Running title: Automatic Detection of Vessel Signatures in Audio Recordings with Spectral Amplitude Variation Signature

Number of words: ∼6105

Number of tables: 3

Number of figures: 14

Number of references: 18

Automatic Detection of Vessel Signatures in Audio Recordings with Spectral Amplitude Variation Signature.

Clausius Duque Gonçalves Reis[1,2,*],
Linilson Rodrigues Padovese[3],
Maria Cristina Ferreira de Oliveira[2].

1. Federal University of Viçosa (UFV-CRP) - Rio Paranaíba, MG.
2. University of São Paulo (ICMC-USP) - São Carlos, SP.
3. University of São Paulo (POLI-USP) - São Paulo, SP.

*Correspondence author. clausius@ufv.br

Date of submission: 3rd February 2020

Summary

1. Sound emissions by ships and boats can strongly impact marine life, with potential to affect communications, breeding and prey and predator relationships. Automatic detection of boat signatures in underwater audio recordings is thus an important task. Automated solutions are particularly relevant for monitoring preservation areas where the presence of watercrafts is usually regulated. The task is particularly challenging because it requires distinguishing multiple overlapping acoustic events in typically noisy audio recordings.

2. In this paper we introduce an algorithm for boat and ship detection which computes an acoustic signature that captures the variance in the frequency amplitudes observed over the duration of the signal.

3. We evaluated the algorithm on a database of underwater recordings collected at two conservation areas in the State of São Paulo, Brazil, with very good results, and also compared it with an existing solution.

4. Besides being effective, the algorithm requires limited user input and no parameter fine tuning to handle diverse situations. It thus provides a solution to automate the detection of vessels, with potential applications for monitoring marine preservation areas.

Keywords: Acoustic Ecology, Boat, DEMON, Detection, Signature, Sound-

scape, Spectrogram, Underwater

Introduction

Audio signals emitted by boats, ships and water crafts in general can strongly impact marine life. They may interfere with species communication (Wittekind and Schuster, 2016), echolocation mechanisms (Veirs et al., 2016) and threaten fish population growth (Jain-Schlaepfer et al., 2018).

Ecological audio recorders and hydrophones offer a cost effective method for ocean monitoring in order to identify undesirable or threatening scenarios. However, manual inspection of collections of audio recordings obtained over extensive time periods is not feasible, in view of the time and human effort it demands. This has motivated many research efforts towards devising methods to automatically detect the presence of vessels in underwater acoustic recordings.

While some solutions have employed classification approaches derived from the analysis of acoustic features (Leal et al., 2015), others have focused on the specific spectral signatures of the sounds emitted by the different types of vessels, e.g. training neural networks to detect a set of known signatures (Pollara et al., 2017; Slamnoiu et al., 2016; Chung et al., 2011; Hanson et al., 2008). Nonetheless, improving detection accuracy in noisy conditions and reducing the rate of false positives remain as challenges. A related relevant issue is automatic classification, e.g., categorizing the vessels detected, for example, according to their size (small, medium, large).

The *Detection of Envelope Modulation On Noise* (DEMON) is one of the most reliable methods for ship detection and classification (Chung et al., 2011). However, it is reported to yield poor results on noisy signals. Moreover,

it is highly dependable on a manual selection of the relevant frequency band pass filters for analysis, which must be tuned according to the type of target vessel (Hanson et al., 2008).

In this work we introduce a novel method to automatically detect the presence of boats or ships in acoustic underwater recordings. Similarly to the DEMON method, ours relies on computing acoustic signatures from the input signal. However, our method has lower computational cost and requires no calibration or human intervention. Moreover, we show that it can successfully detect vessels even in challenging conditions posed by very noisy signals, as in recordings that include crustacean acoustic activity.

In the remainder of this paper, first we we discuss related research on automatic vessel detection and characterization. Then we describe the technique developed, explaining its underlying rationale and presenting the algorithm for signature computation and vessel detection. We present results from employing the proposed algorithm to detect ships and small boats in a database of underwater recordings collected at two protected sites on the Brazilian coast. We also present results comparing the performance of our method with an implementation of a DEMON-type signature, in terms of quality and computational cost. Finally, we present conclusions and discuss future work.

Related Work

A significant increase in the number of large vessels navigating the oceans has motivated investigations on the effects of noise levels on marine life (Wit-

tekind and Schuster, 2016; Merchant et al., 2014).

Large vessels emit acoustic signals in the low frequency spectrum, the same frequency bands used for communication by certain groups of animals, such as baleen whales (Wittekind and Schuster, 2016). Noise due to ships has also been found to disrupt the echolocation mechanism of endangered species such as killer whales (Veirs et al., 2016).

A recent study suggested motorboats cause a stress response in fish embryos (Jain-Schlaepfer et al., 2018), with an increase in their heart rate and even stronger effects, depending on the engine type, e.g., two or four strokes.

The potential impact of underwater noise on marine mammals is widely recognized. Merchant et al. (2014) characterized the natural and anthropogenic contributors to underwater noise, specifically in a conservation area with increased shipping activity associated with offshore energy developments. They monitored ship activity using *Automatic Identification System* (AIS), in which transponders send information about each ship present in the area, in addition to video surveillance and automated sound recorders to monitor noise levels. They correlated the data collected from those sources to investigate the relationship between broadband sound exposure and indicators proposed by the EU *Marine Strategy Framework Directive* to access the effects of noise exposure over a locally protected bottlenose dolphin population. The authors determined that the bottlenose dolphin population is not under current danger due to the current noise levels they are exposed, however their study will serve as a baseline for future investigations on the effect of noise levels on the local marine life.

The combination of low-cost automated acoustic recorders and algorithms

capable of detecting boats in long duration recordings can provide an effective solution to monitoring conservation areas for the presence of vessels. Leal et al. (2015), for instance, introduced a method for boat classification based on Fourier transform and signal processing to identify the occurrence of vessels in the recordings.

Features extracted from the audio spectrum and fed to *Artificial Neural Networks* (ANNs) and *Support Vector Machines* (SVMs) have been successfully employed to classify ships according to their type (Leal et al., 2015). A 10-dimensional feature array was created, given by the highest peak value in the original audio signal plus eight values corresponding to the energies of the first eight 500-Hz segments, and the central frequency of the smoothed signal. Leal et al. (2015) argue that acoustic emissions by vessels are too complex to be analyzed directly from the raw form of the signal's discrete Fourier transform.

Averbuch et al. (2011) rely on wavelet packet coefficients to detect certain types of vessel in recordings with background noise. They derive acoustic signatures using a combination of a Linear Discriminant Analysis (LDA) classifier with Classification and Regression Trees (CART), plus an additional component to reduce false alarms. Automatic real-time detection requires a training step with a set of pre-registered signatures of interest, where the signatures are generated from "energy maps" derived from the blocks of "wavelet packet coefficients".

The method by Yan et al. (2017) uses a combination of resonance-based sparse signal decomposition (RSSD) and Hilbert marginal spectrum (HMS) analysis to recognize certain types of vessels, defined *a priori*. The authors

use an RSSD decomposition of the original audio signal to separate it into high and low resonance signals, corresponding respectively to the target and background noise. Hilbert-Huang transform (HHT) is applied to extract characteristics from the purified audio and SVM is employed for signal classification. Gaussian white noise was added to the recordings in different proportions to perform the decomposition tests and obtain the boat signatures.

Both previous solutions demand transformations of the original signal, e.g., employing RSSD or waveletets, and are targeted at the problem of categorizing the vessel signatures against a predefined set of classes. The solution proposed in this paper to compute vessel signatures does not require complex transformations of the original signal, as it relies directly on its FFT frequency spectrum. Moreover, it is aimed at detecting the presence of acoustic events introduced by arbitrary types of vessels in the audio recordings, without performing classification.

A well-known method for vessel detection in audio recordings is the *Detection of Envelope Modulation On Noise* (DEMON) algorithm. By enhancing audio frequencies characteristic of vessel emissions it creates an acoustic signature that can be employed for classification (Pollara et al., 2017, 2016; Chung et al., 2011). Introduced over 50 years ago (Tuteur, 1963), the DEMON signature has inspired several novel ship classification algorithms (Chung et al., 2011).

A recent study compared four DEMON-type algorithms (Slamnoiu et al., 2016) regarding their ability to detect small ships and divers, as well as their robustness to acoustic noise. The same authors (Slamnoiu et al., 2016) in-

troduced a variation of the classic algorithms that was shown to yield similar results with less computational effort, also arguing that their modified DEMON algorithm is more robust to noise.

Chung et al. (2011) described an approach for identification and classification of multiple ships in busy harbor conditions using the DEMON algorithm. Employing a system consisting of four hydrophones, the authors used cross-correlation to compute the delay between the corresponding recordings and could even estimate the relative positions of the boats. Chung et al. (2011) state that ship detection and classification are governed by the propeller and by engine parameters such as number of blades and rotational speed. As such, the audio spectrum is one of the most reliable acoustic parameters to accomplish such tasks. A major limitation of approaches that rely on the standard DEMON algorithm is that human intervention is required to obtain good results. For instance, the user must select a suitable band pass filter to recognize a specific signature.

In order to improve the resolution of the resulting DEMON algorithm and increase the capability of detecting subtle details within the recordings, Hanson et al. (2008) used cyclostationary signal processing, exploiting the spectral redundancy inherent to the propeller signal. This enhancement enabled blind identification of the shaft speed and number of propeller blades, even in noisy signal conditions. This modification improved signal resolution and allowed identifying the presence of multiple vessels. Emission by snapping shrimps is, according to the authors, a major source of noise that can render methods based on the DEMON signature innefective.

