Automatic detection of echolocation clicks based on a Gabor model of their waveform

Shyam Madhusudhana,[a] Alexander Gavrilov, and Christine Erbe
Centre for Marine Science and Technology, Curtin University, Perth, Western Australia, Australia

(Received 2 October 2014; revised 29 March 2015; accepted 8 May 2015)

Prior research has shown that echolocation clicks of several species of terrestrial and marine fauna can be modelled as Gabor-like functions. Here, a system is proposed for the automatic detection of a variety of such signals. By means of mathematical formulation, it is shown that the output of the Teager–Kaiser Energy Operator (TKEO) applied to Gabor-like signals can be approximated by a Gaussian function. Based on the inferences, a detection algorithm involving the post-processing of the TKEO outputs is presented. The ratio of the outputs of two moving-average filters, a Gaussian and a rectangular filter, is shown to be an effective detection parameter. Detector performance is assessed using synthetic and real (taken from MobySound database) recordings. The detection method is shown to work readily with a variety of echolocation clicks and in various recording scenarios. The system exhibits low computational complexity and operates several times faster than real-time. Performance comparisons are made to other publicly available detectors including PAMGUARD. © *2015 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4921609]

[AMT] Pages: 3077–3086

I. INTRODUCTION

Passive acoustic monitoring (PAM) is an increasingly common tool in studies of marine, terrestrial, and avian fauna and in environmental impact assessments. This article deals with the analysis and automatic detection of a class of bioacoustic signals, known as echolocation clicks, observed in both terrestrial and underwater soundscapes.

It has been shown that echolocation clicks of several species of marine and terrestrial fauna can be approximated by Gabor-like functions (formulation presented in Sec. II). Examples include odontocetes (Kamminga and Beitsma, 1990; Kamminga *et al.*, 1996; Kamminga *et al.*, 1993; Kamminga and Stuart, 1995) and Egyptian fruit bats (Holland *et al.*, 2004). A Gabor function (Gabor, 1946) is a harmonic function localised by a Gaussian envelope. Several other studies, albeit without using the term "Gabor function" explicitly, acknowledge the presence of a Gaussian-like amplitude envelope resulting in small time-bandwidth products in the biosonar signals. Some of the species covered by these studies include Blainville's beaked whale (*Mesoplodon densirostris*) (Johnson *et al.*, 2006), finless porpoise (*Neophocaena phocaenoides*) (Goold and Jefferson, 2002), Hector's dolphin (*Cephalorhynchus hectori*) (Thorpe and Dawson, 1991), and Mediterranean bottlenose dolphins (*Tursiops truncatus*) (Greco and Gini, 2006). A Gabor wavelet transform (Gabor, 1946) or a Gabor filter (Marčelja, 1980) applied to an acoustic time series could thus help to highlight the underlying clicks. In another study, van der Schaar *et al.* (2007) attempted identification of individual sperm whales (*Physeter macrocephalus*) based on modelling their clicks by Gabor functions. We will show that the application of the Teager–Kaiser Energy Operator (TKEO) (Kaiser, 1990a) to such signals simplifies and enhances their detectability with automatic detectors.

The TKEO has been used by several bioacousticians for automatic detection of underwater echolocation clicks (Kandia and Stylianou, 2006; Roch *et al.*, 2008; Soldevilla *et al.*, 2008; Roch *et al.*, 2011b; Klinck and Mellinger, 2011). Several non-TKEO based methods have also been proposed, such as those based on kurtosis (Gervaise *et al.*, 2010), on phase slopes (Kandia and Stylianou, 2008), on spectrogram correlation (Harland, 2008; Dobbins, 2009) and thresholding (Morrissey *et al.*, 2006), on stochastic matched filtering (Caudal and Glotin, 2008), on amplitude envelope levels (DeRuiter *et al.*, 2009), and on the use of support vector machines (Jarvis *et al.*, 2008). Most of the existing click-detection algorithms based on the TKEO either use a simple moving-average filter comparing the outputs to a fixed threshold, rely on a noise floor that is pre-computed over a large time interval or perform some form of forward-backward peak selection operation within large audio segments (Kandia and Stylianou, 2006; Roch *et al.*, 2008; Soldevilla *et al.*, 2008; Roch *et al.*, 2011b; Klinck and Mellinger, 2011). Some of the approaches that avoid the pitfalls of employing a fixed threshold perform multi-pass processing over large segments of recordings with an inherent assumption that spikes of echolocation clicks do not constitute a majority of the considered segment. The threshold is computed in an initial pass, and then the spike locations corresponding to clicks in the segment are identified over one or more subsequent passes over the entire segment in consideration. The dependence of a detector on the assessment of certain signal statistics over long durations not only affects its response time, but also bears an impact on the consistency of its performance when employed in highly dynamic noise environments. Hence, such methods are not ideal for application in an online scenario. They also run the risk of discarding weaker clicks in a temporal neighbourhood of

[a] Electronic mail: s.madhusudhana@postgrad.curtin.edu.au

multiple higher energy clicks. The method proposed by Kandia and Stylianou (2006) is targeted at detecting sperm whale clicks and is based on measuring the deviation of the distribution of the TKEO output from a Gaussian shape. Analysis is performed iteratively on short successive frames. Barring the other elements meant for precisely locating the onset of a click, the algorithm would report detections when the deviation exceeds a pre-estimated skewness threshold. The method proposed by Roch *et al.* (2008) also performs operations frame-wise. The 40th percentile of the TKEO outputs in a frame is taken as the "noise floor" and parts of the TKEO output that lie over 50 times this noise floor are considered to represent clicks. Similar approaches are employed in Roch *et al.* (2011b) and Soldevilla *et al.* (2008). Contrary to the usual practice of applying the TKEO directly to audio signals, Klinck and Mellinger (2011) apply the TKEO to the ratio of the outputs of two different band-pass filters and compare the result to a dynamic detection threshold. The threshold also relies on measurements from frames of 60 s duration.

In this article, we present a new algorithm that employs two short moving-average filters to provide near-instantaneous spike detection in the TKEO output and that is well suited for processing continuous input audio samples.

The next section presents an analysis of applying the TKEO to Gabor signals. Then the inferences made from the analysis are verified with a case study. The subsequent sections describe the detection algorithm and discuss its performance.

II. APPLYING THE TKEO ON A GABOR-LIKE SIGNAL

A. Theoretical analysis

The TKEO output of an arbitrary continuous signal $x(t)$ is given by (Kaiser, 1990b)

$$\Psi_c[x(t)] = \dot{x}^2(t) - x(t)\ddot{x}(t), \tag{1a}$$

where the operators $\dot{}$ and $\ddot{}$ denote the first and second derivatives, respectively. The TKEO output of an arbitrary discrete signal x_n is given by (Kaiser, 1990a)

$$\Psi_d[x_n] = x_n^2 - x_{n-1}x_{n+1}. \tag{1b}$$

For a Gabor function, there are several equivalent ways of mathematically expressing its Gaussian amplitude envelope (e.g., Kamminga and Beitsma, 1990; Holland *et al.*, 2004). For ease of establishing a relationship with the width of an echolocation click, we chose the following representations for continuous and discrete Gabor signals:

$$G(t) = Ae^{-(t-t_0)^2/2\sigma^2}\cos\{\omega(t-t_0)+\phi\}, \tag{2a}$$

$$G_n = Ae^{-(nT_s-t_0)^2/2\sigma^2}\cos\{\omega(nT_s-t_0)+\phi\}, \tag{2b}$$

where A is the signal amplitude, t_o and σ are the mid-epoch and standard deviation of the Gaussian envelope, respectively, and T_s is the sampling interval in the discrete case. The cosine term represents the carrier signal with phase ϕ and angular frequency $\omega = 2\pi/T_c$, where T_c is the period of the carrier wave.

