



Signal processing basics applied to ecoacoustics

Ignacio Sanchez-Gendriz

Federal University of Rio Grande do Norte, Natal, Brazil



ARTICLE INFO

Keywords:

Digital signal processing
Ecoacoustics
Soundscape
Long-term sound recordings
Passive acoustic monitoring

ABSTRACT

Ecoacoustics is a research field that has attracted attention of researchers from areas as diverse as ecology, biology, engineering, and human sciences, to cite a few. Ecoacoustics studies the sounds that emanates from the environments, by gaining insights of landscape dynamics from acoustic patterns and particularities at different places. With recent advance in technology, it is common to obtain sound datasets recorded 24-h a day for several months. The analysis of these long-term sound recordings represents several challenges. Investigators already involved in acoustic are familiar with signal processing methods, which are essential tools for sound analysis. However, in general beginners interested in Ecoacoustics do not share this background formation. The present work illustrates basic topics of digital signal processing in a comprehensible style, and an effective pipeline for long-term sound explorative data analysis (EDA) is presented. Finally, the described signal processing fundamentals are applied for EDA of 1-month underwater acoustic data recorded in a Brazilian marine protected area. The Matlab codes used for the analysis will be available as supplementary material.

1. Introduction

Ecoacoustics considers soundscape (Pijanowski et al., 2011a; Pijanowski et al., 2011b) or sounds that emanates from environments as a key ecological attribute (Farina and Gage, 2017). Sounds can give insights on dynamics of environments and they are recognized as an indicator of ecological processes (Sueur and Farina, 2015). Recent technological advance allows to collect sounds 24-h continuously for several months, which represents an opportunity for new discoveries, but also it is a challenge to overcome from the analysis point of view. The main objective of the present work is to illustrate how Digital Signal Processing (DSP) basics can be useful in the context of Ecoacoustics analysis.

Here we describe a framework that can be implemented for Explorative Data Analysis (EDA) of long-term sound recordings. The presented EDA framework is applied for 1-month underwater sound dataset.

The paper was structured as follows. Section 2 describes basic DSP concepts, exemplified by using simulated or real sound signals; at the end of the section the framework for EDA of long-term sound recordings is presented. In section 3 the application of the proposed framework is used for the analysis of 1-month of continuous underwater sound recordings. Section 4 discusses the applicability and possible extensions of the illustrated framework. The Matlab codes used for generating all the figures as well as for implementing the presented analysis will be available as supplementary material. For researchers interested in open

coding, Octave is a free source alternative to Matlab (Eaton, 2012; Nagar, 2018). Octave is mostly compatible with Matlab (About GNU Octave, n.d.) and can be considered for using the available scripts.

2. Methods

For beginners in areas like Ecoacoustics that deals with sound signal analysis, there are several new terms and concepts that need to be understood. A word cloud representation could be used to illustrate the feeling of overload when newcomers read technical papers for the first time, see Fig. 1A. However, an explanation using a logical flow of topics could contribute with a clear understanding of relevant concepts, see Fig. 1B as an illustrative example of major topics covered in the manuscript.

Methods here described are divided into five subsections, which deal respectively with signal fundamentals, frequency domain methods, filtering basics, some measurements relevant for acoustic analysis and finally the description about the proposed framework for EDA of long-term sound recordings.

2.1. Signal fundamentals: continuous, discrete, and sampling, period and frequency

The concept of signal as used in the present text, will refer to the sequence of values related to the magnitude of a physical variable

E-mail address: ignacio.gendriz@lais.huol.ufrn.br.

measured by a sensor. Specifically, sound signal refers to the signal captured by a microphone (or hydrophone in case of underwater recordings). Physical variables in the nature are continuous (analog signals), they can be evaluated at every time instant, thus resulting in infinite possible values for a finite time interval measurement.

Furthermore, a signal $x(t)$ can be classified based on its periodicity as periodic or aperiodic. Periodic signals repeat its values every T time intervals, which can be formally written as $x(t) = x(t + T)$. An example of a continuous periodic signal is illustrated in Fig. 2A; for sinusoidal signals like this, the frequency of the signal (f) is the rate of cycles/s (Hz), and the relation $T = 1/f$ is valid.

Computers cannot manage infinite sequence of values; therefore, continuous signals must be sampled to be stored and analyzed. After the sampling process, finite precision values measured at specific time instants are obtained, i.e., a digital signal, see Fig. 2B for a representation of a digital signal. The sampling process implies that not all signal values are stored, so it is correct to think that sampling could imply some loss of information, thus some precautions must be considered. The sampling frequency (f_s) parameter defines the number of samples taken every second. The Nyquist theorem states that to perform a sampling without loss of information the sampling frequency must be selected to be superior to 2 times the highest frequency in the continuous signal being sampled (Cabella et al., 2019). Fig. 2B shows a signal sampled in accordance with the sampling theorem. It can be observed from Fig. 2B that the underlying (continuous) signal it is well represented by samples. In contrast, Fig. 2C represents a sampling with a f_s below the ideal limit established by the Nyquist theorem, this phenomenon is known as aliasing. When aliasing occurs, the samples could represent signals with different frequencies.

Obtaining the frequency for sinusoidal sequences it is not a complex task, however, for dealing with acoustic signals a more elaborated mathematical tools such as Discrete Fourier Transform (DFT) needs to be used.

2.2. DFT, signals from another perspective: The frequency domain

The time domain representation of a given signal illustrates how its amplitude changes over time, this representation was used in Fig. 2. Based on the fact that signals can be represented as a sum of sine and cosines (Oppenheim, 2011; Platt and Denman, 1975), if we know the amplitude and frequency of the sinusoidal ‘units’ that compose a signal,

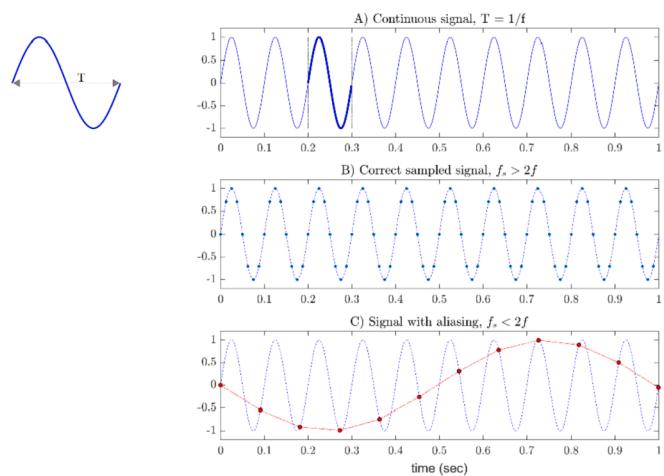


Fig. 2. Representation of continuous and discrete periodic sinusoidal signals. A) Continuous sinusoidal signal, one cycle of the signal is highlighted in dark blue, the duration of each cycle of the signal is measured by the parameter T , which denotes the period of the signal; the number of cycles in 1 s is the frequency ($f = 10$ Hz) of the signal. B) Discrete signal sampled in accordance with Nyquist's theorem ($f_s > 2f$), each point represents a sample from original continuous signal (dashed line). C) Discrete signal sampled not complying Nyquist's criterium ($f_s < 2f$), the samples could represent either the original signal ($f = 10$ Hz) or another signal with different frequency ($f = 1$ Hz). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

we can then represent it in another perspective: the frequency domain. Fig. 3 shows the time and frequency representation for a signal $x(t)$, obtained by the sum of five sinusoidal signals (Eq. 1).

