou^{a,b}

^a *Foundation for Research and Technology Hellas (FORTH), Heraklion Crete, Greece*

^b *University of Crete, Computer Science Department, Heraklion Crete, Greece*

Received 27 January 2006; received in revised form 13 April 2006; accepted 21 May 2006

Available online 13 July 2006

Abstract

The development of an algorithm for automatic detection of sperm whale clicks in large recordings is described. It is based on the Teager–Kaiser (TK) energy operator and it is able to detect efficiently creaks as well as regular clicks. A matching filter is first used as a pre-processor in order to overcome the difficulties caused by the multi-pulse structure of regular clicks. Next, the TK energy operator is applied to the output of the matching filter. A first selection of clicks is performed based on statistical measurements on the TK output, while the final selection is carried out by a forward–backward search algorithm. The proposed system has been tested on a total duration of 25 min of data containing regular clicks as well as creak clicks, where the location of clicks has been marked by hand. An average rate of 94.05% of correct detections was achieved by comparison with the hand-labeled data created from the tested files. Comparing to a standard method used for the same task, the proposed algorithm is more effective in detection rate by 30% and much more accurate and robust.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Sperm whale; Click; Teager–Kaiser energy operator; Detection; Enhancement

1. Introduction

Sperm whales (*Physeter macrocephalus*) are among the most vocally active marine mammals, therefore are well suited for passive acoustic surveys. Their distinct vocalizations are made up mostly of impulsive click-type sounds (Chapter 5 in [1], [2–4], although

* Corresponding author.

E-mail addresses: vkandia@ics.forth.gr (V. Kandia), yannis@csd.uoc.gr (Y. Stylianou).

there is strong evidence that they produce other kinds of sounds as well [5–7]. Sperm whale recordings, once quantified, may have a range of uses: animals localization [8,9], study of animals behavior [1,10], abundance estimation [11,3,6], length estimation [12,13]. They can also help us to understand their sound production mechanism [2,14–18] as well as to identify sperm whales acoustically in an automatic way [19]. Many hours of vocalizing sperm whales have been recorded using various configurations: two (or more) towed hydrophones a few meters apart deployed behind a monitoring vessel [8,20], towed-array systems [21–23], or even tags [14,24]. In order to extract and evaluate the information contained in these recordings one major step should be done: find the clicks amid the hours of data. One can visually inspect the waveforms or spectrograms of the signals but a more practical solution consists of using an automatic detection algorithm. The advantages of the latter are well described in [25]. Given the importance of the detection of clicks, several softwares for automatic detection of sperm whale sounds have been developed (i.e., Rainbow Click [26], Moby Click [27,3], Ishmael [28]). Most of the existing methods have a high rate of false detection especially when the recorded clicks do not conform to the expected click pattern or when the recordings have a low signal-to-noise ratio (SNR). Usually this is the case when creak clicks are analyzed. Moreover, the user has to deal with a great number of parameters when using these methods, and this constitutes a drawback.

In this paper we present the application of the Teager–Kaiser (TK) energy operator [29] for the automatic detection of clicks from sperm whales. In [30], the TK energy operator has been tested on synthetic data as well as on real recordings. The TK energy operator is characterized by instantaneous tracking capability by using only three consecutive signal samples. Most of our attention in [30] was on the clicks from the creak sounds since the characteristic low sound level of this type of clicks and their short inter-clicks intervals, makes their detection harder. Moreover, it has been applied on isolated windows of analysis. In this work, we extend the use of the energy operator on the automatic detection of clicks in large recordings. Extended tests were conducted using data provided by the Naval Undersea Warfare Center (NUWC) and collected at the Atlantic Undersea Test and Evaluation Center (AUTECE). The proposed automatic system is able to detect clicks from creak areas as well as from regular clicks. It has been observed that the multi-pulse structure of regular clicks often combined with various reflections causes problems in the detection task. Therefore, we first used an echo cancelation algorithm implemented as a matching filter and then, we applied the TK energy operator to the output of the filter. The selection of clicks was made in three steps. In the first one, a selection was performed based on the statistics of samples of the new signal (TK output). In the second one, we applied a peak picking algorithm and in the last step, a forward–backward search algorithm was applied with respect to the time instant of the highest peak which was assumed to be a click sound. To test the efficiency of our system we have manually marked the clicks on some of the recordings (for regular and creak clicks). Moreover, the performance of the proposed algorithm is compared to the output produced by the Rainbow Click software [26].

A short review of the Rainbow Click detector and the TK energy operator is described in Section 2. Section 3 details the steps of the proposed algorithm. In Section 4, the data set and the hand-labeled data used for the evaluation of the detection algorithm are presented along with the results from the comparison between the automatic detection made by different algorithms and the detection made by hand. Discussion on these results and future work concludes the paper.

2. Background

2.1. Rainbow click detector

Rainbow click performs the detection task in mainly, two steps. In the first step, referred to as “First Level Trigger”, the input signal is rectified and then is passed through a low-pass filter of first order (realized as an all-pole model) of Z -transform:

$$H(z) = \frac{a}{1 - (1 - a)z^{-1}} \quad (1)$$

Coefficient a is chosen accordingly ($0 < a < 1$); for the transient clicks it is usually close to 1, while if $a \ll 1$, the filter provides a measure of the background noise level. At every time instant two signals are obtained: the estimated click signal, $g_1(n)$ with $a \approx 1$, and an estimation of the noise signal, $g_2(n)$, with $a \ll 1$. An estimation of the mean noise level is obtained by $g_2(n)$. The enhanced signal is obtained by subtracting the estimated noise level, $g_2(n)$, from the click signal $g_1(n)$:

$$g(n) = g_1(n) - g_2(n) \quad (2)$$

The start of a click signal is considered the time instant when the amplitude of the enhanced signal passes over a given threshold (referred to as “On Threshold”), while a click ends when the amplitude drops below another second threshold (referred to as “Off Threshold”). These thresholds are set to a number of standard deviations above the measured mean noise level.

Since clicks often contain several pulses, a minimum number of samples of click separation is set, while other rules are used for splitting or merging (i.e., the “Maximum Click Length”) the detected click sounds.

The second step is a selection process of the clicks detected in the first step. Clicks are first filtered by an user-defined bandpass filter. Filtered clicks with an energy greater than a threshold are accepted as clicks, while the others are rejected.

Two mechanisms are also available for eliminating noise from the propeller of a vessel (“Forward Veto”) and loud signals from other sources (i.e., seismic survey vessels). Note that all the parameters can be controlled by the user.

In the Rainbow Click detector the first step is very important. From the above description it follows easily that the two filtering procedures can be combined into one filter with Z -transform:

$$H(z) = (a_c - a_n) \frac{1 - z^{-1}}{1 - (2 - a_c - a_n)z^{-1} + (1 - a_c)(1 - a_n)z^{-2}} \quad (3)$$

where a_c and a_n denote the coefficient a used for the transient and the noise signal, respectively. Fig. 1 shows the magnitude of the frequency response of the combined filter for $a_c = 0.9$ and $a_n = 0.1$. Obviously, this is a high-pass filter, which is, however expected since the two low-pass filters (the one for the click sounds and the other for the estimation of the noise level) differ only in bandwidth which is controlled by a . Hence, a high value of a results in a low-pass filter of wide bandwidth while a low value of a provides a low-pass filter with a narrow bandwidth.