Several relevant contributions previously discussed relied on the DEMON

algorithm and variations therein. Nonetheless, besides its sensitivity to certain types of noise, computing a DEMON signature from raw audio signals requires non-trivial levels of human intervention. The solution introduced in this paper adopts a different strategy to obtain an acoustic signature, which is simpler than the DEMON signature, practically eliminating the need of user intervention during the extraction and detection processes. Our tests, performed on audio recordings collected at marine ecological conservation areas with intense crustacean acoustic activity, indicate that the technique can successfully detect the presence of different types of boats and also identify their corresponding acoustic signatures.

Materials and Methods - Spectral Amplitude Variation Signature for Vessel Detection

A first step in automatic detection of boats, ships and other types of water crafts in underwater acoustic recordings is to distinguish their acoustic signatures from those due to natural elements such as fish choruses or crustacean acoustic activity. Audio signals emitted by vessels can be characterized by narrow stationary frequency signatures, as the engine, propellers and cavitation produce sounds that usually remain stable along the duration of a single recording.

The audio spectrograms shown in Figure 1 illustrate occurrences of three acoustic events captured underwater, namely boats, indicated by the red squares, fish choruses, indicated by the blue squares and background noise

from crustaceans, visually perceptible as vertical lines in both spectrograms. The spectral signatures of both the fish choruses and the boats are characterized by frequency peaks that persist along extensive time periods. Even though they may appear similar, the frequency peaks due to boats appear over narrow frequency bands, unlike the fish choruses, which are observed over wider bands. The distinction poses a challenging scenario for automatic boat detection, as boats and fish choruses can be easily mistaken, and boat events may be totally masked by strong background noise.

Our proposed method for automatic vessel detection in underwater ecological recordings considers certain unique features of the sounds emitted by vessels to discern their characteristic frequency peaks from those produced by other acoustic events. The algorithm considers as input the dB frequency spectrum of the audio signal segmented into one-minute audio samples. It initially computes a so-called **frequency amplitude variation** (FAV) signature from each one-minute segment of the audio spectrum.

The overall process, as well as the motivation for our choices in deriving the proposed algorithm are detailed next. Unlike the DEMON method, which considers the audio envelope of the signal, our algorithm relies on information obtained from the audio spectrogram matrix. The audio spectrogram is computed with the *Short-Time Fourier Transform* (STFT), using as input the audio signal and an *FFT size*, which was established as 11025, equal to the audio sampling rate, in order to yield a nominal frequency resolution of 1Hz while resulting in one FFT sample per second. The resulting spectrogram matrix thus has $F = (11025/2)$ frequencies and $S = (N \times 60)$ samples, or one sample for each second of the input audio, where N is the audio duration, in

minutes. Our analyses have been conducted on 15-minute recordings.

Algorithm 1 describes the necessary steps to obtain the FAV signature and the corresponding boat detection signature.

Algorithm 1: FAV signature computation

Input:
- **Sample[0..60][0..Rate/2]** ▷ 1-Minute Spectrogram matrix
- **Rate** ▷ Audio sampling rate
- **MaxFreq** ▷ Maximum frequency of analysis
- **SDMultiplier** ▷ SD threshold multiplier

Result:
- **FAV** ▷ Frequency Amplitude Variation signature
- **BoatDetection** ▷ Boat detection signature

$Mean[0..Rate/2] \leftarrow frequencyMean(Sample)$
$SmoothMean[0..Rate/2] \leftarrow smooth(Mean, blackman, 8)$
$SD \leftarrow StandardDeviation(SmoothMean)$
$Mean \leftarrow Mean[0..MaxFreq]$
$SmoothMean \leftarrow SmoothMean[0..MaxFreq]$
$FAV[0..MaxFreq] \leftarrow 0$
$BoatDetection[0..MaxFreq] \leftarrow 0$
for $i \leftarrow 0$ **to** $MaxFreq$ **do**
  $variation \leftarrow SmoothMean[i] - SmoothMean[i-1]$
  if $variation > 0$ **then**
    $variation \leftarrow 0$ ▷ Post frequency peak
  end
  $FAV[i] \leftarrow |(variation)^3|$
  if $FAV[i] > (SD \times SDMultiplier)$ **then**
    $BoatDetection[i] \leftarrow 1$
  end
end
$BoatDetection[0..20] \leftarrow 0$
$BoatDetection[(MaxFreq - 10)..MaxFreq] \leftarrow 0$
return $FAV, BoatDetection$

As described, the spectrogram matrix covers N one-minute samples of the audio recording. The frequency amplitudes of each one-minute sample

are averaged in order to obtain a single representative frequency amplitude. The result is a spectrum signature of the one-minute audio segment stored in an array of size F, where each entry corresponds to an average frequency amplitude (Algorithm 1, lines 6-7). Computing the average spectrum signature of one-minute audio samples, instead of the entire audio, avoids missing short duration events. It also allows precisely locating the events along the audio duration, which is useful for analysis and visualization purposes.

The spectrum signature is the input to the FAV signature computation, which is at the core of the boat detection procedure. The size of the spectrum signature is equal to the maximum frequency component in the recording, as it has been obtained considering F, however it can be set to a maximum frequency of analysis (Algorithm 1, lines 9-10). The spectrum signature is thus stored in an array in which each entry corresponds to a frequency window of 1Hz. Let $SmoothMean[0...F-1]$ denote the array storing the smoothed spectrum signature, with F values. It is possible to compute the differences between each pair of frequency amplitudes in the signature, applying Equation 1 to each element in this array (Algorithm 1, lines 14-16).

$$FAV_f = \left|(SmoothMean_f - SmoothMean_{f+1})^3\right| \tag{1}$$

Equation 1 yields the unsigned differences between consecutive frequency pairs of the input signal (Algorithm 1, line 18). The resulting differences, raised to power 3, are stored as elements of a new array FAV, of size $F-1$, which stores the FAV signature. The reason for raising the resulting differences to power 3 is to emphasize the frequency peaks due to boats against the

peaks due to other events that also introduce frequency amplitude variations. Raising to other power values could yield a similar effect, but empirical investigations led us to this choice.

The FAV signature captures and emphasizes persistent variations in the frequency amplitudes contained in the input spectrum. Such variations can yield a single peak or multiple frequency peaks, resulting in different patterns characteristic of certain acoustic events, including boats. Our goal is thus to identify the peak patterns characteristic of acoustic emissions by boats.

A threshold of $\alpha = 1.5 \times Standard\ Deviation$ of the original spectrum has been empirically found suitable to identify the frequency peak patterns due to boats, even in noisy conditions. The algorithm applies this threshold to the FAV signature (Algorithm 1, lines 19-20) to yield a binary boat detection signature array, also of size $F - 1$, applying Equation 2:

$$BoatDetection_f = h(FAV_f, \alpha) \tag{2}$$

where

$$h(FAV_f, \alpha) = \begin{cases} 1, & \text{if } FAV_f >= \alpha \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

$$\alpha = \left(\sqrt{\frac{1}{n} \sum_{i=1}^{n} (SmoothMean_i - \mu)^2} \right) \times 1.5 \tag{4}$$

$$\mu = \left| \frac{1}{n} \sum_{i=1}^{n} SmoothMean_i \right| \tag{5}$$

The result is a binary array $BoatDetection$, in which each entry is flagged to 1 whenever the corresponding entry in the FAV signature array is above the threshold ($SmoothMean_n > \alpha$), signalling this as a peak indicative of the presence of a boat, otherwise the corresponding array entry remains set as 0.

The acoustic events introduce sound amplitude variations in the audio signal spectrum, which can be observed in the form of characteristic patterns of peak amplitudes. The underlying rationale in computing the FAV signature is to identify the peak patterns due to boats and ignore those related to other acoustic events captured in the recordings. Our method identifies a characteristic signature of each one-minute audio sample. It is thus possible to distinguish the patterns due to boats from those originated from overlapping acoustic events. Once the frequency peaks have been identified in the FAV signature, a thresholding operation can be applied to obtain a binary array with the boat detection signature.

Since our algorithm measures the differences in amplitude between neighboring frequencies, it performs better with a standard spectrum in dB scale, in which the amplitude variations are smoother. Measuring this frequency variation along the audio spectrum will result in a characteristic signature with amplitude peaks due to noticeable acoustic events. Operating on the dB scale audio spectrum (Figure 2-A) facilitates distinguishing the peaks due to boats from those introduced by other acoustic events.

Figure 3 illustrates two signatures computed with this approach for two one-minute samples from the recordings depicted in the spectrograms in Figure 1. The charts show the spectrum signature (blue line), the FAV signature

(green line) and the boat detection signature (red vertical watermarks).

The signatures in Figure 3-A refer to the first minute of the recording depicted in Figure 1-A, which includes a fish chorus overlapped with a ship or large boat, whereas those in Figure 3-B refer to minute 11 in the recording depicted in Figure 1-B, which includes a small boat and heavy noise due to crustacean acoustic emission. One observes in the FAV signature in Figure 3-A that only frequency peaks due to boats stand out, whereas other acoustic events are largely ignored. The FAV and boat detection signatures depicted in Figure 3-B show that the algorithm detected the faint signal of the small boat even in the presence of heavy background noise.

Our method to process an input audio database consists of the following steps. Given a root directory path containing the audio files to be processed, first the signatures are computed for each individual recording applying Algorithm 1, i.e., the spectrum, the frequency amplitude variation (FAV) and the boat detection signatures. The resulting signatures for each file are stored in a corresponding CSV file that preserves the source file name. The algorithm also saves summary information on each boat detection signature, namely the number of peaks identified, the shortest distance between the peaks, and the lowest and highest amplitude peak values – this information may be useful in further vessel classification procedures. The CSV signature files are saved in a child folder of the root directory; the summary information is stored in a single CSV file saved in the root directory.