$$x(t) = \frac{4}{\pi} \sum_{k=odd}^1 \frac{1}{k} \sin(2\pi kft), \text{ odd } = 1, 3, 5, 7, 9 \quad (1)$$

Specifically, the spectrum in frequency domain illustrates the amplitude and frequency for sinusoids that composes a given signal. One way to transform from time domain to frequency domain and vice versa is by means of Fourier Analysis, which have been applied in ecological studies (Bush et al., 2017; Scharlemann et al., 2008). Particularly for discrete signals Discrete Fourier Transform (DFT) and

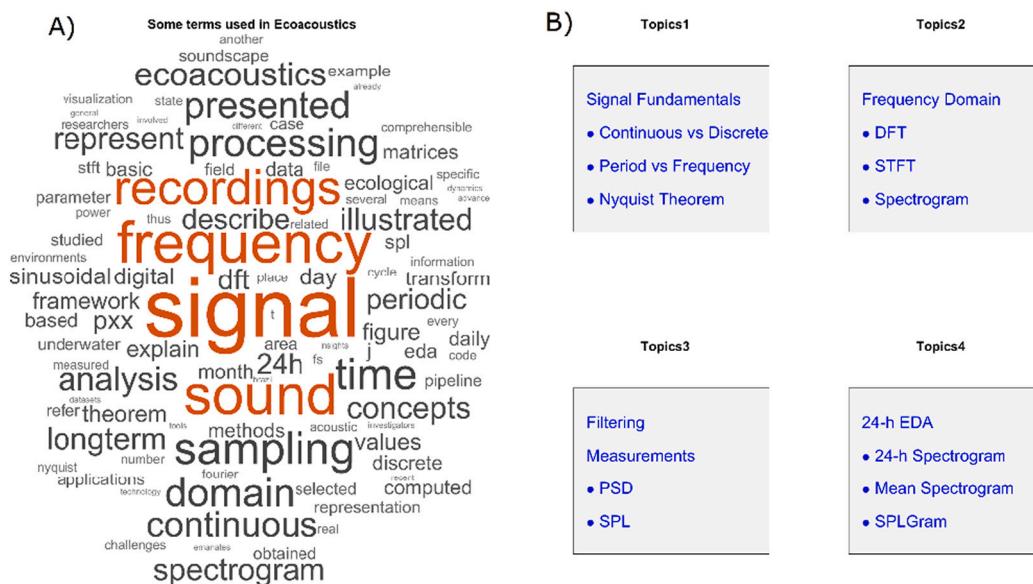


Fig. 1. Representation of terms, concepts and topics that can be found in Ecoacoustics literature. A) Word cloud from the present manuscript. B) Some of the principal topics covered in Methods section.

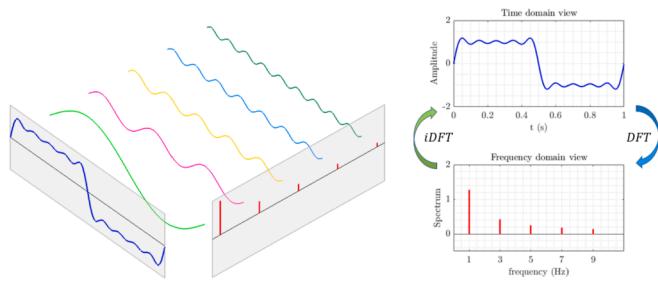


Fig. 3. Time and frequency domain view of a signal composed by the sum of basic sinusoidal signals. DFT can be used to transform from time domain to frequency domain, in contrast inverse DFT (iDFT) allows to convert from frequency domain to time domain.

inverse DFT (iDFT) can be used, both DFT and IDFT are part of the Fourier Analysis methods. It is worth noting that signals in time or frequency domain contain the same information. In general terms, DFT returns complex values, the magnitude and phase¹ of those complex numbers are defined as Magnitude and Phase spectrums, respectively. Magnitude (or Phase) Spectrum gives the magnitude (or phase) for each frequency component (Lyons, 2004). In the manuscript, the term Spectrum (without any other indication) refers to the Magnitude Spectrum, which is common in Ecoacoustics literature.

The DFT (denoted as $X[k]$) of a discrete signal $x[n]$ can be computed as in Eq. 2, and the iDFT of $X[k]$ can be computed as in Eq. 3, where W_n is a complex number see (Oppenheim, 2011).

$$X[k] = \sum_{n=0}^{N-1} x[n] W_N^{kn} \quad (2)$$

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] W_N^{-kn} \quad (3)$$

The implementation of DFT directly from its mathematical definition is computationally inefficient. For DFT to be used for practical applications, principally for massive data analysis, computational efficiency is one of the main requirements to be met. The methods designed for efficient computation of DFT are referenced in the literature as Fast Fourier Transform (FFT) algorithms (Cooley et al., 1967; Oppenheim, 2011). Thus, several software packages such as Matlab, Python and R will compute DFT by calling FFT inbuilt functions.

As we saw up to this point, DFT allows to get the frequency content for a signal. Nevertheless, when signals change frequency content over time, the DFT will not allow to distinguish the frequency variation in time. In cases that it is necessary to compute the frequency variation of signals over time, methods in time-frequency domain such as Short Time Fourier Transform (STFT) should be used (Smith, 2011).

2.2.1. STFT and spectrogram

The STFT of a signal is constructed by calculating the DFTs of consecutive windowed segments of the signal, for more details see the steps for STFT computation in (Smith, 2011). Fig. 4 illustrates the STFT computation process for a simulated signal (chirp) that changes its frequency over time from 5 Hz to 15 Hz, $f_s = 100$ Hz. Consecutive sliding windows allow a time overlap between them (Fig. 4B), e.g., for a sliding window of 100 samples, a 20% overlap means that the last 20 samples of the current window are the same as the first 20 samples of the next window. For each step of the STFT calculation, the selected signal segment is multiplied by a window function, which produces a signal

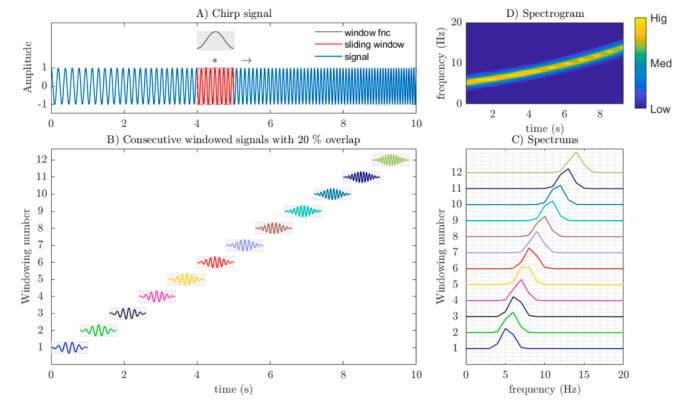


Fig. 4. Representation of the procedure of windowing, STFT computation and Spectrogram visualization for a chirp signal with frequency varying from 5 Hz to 15 Hz. A) Nonstationary signal changing frequency content over time, a sliding window and the Hamming function are highlighted. B) Windowing procedure used to compute the STFT, consecutive windowed signals share 20% of its samples, i.e., they have 20% overlap. C) The spectrums magnitude for each sliding window, all the computed spectrums compose the STFT for the analyzed signal. D) Spectrogram for the presented signal.

segment that we call a windowed signal. The multiplication by the window function diminishes possible discontinuities between beginning and ending of the windowed signals, since DFT assumes that analyzed signals are periodic. This intermediary step minimizes an undesired phenomenon known as spectral leakage (Oppenheim, 2011), characterized by an ‘energy leakage’ to frequencies which are not present at the original signal (Proakis and Manolakis, 2007).