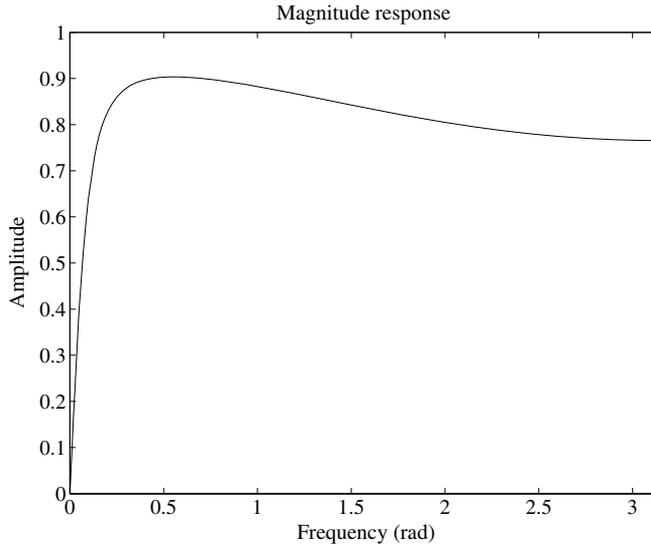


Fig. 1. Magnitude of the frequency response of the combined filter used in the first step of the Rainbow click detector, for $a_c = 0.9$ and $a_n = 0.1$.

2.2. The Teager–Kaiser energy operator

The Teager–Kaiser (TK) energy operator is defined in the continuous domain as:

$$\Psi[x(t)] = \left(\frac{dx(t)}{dt}\right)^2 - x(t) \frac{d^2x(t)}{dt^2} = \dot{x}^2 - x\ddot{x} \tag{4}$$

where \dot{x} and \ddot{x} denotes the first and second derivative over time, respectively. For a discrete time signal, it is shown in [29] that the TK energy operator is given by

$$\Psi[x(n)] = x^2(n) - x(n+1)x(n-1) \tag{5}$$

where n denotes the sample number. An important property of the TK energy operator in (5) is that it is nearly *instantaneous* given that only three samples are required in the energy computation at each time instant: $x(n-1)$, $x(n)$, and $x(n+1)$.

The operator is referred to as *energy* operator since it is related to the concept of energy in the generation of acoustic waves [29]. In that context, by energy it does not mean simply the mean square value of the signal but is referring to the actual physical energy required to produce the signal [31]. The TK energy operator can be seen as a special case of a family of quadratic energy operators defined by:

$$E(n) = \sum_{m=0}^{N-1} x(n+m)x(n-m)h(m) \tag{6}$$

where $h(m)$ is an N -point set of quadratic filter coefficients. It is easily seen from (5) and (6) that the TK energy operator is obtained when:

$$h(m) = \begin{cases} 1 & m = 0 \\ -1 & m = 1 \\ 0 & \text{otherwise} \end{cases}$$

The frequency response of the impulse response is given then by:

$$H(e^{j\Omega}) = e^{-j\Omega/2} \sqrt{2(1 - \cos \Omega)}$$

In Fig. 2 the magnitude of the frequency response of the TK energy operator filter is plotted. This is a high-pass filter and therefore, the output of this filter will mainly contain the high frequencies of the input.

Despite the similarities of the two filters shown in Figs. 1 and 2, we must note that none of them represents a simple and usual linear filter, and therefore, a direct comparison between them is not possible. Indeed, the one used in the Rainbow click is applied on the rectified input signal (i.e., the absolute values of the input signal which a non-linear operator) while the filter used in the TK-energy operator is a quadratic filter.

In the case where the input is a random signal, the statistical properties of the TK energy operator should be studied. If an input signal, $u(t)$, is a realization from a wide sense stationary Gaussian process with a power spectrum $S_{uu}(\Omega)$

$$S_{uu}(\Omega) = \int_{-\infty}^{\infty} R_{uu}(\tau) e^{-j\Omega\tau} d\tau$$

where $R_{uu}(\tau)$ denotes the autocorrelation function of the process, it is shown in [30] that the output of the TK energy operator will also follow a Gaussian distribution:

$$\Psi[u(t)] \sim \mathcal{N}(\mu_{\Psi}, \sigma_{\Psi}^2)$$

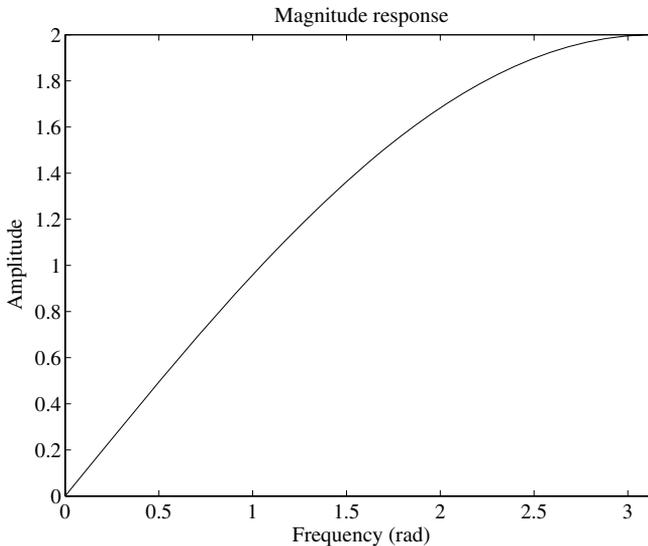


Fig. 2. Magnitude of the frequency response of the TK-energy operator filter.

where the mean value, μ_Ψ , is given by:

$$\mu_\Psi = \mathcal{E}\{\Psi[u(t)]\} = \frac{1}{\pi} \int_{-\infty}^{\infty} \Omega^2 S_{uu}(\Omega) d\Omega \tag{7}$$

and the variance, σ_Ψ^2 by:

$$\sigma_\Psi^2 = 3 \left(\frac{d^2 R_{uu}(0)}{d\tau^2} \right)^2 + \sigma_u^2 \sigma_{\ddot{u}}^2 \tag{8}$$

where $\sigma_{\ddot{u}}^2$ denotes the variance of the second derivative over time of $u(t)$ and is obtained by:

$$\sigma_{\ddot{u}}^2 = \frac{d^4 R_{uu}(0)}{d\tau^4} \tag{9}$$

which is the fourth derivative of the autocorrelation function at $\tau = 0$. Details on the derivation of the above formulas can be found in [30].

