On the Detection of Impulsive and Tonal Events in Passive Acoustics Monitoring

Ramón Miralles, Guillermo Lara, Alicia Carrión, Jorge Gosalbez, Ignacio Bosch
Instituto de Telecomunicaciones y Aplicaciones Multimedia (iTEAM)
Universitat Politècnica de València
Camino de Vera S/N, 46022, Valencia, Spain
web: www.gts.upv.es
Email: rmiralle@dcom.upv.es

***Abstract*—Anthropogenic underwater sound is now recognized as a world-wide problem. Shipping, seismic surveys and pile driving, among some other human activities, have shown a broad range of negative effects in a variety of species. It is thus important to control, register and characterize these activities in order to establish policies that mitigate its effects. For this purpose, automatic detectors of the major anthropogenic sound categories (impulsive and tonal sounds) are needed. An efficient algorithm capable of detecting impulsive and tonal sound categories is proposed. The detection algorithm is based on the Pulsed to Tonal Ratio (*PTR*) and, in this work, we show not only its simplicity but also how well it performs in a variety of situations. In contrast to what happens with some other techniques, the proposed detector generally succeeds in detecting pulsed and tonal events with additive white and pink noise. This makes the *PTR* detector particularly appropriate for passive acoustic monitoring.**

I. Introduction

The presence of anthropogenic sounds in our oceans and seas is a topic of major concern and has been recognized as a world-wide problem. Recent studies have shown a broad range of negative effects that exposure to high levels of this sounds can have on marine species and ecosystems [1], [2]. Commonly, anthropogenic sounds are divided into pulsed (or impulsive) sounds and continuous sounds. Pulsed sounds can be originated, for example, by marine pile driving, airgun seismic explorations, explosions, etc. Examples of continuous sound include ship noise, operational noise from machinery including marine renewable energy devices, etc. These two types of sounds have specific characteristics which highlight the importance of choosing the most appropriate method for its measurement and characterization. Whereas the most suitable metric for continuous sounds is the Sound Pressure Level (SPL), the measurement of impulsive sounds is best characterized by the Sound Exposure Level (SEL) or the peak sound pressure level. It is thus important to be able, not only to detect, but also to distinguish between these two events in a real application.

The detection of impulsive events has been broadly studied in some fields such us geophysics [3]. One of the most basic and commonly used detection technique is the Short Term Average / Long Term Average (STA/LTA). This is mainly due to its ease of implementation, but also to its surprisingly good performance on many data sets. Some authors have proposed improvements over the basic method such as those based on taking the marginals of the Time-Frequency Representation (TFR) [4]. Regarding the detection of tonal components, if the fundamental frequency is unknown, most of the algorithms work by tracking the spectral peaks over time, grouping together peaks in successive time slices in a spectrogram if the central frequency of the peak does not change. The same technique is employed for the detection of marine mammals whistles [5], [6], [7].

Having a single algorithm capable of efficiently combining the detection of such different events (impulsive and tonal) is a challenging task. A trade-off between temporal and frequencial resolution exists and there is no easy way to find the optimum approach to create that kind of algorithm. Fortunately, we can use a measurement named Pulsed to Tonal Ratio (*PTR*) as described in [8]. This ratio characterizes the amount of energy distributed in the tonal and pulsed components and, as we will see in the next section, can be used to devise a simple and robust detector of impulsive and tonal sounds.

II. The Pulsed to Tonal Ratio detector

The *PTR* of a fragment of a discrete time signal $x_i(n)$ can be defined as the mean power of the pulsed components of the sound P_{pulsed}, divided by the mean power of tonal components of the sound P_{tonal}. The *PTR* can be calculated as shown in equation (1).

$$PTR\ (dB) = 10 \cdot log\left[\frac{P_{pulsed}}{P_{tonal}}\right] \tag{1}$$

From the definition, it is clear that the *PTR* can play an important role when detecting and classifying pulsed and tonal events. As demonstrated in [8], a feasible way to obtain the *PTR* is by computing the two-dimensional Discrete Cosine Transform (DCT_{2D}) of any TFR and taking its marginals (see equation (2)),

$$PTR\ (dB) = 10 \cdot log\left[\frac{\sum_{u=0}^{N_1-1} F(u,0)^2}{\sum_{v=0}^{N_2-1} F(0,v)^2}\right] \tag{2}$$

where $F(u,v)$ is the DCT_{2D} of the TFR and N_1, N_2 are related respectively to the number of time and frequency bins of the TFR. Figure 1 illustrates the idea behind the computation of the marginals of the DCT_{2D} to get the *PTR*.