For the example illustrated in Fig. 4 each sliding window have 1 s duration, consecutive segments have 20% overlap (0.2 s), and the Hamming² function was selected as the window function. This setup generates 12 windowed signals and 12 spectrums, see Fig. 4B and Fig. 4C, respectively. The visualization of the amplitude of the spectrums that compose the STFT is known as the spectrogram, see Fig. 4D. The spectrogram represents time along the x-axis, frequency along y-axis and amplitude of the time-frequency bins as gradient of colors, as can be observed in Fig. 4D.

2.2.2. Parameters and considerations about practical computation of STFT

2.2.2.1. Time vs frequency resolution trade off (Uncertainty principle). When calculating DFT for each windowed signal, the frequency resolution for the respective spectrum is inversely proportional to window duration (Proakis and Manolakis, 2007). A simple relation for the frequency resolution (f_{Res}) is $f_{Res} = f_s/N$, where N represents the number of samples contained in the analyzed window segments (Semmlow, 2018). For the example shown in Fig. 4B, where the length of the windows represents 1 s (100 samples), the frequency resolution is 1 Hz, which means that frequency components separated by less than 1 Hz cannot be differentiated. To improve frequency resolution, the window duration should be increased, thus decreasing time resolution. That is, there is a compromise between frequency resolution and temporal resolution. If we improve the resolution in one domain, we get worse in the other, meaning that we cannot simultaneously localize exactly a signal both in frequency and time domains. This trade-off between frequency and time resolution is known as Heisenberg–Gabor uncertainty principle or Gabor limit (Moca et al., 2021). A processing trick that can be used to improve

¹ The magnitude $|c|$ of a complex number $c = a + bi$ is computed as $|c| = \sqrt{a^2 + b^2}$, the phase is computed as $\phi = \tan^{-1}\left(\frac{b}{a}\right)$.

² Hamming function (as well as Hann, Bartlett and others possible windows) has values close to zero at the extremes and close to one at the middle. This justifies why the multiplication by the window function diminishes possible discontinuities between beginning and ending of the windowed signals.

the visualization of time resolution without a necessary detriment of frequency discrimination is by allowing overlaps between consecutive windows.

2.2.2.2. Visualization in dB scale. Decibel (dB) is a logarithmic scale that is commonly employed for quantifying a physical variable with large dynamic range, such is the case of sound (Erbe, 2011). For visualization purposes the logarithmic transformation $20 \log(X)$ is employed for illustrating the spectrogram obtained from STFT computation. The amplitude scaling $20\log(X)$ is commonly used in signal processing and particularly in Ecoacoustics. This transformation enables the visualization of low amplitude spectral components even in the presence of large amplitude components. Also, metrics expressed in dB relative to a reference pressure (P_{ref}) are referred to as "levels" (Erbe, 2011), see for example the definition of Sound Pressure Levels introduced later in section 2.4. For underwater measurements $P_{ref} = 1\mu\text{Pa}$, and for terrestrial applications $P_{ref} = 20\mu\text{Pa}$ (Merchant et al., 2015).

As we saw up to this point, signals can be characterized by its spectral content, next we will illustrate how to select desired spectral components for a given signal, while suppressing undesired frequency components. The procedure that allows such spectral separation is known as filtering.

2.3. Filtering basics

Most common filters used in Ecoacoustic applications can be represented by linear mathematical operations, as represented in subsection 2.3.1. Thus, a filter can be viewed as a mathematical operation, in which a signal $x[n]$ is feed as input and a signal $y[n]$ is produced as output. In this section we are going to clarify some points of the filtering procedure. Since we are dealing with digital signals in the present manuscript, we will describe digital filters, thus when referring to filtering process we will be describing numerical operations performed on digital samples.

2.3.1. Finite impulse response and infinite impulse response filters

The Eq. 4 illustrates the filtering operation, the output at the current sample can be computed from a linear combination of current and past input samples, and past output samples. The numbers a_i and b_j are known as filter coefficients. The filter order is the maximum value between N and M , where N and M indicate the number of input and output past samples that can be used for computing the current output, respectively.

$$y[n] = \sum_{i=0}^N a_i x[n-i] - \sum_{j=1}^M b_j y[n-j] \quad (4)$$

If all the b_j ($j = 1, 2, \dots, M$) coefficients are zero, then the filter is classified as Finite Impulse Response (FIR), and the output at the current sample only depends on current and past inputs $x[n]$. In the other case, if at least one value of b_j is different from zero the filter is classified as Infinite Impulse Response (IIR), and the process include also past samples from the output.

Frequency response is one of the most important criteria for designing digital filters. Frequency response indicates how the filter select (whether it attenuates or not) specific spectral components contained in the signal $x[n]$. Filters can be classified based on frequency response characteristics, and the four most common categories are Low-Pass, High-Pass, Band-Pass and Band-Reject (Smith, 2002).

Fig. 5 exemplifies a low-pass and high-pass filtering on simulated signal composed by the sum of sinusoids with frequencies at 10 Hz, 30 Hz, 50 Hz, 70 Hz and 90 Hz. Since filters selects which spectral components pass or are attenuated, they can be easily understood in the frequency domain. When a signal is filtered, basically its spectrum is multiplied by the filter frequency response. See Fig. 5 for an illustration of filtering functioning, for example, when a low-pass is applied on the

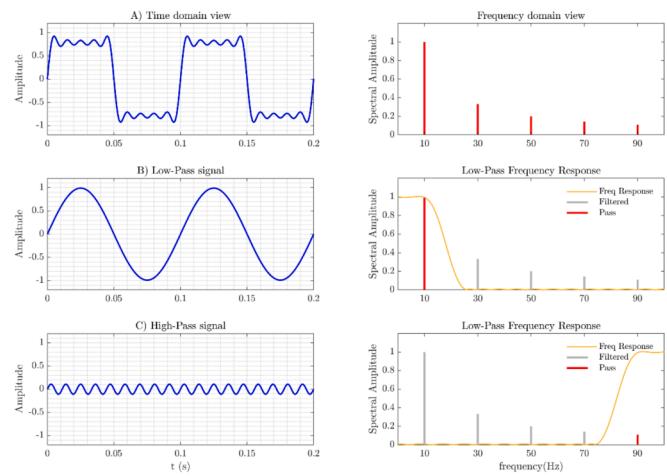


Fig. 5. Exemplification of the filtering process. A) Top panel represents a signal composed by the sum of sinusoids and its spectral content. B) Middle panel illustrates the result of low-pass filtering, the frequency response allows the pass of spectral components below 10 Hz and attenuates spectral components above 25 Hz. C) The bottom panel shows the result of high-pass filtering, spectral components below 75 Hz are attenuated, since frequencies greater than or equal to 90 Hz are practically unchanged.

signal, low-frequencies in the pass band are multiplied by values close to one, thus spectral content of the pass band is practically unchanged. On the other hand, high-frequency components in the stop band are multiplied by values near to zero, therefore higher frequencies are almost eliminated after low-pass filtering.