In this paper, the recorded signal, $s(n)$, is assumed to have three components: an interference signal (usually this is considered to be a low frequency signal), $x(n)$, a transient signal, $y(n)$, and background noise, $u(n)$:

$$s(n) = x(n) + y(n) + u(n) \tag{10}$$

Applying the TK energy operator on $s(n)$ we obtain [30]:

$$\Psi[s(n)] = \Psi[x(n)] + \Psi[y(n)] + \Psi[u(n)] + T(n) \tag{11}$$

where $T(n)$ denotes the sum of all cross terms between the possible pairs of the input components ($x(n)$, $y(n)$, and $u(n)$). For example, the cross term, Ψ_c , between $x(n)$ and $y(n)$ is defined by:

$$\Psi_c[x(n), y(n)] = x(n)y(n) - x(n+1)y(n-1) \tag{12}$$

It is worth to note that when the transient signal is modeled by a periodic, with period P , train of pulses:

$$y(n) = \sum_{k=-\infty}^{\infty} \delta(n - kP) \tag{13}$$

where

$$\delta(n - l) = \begin{cases} 1 & n = l \\ 0 & n \neq l \end{cases} \tag{14}$$

and taking into account the high-pass character as well as the statistical behavior of the operator, it can be shown that the output of the TK operator in (11) is approximately given by [30]:

$$\Psi[s(n)] \approx \Psi[y(n)] + w(n) \tag{15}$$

where $w(n)$ is a Gaussian random signal with characteristics given by (7) and (8).

From (15) it follows that when a transient signal, $y(n)$, is present, the probability density function (pdf) of $\Psi[s(n)]$ will deviate from a typical Gaussian curve (bell-shaped). Indeed, in this case the output pdf will be better approximated by heavy tail distributions (e.g., Laplacian density). Following this observation, a simple criterion based on measurements

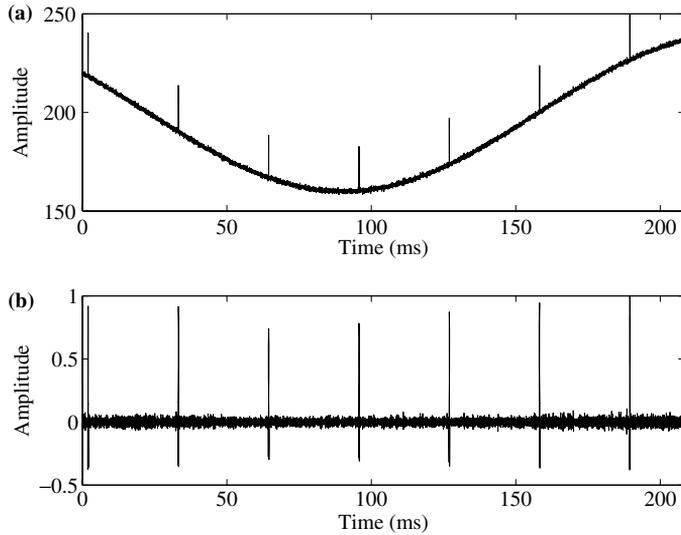


Fig. 3. (a) The global input to the TK-energy operator and (b) the global output from the TK-energy operator. Note that the maximum signal amplitude in (b) has been normalized to unity.

like skewness can be used for detecting the presence of transient components in the signal. Moreover, the above comment implies that the output of the TK energy operator will be dominated by the transient signal (even for low transient-to-noise ratios). A typical example of a synthetic signal (low-pass interference signal, periodic train of pulses and Gaussian noise) and the output from the TK energy operator is depicted in Fig. 3. As it is expected the output is dominated by the ideal transient signal.

Although a real click sound is not an ideal pulse as in (14), it is still a wideband signal (narrow support in the time domain); therefore the application of the TK energy operator on a signal containing all these components will produce an output mainly dominated by the energy of the transient signal.

3. Detection algorithm

Although sperm whale clicks are mainly characterized by high regularity in their production (i.e. periodicity) they are non-stationary signals; their periodicity (usually referred to as Inter-Click interval, ICI) changes over time. However, considering short window of analysis with a few number of clicks inside the window (about 4–6 clicks) the stationary hypothesis for the windowed signal may hold (in general). Moreover, short analysis windows simplifies the detection algorithm while reducing the computation time and memory allocation needs for processing the input audio files.

The proposed detection system is depicted in Fig. 4. We carried out a frame by frame analysis where the window size was determined using the previous estimated ICI. For initialization of the algorithm an initial estimated value for ICI of 0.8 s was used. The application of TK energy operator on a complex signal like the multi-pulse sperm whale regular click raises some serious difficulties. According to the existing theory for sound generation

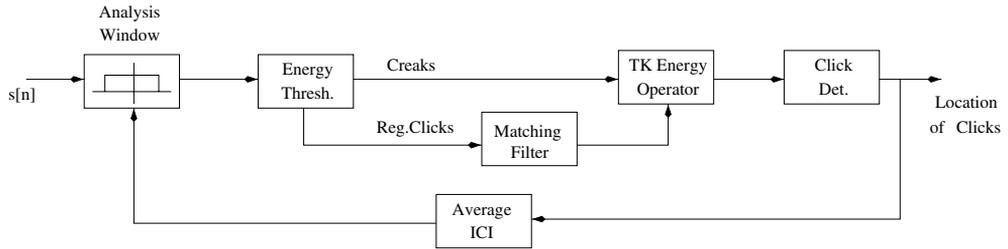


Fig. 4. Block diagram of the proposed detection system based on the Teager–Kaiser energy operator.

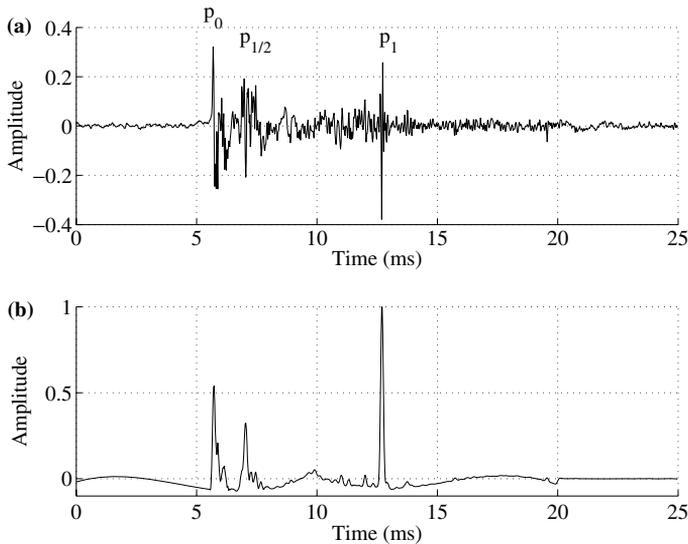


Fig. 5. (a) Waveform of a regular (although not a typical one) sperm whale click and (b) its TK output (maximum amplitude has been normalized to unity).

of sperm whale clicks, the main pulse¹ (referred to as p_1) contains most of the signal energy [15]. However, this is not always the case [18,17]; the directionality of regular clicks combined with the various conditions under which the recordings were made, may result in a different pattern of pulses than the usual one, where p_0 is the dominated pulse. According to Zimmer et al. [17], this is the case when the recording aspect is close to the acoustic axis of p_0 and off the axis of the p_1 pulse (recordings made behind the animal). Furthermore, when recordings are made on the off-axis of a sperm whale a quite complex (noisy) waveform is obtained [17]. In Fig. 5(a) an example of a regular click is shown where both pulses (p_0 and p_1) are quite strong. Obviously, this signal does not conform to the usual case of clicks discussed in [21] or in [18,17].