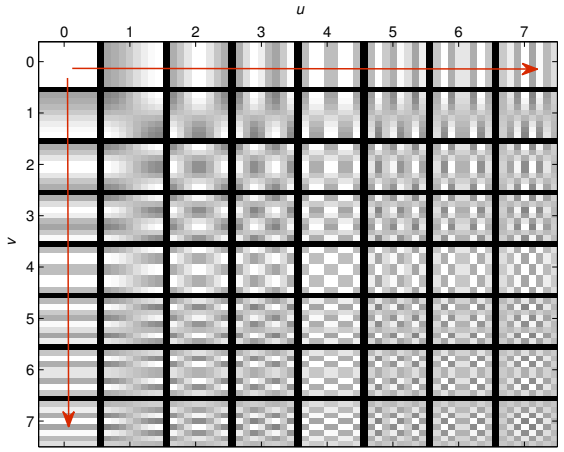

Fig. 1. The DCT_{2D} basis for a transformation of size 8x8. Taking the marginals of $F(u, v)$, indicated by the red arrows, is related to the energy in the pulsed and tonal components.

What makes the *PTR* a powerful tool when working as a detector is the ability to obtain the *PTR* without needing to calculate any kind of TFR. As shown in [8], the computation of the ratio can be greatly simplified if the associated discrete TFR satisfies marginal properties such as discrete Wigner Ville Distribution or discrete positive distributions: Coehn's class type II, etc. For these distributions, the computation of the marginals is greatly simplified. Thus, we can obtain the *PTR* using equations (3) and (4) in equation (2)

$$F(u,0) = DCT_{1D}[K_1 \cdot |x_i(n)|^2] \tag{3}$$
$$F(0,v) = DCT_{1D}[K_2 \cdot |X_i(k)|^2] \tag{4}$$

where $X_i(k)$ is the Fourier Transform of the audio chunk ($X_i(k) = DFT_{N_2}[x_i(n)]$). The constants K_1 and K_2 depend on how the particular positive distribution is extended to discrete-time signals. For example, the Discrete Wigner Ville defined in [9] has $K_1 = N_2$ and $K_2 = N_1$.

A detector of pulsed and tonal events based on the *PTR* can be as follows:

1) For a given audio signal $x(n)$ and analysis window of length $T_{analysis}$ s. (or $T_{analysis} \cdot f_s$ samples), obtain non-overlapping audio fragments $x_i(n)$.
2) Compute the *PTR* of each one of the audio fragments $\{PTR_i\}$ using equations (3), (4) and (2) with $K_1 = N_2$ and $K_2 = 1$.
3) Compute the median of the set of *PTR*s and subtract it ($\widehat{PTR}_i = PTR_i - median[PTR_i]$).
4) Use a threshold β_1 to classify the fragments as pulsed if $\widehat{PTR}_i$ is above the threshold $\{i_{pulsed} : \widehat{PTR}_i > \beta_1\}$ or as tonals if $\widehat{PTR}_i$ is below the threshold $\{i_{tonals} : \widehat{PTR}_i < -\beta_1\}$.

Figure 2 shows an example of the proposed *PTR* detector when applied to a synthetic audio signal composed of two tones and two impulses. The *PTR* was computed in this example for a window of length $T_{analysis} = 0.31$ s. and a threshold $\beta_1 = 3$ dB.

