Higher-order statistics for bioacoustic click detection

Sylvain Busson¹, Cédric Gervaise¹, Amélie Barazzutti¹, Bazile Kinda¹,
Virgine Jaud¹, Laurent Chauvaud², Anthony Robson²

¹ ENSIETA, DTN, 2 rue F. Verny, 29206 Brest, sylvain.busson@ensieta.fr
² LEMAR, IUEM Technopôle Brest-Iroise, Place N. Copernic, 29280 Plouzané, laurent.chauvaud@univ-brest.fr

The passive acoustic monitoring (PAM) is a tool of choice for non-intrusive study of aquatic organisms in the wild. Anthropogenic disturbance may affect animal behavior and should generally be minimised. In this study, a PAM system was used to record the bioacoustic sound produced by the valve movements of the great scallop (Pecten maximus). The primary function of PAM system is to detect bioacoustic emissions embedded in the background noise. Sounds produced by the great scallop are transients. Detecting them with power based detector could lead to misses particularly when signal to noise ratio is weak. In this paper, a click detection scheme based on higher-order statistics is applied to the monitoring of the great scallop in its habitat. The detector is effective in tracking the sound produced by the valve movements of the great scallop. The PAM system in association with the click detector represent a promising tool for monitoring, with minimal anthropogenic disturbance, the behavior of minimally mobile aquatic organisms in the wild.

1 Introduction

Passive acoustic monitoring (PAM) [1] is effective for underwater sound surveys. It has been successfully deployed in sea observatories to study the impacts of anthropogenic sounds on marine mammals and for the mitigation of mass stranding due to navy sonar or seismic airgun impulses [2, 3, 4]. One of the major benefits of PAM is that it provides a wide range [5] non-intrusive and non-destructive monitoring.

Since it has been shown that biogenic marine carbonate could be indicators of pollution and climate variation [6], needs of methods for monitoring them are of growing importance. In particular, the relation between the bivalve movement and the growth of the great scallop (Pecten maximus) is of great interest. Anthropogenic disturbance may affect animal behavior and should generally be minimised [7]. PAM may be an effective non-intrusive method for monitoring bivalve movements.

Prior laboratory experiments with acoustic and video monitoring showed that the sound production of the scallop is composed of three types of sounds associated with scallop valve movements that result in the "cough", the "jump" and the "swim" [8]. The "breath sound" (see top panel of Fig.5) induced by a single valve adduction due to either cough or jump depending on the speed and magnitude of the adduction, the "creak sound" (see bottom panel of Fig.5) related to a valve abduction (creak sounds occur often just before a breath sound) and the "swim sound" which consists in sequences of jumps that enable the scallop to move away from its current location. In this paper we applied an impulse sound dedicated PAM algorithm [9] to the issue of the non-intrusive monitoring of the great scallop breath sound production.

The first function of PAM is the detection of bioacoustic signals in a mixture of ambient marine noises and bioacoustic productions. Detector effectiveness and detection range depend on the environmental conditions, the sea state and the hydrophone implementation. Ambient noise is low and stationary compared to bioacoustic signals in deep-sea observatories while it is loud, widely fluctuating and subject to spikes in coastal observatories. Common implementations of a detection function are based on the power of the signal. A signal to noise ratio (SNR) of about 14 dB is required for reliable detection [5]. The common estimation of noise level by signal averaging over a time interval is not robust to abrupt changes in the signal level as encountered in stochastic noise conditions. Applying a power detector in non-stationary noise conditions will result in false alarms and misses.

The Teager-Kaiser (TK) operator has been proposed as an alternative to power detectors for transient sounds. It has been shown effective in tracking clicks when applied on already isolated data segments of interest [10] but its automated implementation suffers from false alarms under noise-only conditions [11].

Kurtosis as well as higher order statistics have been widely used in signal processing [12, 13, 14, 15]. In particular, the signal kurtosis and the higher order moments are tools of choice for the detection of transients in underwater sounds [16] the investigation of stationarity and gaussianity of underwater noise [17]. Performances of high order detectors are improved via pre-filtering of the signal [18]. In the work related here, we use a preliminary data conditioning step using an automated processing filtering scheme prior proceeding with click detection. The outcomes of this filtering step are time window in the sound recordings, as well as optimal frequency bands focusing on targeted impulsive sounds. The assumption made in [9] is that click trains
are embedded in stochastic but Gaussian noise. Click detection is achieved by normal distribution tests, using the kurtosis of signal segments as a non-Gaussianity test. Applications of the click detector on field data show that the SNR of beaked whales clicks can be upgrade from -4.6 dB for the raw data up to 14 dB in the optimal frequency band [9].