To deal with the aforementioned variability in the distribution of energy inside a click sound, we define as *click instant* the onset time of the click, in other words the time instant

¹ Our notation on pulses follows that of [21,17].

of the p_0 pulse, independently of its power (weak or strong). Therefore, in this paper, click detection is synonymous to the detection of the first pulse of the click. Fig. 5(b) depicts the output of the TK energy operator when the input is the signal shown in Fig. 5(a). Here, it is worthwhile to mention the clean pattern of pulses in the output signal. Actually, between the p_0 and p_1 pulses, an additional pulse is noticeable. Given that the system recordings were in the far field and following the analysis of clicks presented by Zimmer et al. [17], this could be the pulse referred to as $p_{1/2}$, which seems to be an orientation dependent delay relative to the p_0 pulse [17].

Energy-based criteria for the detection of clicks are quite common. Despite the clean pattern of pulses obtained in Fig. 5(b), if a criterion of maximum energy were applied to the output of the TK energy operator, a detection error of about 7 ms would be obtained. In an attempt to overcome this problem and increase the signal to noise ratio (SNR), and motivated by the fact that a sudden excitation of a stable system by an impulse signal will produce as output an oscillation of decaying with time amplitude, we suggest the use of a simple matching filter with impulse response given by:

$$s(t) = \cos(2\pi ft)e^{-\alpha t} \text{ for } t > 0 \quad (16)$$

where frequency f was chosen to be 1000 Hz by inspection of the existing oscillation into some waveforms with regular clicks and the damping factor α was given by the formula

$$\alpha = \frac{f}{0.12N_p} \quad (17)$$

where N_p represents the number of periods (typical value used in the paper was 20). It must be emphasized, however, that the chosen value for f does not represent a critical value, neither the constant (0.12) involved in the definition of the decay rate α in (17). Fig. 6 shows the impulse response of the matching filter using the above values for f and N_p .

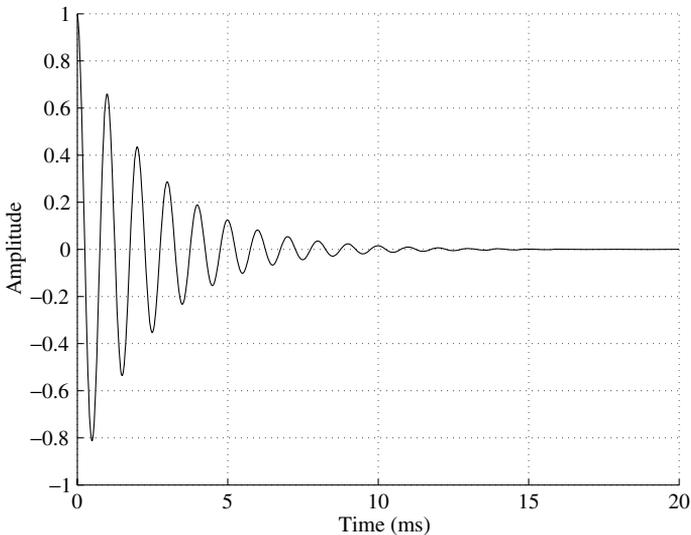


Fig. 6. Impulse response of the matching filter.

Although the matching filter is not optimized according to the current analysis window (adaptation) resulting, therefore, in a lower (comparing to the optimal case of adaptation) SNR, we have found that this simple matching filter provides good results since it is similar to the expected ideal impulse response of a click sound. Fig. 7 elucidates how the application of the matching filter facilitates the detection of the beginning of a regular click. The illustrated regular click is the same with the one shown in Fig. 5(a).

For creak sounds, the presence of oscillation after the pulse (click) is still valid. However, the low energy of creak clicks combined with the power of recorded background noise, results in a very low SNR, making useless the application of the matching filter. Actually, the application of a matching filter under so adverse SNR conditions leads, most of the time, in erroneous detection of click onset times. On the other hand, creak clicks are mainly mono-pulsed signals and so, they do not pose the above-mentioned problems [2]. Since creak clicks have a much lower sound level than regular clicks, the matching filter was applied only when the input signal was above a certain energy threshold in order to prevent its application on creaks.

The existence of clicks in the analyzed signal is not a necessary condition for the algorithm. As it is shown in the previous Section, the presence of a transient (click) signal will modify the distribution of the TK output ($\Psi[u(n)]$) causing a negative skewness. Therefore, by measuring the skewness of the TK output signal, a detector of the presence of the click sound is obtained. The threshold used in the skewness criterion was estimated after the statistical analysis of a selected number of data used in the current work. As a result, the algorithm does not waste time and effort on click detection where only noise is present.

The final steps for the detection of clicks are depicted in Fig. 8. Since the presence of the transient signal modifies the mean value of the noise distribution, we formulate a standard maximum-likelihood decision criterion [32,33] (also known as a test of the mean) in order

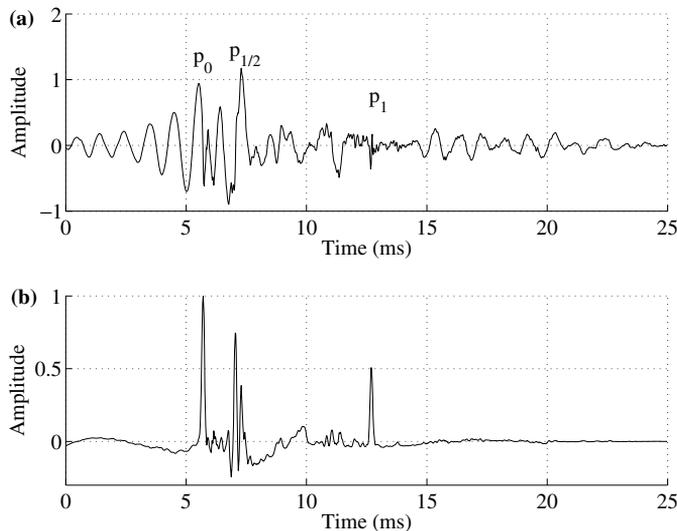


Fig. 7. After the application of the matching filter: (a) regular click and (b) its TK output (maximum amplitude has been normalized to unity).

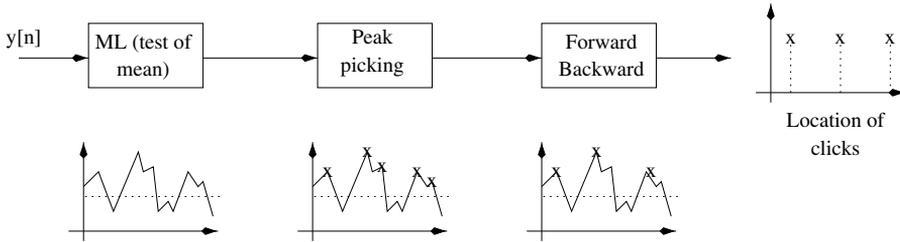


Fig. 8. Block diagram of click detection.

to determine a threshold (decision rule) for possible click values. A peak-picking procedure is then applied so as to separate the peaks of the signal from all the other signal values that passed the threshold. Finally, making the assumption that the highest peak corresponds to a click sound, we applied a forward and backward search with reference to the time instant of the highest peak. Note that periodicity of clicks cannot be used for detecting efficiently the location of clicks (although this could simplify by far the detection algorithm) since periodicity may change quickly even inside an analysis window. This is especially true during the transition from an area of quasi-regular clicks into a creak area. A mathematical description of the forward search algorithm follows.