Harmonic signals localised by a Gaussian envelope can be represented more generally as

$$G(t) = Ae^{-(t-t_0)^2/2\sigma^2}\cos\{\omega_t(t-t_0)+\phi\}, \tag{3}$$

where ω_t describes the angular frequency as a function of time. Of particular interest to us are the cases with constant frequency carrier waves (CFCW) and those with linearly chirped carrier waves (LCCW) due to their similarity to commonly encountered echolocation clicks. The term "Gabor-like" used in the article refers to these two types of signals. Signals of the latter form are commonly known as Gabor chirps (Mann and Haykin, 1991). The time dependence of their carrier frequency can be expressed as

$$\omega_t = \omega_0 + \dot{\omega}_t(t-t_0). \tag{4}$$

Note that in this form, ω_o corresponds to the carrier wave's central frequency, which is its instantaneous frequency at t_o. We will denote the carrier's instantaneous period corresponding to the central frequency as T_o. For Gabor-like signals of CFCW type, $\dot{\omega}_t = 0$ in Eq. (4). The carrier wave's effective instantaneous frequency resulting from Eq. (4) must remain positive and finite within the full width of the Gaussian envelope, which can be defined as 6σ. This constrains the values of $\dot{\omega}_t$ to the range $0 \le |\dot{\omega}_t| < (\omega_0/3\sigma)$.

Substituting $G(t)$ in Eq. (3) for $x(t)$ in Eq. (1) and simplifying the result using trigonometric identities, we arrive at the following form of the TKEO output for Gabor-like signals:

$$\begin{aligned}\Psi_c[G(t)] = A^2e^{-(t-t_0)^2/\sigma^2}\Big\{&[\omega_t + \dot{\omega}_t(t-t_0)]^2 \\ &+\frac{1}{2}[2\dot{\omega}_t + \ddot{\omega}_t(t-t_0)]\sin 2\theta + \frac{1}{\sigma^2}\cos^2\theta\Big\}, \\ \theta = \omega_t(t-t_0)+\phi.&\end{aligned} \tag{5}$$

Ψ_c consists (in order of appearance) of a constant (A^2), a Gaussian component, and a component comprising three additive terms that affect the shape of the Gaussian component. For convenience, we will refer to the three additive terms as T1, T2, and T3 in the order they appear in Eq. (5). By denoting the standard deviation of the Gaussian curve component in Ψ as σ_{TK}, we can express its relationship to the Gaussian envelope of $G(t)$ as

$$\sigma_{TK} = \frac{\sigma}{\sqrt{2}}. \tag{6}$$

Using Eq. (4), Eq. (5) can be rewritten for Gabor-like signals as

$$\begin{aligned}\Psi_c[G(t)] = A^2e^{-(t-t_0)^2/\sigma^2}\Big\{&[\omega_0 + 2\dot{\omega}_t(t-t_0)]^2 \\ &+\dot{\omega}_t\sin 2\theta + \frac{1}{\sigma^2}\cos^2\theta\Big\}, \\ \theta = [\omega_0 + \dot{\omega}_t(t-t_0)](t-t_0)+\phi.&\end{aligned} \tag{7}$$

Let us consider separately the effect of T1, T2, and T3 on Ψ. The term T1 is a quadratic quantity, and its minimum occurs at $-\omega_0/2\dot{\omega}_t$ relative to the Gaussian component's maximum. The magnitude of this temporal offset at its minimum is $3\sigma_{TK}/\sqrt{2}$ at the maximum $|\dot{\omega}_t| = \omega_o/3\sigma$, and it increases with decreasing $|\dot{\omega}_t|$. With its minimum occurring sufficiently away from t_o, the term T1 introduces a skew in the Gaussian component of Ψ. Notice that T1 is a constant ($\mathrm{T1} = \omega_o^2$) for Gabor-like signals of CFCW type and, consequently, the Gaussian shape of Ψ is not skewed. The effects of T2 and T3 on Ψ can be examined by considering their values at the limits of $\dot{\omega}_t$. For the maximum value of $\dot{\omega}_t$, Eq. (7) can be rewritten as

$$\Psi_c[G(t)] = A^2\omega_o^2 e^{-(t-t_0)^2/\sigma^2}\left\{\left[1+\frac{2}{3\sigma}(t-t_0)\right]^2 + \frac{1}{\pi\lambda}\sin 2\theta + \frac{9}{\pi^2\lambda^2}\cos^2\theta\right\}, \qquad (8)$$

where $\lambda = 6\sigma/T_o$ is the number of periods of the carrier wave's central frequency contained within the full width (6σ) of the Gaussian envelope of $G(t)$. The harmonic elements of T2 and T3 introduce distortions in an otherwise smooth curve of Ψ. The scaling of these distortions, viz., $1/\pi\lambda$ and $9/\pi^2\lambda^2$ (hereafter referred to as distortion scaling factors), are driven by λ. These terms are, however, small relative to unity when the Gabor-like signal is well-formed, i.e., contains at least a few periods of the carrier. Figure 1 shows the variation of the distortion scaling factors in T2 and T3 for a few values of λ at $\dot{\omega}_t = \omega_0/3\sigma$. Because T1 approaches unity at t_o in Eq. (8), the maximum cumulative distortion produced by T2 and T3 can be seen from Fig. 1 as being small relative to T1 in the region around t_o for well-formed signals. For any particular value of λ, the maximum distortion of the Gaussian in Ψ occurs at maximum $\dot{\omega}_t$ and, as $\dot{\omega}_t$ approaches 0, the distortion results only from T3. So we can infer in general that for well-formed Gabor-like signals, the magnitude of the distortions caused by T2 and T3 are small compared to the scaling and skewing caused by T1 over a significant extent of the Gaussian component of Ψ in the vicinity of t_o. Hence the resulting nature of Ψ is largely dominated by a Gaussian. This is demonstrated in Fig. 2 for a synthetic signal with a reasonably high rate of $\dot{\omega}_t$. Similarly high rates of frequency change in echolocation signals have been observed only in some subspecies of beaked whales (Zimmer *et al.*, 2005; Rankin *et al.*, 2011). Although the distortion of Ψ is visible at large $|\dot{\omega}_t|$, it is not significant compared to the non-skewed Gaussian output of the TKEO.

Thus far we have shown that applying the TKEO to Gabor-like signals suppresses the harmonic component and that its output is well approximated by a scaled Gaussian impulse that is narrower than the amplitude envelope of the input signal by a factor of $1/\sqrt{2}$.

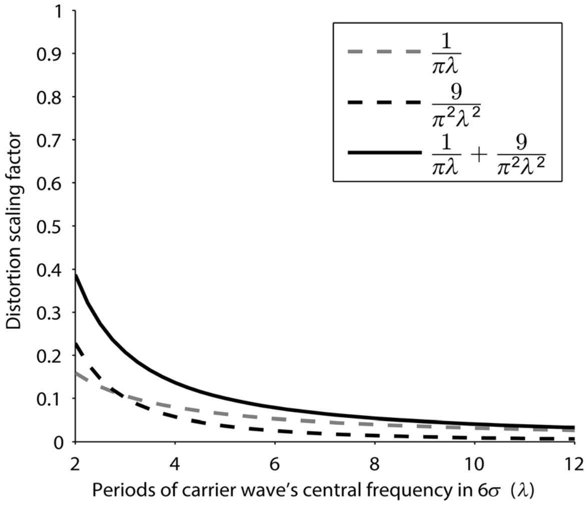

FIG. 1. Scaling (dashed lines) of the distortion produced by the harmonic elements of T2 and T3 in Eq. (8), shown for a few values of λ. The solid line is indicative of the upper limit on the magnitude of distortion as a cumulative effect of T2 and T3.