2.3.1.1. Considerations about selecting FIR or IIR filter. There are some differences between FIR and IIR filters that should be considered for deciding which type of filter to select for a specific application (Oppenheim, 2011). Both types of filters have advantages and disadvantages. For example, FIR filters have the advantage that they do not distort the filtered signal, and only introduce a time delay in the processed signal, which can be easily removed. However, IIR filters can achieve a desired frequency response with less coefficients than its FIR counterparts. In contrast, the disadvantage of IIR filters is that they could introduce distortion on the processed signal, which could be not tolerated for specific applications. In general, the selection of the filter class will depend on the context of applications. For example, for the most common cases in Ecoacoustic applications, the signal processing is executed off-line by using personal computers. The processing power of modern computers allows that the number of coefficients needed for filtering with precise frequency response are not a real concern. Therefore, FIR filters are good choices for applications in most Ecoacoustic research. Although less common, IIR filters might be the preferred option for implementations on embedded hardware with restricted computational power.

2.3.2. Moving average filters

Moving average filters can be considered as special cases of FIR filters, for example Eq. 5 represents a formulation of a moving average filter. If we compare Eq. 4 and Eq. 5, can be observed that filter coefficients $a_i = 1/N$ and b_i are zero. These filters compute the average of the N most recent samples, e.g., for $N = 5$ moving average sum the latest 5 points and divide the result by 5. For higher values of N , the filtering produces smoother signals and for lower values the noise is less attenuated, this consideration must be considered for practical applications. Since slow oscillations are less affected and high oscillations with low amplitude are eliminated, in the frequency domain the moving average works as low pass filters. Moving average filters are simple for implementation and an effective choice for smoothing signals and for

removing random noise, while maintaining signal trends (Hunter, 1981; Smith, 2002).

$$y[n] = \frac{1}{N} \sum_{i=0}^{N-1} x[n-i] \quad (5)$$

Fig. 6 shows the application of moving average filter on sinusoidal signal containing random noise. After moving average filter, the resulting signal is similar to the original signal without noise, see **Fig. 6 B).**

2.4. Parseval relation, power spectral density and sound pressure level

The Parseval's relation states that the energy of a signal $x[n]$ with N samples can be computed from the Fourier Transform. Specifically, if $X[k]$ is the DFT of $x[n]$, then the Parseval's relation (Oppenheim, 2011) can be written as in Eq. 6:

$$\sum_{n=0}^{N-1} |x[n]|^2 = \frac{1}{N} \sum_{n=0}^{N-1} |X[k]|^2 \quad (6)$$

The Parseval's relation is helpful in Ecoacoustics, since it can be used to compute energy variations in time signal through the frequency domain. Measurements related to signal energy variations could be used for acoustic events detection. Here we illustrate how to compute signal energy variation from frequency domain, but the main goal is not to detail the complete procedure to obtain calibrated measurements, for a more detailed explanation see (Merchant et al., 2015).

The Mean Square Pressure (MSP) is related to the mean square value of a sound pressure signal $x[n]$, and can be computed as in Eq. 7:

$$MSP = \frac{1}{N} \sum_{n=1}^N |x[n]|^2 \quad (7)$$

Power Spectral Density (PSD) is the contribution of power per unit of bandwidth (Proakis and Manolakis, 2007), and the power in selected frequency bands can be used to estimate the contribution of specific spectral components. PSD can be computed as the magnitude square of DFT $X[k]$ (Alkin, 2016), thus it is derived from the Eq. 6 that we can compute energy from PSD based on Parseval's relation.

In the same way, Parseval relation can be used to obtain MSP from frequency domain, therefore, we can compute broadband MSP or MSP in selected frequency bands. Based on a simulated signal $x[n]$, Fig. 7 illustrates that computing MSP from time domain or from frequency domain results in the same quantity (see **Fig. 7C**). The simulated signal $x[n]$ is composed by the sum of five sinusoidal waveforms, $x_i[n]$, $i = 1, 2,$

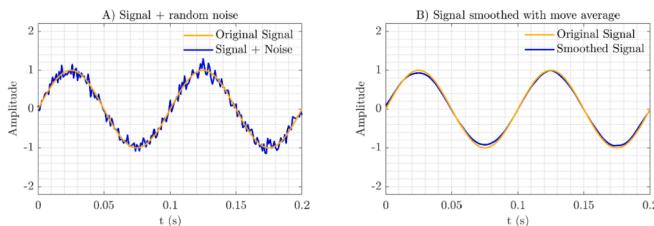


Fig. 6. Application of moving average filter for smoothing a sinusoidal signal containing random noise. A) Noisy signal and original sinusoidal signal, noisy signal contains abrupt oscillations generated by random noise. B) Smoothed signal and original sinusoidal signal, smoothed signal after moving average filtering it is close to the original sinusoidal signal.

...5. In **Fig. 7B**, three frequency bands are delimited, which were named: FB1 (5–20 Hz), FB2 (25–60 Hz) and FB3 (65–95 Hz) respectively.

Taken as an example, to compute the MSP for FB2 in time domain both sinusoidal signals $x_2[n]$ and $x_3[n]$ were summed up, then Eq. 7 was applied.³ On the other hand, from the frequency domain, the PSD values corresponding to spectral components comprising FB2 were summed. The advantage for computations from frequency domain is that once PSD is computed it serves to quantify spectral content as well as for MSP calculations in desired spectral bands.

A useful application of Parseval's relation is that we can compute energy or power in selected frequency bands, without the need to do the calculations with the signal in the time domain, see **Fig. 10** for a real energy detector.

Sound Pressure Level (SPL) is a common metric used in acoustics analyses (Merchant et al., 2015), SPL can be expressed as in Eq. 8, where $P_{ref} = 1\mu Pa$ is the reference pressure underwater. The SPL metric can also be computed from frequency domain, once calculated PSD matrices. The calculation and usefulness of PSD matrices will be illustrated in following sections.

$$SPL = 10 \log \frac{MSP}{P_{ref}^2} [dB \text{ re } 1\mu Pa] \quad (8)$$

2.5. Pipeline for 24-h EDA

The procedure used to calculate daily PSD matrices, named here as P_{xx} matrices, are explained below and exemplified in **Fig. 8**. For each wav file, recorded for day j (j varying from February 4th to March 4th) a respective p_{xx} ⁴ matrix was computed, by means of the Welch method (Welch, 1967), with 1-s Hamming window, 1025 frequency points, 50% of overlap, and with 60-s temporal signal segments. Then, all p_{xx} matrices referents to day j were merge in to a daily P_{xx} matrix and, finally, by visualizing P_{xx} matrices the 24-h spectrograms were obtained. Based on the system recording settings (see Section 3 for details), the numbers of recorded wav files by days, was $N = 96$. Aiming to analyses the whole collected dataset, 24-h spectrograms were constructed for each monitored day. These 24-h spectrograms were obtained by visualization of daily P_{xx} matrices.