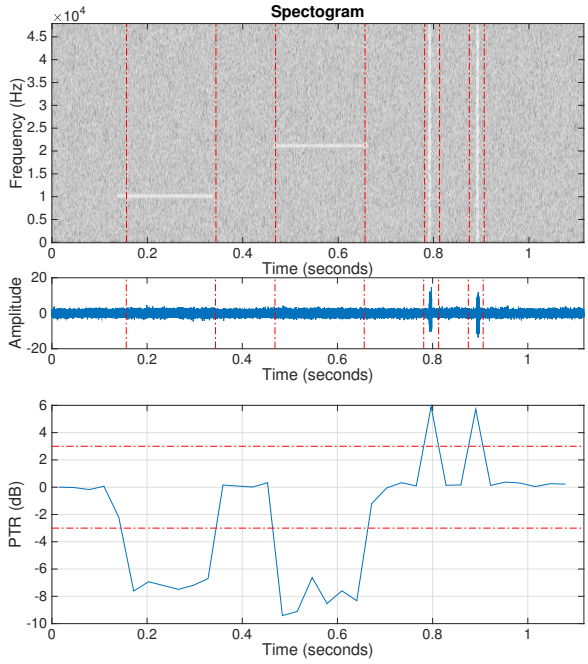

Fig. 2. Example of the proposed *PTR* detector. Top panel: Time-frequency representation of a noise signal containing two tonal events and two impulsive events (the beginning and the end of these events is indicated by the red-dashed vertical lines). Middle panel: Time representation of the noise signal. Bottom panel: Computing the *PTR* of the noise signal and establishing a threshold (± 3 dB in this example indicated by the red-dashed horizontal lines) allows the detection of the tonal and impulsive events.

III. Comparison with the STA/LTA impulse detector.

We are going to compare the ability of detecting pulses of the proposed *PTR* detector with that of the STA/LTA impulse detector. The STA/LTA is easily implemented as ratio of the mean square value of the signal in a short time window divided by a the mean square value of the signal in a long time window. In the implementation used in this work the short window operates on more recent data than the long window, and there is no overlap or delay. The mean square value of the short time window is given by

$$STA_i = \frac{1}{N_S} \sum_{n=i}^{i+N_S-1} x^2(n) \tag{5}$$

, whereas the mean square value of the long time window is given by

$$LTA_i = \frac{1}{N_L} \sum_{n=i-N_L}^{i-1} x^2(n) \tag{6}$$

with $i = \{..., -2N_S, -N_S, 0, N_S, 2N_S, ...\}$. We can obtain the ratio STA_i/LTA_i dividing equation (5) and (6). In order to make it easy to compare the *PTR* and the *STA/LTA* detector we have chosen N_S (the length of the short time window) to be the same as the length of the analysis window in the *PTR* ($N_S = T_{analysis} \cdot f_s$). The length of the long time window was $N_L = 4 \cdot N_S$.

A. Simulations with synthetic data.

An additive model, described in equation (7), was used to simulate the effect of an impulsive sound $p(t)$ in a noisy environment $n(t)$.

$$x(t) = p(t) + n(t) \tag{7}$$

The impulsive sound $p(t)$ of bandwidth BW can be simulated by adding a sufficient number of sinusoids with frequencies ranging from f_{low} to f_{high} ($BW = f_{high} - f_{low}$). Equation (8) shows how $p(t)$ was obtained.

$$p(t) = A \cdot \left(\sum_{f_i = f_{low}}^{f_{high}} cos(2\pi f_i t) \right) \cdot w(t) \tag{8}$$

The term $w(t)$ is a Hamming window of length T_{pulse} that controls the duration of the transient, whereas the term A controls the amplitude. In the simulations presented here, A was adjusted according to the desired Signal to Noise Ratio (SNR).

In equation (7), $n(t)$ represents the underwater background noise. Underwater noise in some frequency bands can be modeled as white noise, whereas in other frequency bands it can be modeled as continuous and negatively sloping or pink noise [10]. As a result of that, we tested each detector with both: white and pink noise. The pink noise spectrum was modeled as $S_n(f) = 1/f^{0.35}$.

