Advanced signal processing methods for the analysis of transient radiated noise from submarines

Thomas LEISSING[1]; Christian AUDOLY[1]; Hélène LACHAMBRE[2]; Guillaume STEMPFEL[2]

[1] DCNS Research, Toulon, France

[2] Genesis, Aix-en-Provence, France

ABSTRACT

Acoustic discretion is a key performance for many naval vessels such as submarines, in order to prevent underwater detection by adverse sonar systems. In the past decades, acoustic discretion of submarines has been greatly improved, as far as stationary noise is concerned. With the development of new sonar systems, with a capability to detect transient noises, it is now important to treat transient radiated noise as well. As an example, events such as the maneuver of a periscope mast can produce transient underwater noise, in the form of sequence of shocks and other type of noise. The first step toward the evaluation of the risk associated with a transient noise emission is the characterization of these signals. In that context, DCNS has tested some advanced signal processing tools from Genesis: LEA (Logiciel d'Expertise Acoustique), with its new Xtract module dedicated to the automatic separation of the different signal components. Some tests have been performed on signal samples. Several examples involving manual signal edition, automatic denoising and automatic separation of transient and tonal components of a signal are presented. For example, it can be shown that the suppression of background noise allows a better assessment of the spectral analysis of the signal of interest. Moreover, the different signal components, e.g. shocks and tonal components can be separated and treated separately. It is shown that the tools provide a very powerful chain for the treatment of complex signals. Eventually, the use of these tools will allow a better understanding of the transient underwater noise phenomena, allowing defining noise mitigation solutions.

Keywords: Transient noise, underwater noise, signal processing I-INCE Classification of Subjects Number(s): 54.3

1. INTRODUCTION

Transient sonar detection systems have been developed over the last decades, and hence transient noise emissions from submarines have become a matter of concern for acoustic discretion. As an example, events such as the maneuver of a periscope mast can produce transient underwater noise, in the form of sequence of shocks and other type of noise. In order to analyze these transient emissions and evaluate the associated risk it is necessary to be able to characterize these signals. Because of the low radiated noise of submarines measurement of the transient events at sea generally involves low signal-to-noise ratio and most of the time, several events happen in the same time frame, and thus these events must be separated. It is hence necessary to post-process the recorded signals to improve the characterization of these transient events.

This paper is dedicated to the presentation of advanced signal-processing methods and tools developed by Genesis. Two different modules have been tested on transient signals: (i) the first one allows to edit the time-frequency representation of a time signal; (ii) the second one can automatically separate the different components of a complex sequence: noise, tonals and impulsive noises. The different methods and tools used throughout this paper are presented more in detail in Section 2.

These signal-processing tools are used on transient signals originating from a diesel-electric submarine recorded at sea. The measurements performed and the different transient signals selected are described in Section 3. The use of the signal-precessing tools on transient emissions is described in sections 3.3 to 3.6. Finally, conclusions and perspectives are given in Section 4.

[1]thomas.leissing@dcnsgroup.com

[2]helene.lachambre@genesis.fr

2. PRESENTATION OF ALGORITHMS AND TOOLS

2.1 Manual edition of time-frequency representations

LEA is an expert software dedicated to acoustic measurements analysis, sound quality and industrial sound design. Among its numerous functionalities, we will essentially focus on the time-frequency module. It consists in the following analysis/modification/synthesis loop (1):

- Analysis: computation of the time-frequency transform of an audio signal (more precisely a discrete Gabor transform: DGT).
- Modification: done directly on the transform amplitude coefficients. A graphical representation is displayed to the user who can act on the transform with mouse-controlled tools.
- Synthesis: reconstruction of the signal from the modified transform.

After each modification, the user can listen to the reconstructed signal to evaluate the transformation result. If he is satisfied, the user keeps the modification, otherwise, he can cancel it and work again from the pre-modification signal. This try-and-correct process allows to converge quickly to a satisfying result.

Different edition tools are available (the list is non-exhaustive) :

- Isolating or deleting one or several time-frequency areas,
- Add a gain on some time-frequency areas
- Rubbing out some time-frequency areas

In order to improve and facilitate the process of manual time-frequency edition, some additional tools are available. We present here the ones used in this study.

Automatic computation of optimal window size: Based on the selection by the user of an area of interest in the time-frequency representation, this tool automatically computes the optimal window size for a maximization of the energy concentration in this area. Thus, it allows to reduce the spreading of energetic components in the time-frequency representation.

Magic wand: This tool allows the automatic selection of an area of interest around a given time-frequency bins. The region is determined based on relative power levels between selected bin and adjacent ones.

2.2 Automatic separation of components

As an alternative to manual signal processing, the LEA software offers a new module called Xtract. Following previous work (2), this module aims at automatically separating components in sounds, and focuses on tonals and impulsive noises. Two dedicated algorithms are used (one for each component type) which are presented bellow.

After having led a large set of experiments, it turned out that the component separation algorithm performs better on signals without (or only with a weak) stationary background noise. Consequently, we decided to add to our work-flow, as a pre-processing stage, a denoising procedure in order to remove the stationary noise from the input signal. The use of this third algorithm (also presented below) results in the extraction of a third component: stationary noise.