Let $t_i, i = 1, \dots, N$ be the position (location) of all peaks selected by the peak-picking procedure and let t_m be the position of the peak with the highest amplitude among them. The observation that regular clicks contain a certain number of pulses in a close distance, led us to the following steps. We calculated the differences

$$d_{i,m} = t_i - t_m \tag{18}$$

and set a threshold (Thr) in order to separate the various peak groups

$$\text{Thr} = 0.1 \max_i \{\dot{d}_i\} \tag{19}$$

where

$$\dot{d}_i = d_{i,m} - d_{i-1,m} \tag{20}$$

is the first-order forward discrete-time derivative over consecutive time peak locations. The first peak of the next group will be positioned at

$$t_j = \min\{d_{i,m} > \text{Thr}\} \tag{21}$$

The differences

$$d_{i,j}^* = t_i - t_j, i = 1, \dots, N \tag{22}$$

help us isolate the next group of peaks which will contain all peaks for which the condition $0 \leq d_{i,j}^* < \text{Thr}$ is true. Among them, the peak with the highest amplitude will be considered as being the next click (the time instant of p_0 pulse or, as mentioned previously, the click onset time).

After estimating the location of a click, the next group of peaks is detected and the position of the next click is estimated. This process continues until the end of the analysis window is reached. The backward search algorithm is applied in exactly the same way but in the opposite direction; from the time instant of the first estimated click (t_m) to the beginning of the analysis window.

4. Application

4.1. Data Set

The data on which the above detection system was applied were provided by the Naval Undersea Warfare Center (NUWC) and collected at the Atlantic Undersea Test and Evaluation Center (AUTECE), Andros Island, Bahamas. We have used 25 min of recordings made with Hydrophone G, one of the five hydrophones belonging to the Data set # 2. The constellation of hydrophones can be seen in [34]. The audio files (five files with 5 min duration each) were sampled at 48 kHz with accuracy of 16 bits/sample and contained clicks from only one sperm whale. Reverberation was often noticed in the available recordings.

4.2. Hand labeled data

All five files were scanned manually and the starting point of each click sound was marked with the aid of Sound Forge 6.0 (Sonic Foundry). Then, the position of the markers was extracted in milliseconds and saved in a text format. Fig. 9 illustrates what was considered to be the start of (a) a regular click and (b) a creak click, respectively. To increase accuracy in labeling clicks for creak sounds the output sampling frequency was heavily reduced (i.e. from 48 kHz input sampling frequency to 2 kHz).²

4.3. Results

It is worth to note that the proposed detection algorithm is completely automatic while the values of the design (input) parameters like α in (17) or Thr in (19) are not really critical in the efficiency of the detection algorithm. On the other hand, the algorithm relies on a strong assumption; the highest peak in the output of the TK energy operator corresponds to the p_0 pulse of a click sound. Although this is true in general, especially after the application of the matching filter, it may result in erroneous click location if other sources of impulsive noise are present in the signal.

For comparison reasons, the audio files were also processed by the Rainbow click detector. In this test the version 3.01.0013 of the click detector was used. Design parameters were chosen by trials and errors in order to get the best results in terms of detection accuracy. The best performance was obtained by the following parameters. For all the files, coefficient a is taken as 0.99 and 0.001 for the “Signal Filter” and the “Noise Filter”, respectively. To ensure that the whole click is extracted, Rainbow click requires two parameters, referred to as “Presample” and “Postsample”, that are the number of samples added to the detected click area. In addition, the expected maximum length of a click should be given. For all analyzed files, “Presample”, “Postsample”, and expected “Maximum Length” of a click, were taken as 0.1 ms, 0.2 ms, and 21 ms, respectively. Parameters “On Threshold” and “Off Threshold” were set to different values for the first two audio

² Lowering the output sampling frequency (this is not a downsampling, however) of an audio signal has the effect of playing the sound in a lower than the original rate, since the original time is expanded. This facilitates the auditory and visual inspection of fast moving acoustic events. For the case of creaks, this is very important for as an accurate as possible manual labeling.

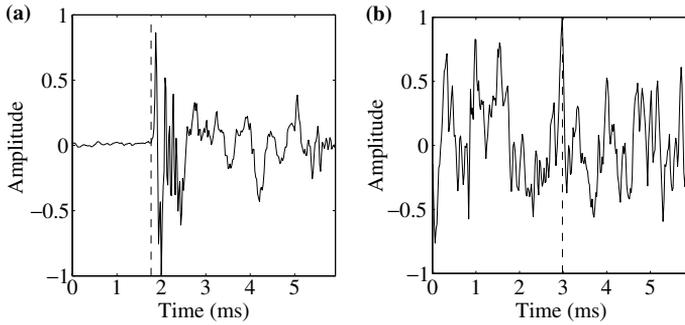


Fig. 9. Hand labeled clicks. Labels are indicated by a dashed line: (a) regular click and (b) creak click.

Table 1
Percentage (%) of correctly identified clicks per file

File name	Clicks	TK		RB	
		Score (%)	Clicks	Score (%)	Clicks
10M_ch2_0–5.wav	266 (0)	100	268	94.74	265
10M_ch2_5–10.wav	944 (549)	60.17	986	15.68	781
10M_ch2_10–15.wav	689 (414)	94.05	732	71.12	622
10M_ch2_15–20.wav	529 (242)	99.81	528	75.05	440
10M_ch2_20–25.wav	435 (155)	75.17	387	69.20	347

Tolerance of 2 ms.

files; “On Threshold” was set to 760 and 300 and “Off Threshold” to 380 and 150 for the first and the second audio file, respectively. For the other three files, the thresholds were the same; 510 for the “On Threshold” and 200 for the “Off Threshold”. Because the number of automatic detected clicks was less than the number of the manually marked clicks (especially in the creak areas), the “Second Level Trigger” was not used in the current testing.