Monte Carlo simulations were done to evaluate the detection and false alarm probabilities when varying SNR. We performed 400 runs for each one of the SNR values computed as $SNR = 10log\left(\frac{\sigma_p^2}{\sigma_n^2}\right)$. The central frequency of the impulsive sound $f_0 = (f_{high} + f_{low})/2$ was randomly changed in each of the runs in the frequency range $[BW/2, f_s/2 - BW/2]$, f_s being the sampling frequency. The impulsive sound duration was set to $T_{pulse} = 0.1$ s. with t varying in the interval $[-50, 50]$ ms. and its position within the 10 seconds noise register was randomly changed in each Monte Carlo run. The resulting synthetic audio was used for both detectors: the STA/LTA and the *PTR*. For the *PTR* detector the parameter $T_{analysis}$, was set to 0.25 s. Thresholds were adjusted to obtain similar False Alarm probabilities in both detectors. The results are shown in Figures 3 and 4.

Figure 3 shows that the STA/LTA detector outperforms the *PTR* detector for white noise by approximately 10%. However, when it comes to pink noise, the performance of the STA/LTA is reduced drastically and high SNR is needed to detect the presence of impulsive sounds (see Figure 4). This is due to the slow oscillations appearing in the pink noise registers that strongly affect the mean square value of the long time window.

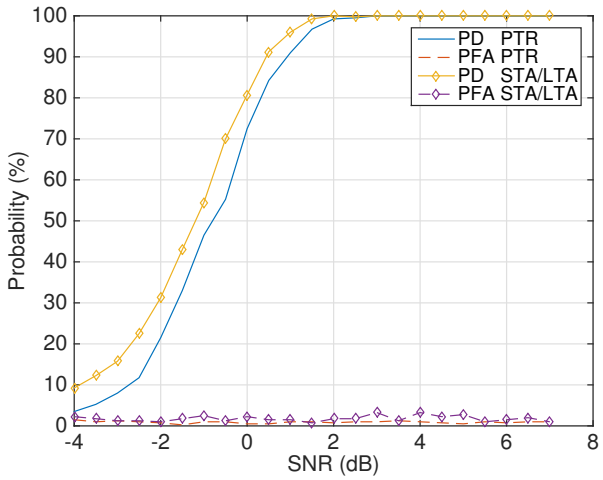

Fig. 3. Monte Carlo simulation of the *PTR* and STA/LTA detectors in presence of white noise. The detection threshold was chosen to obtain a constant PFA=1.5 % in both detectors.

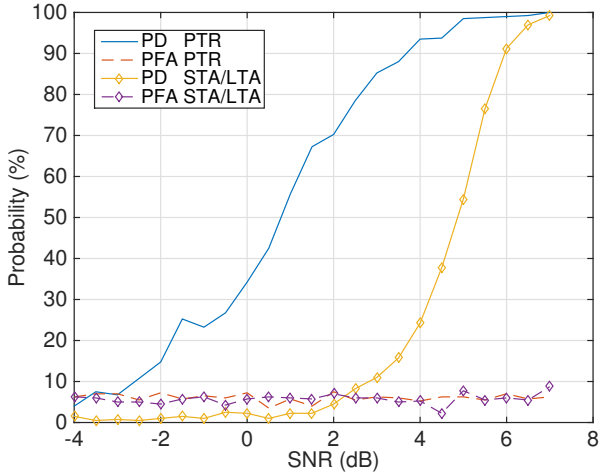

Fig. 4. Monte Carlo simulation of the *PTR* and STA/LTA detectors in presence of pink noise. The detection threshold was chosen to obtain a constant PFA=7% in both detectors.

IV. Comparison with peak tracking detectors.

Regarding the detection of tonal sounds and with the aim of establishing a comparison with the *PTR*, we have chosen a spectrogram-based algorithm that tracks spectral peaks over time. The employed Peak Tracking detector (PEAKT) can be summarized as follows:

1) For a given audio signal $x(n)$, obtain the short time Fourier Transform $P(n, k)$ with a spectrogram window size of 256 points and a 85% of overlapping samples.
2) Use morphological image processing to find connected points (with a connectivity of 8) and look for horizontal lines in the spectrogram.