This method provides as outputs three components plus a remainder defined as follows:

- Noise component: noise is the broadband stationary component of the sound.
- Tonal component: tonals are the sinusoidal components of the sound.
- Impulsive component: short and time-localized components of the sound.
- Remainder: the remainder is made of components which do not fit any of the previous models.

2.2.1 Noise extraction/denoising

For this step, we base our approach on the the state-of-art algorithm of (3) which proposes to take into account the neighborhood of the coefficients by using block-thresholding (introduced by (4) in statistics field) directly on time-frequency plane. In the following, we denote $Y[l,k]$ the coefficients of the time-frequency representation of the signal with l the time index and k the frequency index. This strategy computes a single attenuation factor over time-frequency blocks. The time-frequency plane is segmented in a set of B blocks termed $B_i, \in \mathbb{N}, i \leq B$. The segmentation is chosen arbitrarily. The denoised signal estimator is computed from noisy data with a constant attenuation factor a_i over each block B_i.

The attenuation is computed with a power subtraction estimator in a block B_i:

$$a_i = \max(0, 1 - \frac{\lambda}{1+\hat{\xi}_i}) \quad (1)$$

where λ is the power subtraction parameter (defined later) and ξ_i is the estimated average SNR in B_i, which corresponds to the ratio between the average signal energy and the average noise energy in B_i:

$$\hat{\xi}_i = \frac{\sum_{(l,k)\in B_i} |Y[l,k]|^2}{\sum_{(l,k)\in B_i} \sigma[l,k]^2} - 1 \tag{2}$$

with σ the noise variance.

In order to practically compute the attenuation factors, two parameters must be set: the parameter λ, and the blocks segmentation of the time-frequency plane. Both are related to the variance of the estimation risk. Roughly (refer to (3) for more details), the λ parameter is computed statistically in order to limit the probability of keeping pure noise coefficients and then of making the remaining musical noise inaudible (due to isolated energetic time-frequency bins). For choosing the block size among several possibilities, an adaptive strategy is used, in order to minimize a Stein risk estimate (5), which is an estimator of the expectation of the distance between the denoised signal and the real non-noisy signal.

One of the drawbacks of this block thresholding technique is the fact that it nevertheless outputs a time-frequency diagram with some block structures. They propose a post-processing which consists in using this first estimation as an input to compute a second Wiener time-frequency estimation.

Let $\hat{F}$ be the denoised time-frequency output from the block thresholding algorithm. Then, the new attenuation factor computed during the post-processing is equal to:

$$\tilde{a}[l,k] = \frac{|\hat{F}[l,k]|^2}{|\hat{F}[l,k]|^2 + \sigma[l,k]^2} \tag{3}$$

This new attenuation factor is applied on the noisy time-frequency coefficients to reconstruct the final denoised signal.

Automatic Noise Estimation: From equations 1 and 2, it appears that one needs to know the noise variance over each time-frequency bin ($\sigma[l,k]^2$). In fact, if the noise is stationary, one needs to know its variance over each frequency bin ($\sigma[k]^2$).

In the best case, the noise spectrum is known, otherwise it has to be estimated. A noise estimation may easily be done if a noise sample signal exists. However, in many cases, getting an isolated noise sample is not possible. We therefore propose an automatic estimation of the noise, based on the histogram-based noise estimation algorithm of (6).

Considering that the noise is the stationary part of the signal, the most frequent value of energy in a given frequency band corresponds to the noise energy in this band. A drawback of this method appears if the signal contains isolated constant tones lasting long enough, which will be considered as noise. However, in our approach, we consider that these tones are part of the tonal component. In order to avoid this, we apply a median filtering to remove thin peaks from the noise spectrum.

The automatic noise estimation is summarized as follows:

1. Compute a time-frequency representation of the noisy signal.
2. For each frequency bin, find the maximum of the energy histogram (i.e. the most frequent value of energy).
3. Smooth the resulting power spectrum with a median filter to remove isolated tones.

2.2.2 Tonal extraction

As we define tonals as "frequency localized components, lasting long enough with a smooth evolution over time", we use a partial tracking algorithm to perform the extraction, based on the work of Ellis (7).

Partial tracking: The partial tracking processes the spectrogram by

- picking up the emerging frequency peaks in each time frame (peaks picking step),
- linking them over consecutive frames, therefore building "tracks" (linking step). In order to take into account the trajectory of each track, we predict its frequency position in next frame (8).

Finally some conditions are checked to discard dummy or false-positive tracks (post-processing step).

As for many partial tracking algorithms, this method has the huge drawback that it has too many parameters (over 20) to be usable by non-specialist users. The major work we did is the reduction of the number of parameters. Considering various industrial sounds, we end up with a set of four parameters from the original algorithm from Ellis, to which we add a fifth one.

These five parameters are the following ones, with the most important ones listed first:

- Frequency regularity of the tracks: new parameter. During the post-processing step, tracks for which the frequency evolution over time is too erratic are discarded.
- Maximum frequency slope of the tracks: from Ellis algorithm. During the linking step, peaks from consecutive frames are considered to belong to the same track if they are close enough in frequency.