As correct detection we considered an absolute time distance of less or equal to 2 ms between the automatic detection and the detection made by hand. Detection score refers to the percentage of the correctly detected hand labeled clicks:

$$\frac{\text{Number of correctly detected hand labeled clicks}}{\text{Total number of hand labeled clicks}} \times 100 \quad (23)$$

Results are summarized in Table 1. TK stands for the proposed detection method, while RB stands for the Rainbow click detector.³ The second column corresponds to the total number of manually detected clicks (regular and creak) per file. The number of creak clicks is indicated between parentheses. The columns of clicks for both methods, TK and RB, correspond to the clicks suggested by the algorithms. The proposed system achieved an

³ We have chosen to keep the original names of files used in this experiment, since they may serve as a benchmark database. In this case, files should be easily identifiable. Hand labeled data are available upon request (please send email to the first author).

average rate of 94.05% of correct detections while the detection score of the Rainbow click detector was 71.12%. All the average scores were computed via median. Therefore, the proposed algorithm shows 32% improvement in detection score. It is worth to note that RB always detects less clicks than manually detected. After checking, it was found that most of missing clicks were from the creak areas. This is expected, since creak clicks have very low SNR and their detection is not a trivial task. On the other hand, the TK-based detector overestimates the number of clicks into two files (second and third). The extra clicks come from segments of background (sometime impulsive) noise. If we split the set of all the manually detected clicks – independently of source file – into regular and creak clicks the performance of the TK-based detector shows a clear improvement over RB; for regular clicks the detection score is 84.83% for TK and 68.86% for RB (improvement: 23.2%) while for creak clicks the detection score is 78.01% for TK and 39.70% for RB (improvement: 96.5%).

In order to make result evaluation independent of the tolerance, curves similar to the Receiver Operating Characteristics (ROC) curves were produced. ROC curves show the performance of a detector as a trade off between selectivity and sensitivity. Due to the non-statistical nature of the algorithms presented here, we used a simple convention to produce curves that approximate ROC curves, which we will call *Approximate ROC* (AROC). We monitor the detection score by increasing the tolerance threshold of accepted clicks from 1 ms to 6 ms (absolute deviation). Fig. 10 shows the detection performance for the two detection algorithms. TK detection score is always higher than the RB score.

To prove the accuracy of the proposed algorithm, the Average Absolute Deviation (AAD) in milliseconds of time instances of the automatically detected clicks from those of the manually detected ones is shown in Fig. 11. Results correspond to all analyzed files. It worth to note that for 2 ms tolerance, the proposed algorithm is over 40% more accurate

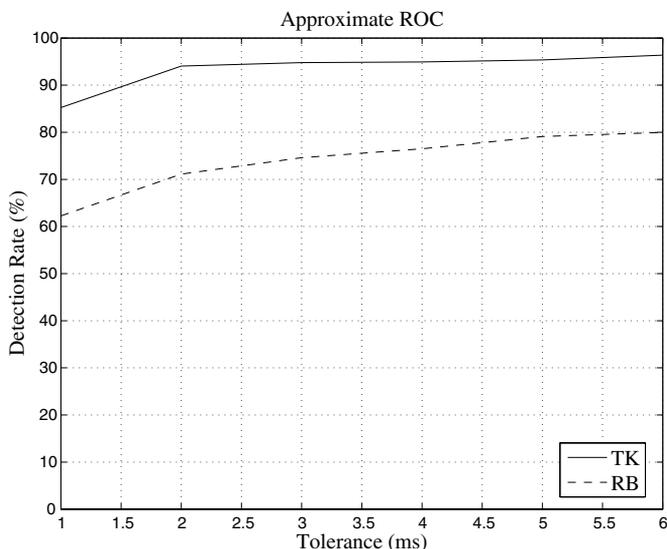


Fig. 10. Approximate ROC curves. Dashed line corresponds to results from the Rainbow click detector and solid line to the proposed (TK-based) one.

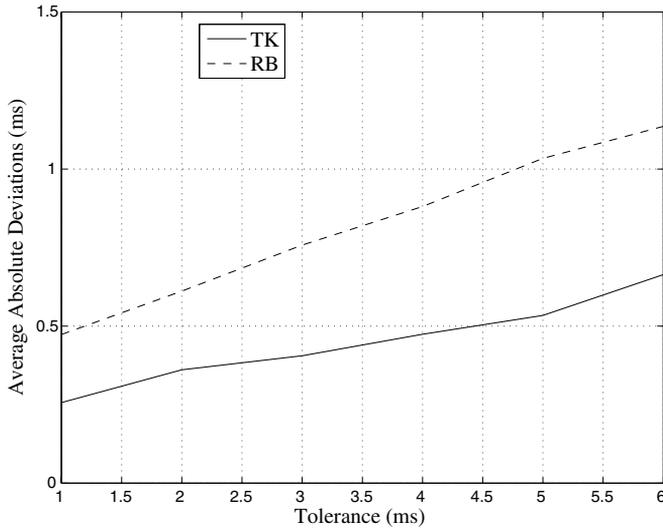


Fig. 11. Average Absolute Deviation from 2863 manually marked clicks. Dashed line corresponds to results from the Rainbow click detector and solid line to the proposed (TK-based) one.

than Rainbow click (0.36 ms for TK, and 0.61 ms for RB). Moreover, the robustness of the proposed algorithm is pronounced; as the tolerance increases, the average absolute deviation for Rainbow click increases faster than for the TK-based detector.

An example of the close agreement between the two kinds of detection used in the current work is given in Figs. 12 and 13. The former refers to regular click detection and the

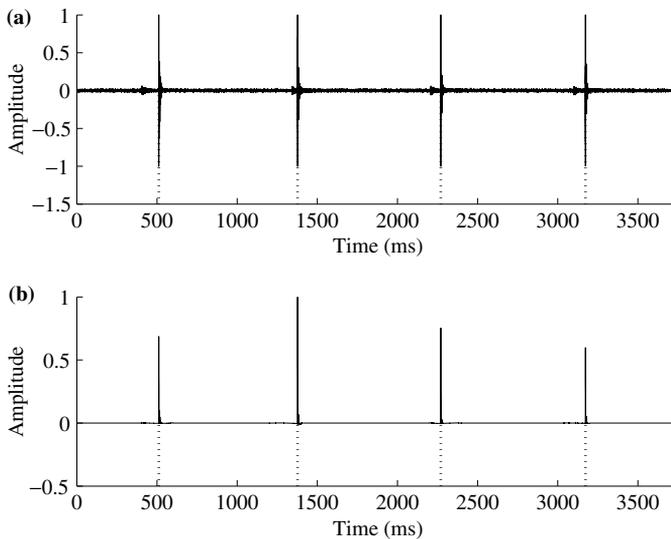


Fig. 12. Comparison of hand labels and automatic click detection: (a) four regular clicks (detection by hand) and (b) its TK output (automatic detection). The close agreement between the two detections is highlighted by dotted lines (AAD = 0.233 ms). Maximum amplitude of the TK-output has been normalized to unity.