B. Case study

To verify the findings from the above analysis for real echolocation clicks, we performed a curve-fitting exercise on 200 handpicked odontocete clicks from a recording made over the Australian Northwest Shelf, sampled at 192 kHz. Gabor curves were fitted to the waveforms of each click, and Gaussians fitted to their corresponding TKEO outputs (see Fig. 3). The Levenberg–Marquardt (LM) algorithm (Gill *et al.*, 1981) is known to perform well in non-linear curve-fitting tasks, and hence it was chosen for this analysis. The averages of the estimated parameters of the individual curve-fits were considered in producing the overlaid (dark)

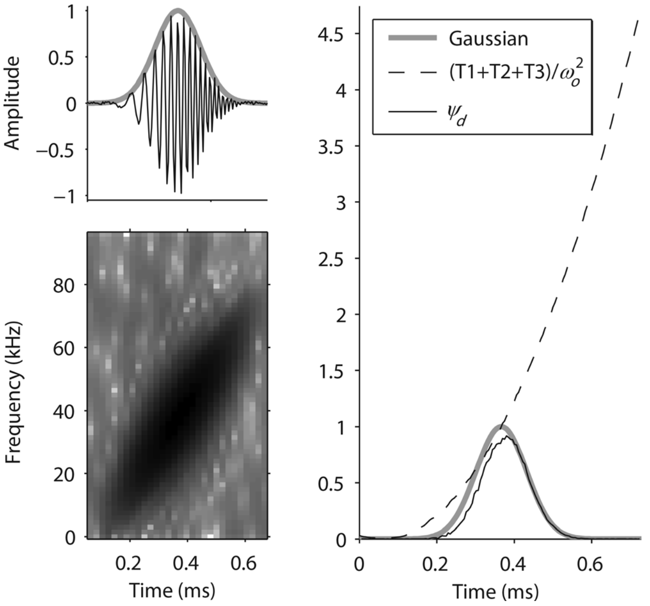

FIG. 2. Waveform (top-left) and spectrogram (bottom-left) of a synthetic Gabor chirp produced with the following values: $A = 1$, $\sigma = 0.091$ ms, $\phi = 0$, and $\dot{\omega}_t$ so chosen as to yield carrier frequency sweep from 21 to 55 kHz over the 6σ duration with centre at 38 kHz. The signal is a simulation of an instance of a real beaked whale click considered later. The gray overlay shows the Gaussian envelope. The right plot shows the corresponding Gaussian and quadratic-approximate ($\mathrm{T1}+\mathrm{T2}+\mathrm{T3}$; scaled here, by $1/\omega_0^2$, to enable comparisons) components of the analytical TKEO output. The discrete TKEO output is overlaid over the pure Gaussian to indicate the introduced skew causing a forward shift of ~0.01 ms in its peak.

Gabor and Gaussian curves. The Gabor fitting of the waveforms yielded parameter estimates of $\sigma = 0.0116$ ms and $T_o = 0.0324$ ms, resulting in $\lambda \approx 2.15$. A σ_{TK} estimate of 0.0079 ms supports the relationship expressed in Eq. (6). For the Gaussian fit of the TKEO outputs, an average *summed square of errors/residuals* value of 0.01 and a *root mean squared error* value of 0.03 confirmed the usefulness of the model for fitting purposes, and an average *adjusted* R^2 value of 0.98 indicated a "good fit."

III. AUTOMATIC DETECTION

So far, we have shown that for signals that can be modelled as Gabor-like functions (e.g., underwater echolocation clicks), the corresponding TKEO values tend to approach a Gaussian shape. Based on the inferences, we will now describe a simple system for the detection of Gabor-like clicks in acoustic recordings.

A. Detector design

A short rectangular moving-average filter produces an averaging or smoothing effect on an input signal. Because the outputs of the TKEO are predominantly non-negative, a longer moving-average filter produces a flattening effect on the TKEO outputs. In contrast, a bell-shaped averaging filter (e.g., Hamming, Hanning, or Gaussian function) has the potential of highlighting short-duration energy surges in TKEO outputs while flattening non-spiked high-energy sections. We chose a scaled Gaussian function for our first moving-average filter (MAF1) as it allows for easy control of the acuteness of the bell shape. Convolution operation with MAF1 can be expressed as

$$h_{MAF1}(n) = \frac{T_s}{\sigma_G\sqrt{2\pi}} \sum_{i=-N}^{N} e^{-(iT_s)^2/2\sigma_G^2} x_{n+i} \tag{9}$$

for a filter of length $2N+1$, where n is the sample index and σ_G is the standard deviation of the Gaussian function. The factor $(T_s/\sigma_G\sqrt{2\pi})$ ensures that the filter gain (area under the curve) approaches unity. The acuteness of the Gaussian can be controlled with σ_G. The choice of values for σ_G and N is discussed in the next sub-section.

Consider a second moving-average filter (MAF2)——a rectangular averaging filter of the same length as MAF1. The amplitude of the filter is chosen such that the filter gains of MAF1 and MAF2 are the same. Similar gains allow for fair comparisons to be made of the two filters' outputs.

For an input unit impulse, $h_{MAF1}(n)$ peaks at the point corresponding to the non-zero element of the impulse and falls off on either side of it. In contrast, the response of MAF2 $[h_{MAF2}(n)]$ is flat. The proposed detection algorithm exploits this difference in characteristics of the responses of the two filters. Consider the difference $[h_{MAF1}(n) - h_{MAF2}(n)]$ expressed as a fraction of $h_{MAF1}(n)$. We denote this quantity filter difference ratio (FDR), which is a normalised measure of the extent of $h_{MAF1}(n)$ over $h_{MAF2}(n)$.

$$FDR(n) = \frac{h_{MAF1}(n) - h_{MAF2}(n)}{h_{MAF1}(n)}. \tag{10}$$

Impulse responses of typical filters and the ensuing FDR are shown in Fig. 4. The dotted horizontal line in the FDR plot highlights the maximum value of FDR (FDR_{peak}). For a chosen combination of MAF1 and MAF2, there are four noteworthy properties of FDR:

(i) The FDR curve and FDR_{peak} remain the same for input impulses of any given amplitude scaling.

(ii) The difference $[h_{MAF1}(n) - h_{MAF2}(n)]$ and the ensuing FDR are maximum when the impulse is at the centre of the filters.

(iii) The value of the numerator never exceeds the denominator. Hence the resulting ratio is less than 1.

(iv) $h_{MAF1}(n)$ is smaller than $h_{MAF2}(n)$ at input samples sufficiently away (in time) from the non-zero element of the impulse. The numerator and hence the ensuing FDR are negative for such points.

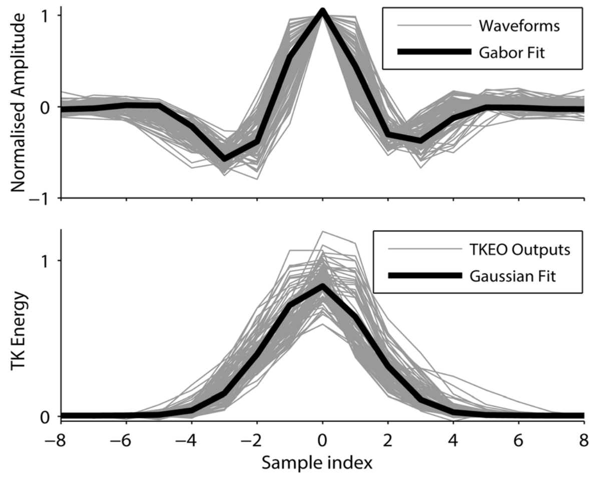

FIG. 3. Curve fitting of echolocation clicks from real recordings with a Gabor function (top) and of their corresponding TKEO outputs with a Gaussian curve (bottom). Gray lines show clicks' waveforms and their corresponding TKEO outputs in the respective plots.