The process requires two parameters: the sampling rate of the audio recordings, which is known *a priori*, and a boat detection threshold. The default threshold, which yielded good results in the databases studied, has

been set empirically to 1.5 times the standard deviation of the amplitude spectrum of the signal. Thus, any peak that surpasses this threshold is interpreted as a boat detection.

Results

The method has been validated on audio recordings collected at two sites located on the West coast of the State of São Paulo, namely the Laje de Santos Marine State Park (LSMSP, located at $24°15'48''S$, $46°12'00''W$) and the Xixová-Japuí State Park (XJSP, located at $24°0'22''S$, $46°23'29''W$). Both sites exhibit boat traffic and are characterized by great marine biodiversity. The LSMSP is 30km (19 miles) off the coast and serves as a protected area for reposition of fish shoals and safe reproduction of species. The XJSP conservation unit encompasses a marine area SW of the Santos Bay and an adjacent inland region of tropical forest. This is an area with strong human presence, close to a very busy port in Brazil (Sánchez-Gendriz and Padovese, 2017; Sánchez-gendriz and Padovese, 2015).

The datasets consist of 15-minute monaural audio files recorded at a sampling rate of 24 kHz. Recordings were obtained underwater with a custom autonomous hydrophone recorder developed at LACMAN (Laboratory of Acoustics and Environment of the Polytechnic School of the University of São Paulo)[1].

In total, we analyzed 2,675 recordings from LSMSP and 2,390 recordings from XJSP. As general observations, we noticed that the sound levels of the

[1] http://lacmam.poli.usp.br

recordings taken at XJSP are, in average, $10dB$ higher than those taken at LSMSP. This is possibly because XJSP is closer to the coast and has intense boat traffic, whereas LSMSP is located in a quieter area farther away from the coast. We also observed in some XJSP audio files a recording failure in the low frequencies up to 100Hz. Both databases include many recordings with multiple occurrences of diverse boats and ships, under different conditions. Our goal was to asssess the effectiveness of the proposed algorithm in automatically detecting the presence of ships and boats on recordings collected from February 1 to February 28, 2015 (LSMSP site) and from February 4 to February 28, 2015 (XJSP site).

The spectrograms in Figures 4 and 5 depict, respectively, recordings taken at two arbitrary days, namely February 17 at LSMP and February 10 at XJSP. These particular days have been selected as illustrative examples of the multiple acoustic events that can be found in both databases. The figures also illustrate the outcome of the proposed boat detection algorithm. Detections are displayed in the form of a line graph at the bottom and temporally aligned with the corresponding audio spectrogram.

The red line indicates the boat detections at their exact location in time along the duration of the audios, and also informs the number of frequency amplitude peaks associated with the boat event, as detected by the algorithm. In the examples, the number of peaks varied from 0 (no boat detected) up to 5 peaks in the recordings taken at the LSMSP site (Figure 4); or from 0 up to 13 peaks in the recordings taken at the XJSP site (Figure 5).

The boat detections identified have been confirmed by listening to the corresponding audio files. In all cases, a manual inspection confirmed the

automatic detection has been successful. In a few and very specific cases false negatives occurred, with the algorithm failing to detect a boat that can be identified in the audio; specific cases are discussed later in this section.

The boat detection line for the recordings from LSMSP (Figure 4) shows that the signatures of the boats detected in the early morning are characterized by having one to three peaks and are not contiguous, whereas the signatures of boats detected from 9:00am onwards are characterized by more peaks which are also very close – these usually correspond to a series of short duration acoustic events.

Observing the recordings from XJSP (Figure 5), one notices that longer duration acoustic events associated with boats are frequent during certain morning and evening periods. The boat detection line chart reveals multiple short duration boat events occurring frequently from nearly 9:00am until dusk. Short-term boat events have also been detected in the early morning, before 9:00am, and during the evening, after 6:00pm. After 6:30pm, when an overlapping fish chorus event is visible in the spectrogram, multiple boat events have been detected with signatures characterized by higher numbers of peaks. From 10:00pm it is possible to observe a long duration boat event that persists until midnight.

Acoustic events typical of the site, such as fish choruses, can mask some of the peaks characteristic of a boat signature, as frequency peaks of both events overlap. They can still be distinguished by further inspecting the differences in the amplitudes of the neighboring frequencies. Nonetheless, the overlap may cause boat peaks at these frequencies to go undetected. In these cases, succesful detection depends on a proper choice of the detection threshold

defined relative to the standard deviation of the spectrum signature, which is the single setting of the automatic detection process that may require user adjustment.

Figure 6 shows an example in which a long duration boat event has been detected. The boat detection line chart appears as a line with a constant number of peaks over a range of frequencies, which persists during nearly the entire duration of the recording. The detection line chart shows two to three peaks along minutes 1 to 9. The peaks overlap with the frequencies of a simultaneous fish chorus event.

The audio of the spectrogram in Figure 7 includes an acoustic event occurring from minutes 7 to 12, approximately. Listening to the audio one observes the sound of (apparently) a motor boat that gradually approaches the hydrophone and then its engine is turned off. The boat signature detected has a varying number of peaks: an initial detection indicates 3 peaks, followed by a short period with no detection; the boat is again detected with a signature of 8 peaks, going up to 12 peaks in minute 11, then in minute 12 only a single peak is detected, when the engine is powered off. The varying number of peaks is due to the non uniform properties of the sound emission by the boat. Its signal amplitude increases as it approaches the hydrophone and fades away when the engine is suddenly turned off. The algorithm successfully detected the boat, despite the unfavourable conditions and overlapping acoustic noise from a fish chorus.

Another difficult case for the algorithm is illustrated in Figure 8, in which the audio depicts a short duration and non-uniform sound emission by a boat, which is masked by extensive background noise due to crustacean acoustic

activity. Nonetheless, the algorithm managed to achieve a partial recognition, as indicated by the detection line graph, in minutes 5 and 6. Only one of the peaks of the boat signature has been identified, due to the short interval between the acoustic events and the overlapping noise.

A particular case in which the algorithm failed to detect the boat signature is shown in Figure 9, which depicts an audio that includes a boat with non-constant acceleration. In this case, the frequencies of the spectral signature do not exceed the default threshold that indicates a positive boat detection. The acoustic event occurred in minute 12 of the recording, going undetected due to the insufficient duration of the peaks in a single frequency.

Many boat occurrences have been detected in the recordings taken at the LSMP site, even though boat visitation is controlled or even prohibited at certain periods. Interestingly, from February 10 to 28 long-term acoustic emissions by boats are detected often in the mornings, mostly boat signatures characterized by up to two peaks. These emissions are low amplitude signals and the boats are apparently located distant from the hydrophone. Yet the algorithm successfully detected such occurrences, even though the recorded signal often includes multiple acoustic events in addition to the boats.

Other cases confirm the effectiveness of the proposed algorithm. Consider, for example, the audio illustrated in the spectrogram in Figure 10. The algorithm detected a boat signature characterized by a single frequency peak in this recording. This was a particularly challenging case, because the boat signal was faint and heavily masked by a loud noise due to crustacean acoustic activity.

A second example is illustrated in Figure 11, which depicts an audio in

which the algorithm detected a boat signature characterized by 35 to 59 peaks along the duration of the recording. Although we cannot estimate the distance of the boat to the hydrophone, its signal is sufficiently loud for its signature to be properly detected even above the 1KHz frequency band.

Yet another challenging example are cases of multiple boats captured in the same recording, possibly as overlapping acoustic events. The audio spectrogram in Figure 12 shows a boat event that persists along the duration of the recording, however a second boat event occurs in minutes 4 to 6. This is difficult to handle due to the overlap of both boat signatures. The signatures computed for the audio samples depicting minutes 2 to 5 are shown in Figure 12 to illustrate how multiple events interfere in the detection process. In the signatures of minutes 2 and 3 one observes the complete and the partial signatures of the persistent boat; in minute 4 the transient boat event partly masks the previous signature, introducing additional peaks. The emission by the transient boat in minute 5 is sufficient to mask the signature of the persistent boat, as it adds noise in the frequency band from 200Hz to 2kHz. The consequence is that only the signature of the transient boat is detected.

Comparison: FAV vs DEMON

A DEMON implementation based on square-law demodulation by Pollara et al. (2016) available online[2] has been used as baseline for a comparison with the proposed FAV signature, regarding detection accuracy and processing

[2] `https://github.com/lxpollara/pyDEMON`

times. The code and audio samples used in the analysis and comparisons can be downloaded from github (Reis, 2019).

We conducted the comparison on 96 audio files selected from our databases (Xixová and Japuí State Parks), corresponding to 24 hours of audio analysed at a one-minute sample resolution, totaling 1,440 one-minute samples representative of typical scenarios found in underwater ecological recordings. We conducted a manual annotation process and identified that from the 1,440 samples considered, 706 do not include boat events, whereas the remaining 734 include diverse observable boat occurrences, varying from very clear to very faint signals. The samples depict distinct situations, including non-uniform vessel emissions near the hydrophone and other difficult scenarios characterized by overlapping with crustacean acoustic emissions or fish choruses.