In the present paper, we first recall how the click detection algorithm works. The algorithm is then applied to a synthetic signal with gaussian noise and to field data (monitoring of the great scallop in the wild) where the noise reveals to be non-gaussian.

2 Method

2.1 The detection theory

Let be a PAM system that gives a set of measurements \( m = \{m(t), t = nT_s, n[0, N - 1]\} \) with \( T_s = 1/f_s \) being the sampling period and \( f_s \) being the sampling frequency. \( m(t) \) can either containing only noise (hypothesis \( H_0 \), signal absent) or noise plus bioacoustic sound (hypothesis \( H_1 \), signal present). Given the measurement \( m \) and some a priori information about noise and signal, the detection task consists of discriminating between two hypotheses \( H_0 \) and \( H_1 \). The detector has two stages. The first stage applies a statistical test \( T \) on \( m \) and the second stage computes a threshold \( \lambda \). The decision between which hypotheses is correct is made as follows:

\[
\begin{cases}
H_0 : T < \lambda \\
H_1 : T > \lambda
\end{cases}
\]

If \( H_1 \) is indicated by the statistical test \( T \) when a true signal is present, it is called a correct detection. Let \( P_{fa} \) be the related probability. If \( H_1 \) is indicated by the statistical test \( T \) when a true signal is absent, it is a false alarm, with \( P_{fa} \) as the related probability.

The choice of test \( T \) and the estimation of probability distribution functions (pdf) for \( m \) and \( T \) are decided by a priori information on noise and signal. Let \( F_0(T) \) and \( F_1(T) \) be the respective cumulative distribution functions (cdf) of \( T \) for hypotheses \( H_0 \) and \( H_1 \), and let \( f_0(T) \) and \( f_1(T) \) be the pdfs of \( T \) for \( H_0 \) and \( H_1 \). For a chosen \( P_{fa} \) level of false alarms compatible with the function of the PAM system, \( \lambda \) and \( P_d \) are assessed:

\[
\lambda = F_0^{-1}(1 - P_{fa})
\]

\[
P_d = F_1(\lambda)
\]

\( T \) is chosen so that it maximizes the contrast between \( H_0 \) and \( H_1 \). Our study uses the kurtosis as the statistical test \( T \).

2.2 Kurtosis as a statistical test \( T \)

When \( X \) be a random variable, the kurtosis is defined as:

\[
K_X = \frac{E[(X - E(X))^4]}{E[(X - E(X))^2]^2}
\]

where \( E \) stands for ensemble expectation. \( K_X \in [1, +\infty[ \) is a measure of the degree of peakedness of a distribution. Pdf of random variables with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have fat tails. Pdfs with \( K_X \) higher than 3 are called leptokurtic and those with \( K_X \) lesser than 3 are called platykurtic. Because of the impulsive nature of clicks, we make the assumption that click sequences embedded in Gaussian noise have a leptokurtic distribution and we use kurtosis as a click detector.

Because the ensemble mean of \( m \) at time \( t \) cannot be assessed for real-data measurements, we use an estimate of \( K_X \). If \( m \) is a \( L \) samples signal, under the assumption of ergodicity and stationarity for \( m \) over \( N \) samples around sample \( n_0 \), we can define a ‘sliding window’ estimate for \( K_N \) around the index \( n_0 \in [1, L - 1] \), by:

\[
K_N = \frac{1}{n} \sum_{i=1}^{n} (m(i) - \bar{m})^4
\]

\[
\bar{m} = \frac{1}{N} \sum_{i=n_0-N/2+1}^{n_0+N/2} m(i)
\]

2.3 Architecture of a kurtosis based click detector

Biological impulse sounds cover a wide range of central frequencies (500 Hz for Right Whale gun shots, 4 kHz for off-axis click of sperm whales and from 10 kHz to 200 kHz for other odontocetes) while underwater ambient noise may be loud and non-stationary. SNR has to be high enough for good detection rate by a kurtosis detector (higher than 6 dB in the click frequency band [9]). Filtering of ambient noise is necessary to achieve proper SNR levels. PAM systems running over long periods require that the detector runs automatically without the need of human operators. The kurtosis based detection scheme we use automatically selects the click bandwidth and adapts its detection threshold to varying background noise.