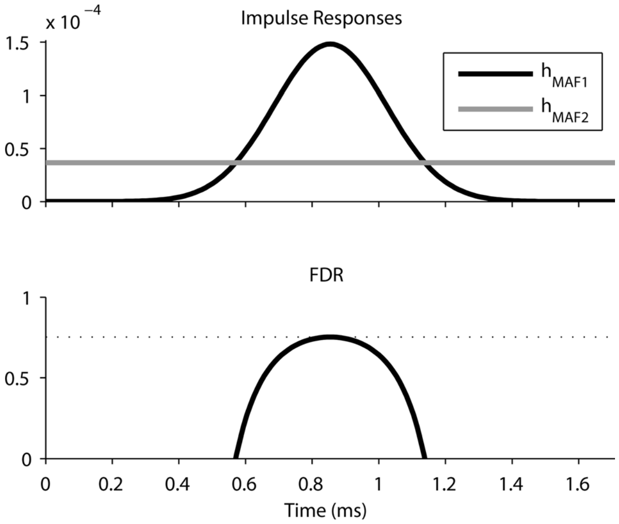

FIG. 4. Impulse responses (top) of filters MAF1 ($\sigma_G = 0.169$ ms) and MAF2 and the corresponding FDR (bottom). FDR plot restricted to the range [0,1]. Dotted line in the FDR plot indicates the peak FDR value.

Similar to a unit impulse, acute Gaussian curves also have a steep rise followed by a steep fall. We can see from Eq. (6) that the Gaussian-like outputs (hereafter referred to as spike) obtained from applying the TKEO to Gabor-like signals also have an acute profile. When the outputs of the TKEO applied to audio recordings containing echolocation clicks are convolved with MAF1 and MAF2, and the FDR is determined, we can expect to see curves similar to those in Fig. 4 at locations corresponding to clicks in the original audio. As with unit impulses of different amplitudes, the FDR curve would remain similar for clicks with different intensities. Hence we chose to set the detector threshold to be a function of FDR_{peak} for the chosen combination of MAF1 and MAF2. However, TKEO outputs of real clicks differ from a unit impulse in two ways. First, a combination of factors (like noise and choice of sampling rate) results in a possibility of bearing small negative values in the neighbourhood of the energy pinnacle of the TKEO output corresponding to an echolocation click. Second, the width of the spike is wider than a unit impulse. As a result of these two factors, the tip of the FDR corresponding to a click would be lower than the FDR_{peak} computed for the chosen filters. Hence the detection threshold can be set as a fraction of the employed filters' FDR_{peak}. Figure 5 demonstrates the outcome of filtering and FDR computation for synthetic data imitating TKEO outputs with different amplitudes. Notice how a fixed threshold, that is 85% of the FDR_{peak}, can serve as a good cut-off for detecting spikes.

Thus far we have established that the output of MAF1 remains high for TKEO values corresponding to echolocation clicks and in turn the FDR value produces a local maximum. However, the TKEO may produce non-positive outputs for sections of input audio that do not correspond to clicks. Depending on the length of MAF1 (and MAF2) and the negative strength of the TKEO output, this may sometimes translate to non-positive outputs from MAF1 and MAF2. This, in turn, would yield FDR values that are not meaningful for our application (e.g., $\pm\infty$). In certain implementations, FDR computation with such values may even cause undesirable exceptions (e.g., divide-by-zero exception). Because we know that a non-positive value in either filters' output does not indicate the presence of a spike in the TKEO output, we can safely bypass calculation of FDR for such values. Considering property (iv) of the FDR, we also bypass computation of FDR when $h_{MAF1}(n) \not> h_{MAF2}(n)$.

Considering property (iii) of FDR and the constraints described in the preceding text [$h_{MAF1}(n) > 0$; $h_{MAF2}(n) > 0$ and $h_{MAF1}(n) > h_{MAF2}(n)$] for the computation of meaningful FDR values, we can see that the usable range of FDR values is effectively reduced to [0,1]. Further, FDR values that are beyond the threshold value (fraction of FDR_{peak}) indicate the presence of Gaussian-like spikes in the TKEO outputs, in turn indicating the presence of echolocation clicks in the input audio.

B. Implementation

The width of a Gaussian at half its peak value, commonly known as full width at half maximum ($\text{FWHM} = 2\sqrt{2\ln(2)}\sigma \approx 2.355\sigma$) provides a better feel for the width of the Gaussian pulse in visual observations. We will denote the FWHM and the standard deviation of the Gaussian envelope in the target click as $FWHM_{EC}$ and σ_{EC}, respectively. The standard deviation, σ_{TK}, of the Gaussian curve resulting from applying the TKEO to echolocation clicks can be derived using Eq. (6) as

$$\sigma_{TK} = \frac{\sigma_{EC}}{\sqrt{2}} = \frac{FWHM_{EC}}{4\sqrt{\ln(2)}}. \tag{11}$$

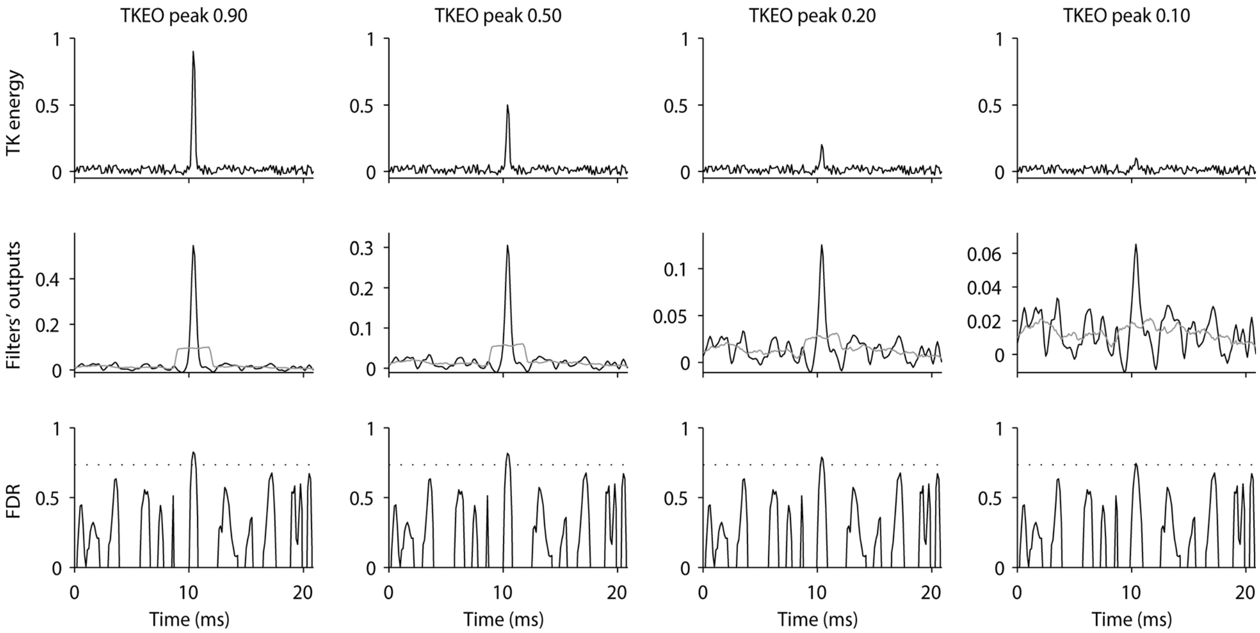

FIG. 5. Demonstration of filtering and FDR computation for synthetic TKEO values with varying strengths for transient surges. First row shows the synthetic TKEO values with spikes ranging from 0.10 to 0.90. Second row shows the result of filtering the TKEO values with MAF1 (black curves) and MAF2 (gray curves). The third row shows the FDR (solid line) and the threshold (dashed line) set as 85% of FDR_{peak}.