- Minimum duration of the tracks: from Ellis algorithm. During the post-processing step, tracks shorter than the minimum duration are discarded.
- Minimum inter-tonal gap between tracks: from Ellis algorithm. During the peak picking step of the partial tracking, the selected peaks are farther one from each other than this parameter.
- FFT size: from Ellis algorithm. This parameter corresponds to the FFT size used for the time-frequency analysis of the signal. Its value should be set so that the tonals show best in the time-frequency representation. Its value is precomputed from the signal frequency sampling, but may still be tuned.

Partial extraction: Once the tracks positions are known in the time-frequency representation, we have to extract them. Two major issues appear: what is the frequency width of each track? How much do they emerge from other components?

For a given track, the width is supposed to be constant over time. At each time step, we estimate the instantaneous width as the distance between the peak and the first local minimum of the power spectrum. The global width of the track is then the median of these values.

A track may not be completely removed (frequency bin set to zero) without creating audible musical noise. We therefore have to estimate the level of adjacent components in order to remove only the emerging part. This level is computed as the mean of the power on the adjacent bins.

2.2.3 Impulsive components extraction

The impulsive part of the signal contains all time-localized, high-amplitude and broadband events, typically impacts or shocks. To perform the extraction of such components, we strongly follow the lines of (9), where a sparse regression strategy is proposed.

The idea is to split the signal into three parts: an impulsive part, a high-amplitude remainder part and a low-amplitude remainder part. To perform this split, the signal is decomposed considering two bases: the first one $\mathscr{D}_1$ is adapted to impulsive components, or transient layer (typically a DGT using a short window in order to capture short events), and the second one $\mathscr{D}_2$ is more adapted to frequency localized events (here, a DGT with a longer analysis window is used). Hence, the impulsive components will be concentrated in the first decomposition, and spread over several time-frames in the second one.

Operations are led in two steps: the first one consists in removing from the signal the high-amplitude remainder components, the second in extracting the impulsive components.

To begin, we isolate high-amplitude remainder elements using $\mathscr{D}_2$ by applying an iterative soft-thresholding algorithm to the time-frequency plane. The thresholding is made by taking into account the magnitude only, while the phase of the signal is ignored. For this step, we turn to the state-of-the-art LASSO algorithm (10) which makes use of ℓ_1 norm for regularization to enforce the sparsity of the extracted signal (which means that this signal contains only a few non-zero coefficients in the time-frequency plane). Then, only the most powerful coefficients of the signal will be removed, and will constitute the high-amplitude remainder. Note that if impulsive elements are far more energetic than other elements in the signal, then this step should be skipped.

Next, the second step is performed starting from the original signal, from which the high amplitude remainder has been removed. Here, the aim is to isolate impulsive elements from low-amplitude remainder, for example low background noise, using the basis $\mathscr{D}_1$. As for the first step, we use a soft-thresholding algorithm. But, in order to take in account the particular structure of impulsive components, instead of using well-known ℓ_1 or ℓ_2, we use mixed-norms (11) for regularization. Mixed-norms (as in Group-LASSO, see (12)), allow to instore group-sparsity. In other words, the thresholding will be made by keeping the integrity of predefined groups in the time-frequency plane. If the groups are chosen with respect to a generic structure, then the thresholding will lead to the isolation of impulsive components. The remaining part will make up the low-amplitude remainder part.

3. APPLICATION TO THE ANALYSIS OF TRANSIENT NOISES FROM SUBMARINES

3.1 Presentation of test signals used

Signals were acquired during tests at sea of a diesel-electric submarine (SSK). Signals were recorded on a set of two omnidirectional hydrophones with the submarine at stabilized immersion and at intermediate speed. Several passes were recorded at two different immersions (periscope and transit immersion) with different events occurring when the submarine passes the closest point of approach (CPA) from the hydrophones. Four different recordings were used in this paper; they are summarized in Table 1 below.

As an example, Figure 1 shows the time-frequency representation of recording A. Except otherwise mentioned, time-frequency representations are obtained with a FFT size of 1024 points, a Hanning window and a 50% overlap. Note that all absolute levels given below are arbitrary.

Table 1 – Identification of recordings used.

Name	Situation	Events
A	Periscopic depth	Periscope hoisting
B	Periscopic depth	Optronic mast hoisting and lowering
C	Transit depth	–
D	Transit depth	Torpedo tubes opening

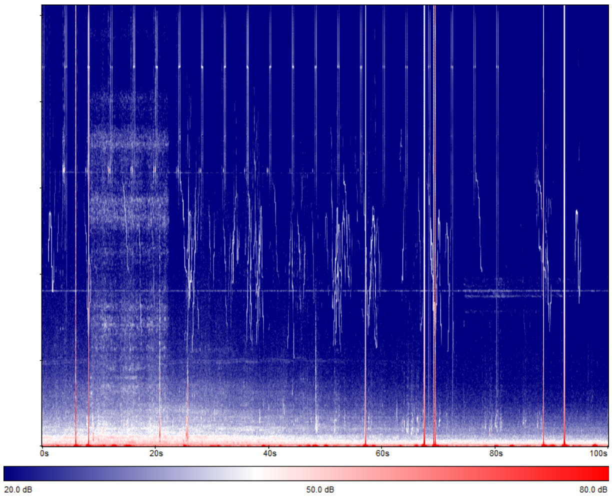


Figure 1 – Example of time-frequency representation of a complete recording (recording A).