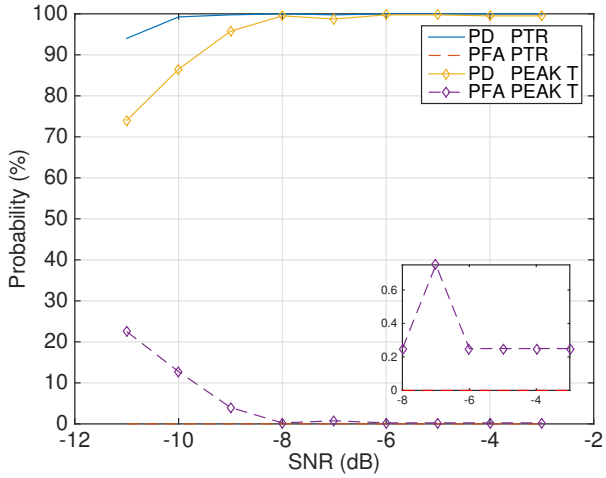

Fig. 5. Monte Carlo simulation of the *PTR* and PEAKT detectors in presence of additive white noise.

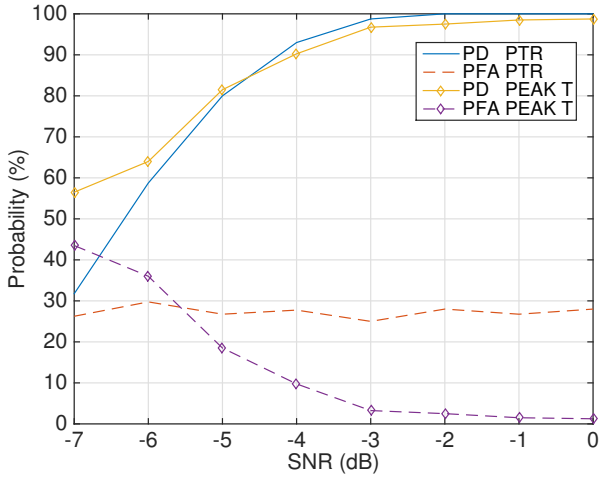

Fig. 6. Monte Carlo simulation of the *PTR* and PEAKT detectors in presence of additive pink noise.

Some authors use click removal, median filtering, and different kinds of spectrogram smoothing-techniques prior to morphological image processing and peak tracking [7]. We have deliberately not used such techniques to maintain a similar computational complexity when comparing the PEAKT and the *PTR* detector.

A. Simulations with synthetic data.

As it was done in subsection III-A, we have simulated the effect of a tonal component $s(t)$ in a noisy environment $n(t)$ by means of an additive model (see equation (9)).

$$x(t) = s(t) + n(t) \tag{9}$$

The tonal sound was modeled, as described in equation (10), with a sinusoidal of central frequency f_0.

$$s(t) = A \cdot cos(2\pi f_0 t) \cdot r(t) \tag{10}$$

The term $r(t)$ is a rectangular window of length T_{tonal} that controls the duration of the tonal component.

Again, 400 Monte Carlo runs were done for each one of the SNR values. The central frequency of the tonal component f_0 was randomly changed in the frequency range $[0, f_s/2]$. The duration of the tonal sound was $T_{tonal} = 0.25$ s and its position within the 10 seconds noise register was also randomly changed in each Monte Carlo run. The resulting synthetic audio was used for both detectors: the PEAKT and the PTR (again with $T_{analysis} = 0.25$ s.).

Figures 5 and 6 show the simulation results for white and pink noise respectively. In contrast to what happens with the rest of the detectors, the PEAKT detector has a PFA that changes with the SNR. Consequently, comparisons are only valid within a certain range of SNR. Figure 5 shows the comparison of the two detectors in white noise. Again, the *PTR* detector gives higher detection percentages and lower probability of false alarms than the PEAKT detector although the performance of both detectors is very good for SNR> -8 dB.

Figure 6 shows the simulation results for tonal events in pink noise. In this case, the proposed *PTR* detector performs generally worse than the PEAKT detector, even though for SNR < -5.5 dB the *PTR* gives higher detection percentages and a lower false alarms ratio.