The kurtosis click detector we use takes raw digital signal as input and gives as output time-frequency intervals where the click sequences are and optimally filtered waveforms of click sequences. Raw signal goes through a series of 6 processing blocks. The first block consists of computing a 4th order Butterworth filter bank. Given a bandwidth \( [f_{min}, f_{max}] \), which can be set to \( [0, f_s/2] \), and a frequency step \( \Delta f \), a grid \( f \) for lower or higher cutoff frequencies is computed:

\[
f : \{f(i) = f_{min} + i\Delta f, i \in \{0, R\}\}
\]

with \( R \) such as

\[
f_{min} + R\Delta f \leq f_{max}
\]

From grid \( f \), \( M = C^2_{LR}/2 \) allowable bandwidths are defined by:

\[
B : B(u) = [f_1 = f(i), f_2 = f(j)]
\]

with

\[
f(i) < f(j) \text{ and } u = i + R(j - 1)
\]
Figure 1: Sketch of the architecture of the automatic detection scheme based on kurtosis.

The raw signal $m(t)$ is passed through a filter bank $B$ to produce $M$ channels $u$ of bandpassed signals $m_f(u,t)$. In Block 2, a sliding window estimator of kurtosis (eq. (5)) is applied to $m_f(u,t)$ to produce $K_{N_{int}}^{m_f}(u,t)$. $N_{int}$ (or equivalently $T_{int} = N_{int}/f_s$, the window duration) is a parameter to set. $K_{N_{int}}^{m_f}(u,t)$ is compared to a decision threshold $\lambda(u,t)$ in Block 4. The threshold estimation made in Block 3 is based on the kurtosis property that for a given set of $m(t)$, lowest values of $K_{N_{int}}^{m_f}(u,t)$ are for noise only. At each $t_0$ time and $u$ bandwidth , Block 3 gathers a vector of $L$ seconds kurtosis values:

$$vK_{N_{int}}^{m_f}(u,t_0,L) = K_{N_{int}}^{m_f}(u,t) , t \in [-\frac{L}{2} + t_0 , \frac{L}{2} + t_0]$$

(11) $vK_{N_{int}}^{m_f}(u,t_0,L)$ are sorted in ascending order and only $p_{noise}$ percents of their maximum values are kept. Let $vK_{N_{int}}^{m_f}(u,t_0,L,p_{noise})$ be this collection of supposed 'noise only' samples of kurtosis. As mentioned in [9] $vK_{N_{int}}^{m_f}(u,t_0,L,p_{noise})$ is supposed to follow a Gaussian distribution $N(\mu,\sigma)$ where $\mu$ and $\sigma$ have to be estimated from $vK_{N_{int}}^{m_f}(u,t_0,L)$, $\mu$ and $\sigma$ cannot be estimated from $vK_{N_{int}}^{m_f}(u,t_0,L)$ via classic expectation and standard deviation formula because $vK_{N_{int}}^{m_f}(u,t_0,L,p_{noise})$ have been chosen from the smallest samples of the distribution. Moment estimation via classic formula would lead to biased estimation. To overcome this difficulty, we estimate the cdf $F_u(u,t_0,L,p_{noise})$ of $vK_{N_{int}}^{m_f}(u,t_0,L,p_{noise})$ instead. A non-linear least square fit (using Levenberg-Marquardt optimization scheme [20]) is applied between $F_u(u,t_0,L,p_{noise})$ and the cdf of a Gaussian random variable $F_G(\mu,\sigma)$. Optimization is initialized with $\mu$ and $\sigma$ obtained from $vK_{N_{int}}^{m_f}(u,t_0,L)$ via standard expectation and standard deviation formula. Let $F_G(\mu_{opt},\sigma_{opt})$ be the best fit, then given a target $P_{fa}$, detection threshold $\lambda(u,t_0)$ is obtained with equation (2). $(T_{int},L,p_{noise})$ are a triplet of parameters to set before running our detection scheme. Due to cumulative effect in equation (5), one click contributes to raise the kurtosis during $T_{int}$. If $T_{int}$ is chosen too long, the real percentage of 'noise only' sample will be too weak. $\lambda(u,t_0)$ will be overestimated and the actual $P_{fa}$ will drop. On the contrary, if $T_{int}$ is chosen too short the running window estimator of kurtosis (eq. 5) may deviate from a Gaussian random variable (because of the few samples taken into account). $\lambda(u,t_0)$ will be underestimated and the actual $P_{fa}$ will raise. $T_{int}$ should be close to the duration of a whole click sequence. Once $T_{int}$ is chosen, the key point to set $L$ and $p_{noise}$ is that $L$ seconds of kurtosis estimation should encompass $p_{noise}$ percents of 'noise only' values.