3.2 Case studies considered

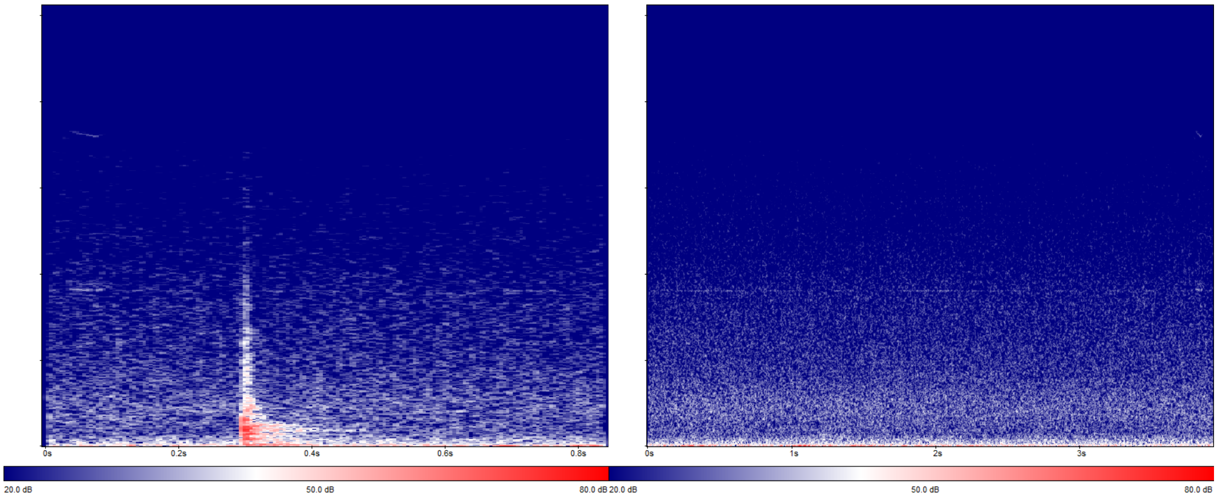
Several events have been extracted from these recordings and are used as case studies for the signal-processing tools presented in this paper. These events vary by their length and frequency content. The first test case (see Section 3.3) presents a single shock in background noise. In the second test case (see Section 3.4) cavitation noise is studied. Cavitation is the formation of vapor cavities in water and occurs for high propeller rotation speed, when the liquid is subjected to rapid changes of pressure that cause the formation of cavities where the pressure is relatively low. Besides damaging the propeller, cavitation noise can be a source of acoustic indiscretion. The third test case presented in Section 3.5 exhibits a low-frequency whistling noise. The fourth and last case studied (see Section 3.6) is a more complex signal, where the sequence shows different mixed events. The sequence length is 5 seconds; it shows two shocks with a constant frequency signal in between the shocks. For each test cases, the use of the different tools presented in Section 2 is illustrated.

3.3 Single shock

Figure 2a shows the time-frequency representation of a shock extracted from recording B. We try to extract this shock from the noisy sequence, by eliminating the background noise as far as possible, without distorting the spectral content of the shock. Two different methods have been used: (i) by using LEA, by selection of zones in the time-frequency representation in a semi-automatic or manual way and (ii) by using the automatic denoising algorithm.

LEA allows processing the signal directly on the time-frequency representation. A "magic wand" tool allows performing automatic selection of zones: the point selected as well as the neighboring points having a level lower than a specified tolerance are selected. It is also possible to perform manual selection of areas in the time-frequency representation. These tools thus allow to isolate an event of a signal. The operation can however turn difficult if the signal to be extracted is of complex structure. In the case of a shock, which clearly emerges from the background noise, the extraction is however relatively simple. Figure 3a shows the manually extracted shock. It can be seen that the manual edition of the time-frequency representation allows to isolate an event in a signal when it clearly emerges from the background noise. The difficulty lies in the extraction of the high-frequency content, which level is of the same order of magnitude as the background noise. It can be seen by comparing Figure 2a and Figure 3a that the high-frequency shock content is lost. This tool thus allows to extract quickly certain emergent events but must be used with caution in the case of low signal-to-noise ratio events.

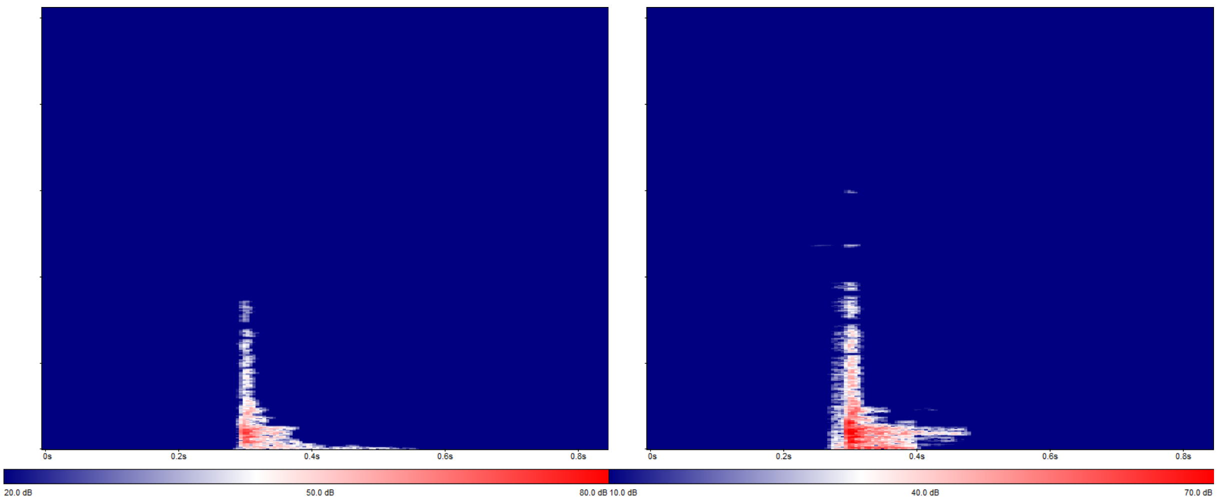
We next perform the same operation by using the automatic denoising tool presented in Section 2.2.1.