V. Testing the detector in a real scenario

In order to show how the proposed *PTR* detector works in real world signals, we have used acoustic recordings obtained with a passive acoustic monitoring system, named SAMARUC. The SAMARUC system has been developed by the authors and it is not only capable of recording the audio events, but it may also be programmed with custom signal processing algorithms. As a result, it provides at recovery time a list of detected events as well as a summary of the whole recording campaign. Recordings were carried out at the Cabrera Marine National Park from March 2014 until September 2015. The SAMARUC system was configured with a sampling frequency of f_s=5200 Hz, and the system was in recording mode for 5 minutes and in standby mode for 10 minutes. The *PTR* detector was used to obtain the detected pulsed and tonal events of the complete data set of audio files. A particular example is given in Figures 7 and 8 for impulsive and tonal events. As it can be seen in Figure 7, although the low SNR of the impulsive events, most of them are detected with a threshold of +2.5 dB. In fact, in this particular example, all events could be detected using a threshold of +2 dB. However, this might lead to false alarms in some other situations. For the tonal events, having a high SNR, all events are properly detected with a threshold of -2.5 dB (Figure 8).

VI. Conclusion

When detecting anthropogenic sounds in passive acoustic monitoring, it is important to distinguish between impulsive

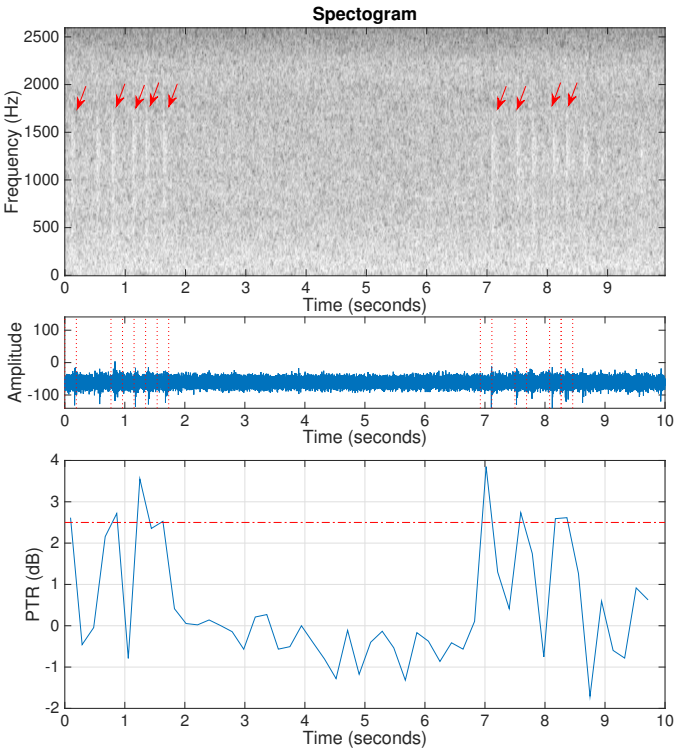

Fig. 7. Example of the proposed *PTR* detector on real passive acoustic recordings. Top panel: Time-frequency representation of a recording with low level impulsive events (the detected events are marked with red arrows). Middle panel: Time representation of the recording. Bottom panel: *PTR* curve and detector when establishing a threshold of $+2.5$ dB.