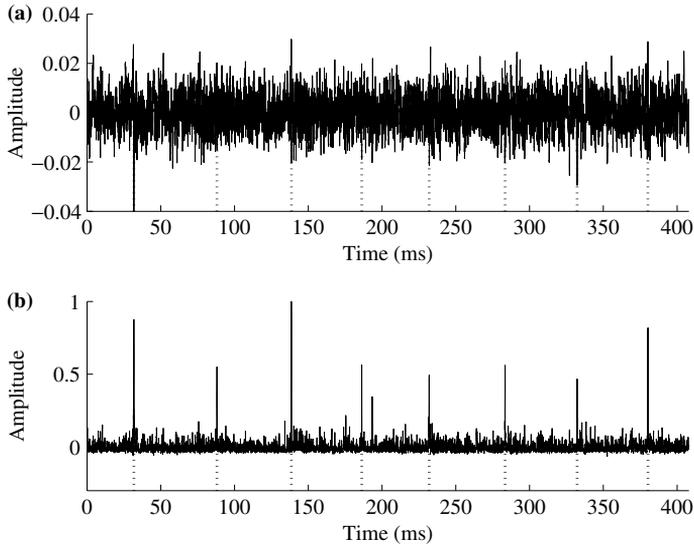


Fig. 13. Comparison of hand label and automatic click detection: (a) segment of a creak (detection by hand) and (b) its TK output (automatic detection). The close agreement between the two detections is highlighted by dotted lines (AAD = 0.0187 ms). Maximum amplitude of the TK-output has been normalized to unity.

latter to creaks. For the example with the regular clicks, AAD is 0.233 ms for the proposed detector and 0.447 ms for the Rainbow click detector. For the example with the creak clicks, AAD is 0.0187 ms for the TK-based detector while no clicks were detected by the Rainbow click detector. Note in Fig. 13 the efficiency of the algorithm: it can reveal the periodicity of the signal even in such bad conditions.

5. Discussion

In general, the proposed algorithm proved effective at detecting automatically sperm whale clicks in large recordings. An attempt to explore the cause behind the low scores of files “10M_ch2_5–10.wav” and “10M_ch2_20–25.wav”, was made by a step-by-step examination. We may notice three source of errors: (a) even after the application of a matching filter, regular sperm whale clicks may contain p_0 pulses with much lower sound level than p_1 (it is reminded that our notation follows that of [21,17]), (b) very faint creaks for which the skewness value is very similar with those taken from segments with absence of click sound, and (c) the length of the analysis window.

According to Zimmer et al. [18,17], p_0 pulse in regular clicks is a backward-directed pulse with low directionality while p_1 pulse is a forward-directed highly directional pulse. Consequently, the presence of a strong or weak p_0 pulse depends heavily on the off-axis aspect of the recorder with respect to the whale for each emitted click. The recordings used for the evaluation of the proposed algorithm were obtained using one of the widely spaced bottom mounted hydrophones at the AUTECH. The movement of the whale with respect to the fixed position of the hydrophone resulted in clicks being recorded with variable p_0 levels, a number of them lying far below p_1 levels. Since the detection algorithm was based on the assumption that the highest peak inside a click corresponds

to its starting point, a weak p_0 , even after the application of the matching filter, may not be detected, leading to a detection error. In Fig. 14(a) we can observe a regular sperm whale click with a weak p_0 pulse. As illustrated in Fig. 14(b), TK energy operator will

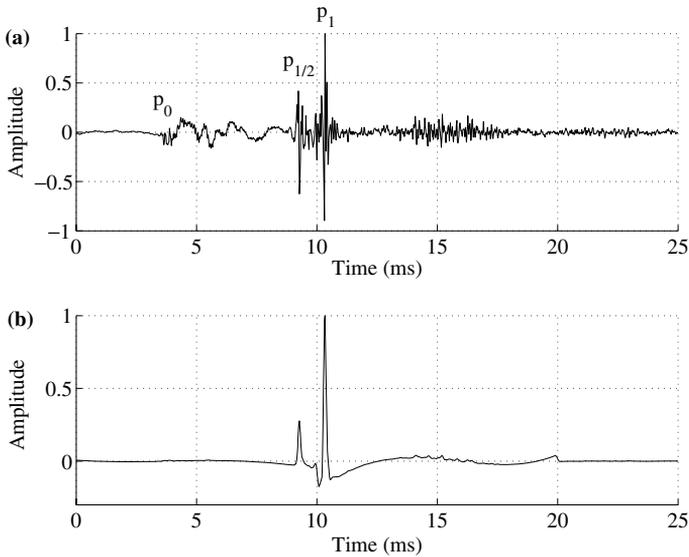


Fig. 14. (a) Waveform of a regular sperm whale click taken from file "10M_ch2_5-10.wav" and (b) its TK output (maximum amplitude has been normalized to unity).

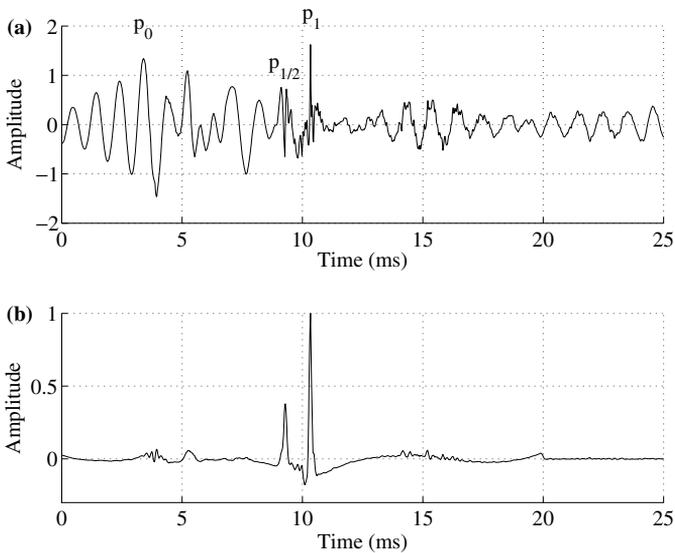


Fig. 15. After the application of the matching filter: (a) waveform of the regular click shown in Fig. 14 and (b) its TK output (maximum amplitude has been normalized to unity).

detect the strongest peak within the click which does not correspond to the beginning of the click. Unfortunately, as shown in Fig. 15, the proposed matching filter does not offer a viable solution in these cases. The design of a new matching filter is under investigation.

Furthermore, although the algorithm proved effective at detecting creak clicks in low SNR conditions (Fig. 13), the skewness criterion that was applied on the TK energy operator output seems to militate against detection of creak clicks. As already mentioned in Section 3, the skewness can be used as a detector of the presence of a click in order to avoid spending effort on the detection of a non-existing sound. Hence, care must be taken in choosing the right threshold value for the presence or the absence of a sound. The fact that segments of the signals analyzed in the current work containing faint creaks gave skewness similar to those segments with no clicks at all is suggestive of detection errors (false negative decisions). Therefore, we need to work out ways of determining a more representative value for the skewness criterion. It is very likely that the above-mentioned sources of errors acted cumulatively to produce the resultant low detection scores.

Adding to the above difficulties was the choice of the analysis window. As we have already stressed, creak clicks have very low periodicity which can change dramatically with time. For example, within 5 s their period may change from 0.3 s to as low as 10 ms. Accordingly, we need to consider short analysis window for the detector. Furthermore, the duration of the window should be selected to be long enough so that the windowed signal contains a few click sounds and in the same time, short enough to avoid analyzing a signal containing a mix of very different periods. We made a compromise of these two conflicting constraints by choosing the window duration to be four times the previously estimated ICI. This was an *ad hoc* selection as we have noticed that the more the window was getting shorter, the more false clicks were detected (false alarms), although we had a higher score of correct detections in a number of cases.