(a) (b)

Figure 2 – Time-frequency representation of the single shock (2a), and noise sample used for denoising (2b).

The automatic denoising uses as an input a noise sample, which must be sufficiently representative of the noise occurring during the event to be denoised. Figure 2b presents the time-frequency representation of the noise sample used; its length is about 4 s and it has been extracted from the same recording (recording B), a few seconds before the shock occurred. Figure 3b shows the time-frequency representation of the automatically denoised shock. Compared to the manual extraction technique (Figure 3a), it can be seen that the high-frequency content of the signal is still present in the denoised shock. This can be observed in Figure 4b, which compares the third-octave band spectrum of the original signal (in blue) with the spectrum of the manually extracted (in green) and automatically denoised signal (in red): the high-frequency information is lost in the case of the manually extracted signal.



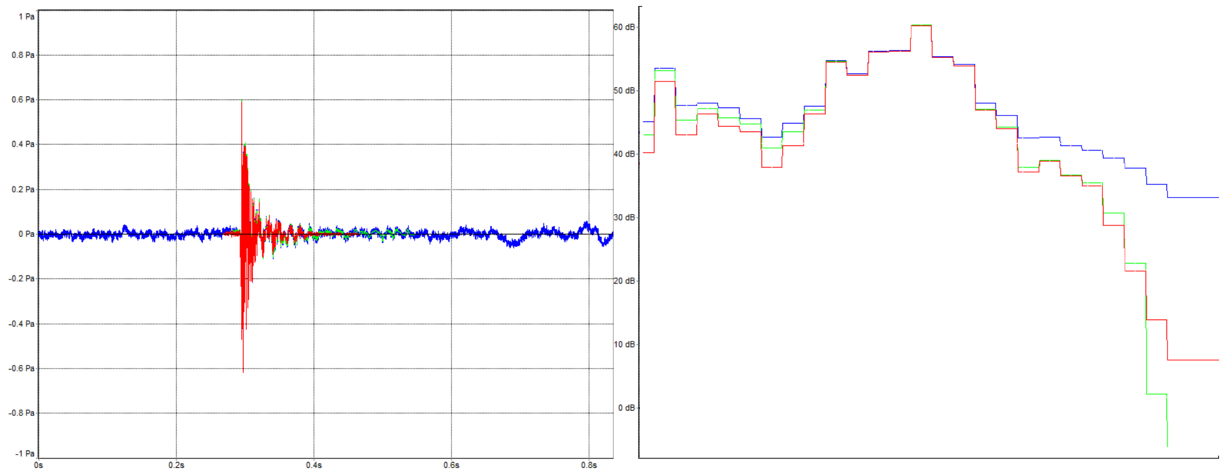
(a) (b)

Figure 3 – Time-frequency representation of the manually extracted shock (3a), and automatically denoised shock (3b).

3.4 Cavitation noise

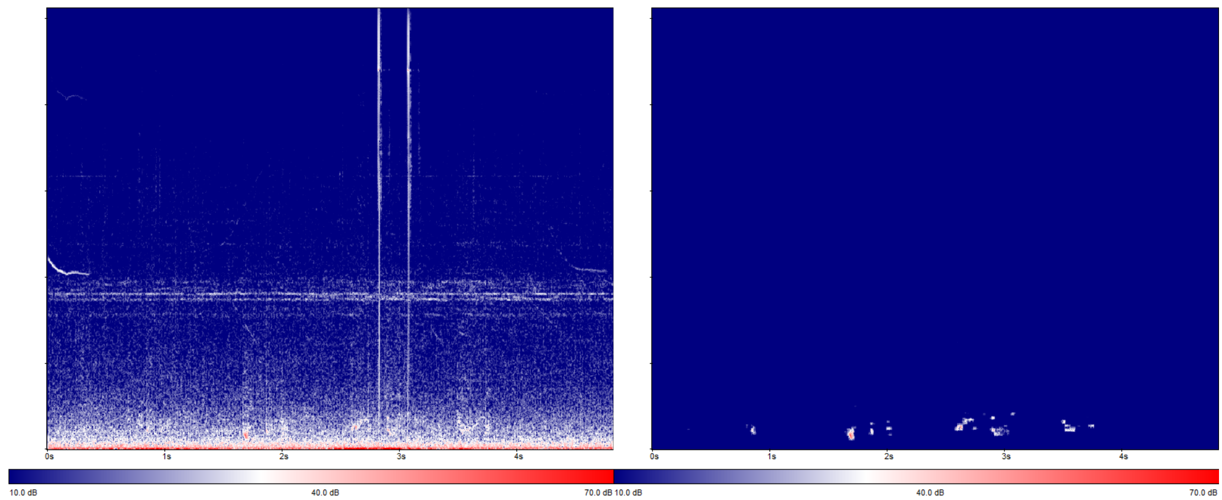
Figure 5a shows the time-frequency representation of a sequence containing cavitation noise extracted from recording A. The sequence length is approximately 4.5 s and shows acoustic events between $t = 1.7$ s and $t = 3.8$ s. These cavitation bursts weakly emerge from the background noise. In this test case it is impossible to use the manual selection technique as used previously, since the level of the events to be extracted is too low.