3. EDA applied to a 1-month underwater sound data

Here we describe as a case study 1-month of continuous underwater sound recorded at Xixová-Japuí State Park (XJSP), located on the SW of Santos Estuarine System (São Paulo State, Brazil), see **Fig. 9**. This area has a rich daily soundscape, but it is also affected by urbanization, industrial and port activities (Sánchez-Gendriz and Padovese, 2016). Fish choruses with daily periodicities were the most important contributor for the XJSP soundscape (Sánchez-Gendriz and Padovese, 2016; Sánchez-Gendriz and Padovese, 2017a). The location for the monitored place is represented in **Fig. 9**. Underwater acoustic data was recorded by means of an autonomous passive monitoring system (Caldas-Morgan et al., 2015). The acoustic signals were acquired at 11.025 kHz sample rate, with 16-bit resolution; individual sound files of 15 min duration were continuously stored in a SD card.

³ In time domain, another way to calculate MSP for FB2 could be by applying a bandpass filter at $x[n]$ that allows the passage of spectral components within FB2, and then using Eq. 7.

⁴ p_{xx} matrix refer to PSD matrix computed for individual sound files, while P_{xx} matrix refers to daily period, thus P_{xx} for specific day is obtained by merging all p_{xx} matrices for that day.

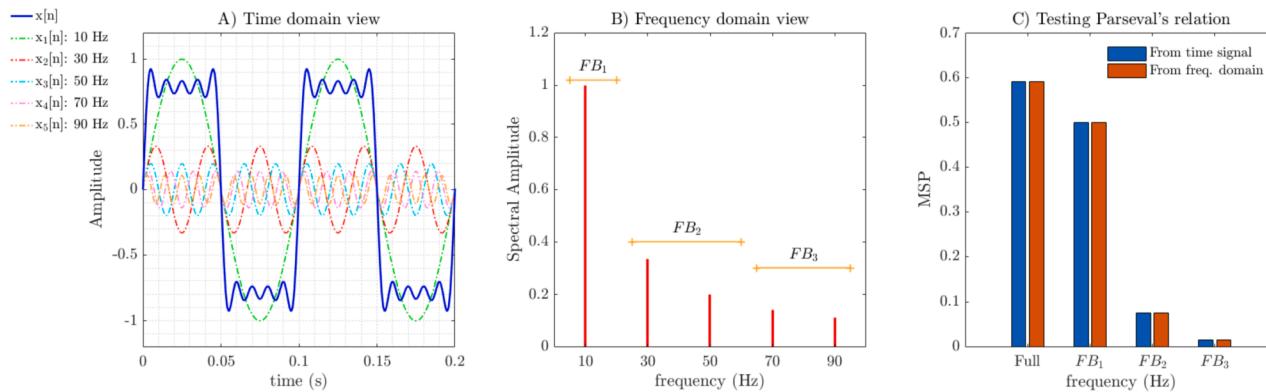


Fig. 7. Testing the Parseval's relation. A) Simulated signal $x[n]$ composed by the sum of five sinusoidal waveforms. B) Spectral representation for the simulated signal, selected Frequency Bands (FB), named FB1, FB2 and FB3 are represented. C) The MSP is computed from the simulated signal (Full FB), and from the time signals representing the pure sinusoids in the respective frequency bands (FB1, FB2 and FB3), bar graph shows that the results are the same regardless of whether MSP is calculated from the time domain or from the frequency domain.

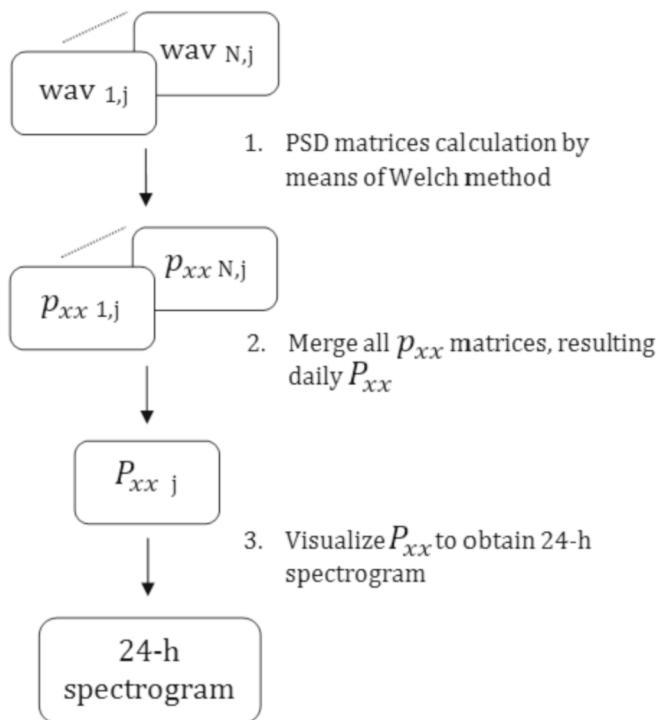


Fig. 8. Flowchart for presented pipeline.

3.1. Short-term spectrogram example

Although the study presents EDA for long-term sound recordings, the techniques here illustrated are also useful for short-term sound analysis. Fig. 10 illustrates a raw signal that contains an acoustic event overlapped with noise (ocean ambient noise and shrimp noise). The sound event is imperceptible from signal in time domain since the noise and the event of interest have comparable magnitude. In contrast, from a visual inspection of the spectrogram the event is clearly detected. Since noise in the raw signal is broadband and that the event is mainly concentrated within 0.4–1.3 kHz, the Parseval's relation (see section 2.4) was used to estimate the energy of the event within the desired frequency band. Next, by applying a threshold (mean + std) on the computed energy was possible to detect the event time occurrence.

If the raw signal is first band-pass filtered, the acoustic event is already notable from filtered time signal. However, the energy detector remains almost immutable. This fact highlights the usefulness of the



Fig. 9. Monitored point located at Xixova marine park.

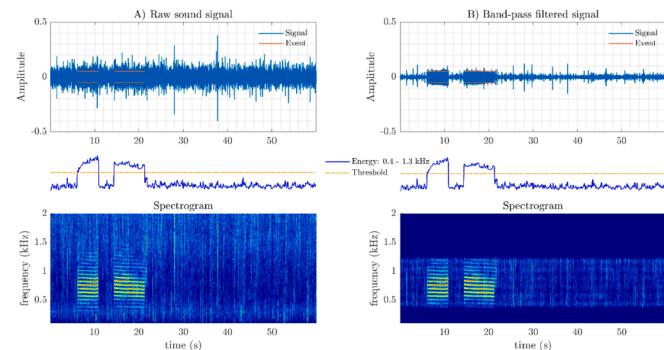


Fig. 10. Energy detector implemented based on Parseval's relation. A) 1 min raw acoustic signal, B) Band-pass filtered signal, frequency band: 0.4–1.3 kHz. Top panels represent time signal. Middle panels illustrate an energy detector implemented by using the Parseval's relation, dashed line represents the threshold for event detection (mean + std). Bottom panels represent the spectrograms of the time signals. The setups used to compute the spectrograms: window duration: 0.08 s (which represents 882 samples, since $f_s = 11.025$ kHz), 80% overlap, window type: Hamming function, the number of points to compute the FFT (nfft) for each window was set to 1024 (i.e., nfft = 1024).

Parseval's relation to implement energy detectors for band-limited acoustic events, even in those challenging cases when the signal is contaminated with broadband noise, as the real case represented in

Fig. 10. The sound event here illustrated shows spectral pattern similar to fish chorus presented in the next section (see Ch₄ annotations in Fig. 12).