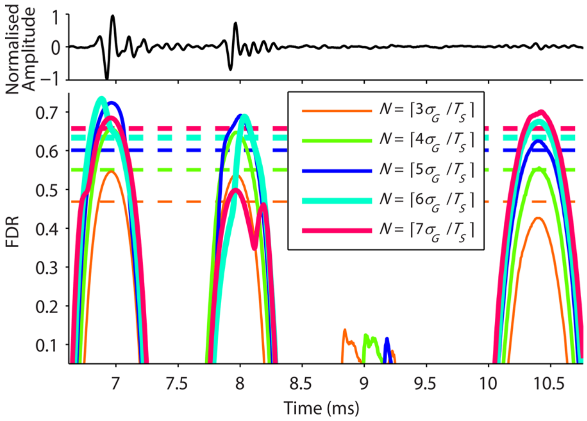

FIG. 6. (Color online) Demonstration of the effect of N on click detection using a segment of real underwater acoustic recording. The top panel shows the waveform of the recording consisting of three distinct pulses. The bottom panel shows the corresponding FDR for different values of N. The range of y axis values is restricted to enable clarity. A detection threshold of 80% of the resulting FDR_{peak} is also shown as dashed lines for each value of N.

The value of σ_{TK} obtained using estimates of $FWHM_{EC}$ made from visual observations of representative clicks' waveforms can be used as a guide in designing the needed filters. We can set the standard deviation of the Gaussian in MAF1 to be the same as σ_{TK} where it would function as a matched filter. We know that 99.7% of the area under a Gaussian curve is contained within a distance of 3σ on either side of its mean. Setting the length of the filter to $6\sigma_G$ would account for contributions only from the bulk of a spike without consideration for the points in its immediate neighbourhood. Extending the filter length would not only weigh the high energy regions, but also appropriately penalise low energy regions, thereby enabling only those sections to stand out that correspond to actual spikes in the TKEO output. However, a very long averaging filter stands the risk of clubbing close lying spikes. This causes smearing in the output thereby affecting their detectability with the FDR. Figure 6 demonstrates the effect N has on FDR and on the subsequent detection. Let us consider the faint pulse occurring at ~10.4 ms. As the energy of the pulse is not significant compared to background noise, a shorter MAF2 produces a larger output resulting in smaller FDR values as compared to the corresponding FDR_{peak}. For the same pulse, the FDR curves corresponding to different N show that larger N yields larger FDR. While increasing N is beneficial for pulses that are temporally well-separated from other high-energy signals, the resulting larger MAF2 increases the risk of accounting for energy from neighbouring signals (including other pulses) for pulses that are not temporally well-isolated. For the pulse occurring at ~8 ms, notice that its FDR is influenced by the preceding pulse for $N = \lceil 6\sigma_G/T_s \rceil$ and is influenced on both sides for $N = \lceil 7\sigma_G/T_s \rceil$. Based on such observations, we have empirically arrived at a value of $N = \lceil 5\sigma_G/T_s \rceil$ for MAF1 (and in turn, for MAF2). Note here that all σ_G values are expressed in time units and may bear non- integer values and hence rounding N up to the next higher integer is necessary. Considering the widths of the different types of echolocation clicks commonly encountered, this value of N does not make the full filter length $(2N+1)$ unwieldy and at the same time enables fair weighting of points both on and in the neighbourhood of a spike. Once the values for σ_G and N are identified as described, MAF1 can be realised as

$$MAF1(n) = \frac{T_s}{\sigma_G\sqrt{2\pi}} e^{-(nT_s)^2/2\sigma_G^2}, \tag{12}$$

where $n = -N, \ldots, -3, -2\ -1, 0, 1, 2, 3, \ldots, N$ is the index of the sampled point in the filter. MAF2 can be realised as

$$MAF2(n) = \frac{\sum_{m=-N}^{N} MAF1(m)}{2N+1}. \tag{13}$$

The value of FDR_{peak} for the combination of MAF1 and MAF2 can be obtained by setting $n = 0$ in Eq. (12) and Eq. (13) and substituting the resulting values in Eq. (10). The product of the obtained FDR_{peak} and a user-controlled value (in the range 0–1) becomes the detection threshold for the system. A schematic of the proposed detection system is presented in Fig. 7.

IV. PERFORMANCE EVALUATION

The performance of the system was evaluated using both synthesised data and real audio recordings. For the latter, we used publicly available underwater audio recordings from MobySound.[1] The recording sets used are listed in Table I. Synthetic data were generated using pieces of real underwater recordings. A 28-s long audio fragment of ambient sea noise free of echolocation clicks was handpicked to serve as background noise. Two sets of 20 short audio clips containing single echolocation clicks were extracted from underwater sound recordings. Clips with sperm whale clicks, representing the CFCW type, constituted one set and clips with beaked whale clicks, representing the LCCW type, constituted the other. Two hundred instances of clicks were randomly drawn (with repetition) from one set and then superimposed at uniformly distributed random points in time across the ambient sea noise recording. The amplitude of

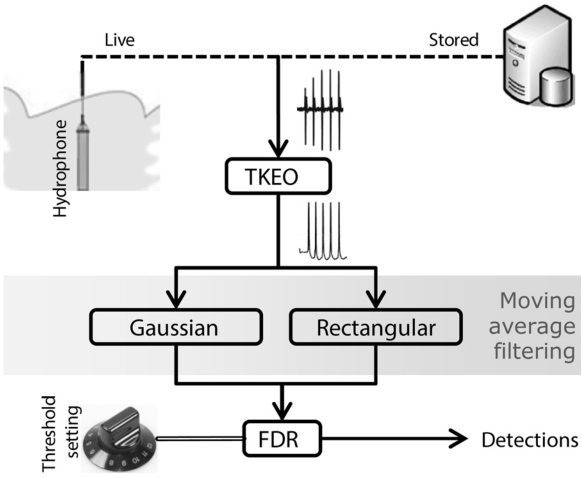

FIG. 7. Schematic of the proposed click-detection system. Dashed lines are used to indicate that the input could either be pre-recorded audio or live real-time inputs.

TABLE I. Datasets obtained from MobySound for testing the proposed detector.

Species	Dataset identifier and audio file(s)
Rough Toothed Dolphins	**RoughToothed_Marianas(MISTC)-Annotated** MISTCS070316-113000.wav
Rissos Dolphins	**Rissos-SCORE-annot** Set1-A2-H17-081406-0000-0030-1225-1255loc.wav
Beaked Whales	**Mesoplodon_CanaryIsles-Annotated** md05_294a10590-11850.wav
Sperm Whales	**Sperm whales_Bahamas(AUTEC)-Annotated** SpermWh_A2_030306-H16_short.wav
Spotted Dolphins	**SpottedDolphin_Bahamas(AUTEC)-Annotated** Set3_A4_042705_CH5_H40_A0600-0630.wav
Striped Dolphins	**StripedDolphin_Marianas(MISTC)-Annotated** MISTCS070309-092000.wav MISTCS070309-083000.wav

each superimposed click was altered to yield a particular signal-to-noise ratio (SNR) value. The SNR values chosen were uniformly distributed within the range from 5 to 30 dB. The SNR value was defined from the energy of the click being superimposed and the energy of background noise, both values determined within the frequency band of interest (3–30 kHz for sperm whales and 20–80 kHz for beaked whales) and integrated over the time interval containing 90% of the click energy. The noise fragment along with the superimposed clicks constitutes a synthetic test input. The start and end times of each superimposition were recorded for later comparison with detection results. Synthesis was repeated 1000 times for each species while generating different insertion points, different clip permutations, and different SNR values at each repetition. To emulate the diversity in click characteristics prevalent in real underwater audio, a certain level of click dissimilarity was ensured within each clip set based on a "by eye" assessment.