The DEMON signature computation requires the definition of a target frequency window for each detection scenario. Inadequate choice of this initial input parameter can render the detection task unfeasible. In order to ensure a proper comparison, the input samples were processed similarly for both methods, setting the higher frequency to $900Hz$ and the low frequency to $400Hz$ as the default parameters for the DEMON signature calculations. Each one-minute sample was smoothed with a size 11 Blackman window convolution, and the corresponding standard deviations were extracted.

In both methods, a detection occurs whenever the corresponding signature exceeds the corresponding detection threshold, where each excess will correspond to a frequency peak in the boat detection signature. We kept the default detection threshold for the FAV signature, i.e., $1.5\times standard\ deviation$.

Setting the detection threshold for DEMON requires a different approach, however, as the base signals vary in amplitude due to interference of noise. Thus, we set an initial threshold as in FAV ($1.5 \times$ *standard deviation*), then obtained a basis amplitude for each frequency by smoothing the DEMON signature with a size 64 Blackman window convolution. The basis amplitudes thus obtained were then added to the initial threshold.

Under normal conditions these threshold settings enabled both methods to correctly detect the presence of boats and create the corresponding boat signatures. An example for a particular one minute sample is illustrated in Figure 13, where a boat with two frequency peaks can be observed, which has been correctly detected in both FAV and DEMON signatures. However, results from the methods differ in more challenging conditions. Detection results for a noisy sample are illustrated in Figure 14, where FAV correctly identified a boat with eight frequency peaks. The DEMON signature, however, detected only one of the peaks and also incurred in a false positive detection of a frequency peak which is not due to a boat.

Considering all 1,440 samples, we computed the detection accuracy and the correctness of the number of peaks identified in the signatures obtained from both methods. Detection accuracy is reported as the numbers and percentages of correct and incorrect boat detections. The signature accuracy is measured as the number of detected frequency peaks, relative to the true number of peaks. We report the average percentages for each method, over all signatures extracted.

Table 1 shows two confusion matrices displaying the detection accuracy results for the FAV and the DEMON implementations, respectively. The

columns labeled **Reality** inform the number of boat occurrences according to the manual annotation, the rows labeled **Algorithm** show the results obtained with FAV or DEMON. Table 2 summarizes the accuracy rates and also the false positive and false negative rates.

FAV correctly detected the presence of boats in 684 out of the 734 samples with boat occurrences, an accuracy rate of 93%. It missed 50 occurrences, and incured in no false positives. DEMON correctly detected 406 boat occurrences, an accuracy rate of 55%. It missed 328 boat events, and incorrectly detected 84 acoustic events as boats. The **false-negative** rates in our analysis of DEMON could be disputed, since the manual annotation identified some very faint boat signals. Possibly different calibrations of DEMON would be required for it to perform better on the database as a whole, as the conditions describing boat events vary considerably. On the other, FAV has shown very robust performance with the settings employed, regardless of the varied conditions.

For the 734 samples including boat occurrences we also computed the correctness of the number of frequency peaks identified, as a measure of the quality of the signatures computed. For instance, given a boat signature with 5 frequency peaks, if a method computes the signature with 3 peaks, its resulting signature accuracy is 60%. As informed in Table 2 (second row), the average signature accuracy of the FAV method over all samples was 80%, whereas for the DEMON method it was 29%.

Table 3 shows the times, in seconds, to extract the DEMON and FAV signatures for a 15 minute audio recording and for a one minute sample of the same recording, considering five executions of the algorithm implementations.

In the 15 minute sampling we observe a speed up of nearly 230, as the average execution times drop from 6.743 seconds to compute the DEMON signature to 0.029 seconds to compute the FAV signature. In the one minute sample the DEMON signature has been computed in 0.416 seconds in average, and the FAV signature in 0.006 seconds, a speed up of nearly 70.

These empirical investigations provide evidence on the improved performance of the FAV algorithm over the DEMON implementation, regarding both computational cost and accuracy, the later measured in terms of detection correctness and signature quality. Moreover, unlike DEMON, FAV requires no specific settings or prior knowledge of the boat frequency signatures. It is a self-contained solution that relies solely on the input signal of the audio spectrum. A single user setting is required to establish the detection threshold, where the default setting as $1.5 \times$ *Standard Deviation* of the original spectrum proved adequate in all cases appearing in the databases considered.

Summary

In this paper we introduced FAV, a method to detect the presence of vessels in underwater ecological acoustic recordings. Results from experiments conducted on recordings collected in February 2015 at two sites in Brazil provide empirical evidence on its effectiveness. The examples discussed show the method can successfully detect the presence of boats even in difficult conditions, such as faint signals and multiple overlapping acoustic events. Manual inspection of all the recordings considered in the experiments re-

vealed no false positives and a small rate of missed detections in very specific situations.

The algorithm can recognize boats under constant acceleration, where the resulting sound signal caused by cavitation generates characteristic peaks at short frequency intervals. The failure cases were of boats with erratic acceleration behavior. This situation shall be addressed in further work, possibly with an analysis of the amplitude variation in time, combined with frequency information and machine learning methods to recognize specific patterns.

Besides detecting emissions by boats and ships, the FAV signature is also susceptible to sound emission by airgun arrays. The reason is that such devices produce acoustic energy with frequency amplitudes mostly under 100Hz (Hermannsen et al., 2015; Landro et al., 2011), the same band of the boat emissions. For boat detection purposes, if an acoustic signal emitted by an airgun array is identified one can assume the consequent detection of a boat signal, since airgun arrays are towed by boats.

In comparison with the classic DEMON method, which also relies on analysis of the acoustic signal for boat detection, our solution provides an alternative that is easy to implement and does not require extensive processing power or complex configuration of input parameters. DEMON, on the other hand, requires human intervention to define the best filters for each type of target signal and is computationally more expensive.

In future work we intend to tackle the problem of boat classification based on their characteristic FAV signatures. We believe classification would be feasible with a combination of machine learning and data visualization

methods integrated into a visual analytics solution that supports user steering. The simultaneous detection of multiple ship signatures is another challenging task that deserves further attention.

Authors' contributions

Clausius Reis conceived the ideas, designed the methodology and implemented the proposed algorithms; Linilson Padovese collected the underwater data; Clausius Reis and Maria Cristina analysed the data and results; Clausius Reis and Maria Cristina led the writing of the manuscript. All authors discussed the results, contributed critically to the drafts and gave final approval for publication.

Acknowledgements

The authors acknowledge the financial support of the São Paulo State Research Foundation (FAPESP grants 2017/05838-3 and 2016/02175-0) and the National Council for Scientific and Technological Development (CNPq grant 301847/2017-7). The views expressed do not reflect the official policy or position of either FAPESP or CNPq.

Data Accessibility

The data and codes used on this paper are available under GPL-3.0 license at `https://doi.org/10.5281/zenodo.3252202` (Reis, 2019).

References

Averbuch A, Zheludev V, Neittaanmäki P, Wartiainen P, Huoman K, Janson K. 2011. Acoustic detection and classification of river boats. *Applied Acoustics* **72**: 22 – 34. ISSN 0003-682X.

Chung KW, Sutin A, Sedunov A, Bruno M. 2011. DEMON acoustic ship signature measurements in an urban harbor. *Advances in Acoustics and Vibration* **2011**. URL `http://doi.org/10.1155/2011/952798`

Hanson D, Antoni J, Brown G, Emslie R. 2008. Cyclostationarity for Passive Underwater Detection of Propeller Craft: A Development of DEMON Processing. *Proceedings of Acoustics* **November**: 1–6.

Hermannsen L, Tougaard J, Beedholm K, Nabe-nielsen J. 2015. Characteristics and Propagation of Airgun Pulses in Shallow Water with Implications for Effects on Small Marine Mammals. *PLOS ONE* **July**: 1–17.
URL `https://doi.org/10.1371/journal.pone.0133436`

Jain-Schlaepfer S, Fakan E, Rummer JL, Simpson SD, Mccormick MI, Cooke S. 2018. Impact of motorboats on fish embryos depends on engine type. *Conserv Physiol* **6**: 1410–1093.
URL `https://doi.org/10.1093/conphys/coy014`

Landro M, Amundsen L, Barker D. 2011. High-frequency signals from air-gun arrays. *Geophysics* **76 NO. 4**: Q19–Q27.
URL `http://dx.doi.org/10.1190/1.3590215`

Leal N, Leal E, Sanchez G. 2015. Marine vessel recognition by acoustic signature. *ARPN Journal of Engineering and Applied Sciences* **10**: 9633–9639.

Merchant ND, Pirotta E, Barton TR, Thompson PM. 2014. Monitoring ship noise to assess the impact of coastal developments on marine mammals. *Marine Pollution Bulletin* **78**: 85–95.
URL `https://doi.org/10.1016/j.marpolbul.2013.10.058`

Pollara AS, Lignan G, Boulange L, Sutin A, Salloum H. 2017. Specifics of demon acoustic signatures for large and small boats. *The Journal of the Acoustical Society of America* **141**: 3991–3991.
URL `https://doi.org/10.1121/1.4989137`

Pollara AS, Sutin A, Salloum H. 2016. Improvement of the Detection of Envelope Modulation on Noise (DEMON) and its application to small boats. *OCEANS 2016 MTS/IEEE Monterey, OCE 2016* : 1–10.
URL `http://doi.org/10.1109/OCEANS.2016.7761197`

Reis CDG. 2019. FAV-Signature. https://github.com/clausiusreis/FAV-Signature. [Online; accessed 24-March-2019].
URL https://doi.org/10.5281/zenodo.3252202

Sánchez-gendriz I, Padovese LR. 2015. Underwater soundscape of marine protected areas in the coastal of Sao Paulo, Brazil. *Marine Pollution Bulletin* **105**: 65–72.
URL http://doi.org/10.1016/j.marpolbul.2016.02.055

Sánchez-Gendriz I, Padovese LR. 2017. Temporal and spectral patterns of fish choruses in two protected areas in southern Atlantic. *Ecological Informatics* **38**: 31–38.
URL https://doi.org/10.1016/j.ecoinf.2017.01.003

Slamnoiu G, Radu O, Rosca V, Pascu C, Damian R, Surdu G, Curca E, Radulescu A. 2016. DEMON-type algorithms for determination of hydro-acoustic signatures of surface ships and of divers. *IOP Conference Series: Materials Science and Engineering* **145**.
URL https://doi.org/10.1088/1757-899X/145/8/082013

Tuteur FB. 1963. *Detection of wide-band signals modulated by a low-frequency sinusoid*. Appendix A-4, New Haven, Conn, USA: Yale University.