Once the original signal has been automatically denoised, it is then possible to use the manual edition of time-frequency representation tool to suppress unwanted events. Figure 5b shows the time-frequency representation



(a) (b)

Figure 4 – Time-domain (4a) and third-octave band spectrum (4b) representations of the original signal (blue), the manually extracted signal (green) and the automatically denoised signal (red).



(a) (b)

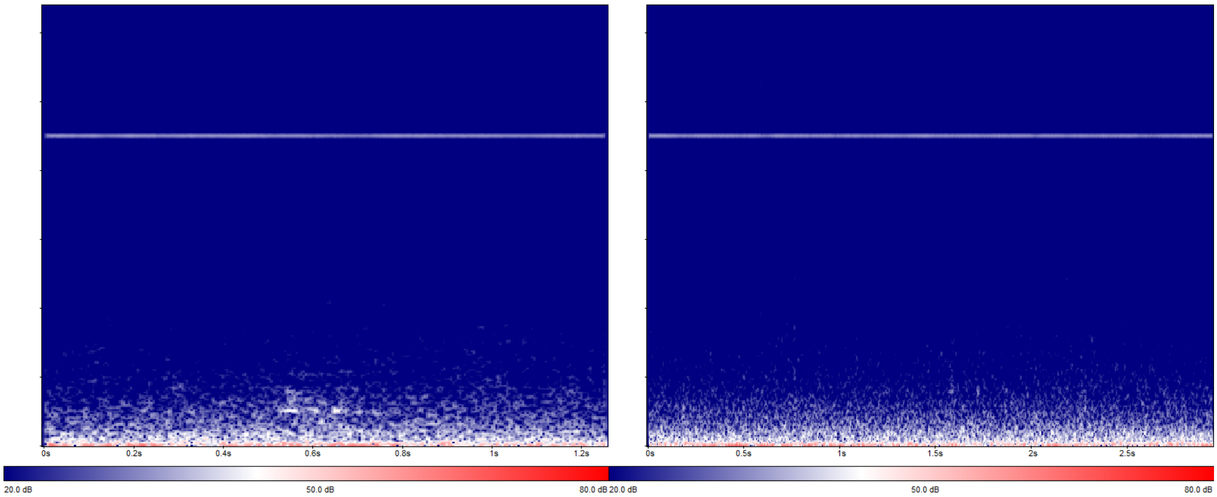
Figure 5 – Time-frequency representation of the original cavitation noise (5a), and automatically extracted events (5b).

of cavitation noise, after automatic denoising and manual edition of the time-frequency representation. Cavitation bursts are clearly visible, and have been successfully isolated from the background noise and other acoustic events.

3.5 Low-frequency whistling

Figure 6a shows the time-frequency representation of a low-frequency whistling noise extracted from recording C. This whistling is short and contains a rather stable frequency content of relatively low frequency. Note that the level of the whistling is of the same order of magnitude as the background noise. We use a two step procedure to extract the whistling from the signal. We first use a noise sample extracted from the same recording a few seconds before the whistling as an input to the automatic denoising algorithm presented in Section 2.2.1. The noise sample is presented in Figure 6b. The automatic tonal components extraction method presented in Section 2.2.2 is then used on the denoised signal.