The FWHM (and in turn σ_G) of MAF1 can be tuned as described in Sec. III B to achieve optimal performance in each of the aforementioned tests, i.e., for each species. However, we chose to use a single setting for all the tests to be able to show that the algorithm is capable of performing detection regardless of the species producing the clicks. The chosen value of FWHM = 0.40 ms translates to a filter length of 329 points for a sampling rate of 192 kHz, and 165 points for a sampling rate of 96 kHz.

TABLE II. Parameter settings used to configure the click detector module in PAMGUARD for tests with synthesised data.

Parameter	Sperm Whale	Beaked Whale
Pre-Filter	High Pass: 200 Hz	High Pass: 10 kHz
Trigger filter	Band Pass: 3–30 kHz	Band Pass: 20–80 kHz
Long filter	0.00001	0.00001
Long filter 2	0.000001	0.000001
Short filter	0.1	0.1
Minimum click separation	100 samples	100 samples
Maximum click length	1024 samples	1024 samples
Pre sample	40 samples	40 samples
Post sample	0 samples	0 samples

For comparative performance analysis, tests with synthesised data were repeated with two other detectors—a TKEO-based detector described in Roch *et al.* (2011b) and PAMGUARD.[2] PAMGUARD is a publicly available software program that provides automatic detection/classification capabilities. The default "click detector" module was employed. It is a non-TKEO based detector that works by comparing signal levels to estimated background noise levels. The detector's various parameters were set as shown in Table II. The latest version of PAMGUARD available at the time of this work, viz., v1.13.02 BETA, was used. For testing the method of Roch *et al.* (2011b), a MATLAB based implementation was employed. The implementation used is available as a part of the *Silbido* (Roch *et al.*, 2011a) package at http://roch.sdsu.edu/software/silbido_JASA2011baseline.zip (accessed on December 13, 2014). The detector's parameters were set as shown in Table III. While some of the parameter values given in Tables II and III were chosen based on *a priori* knowledge, others were arrived at following short trials using a small subset of the test set. While results better than those shown here may be possible for the compared methods, determining the optimal combination of parameter values is a non-trivial task and is beyond the scope of this study.

Tests with synthetic data were repeated for different sensitivity settings for all three methods. For the proposed detector, the threshold settings were varied from 0.4 to 1. In PAMGUARD, the Trigger Threshold parameter of the click detector module was varied from 7 to 14 dB. The method described in Roch *et al.* (2011b) uses different thresholds in the two stages of the detection algorithm. The stage 1 threshold parameter was varied from 2 to 16 dB with the stage 2 threshold set at 5, 10, 25, and 50. Testing was repeated for the proposed detector, with pre-filtered inputs, where the synthesised data were bandpass filtered (with passbands of 3–30 kHz for sperm whales and 20–80 kHz for beaked whales) before being fed to the detector.

With all three methods reporting detections as intervals (start and end times), a click present in input data (real or synthesised) is considered "detected" if any of the following are true:

- The known/recorded interval of the click in the input audio completely envelops the intervals of any reported detections.

TABLE III. Parameter settings used to configure the click detector of Roch *et al.* (2011b).

Parameter	Sperm Whale	Beaked Whale
Ranges (kHz)	3–30	20–80
MinClickSaturation (kHz)	1.5	10
MaxClickSaturation (kHz)	30	60
MeanAve_s (s)	3	3
TransitionBand (kHz)	0.2–3	3–20
FrameLength_s (s)	0.01	0.01
ClickPad_s (s)	0.0075	0.0075
MinClickSep_s (s)	0.5	0.5
ClipThreshold	(Disabled)	(Disabled)

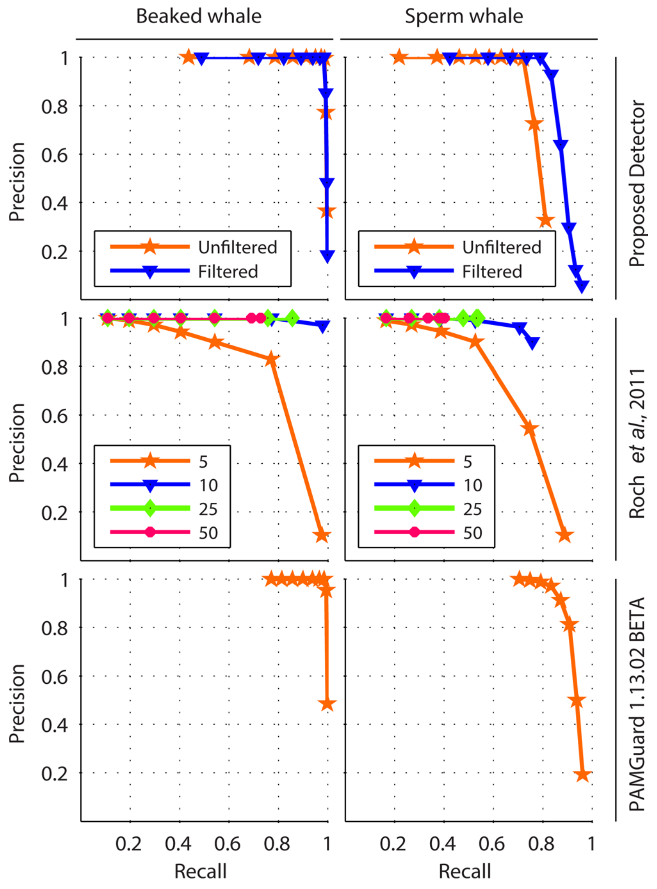

FIG. 8. (Color online) Detector performance on synthesised data—precision-recall trade-off curves.

- A reported detection's interval completely envelops the known/recorded interval of the click.
- The temporal overlap with any reported detection is at least 60% of the known/recorded duration of the click.

In the case of synthesised data, a significant portion of each click occurs around the midpoint of the containing clip. Therefore 60% overlap ensures that the click is appropriately accounted for by any partially overlapping detection. Reported detections that enable any of the preceding three conditions to be satisfied are considered to be "true detections." With these definitions of detected clicks and true detections, performance metric "recall" can be defined as the ratio of the number of detected clicks to the number of clicks present in the test inputs, and the metric "precision" can be defined as the ratio of the number of true detections to the number of reported detections. Figure 8 shows the precision-recall (PR) trade-off characteristics for the three detectors. The various curves in the middle row plots show the PR characteristics for the different stage 2 threshold settings considered. Threshold settings that produced optimal PR trade-off values were identified from Fig. 8 for the three detectors and the variation of the detectors' recall as a function of clicks' SNR were assessed at these thresholds. The corresponding results are shown in Fig. 9. Figure 10 summarises the detector's performance in capturing the pre-annotated clicks of different species in real underwater audio recordings.