Veirs S, Veirs V, Wood JD. 2016. Ship noise extends to frequencies used for echolocation by endangered killer whales. *PeerJ* **4**: e1657.
URL https://doi.org/10.7717/peerj.1657

Wittekind D, Schuster M. 2016. Propeller cavitation noise and background noise in the sea. *Ocean Engineering* **120**: 116–121.
URL http://dx.doi.org/10.1016/j.oceaneng.2015.12.060

Yan J, Sun H, Cheng E, Kuai X, Zhang X. 2017. Ship radiated noise recognition using resonance-based sparse signal decomposition. *Shock and Vibration* **2017**.

Tables

Table 1 Detection accuracy of FAV and DEMON algorithms on a database of 1,440 one minute audio samples.

Confusion matrix FAV			
		Reality	
		NO Boat	**Boat**
Algorithm	**NO Boat**	706	50
	Boat	0	684
	# Samples	706	734

Confusion matrix DEMON			
		Reality	
		No Boat	**Boat**
Algorithm	**NO Boat**	622	328
	Boat	84	406
	# Samples	706	734

Table 2 Boat detection accuracy and signature accuracy results obtained with the FAV and DEMON implementations, as well as false positive and false negative detection rates.

	FAV	**DEMON**
Detection accuracy	93.19%	55.31%
Signature accuracy	80.38%	29.21%
% of false positives	0.00%	10.63%
% of false negatives	20.42%	45.14%

Table 3 Times (in seconds) to compute the boat signatures of a 15 minute audio recording and of a one minute sample extracted from the same audio. Time measures were obtained from five executions of each (DEMON/FAV) implementation on a $Intel© \ Core^{TM} \ i7-7700 \ @ \ 3.60GHz \times 4$ with 16GB memory.

	15 minute audio		**One minute audio**	
Time (Seconds)	**DEMON**	**FAV**	**DEMON**	**FAV**
	6.672	0.031	0.413	0.006
	6.760	0.029	0.421	0.006
	6.743	0.030	0.416	0.006
	6.790	0.027	0.431	0.006
	6.606	0.029	0.409	0.006
Avg	6.743	0.029	0.416	0.006

Figures

Fig. 1 Logarithmic scale spectrograms of two underwater recordings. The red squares mark frequency peaks due to boats and the blue squares mark peaks due to fish choruses. Both recordings include background noise due to acoustic emission by crustaceans, that can be perceived as vertical lines in the spectrograms. (A) 15-minute audio spectrogram with two types of fish chorus and a boat signature characterized by three peaks; (B) 15-minute spectrogram with a single boat signature characterized by 10 peaks and heavy background noise due to crustacean emissions.

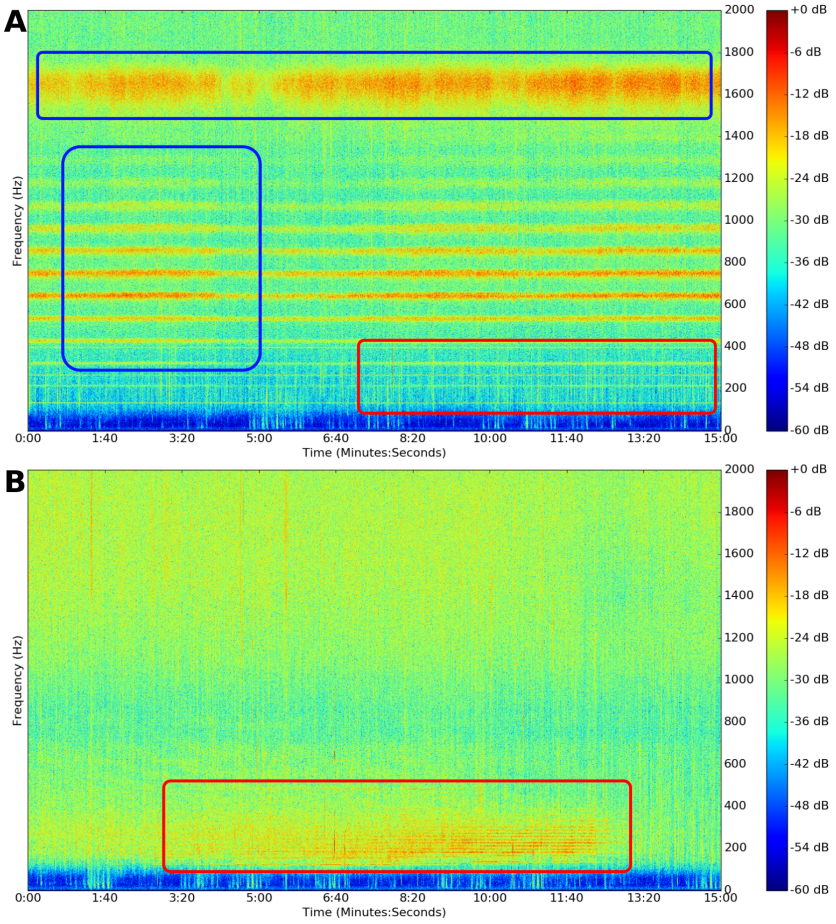

Fig. 2 In both charts the blue lines depict, respectively, the frequency spectra (up to 2.7kHz) of an input signal in dB (A) and in dB^2 (B); the green and red dotted lines depict the spectral signal smoothed by convolution with a size 11 blackman window, scaled respectively by 1.5× and by 3× its standard deviation. The vertical watermarks in red identify the frequency peaks due to boat emissions.

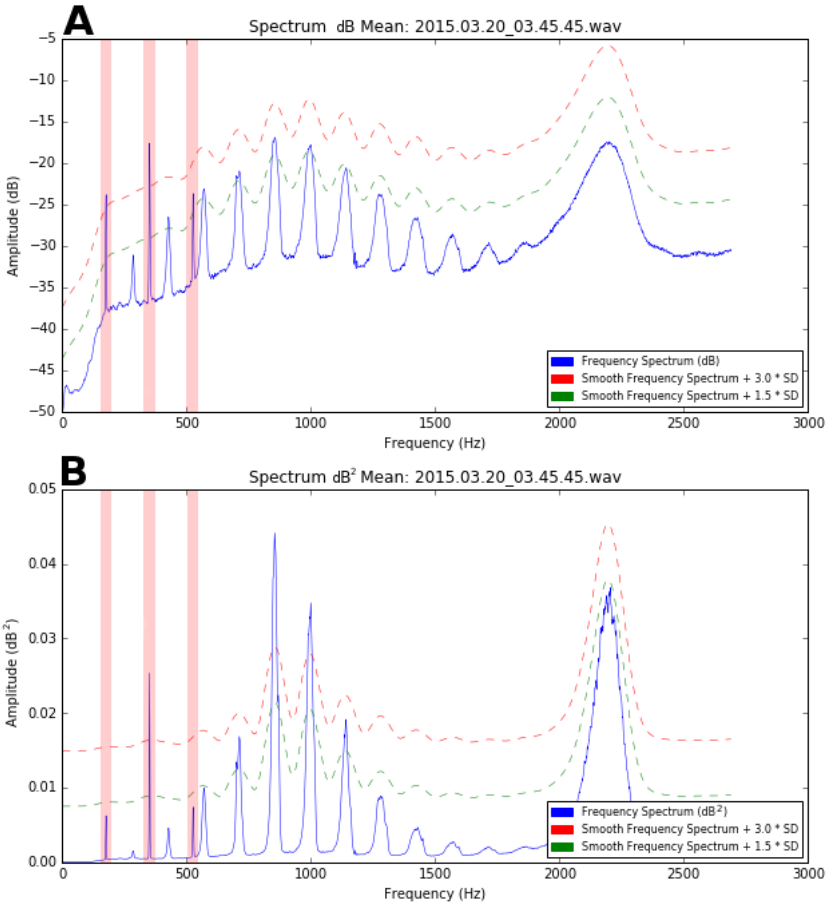

Fig. 3 Spectrum signature (blue line), FAV signature (green line) and binary boat detection signature (red marks) of (A) a recording that includes fish chorus and a boat; and (B) a recording that includes a small boat at minute 11. The horizontal dotted purple lines indicate the threshold defined to associate frequency peaks with acoustic emissions by boats.