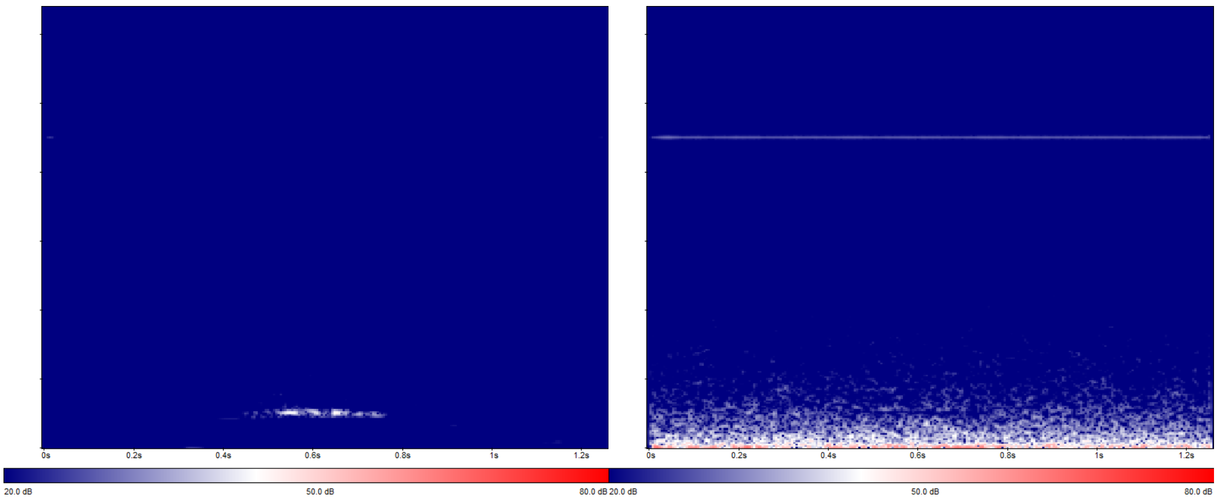
The output from the denoising and tonal extraction algorithms is show on Figure 7: the extracted tonal component is shown on the left (Figure 7a), while the remaining noise is shown on the right (Figure 7b). Note that the continuous high-frequency component on Figures 6a, 6b and 7b is artificial: it comes from the telemetry system, allowing the submarine to determine its exact location with respect to the recording device. It should also be noted that as this component is included in the noise sample used for denoising (see Figure 6b),



(a) (b)

Figure 6 – Time-frequency representation of the low-frequency whistling (6a), and noise sample used for denoising (6b).

it is suppressed from the signal *before* tonal extraction is performed. It hence appears in the remaining noise (Figure 7b) and not in the extracted tonal components (Figure 7a). It can be seen by comparing Figure 6a and Figure 7a that the low-frequency whistling noise is successfully extracted from the background noise.



(a) (b)

Figure 7 – Time-frequency representation of the extracted tonal component (7a), and remaining noise (7b).

3.6 Complex signal

The last test case is a complex signal mixing acoustic events of different types: background noise, two shocks at the beginning and the end of the sequence, and in between the two shocks a harmonic signal. Figure 8 shows the time-frequency representation of this signal extracted from recording D. We seek to separate the impulsive events from the tonal components. We use in this test case the denoising algorithm followed by an automatic separation of tonal and impulsive components (see Section 2.2.1 to Section 2.2.3).

Figure 9 shows the results from the components separation: Figure 9a shows the extracted tonal component, Figure 9b shows the extracted impulsive components and Figure 9c shows the remaining noise. Note that the addition of these three components gives back the original signal. Similar to the previous test case, the harmonic signals appear in the remaining noise. This is due to the fact that the noise sample used for denoising contains these harmonics.

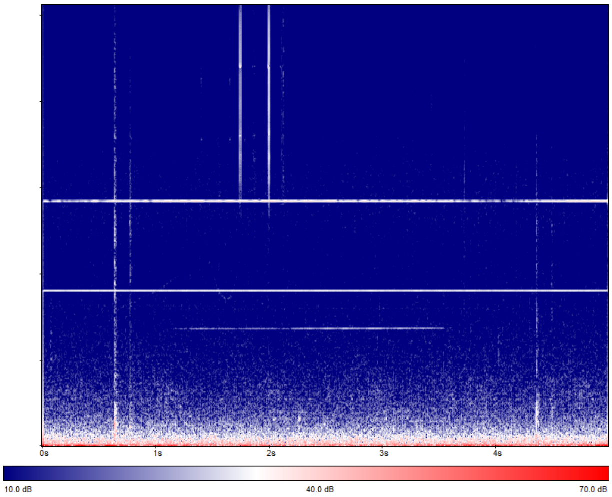
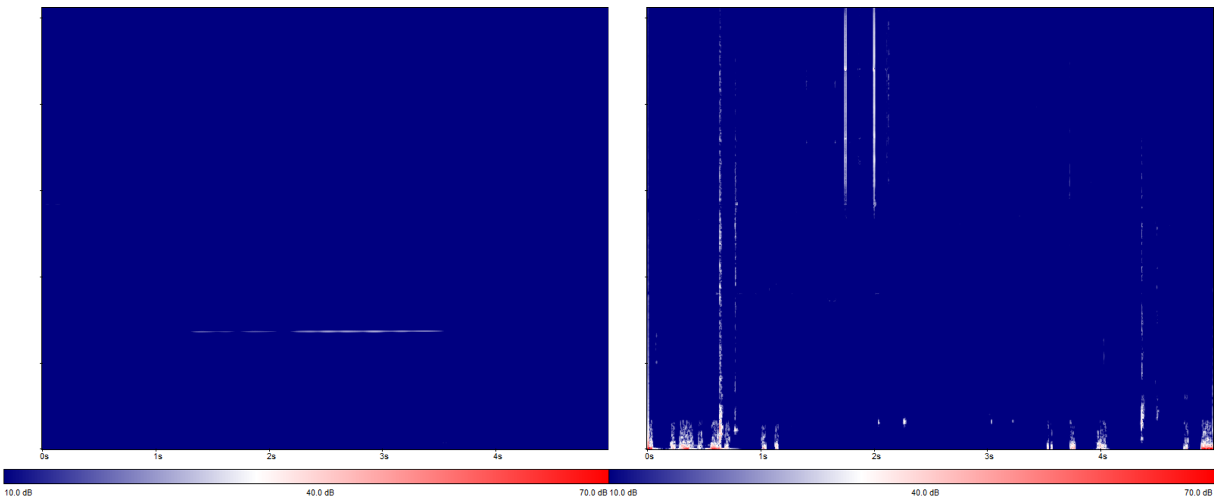
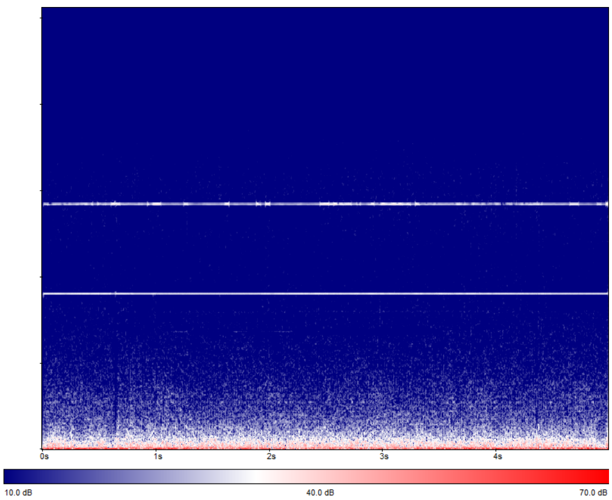