6. Conclusions

An algorithm based on the Teager–Kaiser (TK) energy operator has been developed for the automatic detection of clicks in large recordings from sperm whales. For evaluation purposes, real recordings were initially labeled by hand. Tests performed on recordings of a total duration of 25 min gave high scores of correct detection both for regular clicks and creaks. The use of TK energy operator, which is robust against additive noise and nearly instantaneous, makes the proposed algorithm a promising tool for automatic click detection. In comparison with a widely used click detector, the Rainbow click, the proposed one seems to be more efficient in detection score, more robust and more accurate.

Acknowledgements

The authors thank Naval Undersea Warfare Center (NUWC) for making the data used in this paper publicly available, and the organizers of the second International Workshop on Detection and Localization of Marine Mammals using Passive Acoustics (Monaco 2005), for providing access to these data.

References

- [1] Whitehead H. Sperm whales: social evolution in the ocean. University of Chicago Press; 2003.
- [2] Madsen PT. Sperm whale sound production – in the acoustic realm of the biggest nose on record. PhD thesis, Department of Zoophysiology, University of Aarhus, Denmark; 2002.
- [3] Douglas L. Click counting: an acoustic censusing method for estimating sperm whale abundance. Master Thesis, University of Otago, Dunedin, New Zealand; 2000.
- [4] Gordon JCD. The behavior and ecology of sperm whales of Sri Lanka. PhD thesis, University of Cambridge, Cambridge, England; 1987.
- [5] Goold JC. Behavioural and acoustic observations of sperm whales in Scapa Flow Orkney Islands. *J Mar Biol Assoc UK* 1999;79:541–50.
- [6] Drouot V. Ecology of sperm whale (*Physeter macrocephalus*) in the Mediterranean Sea. PhD thesis, University of Whales, Bangor, UK; 2003.
- [7] Madsen PT, Carder DA, Au WWL, Nachtigall PE, Møhl B, Ridgway SH. Sound production in neonate sperm whales. *J Acoust Soc Am* 2003;113(6):2988–91.
- [8] Thode A. Tracking sperm whale (*Physeter macrocephalus*) dive profiles using a towed passive acoustic array. *J Acoust Soc Am* 2004;116(1):245–53.
- [9] Laplanche C, Adam O, Lopatka M, Motsch JF. Real time sperm whale depth estimation using passive acoustics. In: CD Proc. IEEE Oceans, Brest, France; 2005.
- [10] Laplanche C, Adam O, Lopatka M, Motsch JF. Male sperm whale acoustic behavior observed from multipaths at a single hydrophone. *J Acoust Soc Am* 2005;118(4):2677–87.
- [11] Whitehead H, Weilgart L. Click rates from sperm whales. *J Acoust Soc Am* 1990;87:1798–806.
- [12] Gordon JCD. Evaluation of a method for determining the length of sperm whales (*Physeter macrocephalus*) from their vocalizations. *J Zool Lond* 1991;224:301–14.
- [13] Møhl B. Sound transmission in the nose of the sperm whale *Physeter catodon*. A postmortem study. *J Comp Physiol A* 2001;187:335–40.
- [14] Madsen PT, Payne R, Kristiansen NU, Wahlberg M, Kerr I, Møhl B. Sperm whale sound production studied with ultrasound time/depth-recording tags. *J Exp Biol* 2002;205:1899–906.
- [15] Møhl B, Wahlberg M, Madsen PT, Heerfordt A, Lund A. The monopulsed nature of sperm whale clicks. *J Acoust Soc Am* 2003;114(2):1143–54.
- [16] Thode A, Mellinger DK, Stienessen S, Martinez A, Mullin K. Depth-dependent acoustic features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico. *J Acoust Soc Am* 2002;112(1):308–21.
- [17] Zimmer WMX, Madsen PT, Teloni V, Johnson MP, Tyack PL. Off-axis effects on the multipulse structure of sperm whale usual clicks with implications for sound production. *J Acoust Soc Am* 2005;118(5):3337–45.
- [18] Zimmer WMX, Tyack PL, Johnson MP, Madsen PT. Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis. *J Acoust Soc Am* 2005;117(3):1473–85.
- [19] Dougherty AM. Acoustic identification of individual sperm whales (*Physeter macrocephalus*). Master Thesis, University of Washington, Seattle, USA; 1999.
- [20] Barlow J, Taylor BL. Estimates of sperm whale abundance in the Northeastern Temperate Pacific from a combined acoustic and visual survey. *Mar Mammal Sci* 2005;21(3):429–45.
- [21] Møhl B, Wahlberg M, Madsen PT, Miller LA, Surlykke A. Sperm whale clicks: directionality and source level revisited. *J Acoust Soc Am* 2000;107(1):638–48.
- [22] Madsen PT, Wahlberg M, Møhl B. Male sperm whale (*Physeter macrocephalus*) acoustics in a high-latitude habitat: implications for echolocation and communication. *Behav Ecol Sociobiol* 2002;53:31–41.
- [23] Mellinger DK, Stafford KM, Fox CG. Seasonal occurrence of sperm whale (*Physeter macrocephalus*) sounds in the Gulf of Alaska, 1999–2001. *Mar Mammal Sci* 2004;20(1):48–62.
- [24] Miller PJO, Johnson MP, Tyack PL. Sperm whale behaviour indicates the use of echolocation click buzzes “creaks” in prey capture. *Proc. R. Soc. Lond.*; 2004.
- [25] Mellinger DK, Clark CW. Recognizing transient low-frequency whale sounds by spectrogram correlation. *J Acoust Soc Am* 2000;107(6):3518–29.
- [26] Gillespie D. An acoustic survey for sperm whales in the Southern Ocean sanctuary conducted from the R/V Aurora Australis. *Rep Int Whal Comm* 1997;47:897–908.
- [27] Jäke O. Acoustic Censusing of sperm whales at Kaikoura, New Zealand: An inexpensive method to count clicks and whales automatically. Master Thesis, University of Otago, Dunedin, New Zealand; 1996.
- [28] Mellinger DK. Ishmael 1.0 User’s Guide. NOAA, NOAA/PMEL/OERD, 2115 SE OSU Drive, Newport, OR 97365-5258; 2001. Technical Memorandum OAR PMEL-120.

- [29] Kaiser JF. On a simple algorithm to calculate the “Energy” of a signal. In: Proc. IEEE ICASSP, pages 381–384, Albuquerque, NM, USA; 1990.
- [30] Kandia V, Stylianou Y. Detection of creak clicks of sperm whales in low SNR conditions. In CD Proc. IEEE Oceans, Brest, France; 2005.
- [31] Kaiser JF. On Teager’s energy algorithm and its generalization to continuous signals. In: The 4th IEEE Digital Signal Processing Workshop; 1990, pp. 17–18.
- [32] Melsa JL, Cohn DL. Decision and estimation theory. McGraw-Hill; 1978.
- [33] Duda RO, Stork DG, Hart PE. Pattern classification. second ed. John Wiley and Sons; 2001.
- [34] Jarvis S, Moretti D. Passive detection and localization of transient signals from marine mammals using widely spaced bottom mounted hydrophones in open ocean environments. In: International Workshop on the Application of Passive Acoustics in Fisheries. Massachusetts, USA: MIT Sea Grant; 2002.