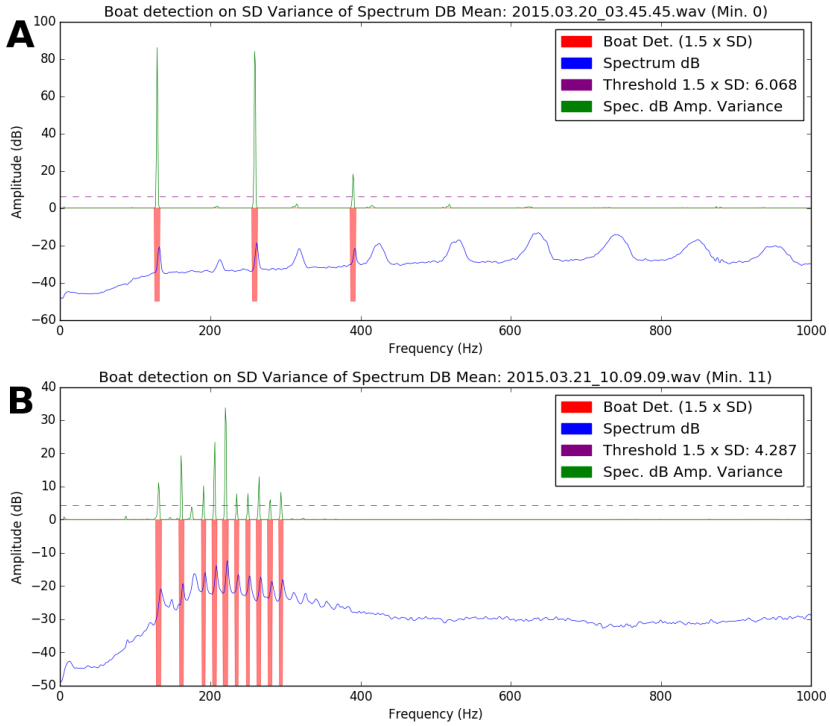

Fig. 4 24-hour spectrogram (frequency band from 0 to 2KHz) of the audio recordings taken at LSMSP in February 17, 2015. The recordings captured three types of fish chorus (A, B and C), as well as long and short duration sound emissions by boats and ships (D). The line chart at the bottom shows the detections obtained with the proposed method minute-by-minute and the number of peaks that characterize the corresponding boat signature.

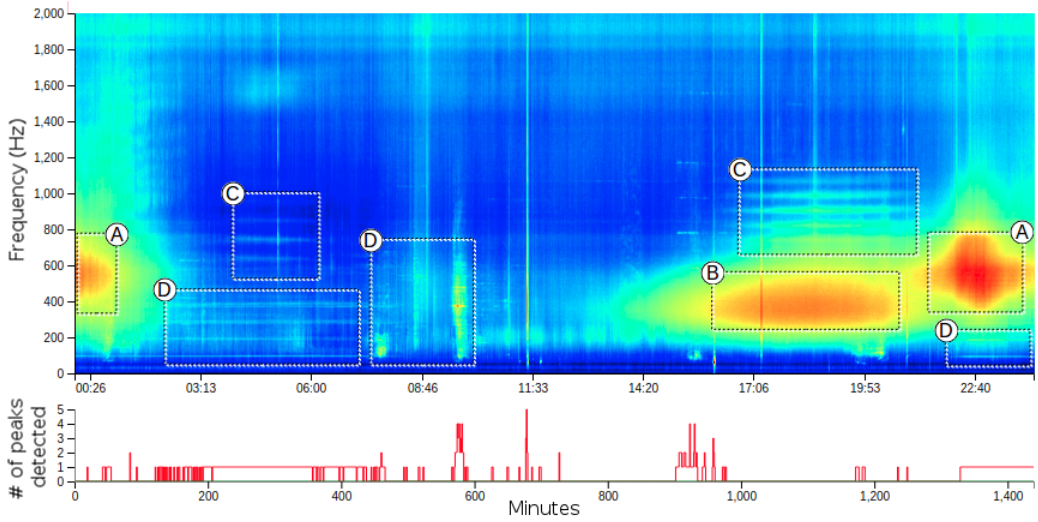

Fig. 5 24-hour spectrogram (frequency band from 0 to 2KHz) of audio recordings taken at XJSP in February 10, 2015. The recordings captured three types of fish choruses (A, B and C) and sound emissions by multiple ships (D) appearing as long and short duratio acoustic events. The detection line chart shown under the spectrogram reveals a concentration of short duration boat events in the period from around 9:00am to 11:00am; their corresponding signatures are characterized by up to 13 frequency peaks.

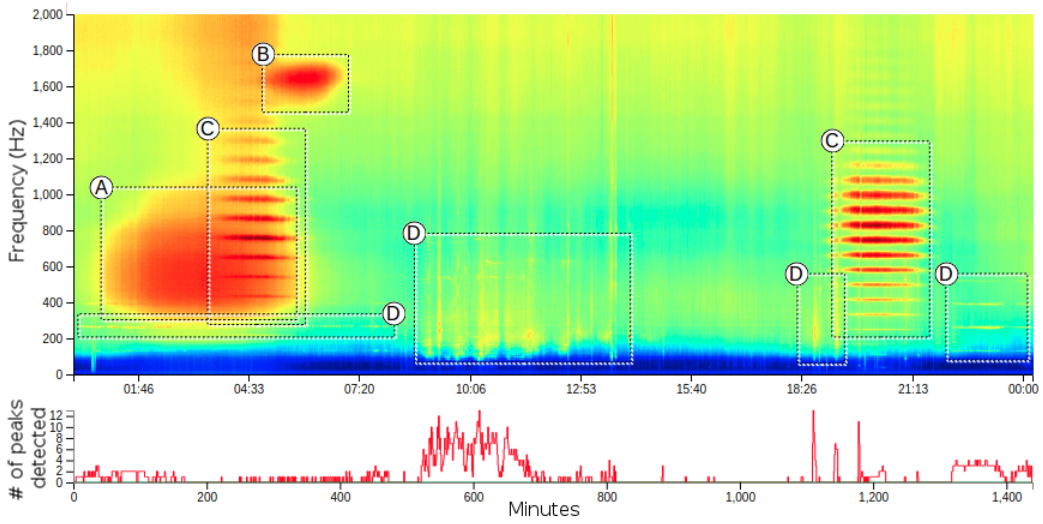

Fig. 6 15-minute spectrogram (frequency band from 0 to 1KHz) of a recording taken at XJSP in February 4, 2015 at 05h:39m:52s. A boat has been detected in the interval from minutes 1 to 9; the corresponding boat detection line chart is shown at the bottom.

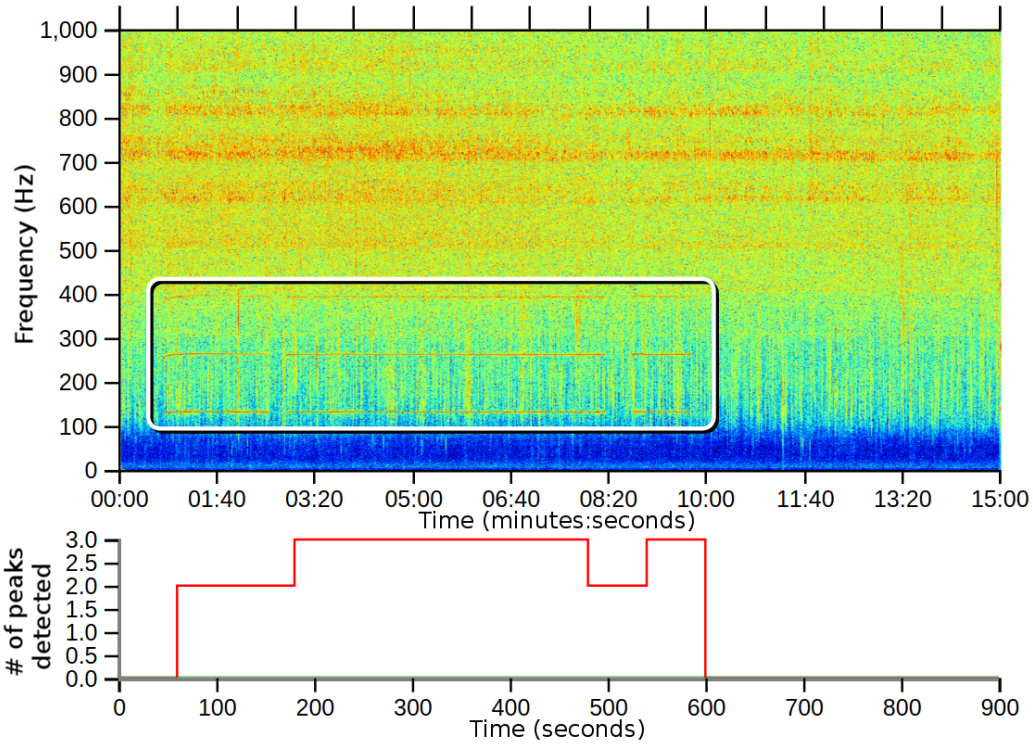

Fig. 7 15-minute spectrogram of a recording taken at XJSP in February 4, 2015 at 21h:12m:44s (frequency band from 0 to 1KHz). A boat occurrence has been detected in the interval from minutes 7 to 12; the corresponding boat detection line chart is shown at the bottom.