Figure 8 – Time-frequency representation of the complex signal.



(a) (b)



(c)

Figure 9 – Time-frequency representation of the extracted tonal (9a) and impulsive (9b) components, and remaining noise (9c).

4. CONCLUSIONS AND PERSPECTIVES

Due to the improvement of transient noise detection means, it is necessary to treat transient radiated noise as a source of indiscretion as well. Events such as shocks or cavitation noise can produce transient underwater emissions that must be mitigated as far as possible. In the objective of the assessment of the risk associated to a particular acoustic event, one has to be able to extract the transient noise under study from the background noise and other acoustic events. This paper has presented and illustrated post-processing tools that help performing this task. The different methods and algorithms used in this paper have been presented in Section 2. Two categories of tools have been used: (i) manual edition of time-frequency representations and (ii) automatic denoising and separation of impulsive and tonal components.

The use of these tools has been illustrated on test cases extracted from four recordings at sea of a diesel-electric submarine. The acoustic events under study vary by their frequency content and their type (shocks, bursts, whistling). In the case of the shock, it has been shown that the manual edition of time-frequency representation provides a quick and efficient tool for signal extraction. This is made possible by the high signal-to-noise ratio of this event. It was shown that the suppression of background noise allows a better assessment of the spectral analysis of the signal of interest. The efficiency of the denoising algorithm was next illustrated on cavitation noise. Cavitation bursts have been successfully extracted from the background noise. Next, the chaining of these post-processing tools has been demonstrated with two additional test-cases: extraction of a low-frequency whistling noise, and extraction of stationary and impulsive components from a complex signal. Eventually, the use of these tools will allow a better understanding of the transient underwater noise phenomena, allowing defining noise mitigation solutions.

ACKNOWLEDGMENTS

Genesis LEA-Xtract algorithms development has been partially funded by the EU FP7 Collaborative project BESST (grant agreement no. 233980), by the FET-Open European project SkAT-VG (grant agreement no. 618067) and by the FET-Open European project UNLocX (grant agreement no. 255931). These algorithms result from a long-term collaboration with Professor Bruno Torresani (I2M, Marseille).

REFERENCES

1. Jaillet F. Représentation et traitement temps-fréquence des signaux audionumériques pour des applications de design sonore. Université de la Méditerranée; 2005. In french.
2. Molla S. Signaux audiophoniques : modélisation hybride et de schéma de codage. Université de Provence; 2003. In french.
3. Yu G, Mallat S, Bacry E. Audio Denoising by Time-Frequency Block Thresholding. IEEE Trans On Signal Processing. 2008;56(5):1830–1839.
4. Cai T, Silvermann BW. Incorporating Information on Neighboring Coefficients into Wavelet Estimation. Sankhya. 2001;63:127–148.
5. Stein C. Estimation of the mean of a multivariate normal distribution. Annals of Statistics. 1981;9(6):1135–1151.
6. Fukane AR, Sahare SL. Noise Estimation Algorithms for Speech Enhancement in highly non-stationnary Environments. Int Journal of Computer Science Issues. 2011;8(2):39–44.
7. Ellis DPW. Sinewave and Sinusoid+Noise Analysis/Synthesis in Matlab; 2003. web resource: http://www.ee.columbia.edu/ dpwe/resources/matlab/sinemodel.
8. Lagrange M, Marchand S, Rault JB. Enhancing the Tracking of Partial for the Sinusoidal Modeling of Polyphonic Sounds. IEEE Trans On Audio Speech and Language Processing. 2007;15(5):1625–1634.
9. Kowalski M. Approximation des signaux: approches variationnelles et modéles aéatoires. Université de Provence; 2008. In french.
10. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society (Series B). 1996;58:267–288.
11. Benedek A, Panzone R. The space ℓ^p with mixed norm. Duke Mathematical Journal. 1961;28:301–324.
12. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society, Series B. 2006;68:49–67.