3.2. 24-h Spectrogram and 24-h SPL

24-h spectrograms are valuable for visualizing daily soundscape variations for a given monitored area. Averaging 24-h spectrogram also can be used for summarizing soundscapes (Sánchez-Gendriz and Padovese, 2016; Sánchez-Gendriz and Padovese, 2017b) for representative periods, as for example 1-month. Daily 24-h spectrograms for complete four weeks are illustrated in Fig. 11. It is worth noting that acoustic events seem to occur every day with a precise time-spectral patterns, concentrating mainly below 2 kHz. For summarizing sound data recorded for 1-month was computed the mean 24-h Spectrogram, illustrated in Fig. 12. From 24-h mean spectrogram, four main acoustic events can be noticed, these acoustic events are originated by fish choruses (Sánchez-Gendriz and Padovese, 2017a) and are denoted as Ch₁, Ch₂, Ch₃ and Ch₄.

SPL computed within the Low-Frequency Band (Low-FB) and High-Frequency Band (Hig-FB) were also used for studying sound levels variations. For the analysis was selected a Low-FB that comprises frequencies from 150 Hz to 1.2 kHz and Hig-FB that includes frequencies from 1.5 kHz to 2.0 kHz. Fig. 13 shows four-week Low-FB SPL time series and Fig. 14 represent four-week Hig-FB SPL time series.

3.3. Finding soundscape periodicities

The autocovariance (Shumway and Stoffer, 2017) and the spectral estimation of the SPL time series were used for quantifying perceptible periodicities in the analyzed data. Fig. 15 shows the autocovariance and the spectrum Low-FB SPL and Hig-FB SPL, respectively. For the purposes of this work were defined strong periodicities and weak periodicities as principal peak and secondary peaks detected on PSD curve. The strong periodicities are annotated in red and weak ones are annotated with green markers, both in time domain (autocovariance) and in frequency domain (spectral content illustrated through the PSD). The steps performed to estimate periodicity values through time domain analysis are

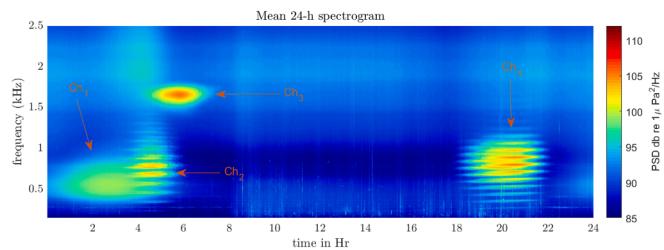


Fig. 12. Mean average (Note that average could refer to either mean, median or mode (Gupta, 2011), thus we explicitly used the term mean average to avoid possible misunderstandings.) for 24-h spectrograms between February 04 to March 04. Four predominant acoustic events (Ch₁, Ch₂, Ch₃ and Ch₄) are indicated by arrows.

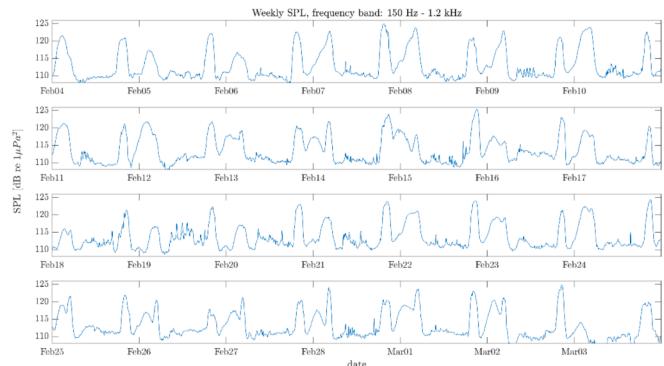


Fig. 13. For week Low-FB SPL, each panel represents one week, respectively.

twofold. First, the mean interval between consecutive peaks (in minutes) was determined. Second, the ratio between the number of minutes for one day (1440) and the mean interval between peaks was computed. As an example, let's use the case of strong periodicity represented in Fig. 15A (see peaks annotated in red). For this case, consecutive peaks are separated in average by 1441.1 min (i.e., mean interval between

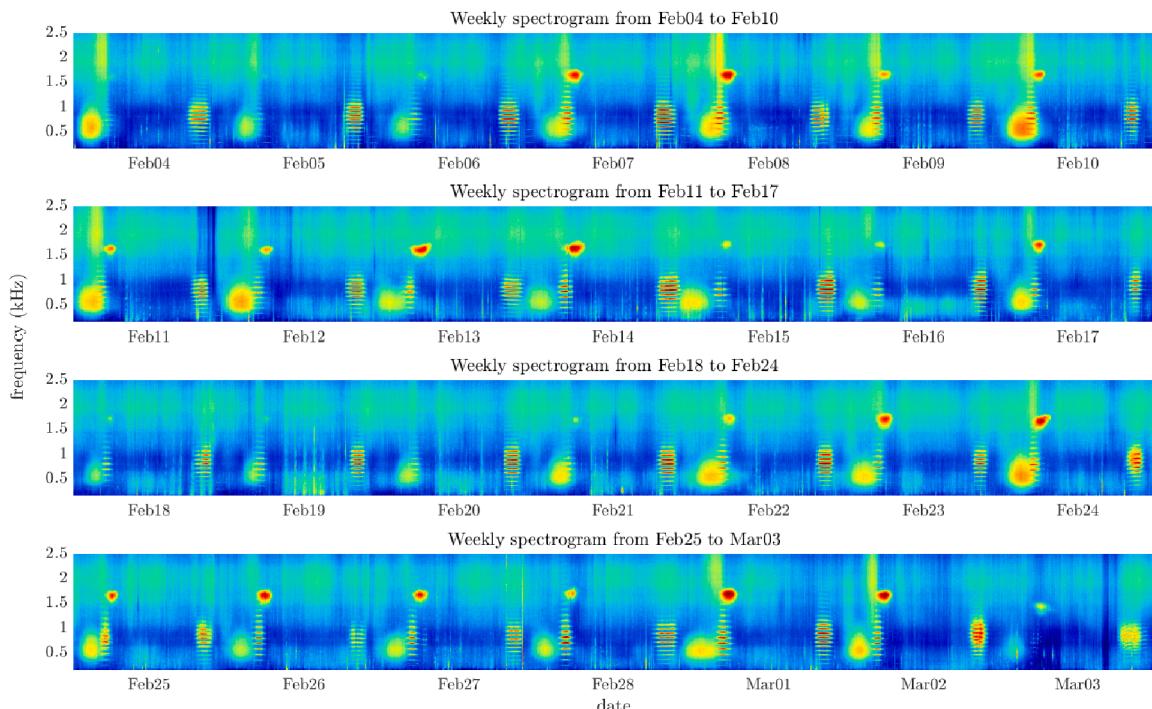


Fig. 11. Four weeks spectrograms. The weekly spectrograms allow to identify acoustic events that suggest a daily pattern.



Fig. 14. For week Hig-FB SPL, each panel represents one week, respectively.