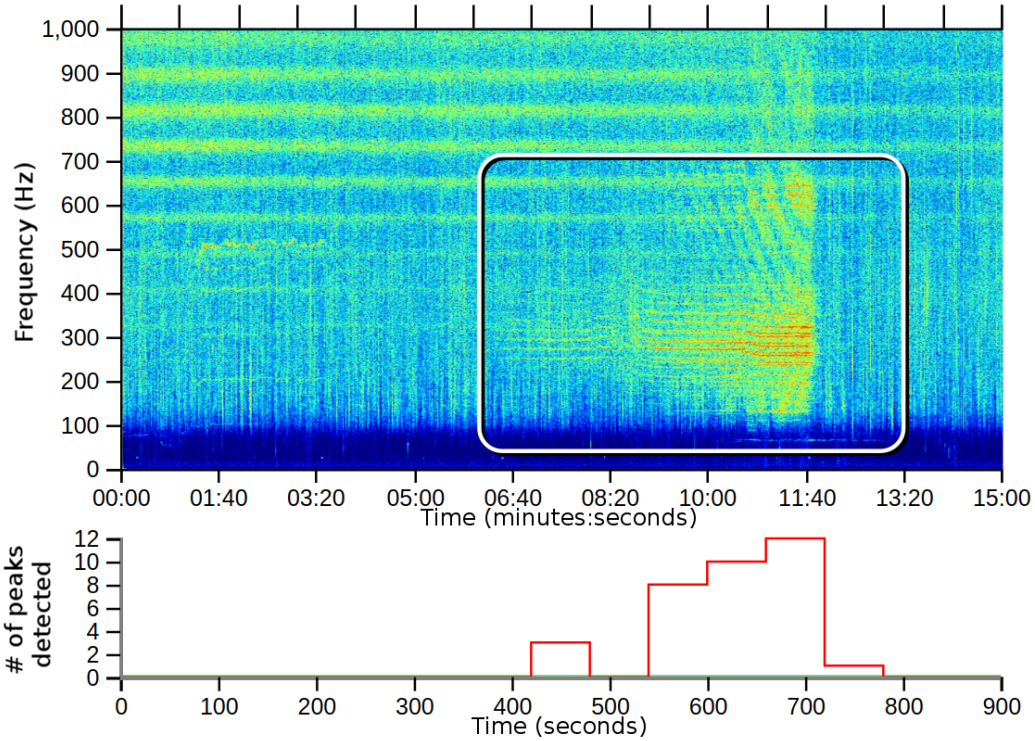

Fig. 8 15-minute spectrogram (frequency band from 0 to 1KHz) of a recording taken at XJSP in February 4, 2015 at 10h:40m:46s, in which a boat has been detected during minutes 5 and 6; the corresponding boat detection line chart is shown at the bottom.

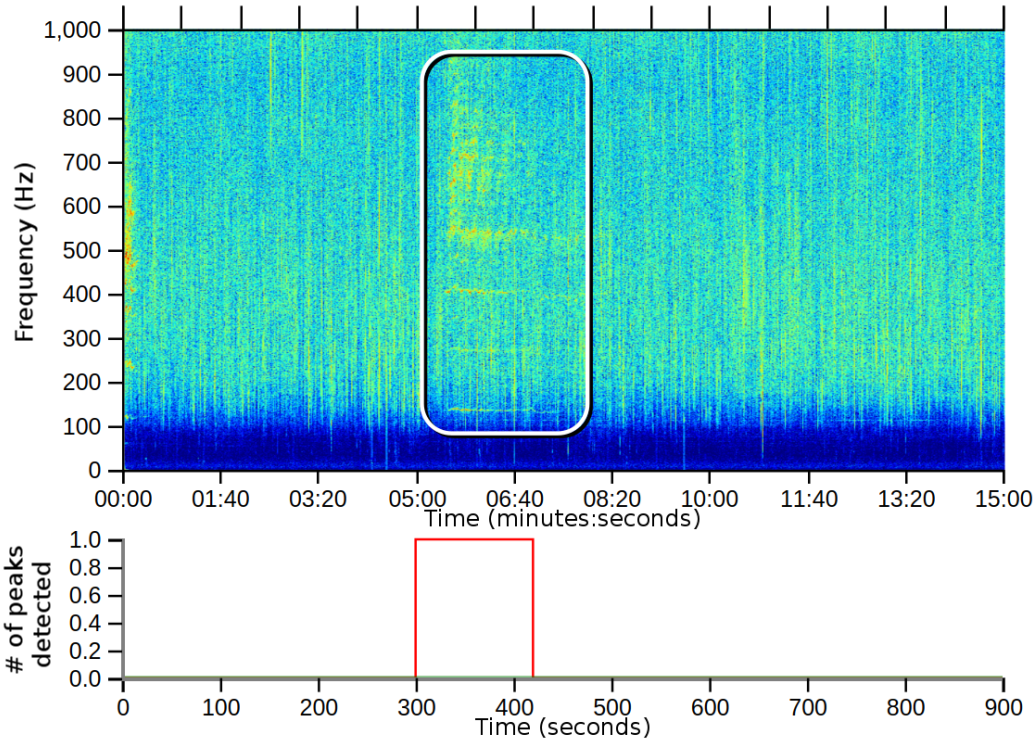

Fig. 9 15-minute spectrogram (frequency band from 0 to 1KHz) of a recording taken at XJSP in February 4, 2015 at 10h:10m:42s. Listening to the audio it is possible to identify the sound of a boat at minute 12, which the algorithm failed to detect, as observed in the corresponding detection line chart at the bottom.

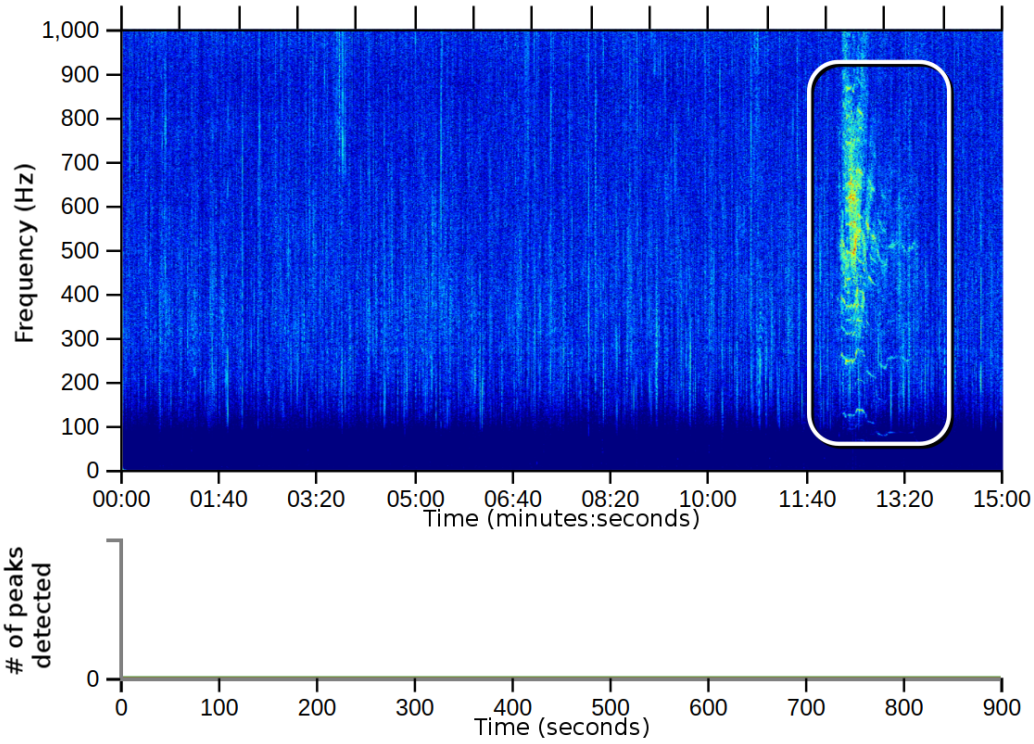

Fig. 10 15-minute spectrogram (frequency band from 0 to 1KHz) of a recording taken at LSMP in February 11, 2015 at 06h:27m:57s. The low amplitude signal from a boat has been detected in the invervals from minutes 0 to 3 and from 8 to 11, even though it is masked by crustacean noise, as observed in the boat detection line chart shown at the bottom.

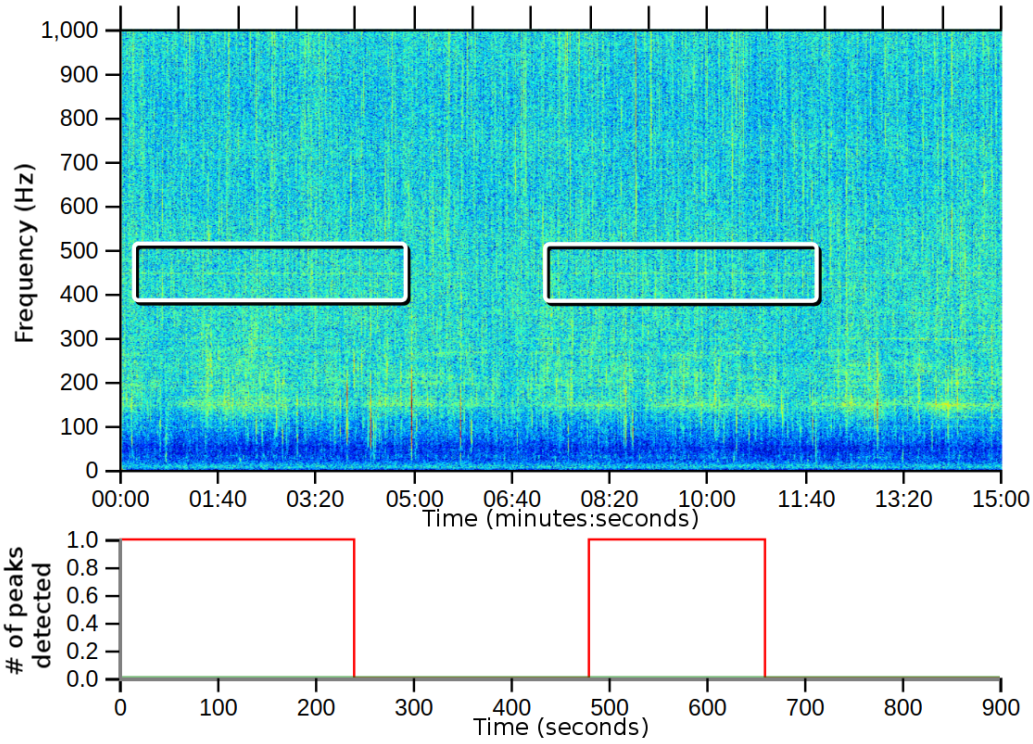

Fig. 11 15-minute spectrogram (frequency band from 0 to 1KHz) of a recording taken at LSMSP in February 11, 2015 at 15h:29m:54s. The signature of a boat has been detected along the duration of the recording, characterized by 35 up to 59 frequency peaks, as observed in the corresponding detection line chart.