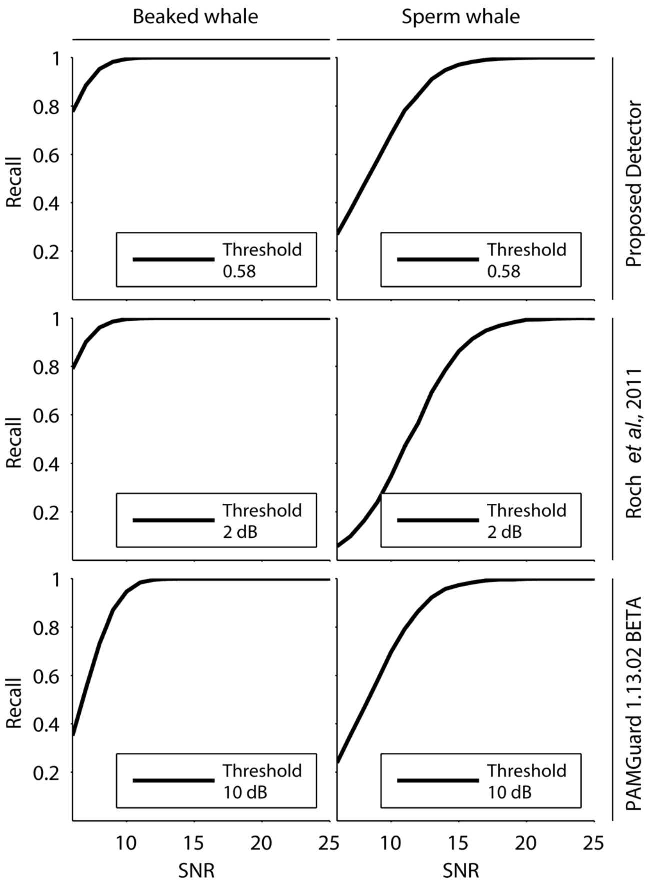

FIG. 9. Detector performance on synthesised data—recall vs SNR. Results for the proposed detector are shown for tests performed with bandpass-filtered inputs. Results for the detector of Roch *et al.* (2011b) are shown for tests performed with a stage 2 threshold of 10 and the plot legend indicates the stage 1 threshold.

For the proposed method, comparing the PR curves for filtered and unfiltered inputs, we can see that improvements in performance can be achieved with appropriate filtering of the input signals. Further improvement in species-targeted detection performance may be possible with an appropriate tuning of FWHM (or σ_G) in MAF1. However, this is a subject for further investigation.

The real-time factor of a detection/classification system is an indicator of its speed/throughput and is defined as the ratio of the time taken by the system for processing a given input to the duration of the input. Smaller the real-time factor, faster is the system. When tested on a desktop computer with an Intel$^{®}$ i7 CPU and 16 GB of RAM (running Microsoft$^{®}$ WINDOWS 7), a MATLAB implementation of the proposed detector exhibited an average (over different thresholds) real-time factor of 0.019 for 192 kHz audio and 0.007 for 96 kHz audio. For the optimal threshold setting identified from Fig. 8, the real-time factor was 0.019 as well. When run on the same computer, PAMGUARD processed the synthesised data with an average real-time factor of 0.058 at the threshold setting of 10 dB. Meaningful real-time factors could not be determined for the implementation of the method of Roch *et al.* (2011b) owing to the serialisation and the subsequent reloading of intermediate results across stages.

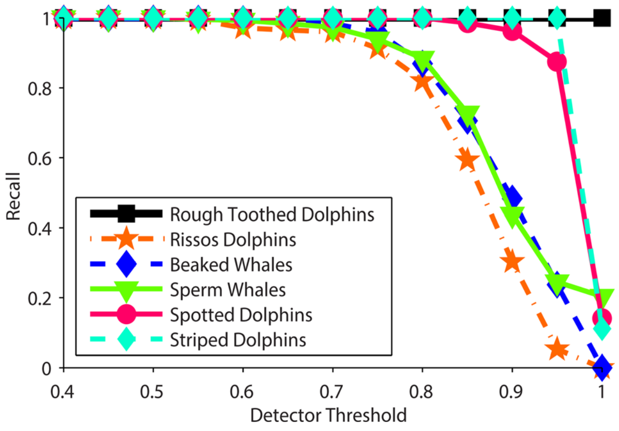

FIG. 10. (Color online) Detector recall (as a function of threshold) on real underwater recordings containing echolocation clicks of different species.

V. DISCUSSION

An automatic detector of echolocation clicks was suggested and tested in this study. As shown with the mathematical formulation, the carrier frequency component of an echolocation click virtually disappears in its TKEO output when the carrier frequency is either constant or varying nearly linearly with time. An additional benefit of this property is that it makes an implementation of the detector immune to species' calling behaviour variations that would affect the clicks' frequency content (Au, 1993, p. 121). This was validated by the performance of the detector on a variety of recorded clicks with no changes in detector settings. The robustness of the system with varying SNRs was demonstrated in the tests with synthesised data. The evaluation with the audio procured from MobySound also showed that the detector worked well with different recording scenarios. The audio recordings were obtained from different geographical locations while the data collection in each set was performed with different recording equipment configurations. The detector exhibited consistency in performance across all recordings used in the tests. Finally, as seen in Fig. 8, the performance of the proposed detector applied to prefiltered data is comparable to the other tested detectors. This shows that the proposed detector can also be used for targeted species' click detection with significant gain in processing speed.

The angle between the direction of a click's direct propagation path to a receiver and the orientation of the individual producing the click has been shown (e.g., Au, 1993; Møhl *et al.*, 2003; Au and Würsig, 2004; Madsen *et al.*, 2004; Au *et al.*, 2012) to have an impact on the waveform of the recorded clicks. While it can be argued that the theoretical signals considered may closely represent on-axis (having little or no relative angles) recorded clicks (Johnson *et al.*, 2006), it can be safely assumed that a majority of the clicks captured in open water recordings were off-axis (having high relative angles). Together, the theoretical proof and the experimental validation show that the detector performs well regardless of the calling species' orientation with respect to the recording equipment. A formal analysis of this sub-topic is a subject for further investigation.

The high processing speed and its simple control-flow make the proposed system feasible for pipelined hardware implementations. The few basic mathematical and logical operations that make up the system would take little processing time on modern hardware. Although there is already noticeable difference in the throughput as compared to PAMGUARD (see real-time factors in the preceding text), an implementation of the proposed system in C/C++ or JAVA has potential in yielding much higher speeds. Also, the response latency of the system is very small involving a one sample delay caused by the TKEO computation followed by a filter group delay of $\lceil 5\sigma_G/T_s \rceil + 1$, resulting in $(N+2)$ samples. Assuming that an implementation performs the two averaging/filtering operations in a parallel fashion, for the settings considered in the preceding tests, it can be shown that the maximum delay in reporting detections would be within $\sim$0.8 ms of the occurrence of the clicks.

ACKNOWLEDGMENTS

The authors and Curtin University would like to thank Chevron, who sponsored this work through their funding of research into Marine Noise under their participation in the Western Australian Energy Research Alliance (WA:ERA). The authors would also like to thank Dr. Aaron Thode of Scripps Institution of Oceanography and Dr. Marie Roch of San Diego State University for their suggestions that led us in improving the scope of this work.

[1] The publicly available underwater audio recordings from MobySound were found at http://www.mobysound.org (Last viewed March 3, 2014).

[2] Information about PAMGUARD is available at http://www.pamguard.org (Last viewed February 4, 2015).

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York), 277 pp.

Au, W. W. L., Branstetter, B., Moore, P. W., and Finneran, J. J. (**2012**). "Dolphin biosonar signals measured at extreme off-axis angles: Insights to sound propagation in the head," J. Acoust. Soc. Am. **132**, 1199–1206.

Au, W. W. L., and Würsig, B. (**2004**). "Echolocation signals of dusky dolphins (*Lagenorhynchus obscurus*) in Kaikoura, New Zealand," J. Acoust. Soc. Am. **115**, 2307–2313.