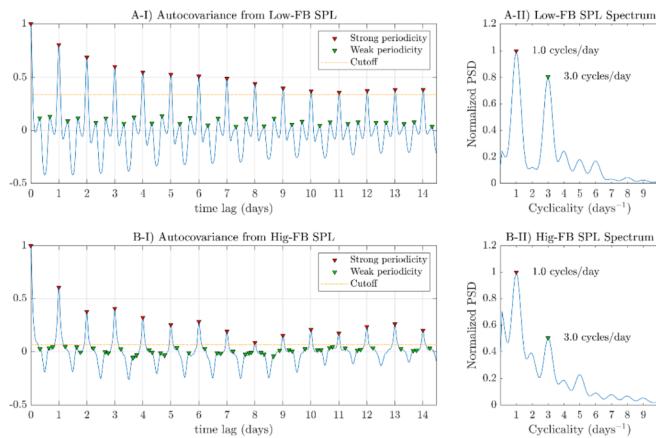


Fig. 15. Determining periodicities in SPL time series. A-I) Autocovariance calculated from Low-FB SPL. A-II) Spectrum of Low-FB SPL. B-I) Autocovariance calculated from Hig-FB SPL. B-II) Spectrum of Hig-FB SPL. For panels A-I and B-I, the peaks in red markers indicate strong periodicities and green markers point out weak periodicities; a cut-off line indicates the minimum value for strong periodicity peak found in autocovariance. Panels A-II and B-II represent the spectrum obtained for SPL, these panels illustrate periodicities (or cyclicity) of SPL time series computed from frequency domain. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

peaks is 1441.1 min). Then, the periodicity is estimated by 1440 / 1441.1, which results in a period of approximately 1 cycle / day. On the other hand, in frequency domain finding periodicities can be done in a simple way, by detecting the prominent spectral peaks.

For the case of Low-FB SPL both the strong periodicities and weak periodicities computed from time domain (autocovariance) or from frequency domain (by means of PSD) results in almost the same value, 1 cycle/day and 3 cycles/day, respectively. Thus, for Low-FB SPL the computed strong periodicity is related to a daily cycle, which agrees with the repetitive 24-h patterns that can be visually detected in Fig. 11. For the case of weak periodicity detected as 3 cycles/day, we can see from mean 24-h spectrogram (Fig. 12) that in Low-FB (150 Hz – 2.0 kHz) there are 3 different major acoustic events, which can explain the occurrence of that weak periodicity.

For the case of Hig-FB SPL the strong periodicity computed from time domain or spectral estimation is the same, but this is not the case for computed weak periodicity. The weak periodicity computed from time domain was approximately 4.1 cycles/day, which does not agree with the value computed from frequency domain (see Fig. 15, bottom panels). The weak periodicity quantified for Hig-FB SPL could be related to weak events in this frequency band or also by the incidence of remaining spectral components for the 3 major events quantified in Low-FB.

3.4. SPL-gram

The SPL-gram representation allows to visualize the SPL variations for long-term acoustic monitoring (Sánchez-Gendriz and Padovese, 2017b). By using the SPL-gram for selected frequency bands a “picture” of the temporal structure of the soundscape can be obtained. To obtain the SPL-gram, daily SPL for specific frequency band is smoothed by applying moving average filter. The moving average eliminate transient oscillations as for example ship noise, while preserving lasting events, such as fish choruses. Once daily smoothed SPL are obtained, the time series are converted to color bars. Finally, by merging color bars in chronological order the SPL-gram representation is obtained. Fig. 16 shows the procedure for obtaining SPL color bars for the case of Low-FB SPL and Hig-FB SPL for a specific day.

Fig. 17A shows the SPL-gram for Low-FB SPL and Hig-FB SPL for underwater sound collected between February 04 to March 04. By plotting on the graph events occurrence such as sunrise and sunset we can obtain a temporal relation between soundscape variation and these events. For example, for the case of Low-FB SPL-gram, the occurrence of sound level increase around the time interval from 19 h – 21 h (UTC time) ends before sunset. This fact is related to the occurrence of a fish chorus (denoted previously as Ch₄, see Fig. 12), which seems to decrease its activity shortly before sunset. Another soundscape trend that can be observed for Low-FB SPL is that the period between 06 h – 18 h appears without substantial lasting acoustic events.

Also, mean values for daily SPL could be used for summarizing daily trends in the studied place. Fig. 17B represents the mean of SPL time series in polar form (bottom panels). For Low-FB SPL can be observed two significant deviations from the baseline levels. Between 02 h - 05 h the first increase in Low-FB is noted, which in turn can be divided into two different events, one achieving the acoustic peak between 02 h – 03 h, and the other attaining the maximum level between 04 h - 05 h. The later deviation in Low-FB SPL it is reached between 20 h – 21 h. For the case of Hig-FB SPL one major deviation from baseline levels appears between 05 h – 06 h. These acoustic peaks agree with the visual information illustrated in mean 24-h spectrogram and the events annotated as Ch₁, Ch₂, Ch₃ and Ch₄ (Fig. 12).

4. Conclusion

Results showed that basic concepts of the field of digital signal processing are valuable tools for Ecoacoustic researchers. To name a few, understanding basic concepts such as filtering, spectral analysis, and visualizations such as spectrogram can be particularly useful for explorative data analysis in Ecoacoustics. Thus, the comprehension of those concepts and the knowledge about how to use it in real applications could contribute to the field.

The concepts presented in the present manuscript are explained in more detail in technical books and papers (Oppenheim, 2011; Platt and

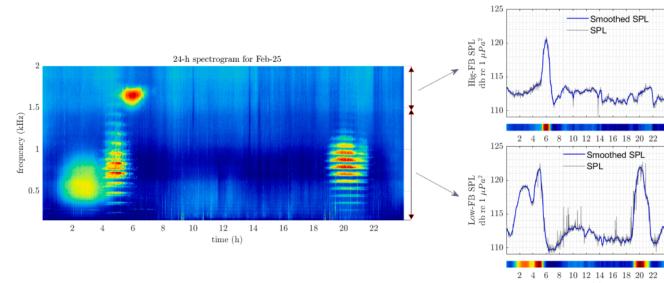


Fig. 16. Left panel shows the 24-h spectrogram for February 25. Right panels represent Low-FB SPL (bottom panel) and High-FB SPL (top panel), also for February 25. For acoustic data recorded over several days, each 24-h SPL is converted to a colorbar, the colorbars for each day can be stacked chronologically, then resulting in a SPL-gram construction.

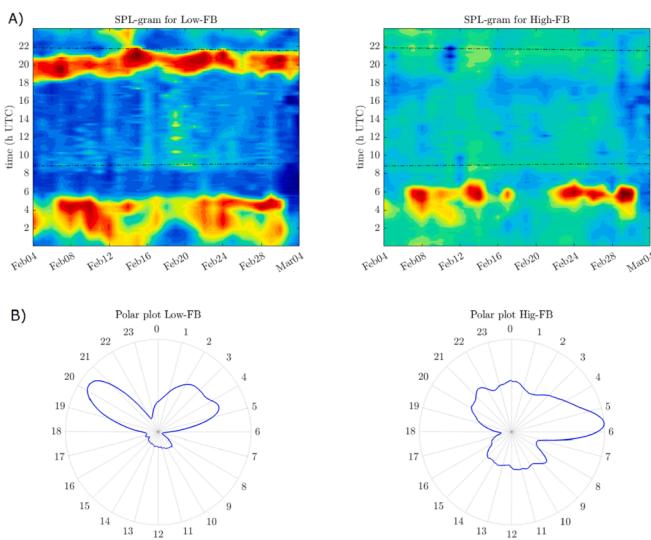


Fig. 17. A) SPL-gram for continuous underwater sound recordings between February 04 to March 04 collected in Xixova State Park, black dashed lines represent sunrise and sunset. B) Polar plot representation for normalized mean Low-FB SPL and Hig-FB SPL.