Block 5 looks over each channel of the filter bank. If there is no detection, hypothesis $H_0$ is decided. If there is detection in one or more channels $H_1$ is decided. In that case, the output of Block 5 is the bandwidth that maximizes the kurtosis $K_{N_{int}}^{m_f}(u,t)$ over $u$.

Block 6 is dedicated to graphical outputs and optimal filtering of clicks sequences. Time-frequency intervals are built by grouping detected contiguous time-frequency samples. A 'click enhanced' composite signal is created by stacking up the time intervals filtered in bandwidths given by Block 5. The 'click-enhanced' composite signal can be further used for dedicated localization species or individual classification processors.

3 Results

3.1 Synthetic data

To illustrate the behavior of the detection scheme (blocks 1 to 6), we apply it to synthetic data. We build 1 minute of a 200 kHz frequency sampled signal composed of:

- a first clicks sequence (duration = 1 second, central frequency = 40 kHz, clicks bandwidth = 20 kHz, Inter Click Interval (ICI) = 0.025 second, magnitude = 5) beginning at 20 seconds,
- a second clicks sequence (duration = 1 second, central frequency = 70 kHz, clicks bandwidth = 20 kHz, ICI=0.025 second, magnitude = 5) beginning at 40 seconds,
- gaussian white noise with unity standard deviation.

We apply our detection scheme with the following settings:

- block 1: filter bank ($f_{min} = 10$ kHz, $f_{max} = 90$ kHz, $\Delta f = 10$ kHz) which gives 36 channels,
- block 2 to 4: $T_{int} = 0.3s$, $L = 5s$, $p = 0.2$.

The figure 2 illustrates behavior of blocks 1 to 4. The top-left plot shows the 36 channels filter bank: the
x-axis is for \( f_{\text{min}} \) and the y-axis is for \( f_{\text{max}} \), the number near each square is the channel index. The top-right figure shows the time and the channel index where detection occurs in red and the green points are for the channel which maximize the kurtosis. The bottom-left plot shows the log-kurtosis versus time and channel index. The bottom-right figure depicts the log-detection threshold over time and channel index.

Figure 2: Outputs of blocks 1 to 4 applied on synthetic data: top-left plot shows the 36 channels filter bank, bottom-left plot shows \( K_{N,p}^m(u,t) \) versus time and channel index, the bottom-right is the log-\( \lambda(u,t_0) \) over time and channel index, and the top-right are the time and channel index where detection occurs in red and the green points are for the channels which maximize the kurtosis.

Figures 3 and 4 are the outputs provided by block 6. We can see that optimal bandwidths are in good agreement with the 2 time-frequency components of the synthetic signal: the first optimal bandwidth is [30 kHz; 50 kHz] and the second is [60 kHz; 80 kHz]. Both time locations and frequency positions are correct. The second click sequence beginning at 40 second is well detected. Optimal filtering increases the SNR of the click sequence (see top-left plot of figure 4).

3.2 Field data

The detection algorithm is applied to the monitoring of the scallop “breath sound” in its natural environment. The figure Fig.5 displays in the top panel a “breath sound” and a “creak sound” in the bottom panel recorded in a quiet and controlled tank. The breath sound has a 500 ms duration whereas the creak sound lasts only for 8 ms. Both sounds are wide band and are composed of multiple oscillations before a main pick and then a release. The challenge for the impulse detector is to discriminate between both type of sounds in a noisy environment.