Caudal, F., and Glotin, H. (**2008**). "Stochastic matched filter outperforms Teager-Kaiser-Mallat for tracking a plurality of sperm whales," in *New Trends for Environmental Monitoring Using Passive Systems* (Hyeres, French Riviera), pp. 1–9.

DeRuiter, S. L., Bahr, A., Blanchet, M. A., Hansen, S. F., Kristensen, J. H., Madsen, P. T., Tyack, P. L., and Wahlberg, M. (**2009**). "Acoustic behaviour of echolocating porpoises during prey capture," J. Exp. Biol. **212**, 3100–3107.

Dobbins, P. (**2009**). "Time and frequency shifted click detection," in *Underwater Acoustic Measurements* (Nafplion, Greece), pp. 1–8.

Gabor, D. (**1946**). "Theory of communication. Part 1: The analysis of information," J. Inst. Electr. Eng. III Radio Commun. Eng. **93**, 429–441.

Gervaise, C., Barazzutti, A., Busson, S., Simard, Y., and Roy, N. (**2010**). "Automatic detection of bioacoustics impulses based on kurtosis under weak signal to noise ratio," Appl. Acoust. **71**, 1020–1026.

Gill, P. E., Murray, W., and Wright, M. H. (**1981**). "The Levenberg-Marquardt method," in *Practical Optimization* (Academic, London), pp. 136–137.

Goold, J. C., and Jefferson, T. A. (**2002**). "Acoustic signals from free-ranging finless porpoises (*Neophocaena phocaenoides*) in the waters around Hong Kong," The Raffles Bulletin of Zoology, pp. 131–139.

Greco, M., and Gini, F. (**2006**). "Analysis and modeling of echolocation signals emitted by Mediterranean bottlenose dolphins," Eurasip J. Appl. Signal Process. **2006**, 1–10.

Harland, E. (**2008**). "Processing the workshop datasets using the TRUD algorithm," Can. Acoust. **36**, 27–33.

Holland, R. A., Waters, D. A., and Rayner, J. M. V. (**2004**). "Echolocation signal structure in the Megachiropteran bat *Rousettus aegyptiacus* Geoffroy 1810," J. Exp. Biol. **207**, 4361–4369.

Jarvis, S., DiMarzio, N., Morrissey, R., and Moretti, D. (**2008**). "A novel multi-class support vector machine classifier for automated classification of beaked whales and other small odontocetes," Can. Acoust. **36**, 34–40.

Johnson, M., Madsen, P. T., Zimmer, W. M. X., de Soto, N. A., and Tyack, P. L. (**2006**). "Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation," J. Exp. Biol. **209**, 5038–5050.

Kaiser, J. F. (**1990a**). "On a simple algorithm to calculate the 'energy' of a signal," in *International Conference on Acoustics, Speech, and Signal Processing, 1990 (ICASSP-90)*, Albuquerque, NM, pp. 381–384.

Kaiser, J. F. (**1990b**). "On Teager's energy algorithm and its generalization to continuous signals," in *Fourth IEEE Digital Signal Processing Workshop*, Mohonk, NY, pp. 17–18.

Kamminga, C., and Beitsma, G. R. (**1990**). "Investigations on cetacean sonar IX: Remarks on dominant sonar frequencies from *Tursiops truncatus*," Aquat. Mamm. **16**, 14–20.

Kamminga, C., and Stuart, A. C. (**1995**). "Wave shape estimation of delphinid sonar signals, a parametric model approach," Acoust. Lett. **19**, 70–76.

Kamminga, C., Stuart, A. C., and Silber, G. K. (**1996**). "Investigations on cetacean sonar XI: Intrinsic comparison of the wave shapes of some members of the Phocoenidae family," Aquat. Mamm. **22**, 45–55.

Kamminga, C., van Hove, M. T., Engelsma, F. J., and Terry, R. P. (**1993**). "Investigations on cetacean sonar X: A comparative analysis of underwater echolocation clicks of Inia spp. and Sotalia spp.," Aquat. Mamm. **19**, 31–43.

Kandia, V., and Stylianou, Y. (**2006**). "Detection of sperm whale clicks based on the Teager-Kaiser energy operator," Appl. Acoust. **67**, 1144–1163.

Kandia, V., and Stylianou, Y. (**2008**). "A phase based detector of whale clicks," in *New Trends for Environmental Monitoring Using Passive Systems* (Hyeres, French Riviera), pp. 1–6.

Klinck, H., and Mellinger, D. K. (**2011**). "The energy ratio mapping algorithm: A tool to improve the energy-based detection of odontocete echolocation clicks," J. Acoust. Soc. Am. **129**, 1807–1812.

Madsen, P. T., Kerr, I., and Payne, R. (**2004**). "Echolocation clicks of two free-ranging, oceanic delphinids with different food preferences: False killer whales *Pseudorca crassidens* and Risso's dolphins *Grampus griseus*," J. Exp. Biol. **207**, 1811–1823.

Mann, S., and Haykin, S. (**1991**). "The chirplet transform: A generalization of Gabor's logon transform," in *Vision Interface'91*, Calgary, Canada, pp. 205–212.

Marčelja, S. (**1980**). "Mathematical description of the responses of simple cortical cells*," J. Opt. Soc. Am. **70**(11), 1297–1300.

Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A., and Lund, A. (**2003**). "The monopulsed nature of sperm whale clicks," J. Acoust. Soc. Am. **114**, 1143–1154.

Morrissey, R. P., Ward, J., DiMarzio, N., Jarvis, S., and Moretti, D. J. (**2006**). "Passive acoustic detection and localization of sperm whales (*Physeter macrocephalus*) in the tongue of the ocean," Appl. Acoust. **67**, 1091–1105.

Rankin, S., Baumann-Pickering, S., Yack, T., and Barlow, J. (**2011**). "Description of sounds recorded from Longman's beaked whale, *Indopacetus pacificus*," J. Acoust. Soc. Am. **130**, EL339–EL344.

Roch, M. A., Brandes, T. S., Patel, B., Barkley, Y., Baumann-Pickering, S., and Soldevilla, M. S. (**2011a**). "Automated extraction of odontocete whistle contours," J. Acoust. Soc. Am. **130**, 2212–2223.

Roch, M. A., Klinck, H., Baumann-Pickering, S., Mellinger, D. K., Qui, S., Soldevilla, M. S., and Hildebrand, J. A. (**2011b**). "Classification of echolocation clicks from odontocetes in the Southern California Bight," J. Acoust. Soc. Am. **129**, 467–475.

Roch, M. A., Soldevilla, M. S., Hoenigman, R., Wiggins, S. M., and Hildebrand, J. A. (**2008**). "Comparison of machine learning techniques for the classification of echolocation clicks from three species of odontocetes," Can. Acoust. **36**, 41–47.

Soldevilla, M. S., Henderson, E. E., Campbell, G. S., Wiggins, S. M., Hildebrand, J. A., and Roch, M. A. (**2008**). "Classification of Risso's and Pacific white-sided dolphins using spectral properties of echolocation clicks," J. Acoust. Soc. Am. **124**, 609–624.

Thorpe, C. W., and Dawson, S. M. (**1991**). "Automatic measurement of descriptive features of Hector's dolphin vocalizations," J. Acoust. Soc. Am. **89**, 435–443.

van der Schaar, M., Delory, E., van der Weide, J., Kamminga, C., Goold, J. C., Jaquet, N., and Andre, M. (**2007**). "A comparison of model and non-model based time-frequency transforms for sperm whale click classification," J. Mar. Biol. Assoc. U.K. **87**, 27–34.

Zimmer, W. M. X., Johnson, M. P., Madsen, P. T., and Tyack, P. L. (**2005**). "Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*)," J. Acoust. Soc. Am. **117**, 3919–3927.