(Denman, 1975). However, the main objective of this work was to present the concepts in a comprehensible way, by means of visualizations and straightforward explanations focused for the specific area of Ecoacoustics. The concepts presented were used for application on real sound data, composed by 1-month underwater recordings, also an example for short-term analysis was illustrated. Here was demonstrated the utility of tools such as 24-h spectrograms and mean 24-h spectrogram for visualizing daily variations and for summarizing soundscape patterns. Also, the use of metrics such as SPL results useful for quantifying periodicities that could be visually perceived on spectrogram representations. The SPL-gram evidenced to be useful as soundscape visualization tool, allowing a graphical representation for acoustic patterns summarization. Finally, normalized SPL and its representation in polar form showed to complement the comprehension and insights obtained from 24-h mean spectrogram and SPL-gram.

Declaration of Competing Interest

None.

Acknowledgments

The author thanks the São Paulo Research Foundation (FAPESP) [grant number 2016/021750] for supporting the research that allowed the recordings of data presented in the study. The author also thanks Linilson Padovese for the relevant discussions and contributions that enabled the development of the current manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecoinf.2021.101445>.

References

- About GNU Octave. <https://www.gnu.org/software/octave/about> (accessed Jul. 17, 2021).
- Alkin, O., 2016. *Signals and Systems*. CRC Press.
- Bush, E.R., et al., May 2017. Fourier analysis to detect phenological cycles using long-term tropical field data and simulations. *Methods Ecol. Evol.* 8 (5), 530–540. <https://doi.org/10.1111/2041-210X.12704>.
- Cabella, B., Meloni, F., Martinez, A.S., Apr. 2019. Inadequate sampling rates can undermine the reliability of ecological interaction estimation. *Math Comput. Appl.* 24 (2), 48. <https://doi.org/10.3390/mca24020048>.
- Caldas-Morgan, M., Alvarez-Rosario, A., Rodrigues Padovese, L., Jun. 2015. An autonomous underwater recorder based on a single board computer. *PLoS One*, vol 10 (6), e0130297. <https://doi.org/10.1371/journal.pone.0130297>.
- Cooley, J., Lewis, P., Welch, P., 1967. Historical notes on the fast Fourier transform. *IEEE Trans. Audio Electroacoust.* 15 (2), 76–79. <https://doi.org/10.1109/TAU.1967.1161903>.
- Eaton, J.W., 2012. GNU octave and reproducible research. *J. Process Control* 22 (8), 1433–1438. <https://doi.org/10.1016/j.jprocont.2012.04.006>.
- Erbe, C., 2011. Underwater Acoustics: Noise and the Effects on Marine Mammals. A Pocket Handbook [Online]. Available. http://oalib.hlsresearch.com/PocketBook_3rded.pdf.
- Farina, A., Gage, S.H., 2017. *Ecoacoustics. The Ecological Role of Sounds*. John Wiley and Sons.
- Gupta, S.N., 2011. Mean, median, mode: an introduction. In: *International Encyclopedia of Statistical Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 788–791.
- Hunter, W.G., 1981. Simple Statistics for Interpreting Environmental Data. *Water Pollut. Control Fed* 53 (2), 167–175 [Online]. Available: <http://www.jstor.org/stable/25041049>.
- Lyons, R.G., 2004. The discrete Fourier transform. In: *Understanding digital signal processing*, 3rd ed. Pearson Education India, pp. 59–135.
- Merchant, N.D., et al., 2015. Measuring acoustic habitats. *Methods Ecol. Evol.* 6 (3), 257–265. <https://doi.org/10.1111/2041-210X.12330>.
- Moca, V.V., Bärzan, H., Nagy-Dábácan, A., Mureşan, R.C., Dec. 2021. Time-frequency super-resolution with superlets. *Nat. Commun* 12 (1), 337. <https://doi.org/10.1038/s41467-020-20539-9>.
- Nagar, S., 2018. *Introduction to Octave*. Apress, Berkeley, CA.
- Oppenheim, A.V., 2011. *Discrete-Time Signal Processing*, 3rd ed. Pearson.
- Pijanowski, B.C., Farina, A., Gage, S.H., Dumyahn, S.L., Krause, B.L., 2011a. What is soundscape ecology? An introduction and overview of an emerging new science. *Landscape Ecol.* 26 (9), 1213–1232. <https://doi.org/10.1007/s10980-011-9600-8>.
- Pijanowski, B.C., et al., 2011b. Soundscape ecology: the science of sound in the landscape. *Bioscience* 61 (3), 203–216. <https://doi.org/10.1525/bio.2011.61.3.6>.
- Platt, T., Denman, K.L., 1975. Spectral analysis in ecology. *Annu. Rev. Ecol. Syst.* 6 (1), 189–210. <https://doi.org/10.1146/annurev.es.06.110175.001201>.
- Proakis, J.G., Manolakis, D.K., 2007. *Digital Signal Processing: Principles, Algorithms, and Applications*, 4th ed.
- Sánchez-Gendriz, I., Padovese, L.R., Apr. 2016. Underwater soundscape of marine protected areas in the south Brazilian coast. *Mar. Pollut. Bull.* 105 (1), 65–72. <https://doi.org/10.1016/j.marpolbul.2016.02.055>.
- Sánchez-Gendriz, I., Padovese, L.R., 2017a. Temporal and spectral patterns of fish choruses in two protected areas in southern Atlantic. *Ecol. Inform.* 38, 31–38. <https://doi.org/10.1016/j.ecoinf.2017.01.003>.
- Sánchez-Gendriz, I., Padovese, L.R., 2017b. A methodology for analyzing biological choruses from long-term passive acoustic monitoring in natural areas. *Ecol. Inform.* 41, 1–10. <https://doi.org/10.1016/j.ecoinf.2017.07.001>.
- Scharlemann, J.P.W., et al., 2008. Global data for ecology and epidemiology: A novel algorithm for temporal fourier processing MODIS data. *PLoS One* 3 (1). <https://doi.org/10.1371/journal.pone.0001408>.
- Semmlow, J., 2018. Signal analysis in the frequency domain—implications and applications. In: *Circuits, Signals and Systems for Bioengineers*. Elsevier, pp. 169–206.
- Shumway, R.H., Stoffer, D.S., 2017. Characteristics of time series. In: *Time Series Analysis and Its Applications*. Springer International Publishing, pp. 1–44.
- Smith, Steven W., 2002. *Digital Signal Processing: A Practical Guide for Engineers and Scientists*. Elsevier.
- Smith, J.O., 2011. *Spectral Audio Signal Processing*. W3K Publishing.
- Sueur, J., Farina, A., Dec. 2015. Ecoacoustics: the ecological investigation and interpretation of environmental sound. *Biosemiotics* 8 (3), 493–502. <https://doi.org/10.1007/s12304-015-9248-x>.
- Welch, P., Jun. 1967. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* 15 (2), 70–73. <https://doi.org/10.1109/TAU.1967.1161903>.