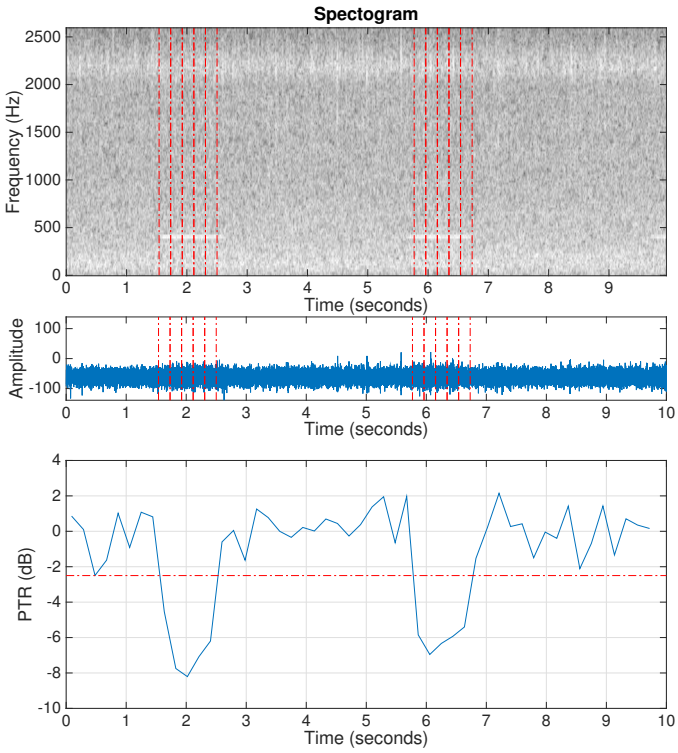

Fig. 8. Example of the proposed *PTR* detector on real passive acoustic recordings. Top panel: Time-frequency representation of a recording containing tonal events. Middle panel: Time representation of the recording. Bottom panel: *PTR* curve and detector when establishing a threshold of -2.5 dB.

and continuous sounds. Having a single algorithm for the detection of such different events (impulsive and tonal) is a challenging task, and most of the time specific detectors are used. In this work we propose using the *PTR* to create a detector that is capable of classifying the acoustic events as impulsive or tonal. Although not originally intended to be used as a detector, the *PTR* has some interesting properties: low computational complexity, simplicity and high sensitivity. This allows us to create a simple algorithm that, without the need to compute the spectrogram, can achieve high detection and classification percentages. Simulations show that the *PTR* produces similar or even superior detection percentages when compared to the STA/LTA impulse detector. The *PTR* detector also outperforms the PEAKT detector when detecting tonal sounds in white noise. However, in the presence of pink noise the *PTR* performs worse than the PEAKT detector. Nevertheless, it is important to emphasize that the proposed *PTR* detector has a constant false alarm probability for a fixed threshold and independently of the SNR. The *PTR* has shown good performance in real world passive acoustic recordings.

Acknowledgment

This work has been partially supported by the Spanish Administration under Grant BIA2014-55311-C2-2-P, by the Fundación Biodiversidad and by European Commision DG Environment (QUIETMED).

References

[1] J.J. Finneran. Noise-induced hearing loss in marine mammals: A review of temporary threshold shift studies from 1996 to 2015. *The Journal of the Acoustical Society of America*, 138(3):1702–1726, September 2015.

[2] H.P. Kunc, K.E. McLaughlin, and R. Schmidt. Aquatic noise pollution: implications for individuals, populations, and ecosystems. *Proceedings of the Royal Society B: Biological Sciences*, 283(1836):20160839, August 2016.

[3] R.V. Allen. Automatic earthquake recognition and timing from single trace. *Bulletin of the Seismological Society of America*, 68(5):1521–1532, October 1978.

[4] S.R. Taylor, S. J. Arrowsmith, and D.N. Anderson. Detection of Short Time Transients from Spectrograms Using Scan Statistics. *Bulletin of the Seismological Society of America*, 100(5A):1940–1951, October 2010.

[5] D.K. Mellinger, S.W. Martin, R.P. Morrissey, L. Thomas, and J.J. Yosco. A method for detecting whistles, moans, and other frequency contour sounds. *The Journal of the Acoustical Society of America*, 129(6):4055–4061, June 2011.

[6] T.H. Lin, L.S. Chou, T. Akamatsu, H.C. Chan, and C.F. Chen. An automatic detection algorithm for extracting the representative frequency of cetacean tonal sounds. *The Journal of the Acoustical Society of America*, 134(3):2477–2485, September 2013.

[7] D. Gillespie, M. Caillat, J. Gordon, and P. White. Automatic detection and classification of odontocete whistles. *The Journal of the Acoustical Society of America*, 134(3):2427–2437, September 2013.

[8] R. Miralles, G. Lara, J.A. Esteban, and A. Rodriguez. The pulsed to tonal strength parameter and its importance in characterizing and classifying Beluga whale sounds. *The Journal of the Acoustical Society of America*, 131(3):2173–2179, March 2012.

[9] M.S. Richman, T.W. Parks, and R.G. Shenoy. Discrete-time, discrete-frequency time-frequency representations. volume 2, pages 1029–1032. IEEE, 1995.

[10] R.J. Urick. *Ambient Noise in the Sea*. Undersea Warfare Technology Office, Naval Sea Systems Command, Department of the Navy, Washington, D.C. 20362, 1984.