A spat of scallops was installed three weeks before the experiments in the Brest (France) harbor (latitude: -4.51 W, longitude: 48.28 N, April 2009). The depth was about 10 m. An autonomous acoustic recorder was deployed over the spat. The sounds were recorded at a sampling frequency of \( f_s = 32768 \) Hz with a 16 bit resolution. The detection scheme was applied with the following settings on a 2 minute sound file:

- block 1: filter bank \((f_{\text{min}} = 100, f_{\text{max}} = 15 \) kHz, \( \Delta f = 3 \) kHz), which gives 10 channels,
- block 2 to 4: \( T_{\text{int}} = 0.003 \) s, \( L = 15 \) s, \( p_{\text{noise}} = 0.5 \).

As the study animals occur in shallow water environment, the background noise is non-stationary due to ship noise, wave noise and weak impulse sounds produced by living organisms. Gaussianity analysis of the underwater noise of a shallow water harbor shows that the noise is mainly Gaussian even though it may be non-stationary [17]. In our experiment, a one-sample Kolmogorov-Smirnov test applied on a 1 minute sound recording does not reveal that the sound was Gaussian.

The figure Fig.6 shows the results of the detection scheme applied to the breath sound of the scallop. The top plot displays in black the original signal and in red the 'clicks enhanced’ composite signal. Each graph represents a 2 seconds segment.
the ‘clicks enhanced’ signal. The bottom panel shows the spectrogram of the signal and the time-frequency area where the kurtosis exceeds the detection threshold. The 2 minute signal contains only one breath sound. The algorithm is able to detect it and to discard all the others impulse sounds (for example, creak sounds at 6 s and 8.5 s in top panel Fig.6). In order to check for the robustness of the algorithm, we apply the detector with the same setting to another 2 minute sound file. In the latter, we identify 9 breath sounds with varying SNR. The detector is able to find 5 of the 9 breath sounds and report 6 false alarms. The optimal bandwidths of 4 of the 5 correct detections have lower frequency superior or equal to 6000 Hz. The figure Fig.7 shows the mean power spectral density (PSD) of the harbor noise and the breath sound computed for 16 occurrences of each type of sound. Up to 6 kHz, the recording are dominated by the ambient noise and both curves are similar. Beyond 6 kHz, the PSD difference between ambient noise and breath sound increases from 2.5 dB up to 10 dB at 15 kHz. The optimal bandwidths are in close agreement with the frequency range of the PSD excess of the breath sound over the background noise. This result shows that the detector is able to detect transients even in noisy environment (the detection range is estimated to be equal to 1 m).

4 Conclusion

A click detection scheme is applied to detect underwater bioacoustic impulsive sounds. It worked under the assumption that clicks are embedded in Gaussian noise. Clicks were detected if the sliding window estimate of kurtosis in a segment of signal was higher than a detection threshold. The kurtosis threshold between noise and impulsive sounds was adaptively computed, hence the detector was adjusted to varying noise conditions.

The architecture of the click detector based on kurtosis was recalled. The application to synthetic data where the noise was Gaussian showed that the detection was effective and the optimal bandwidth correctly adapted to changing central frequency of the clicks.

The click detector was then applied to field data for the purpose of the non-intrusive monitoring of the great scallop in its habitat. Results showed that the click detector was able to detect the breath sound of the scallop and to discriminate between the breath sound and the creak sound. Even when the background noise was loud, non Gaussian and non-stationary, the detector was effective in tracking breath sound optimal bandwidth. The next step of development of the algorithm is to conduct advanced analysis of the scallop sound and the variability of the harbor noise in order to get more robust settings.

We conclude that PAM associated with the kurtosis based click detector represents a promising step in the development of non-intrusive sensor for the monitoring of minimally mobile aquatic organisms such as scallops in the wild.
Acknowledgments

The authors are thankful to G.I.S Europole Mer and French DGA-MRIS for partially funding this work and Erwan Amice from IUEM LEMAR Brest for his precious help during the experiments.

References