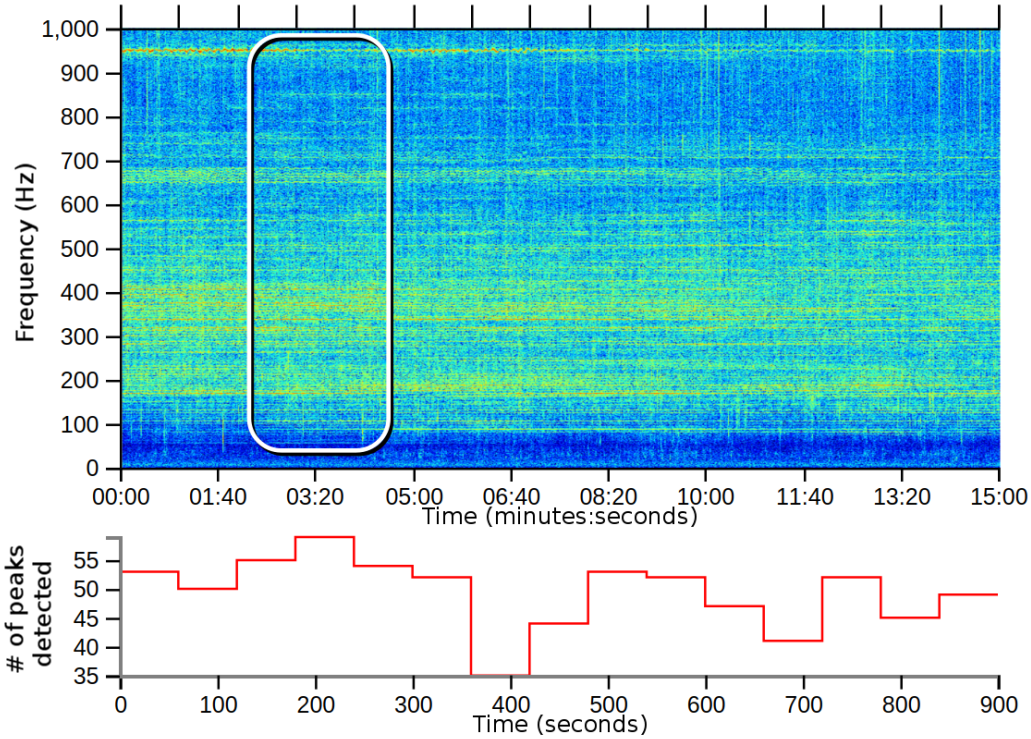

Fig. 12 15-minute spectrogram of a recording taken at LSMSP in February 4, 2015 at 04h:48m:55s, where overlapping sound emissions from two boats can be observed. The charts show the signatures computed of the audio samples corresponding to minutes 2, 3, 4 and 5, displaying the input spectrum signature (blue lines) with the corresponding boat detections highlighted by the red vertical lines.

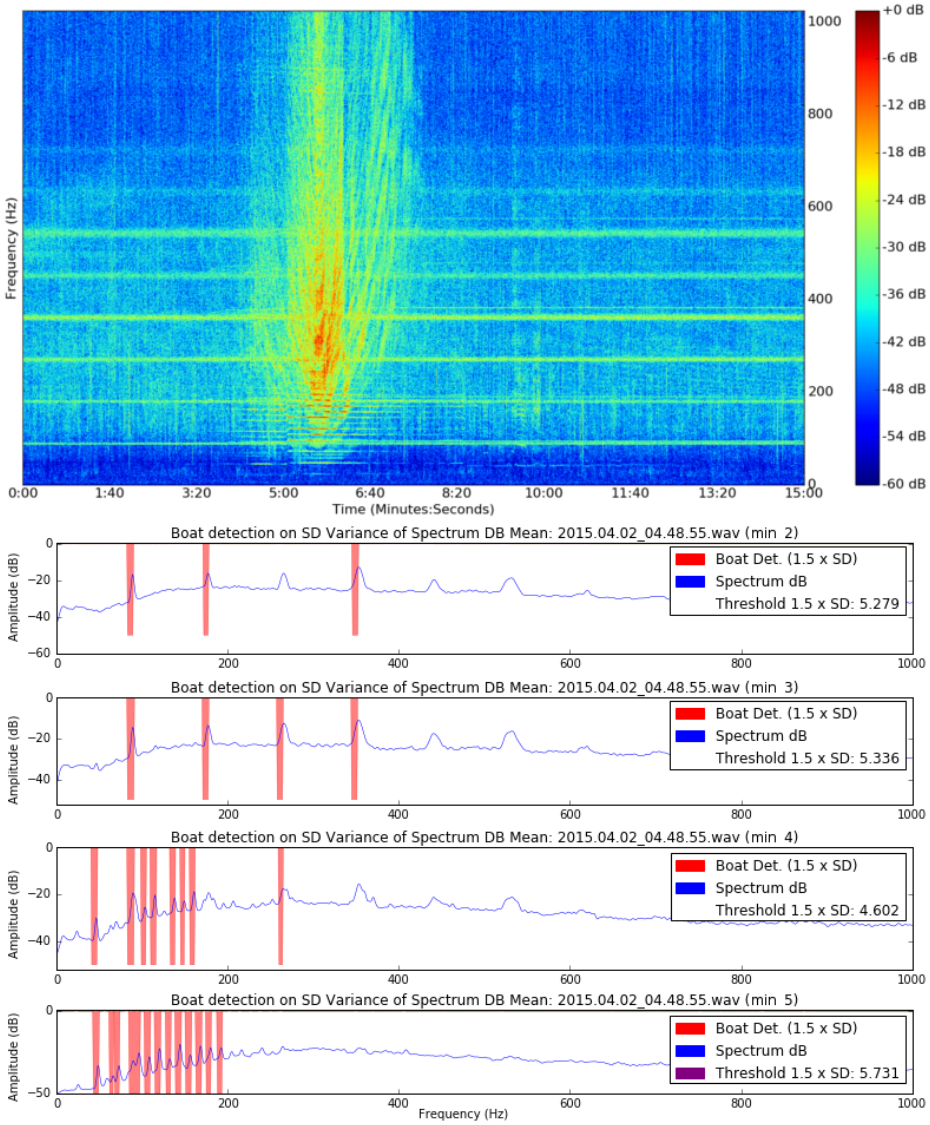

Fig. 13 FAV and DEMON signatures obtained for a one-minute audio sample that includes crustacean acoustic activity overlaped with a persistent acoustic emission by a boat (spectrogram on the right). The dotted gray lines indicate the detection threshold ($1.5 \times$ *Standard Deviation*). Both algorithms correctly detected the boat signature, with two frequency peaks. The red watermarks indicate the peaks detected in the FAV signature (purple line); the black dots indicate the detections in the DEMON signature (green line).

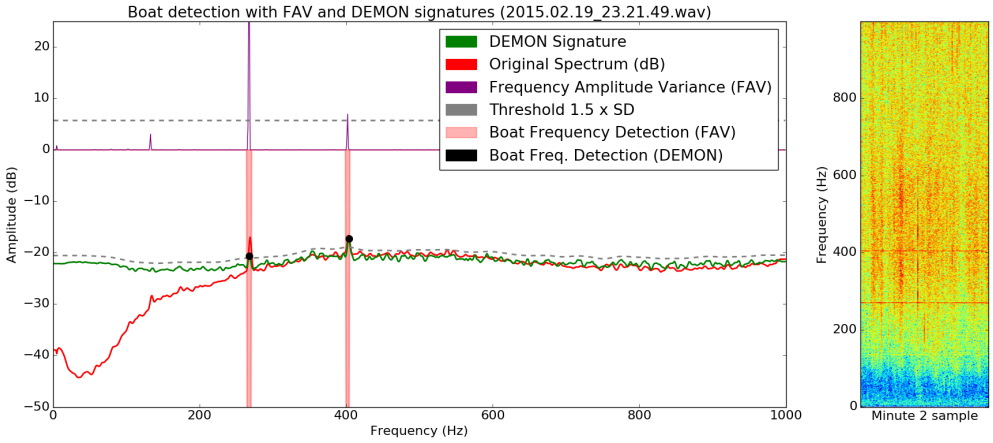

Fig. 14 FAV and DEMON signatures obtained for a one-minute audio sample that includes crustacean acoustic activity overlaped with a persistent acoustic emission by a boat (spectrogram shown on the right). The dotted gray lines indicate the detection threshold ($1.5 \times$ *Standard Deviation*). In this case the boat signature is characterized by eight frequency peaks. The red watermarks indicate the boat peaks correctly detected in the FAV signature (purple line); the black dots indicate the detections in the DEMON signature (green line). The DEMON algorithm identified a single peak due to the boat, the other peak is actually a false-positive, not related to the boat.