A blind denoising process with applications to underwater acoustic signals

Philippe Courmontagne[1]
[1]ISEN Toulon - IM2NP
UMR CNRS 6242
Equipe Signaux et Systèmes
Maison des Technologies, Place George Pompidou
83000 TOULON
FRANCE
Email: <firstname.lastname>@isen.fr

Samir Ouelha[1;2], Ugo Moreaud[1;2], Fabien Chaillan[2]
[2]DCNS/UWS Dpt.
Acoustic Signature R&D
BP 403 Le Mourillon
83055 TOULON Cedex FRANCE
Email: <firstname.lastname>@dcnsgroup.com

Abstract **- In many applications, such as transient classification, underwater communication and positioning system …, it could be of great interest to reduce the noise level of the received signal before processing. In most cases, the algorithms dedicated to this problem either required some *a priori* knowledge about the useful signal and/or noise characteristics, either made some not always verified assumptions about the disturbing term statistics. The purpose of this article is to present a new blind denoising approach, without any considerations about the signal of interest and the noise. Results obtained on real and synthetic data are proposed and discussed.**

I. Introduction

In many applications, such as acoustic communication, detection and localization of surface and subsurface objects, depth sounders, sub-bottom profilers …, the identification and recognition of acoustic signals has become an important issue. But, because underwater acoustic signals are affected by ambient noise, which sources are both natural and human-made, it appears of great interest to remove the noise so as to enhance the useful signal before recognition. Doing so, overall performance of underwater acoustic instruments or dedicated processing (such as underwater transient classifier) can be improved.

In the scientific literature, there are several solutions to reduce the noise level of underwater acoustic signals, depending on the knowledge about the useful signal and/or the disturbing terms. Some algorithms require an *a priori* knowledge about the signal and noise second order statistics (such as the Stochastic Matched Filter [1]), others try to deal with the only knowledge of the noise spectral power density (as an example: the short-time Wiener filtering [2]) or make the assumption of a Gaussian noise (i.e. the ambient noise is assumed to have its probability density function equal to that of the normal distribution), in order to use some wavelet based denoising approaches, that rest on the use of thresholding steps [3]. But what happens, if no information about the signal and the noise are available? Moreover, as denoted in several articles [4-5], natural and manmade sources, such as reverberation, industrial noise …, that cause additive noise distribution, exhibit performances far away from the Gaussian model. For this reason, the disturbing signal has to be considered as a realization of a non-Gaussian process. Some authors have proposed to describe the noise as a Gaussian-Gaussian mixture [6], others [7] have considered the noise distribution as an asymmetric generalized Gaussian probability density function …. In fact, depending on the source of the background noise, several noise models could be considered. Anyway, the denoising methods based on the ideal assumption of Gaussian noise may decay, as the background noise is evidently non-Gaussian. Taking the previous remarks into account, the purpose of this article is to propose a new blind denoising method, which means an automatic denoising process available without any consideration about the signal and noise statistics.

In some previous works [8-9], we have proposed a new time-frequency representation (the *Hearingogram*) of underwater acoustic signals, in order to take into account the human physiology and in the purpose of transient classification. In particular, we have proposed a new process allowing both to denoise and to construct this time-frequency plane. In this paper, we propose to take advantage of this denoised *Hearingogram* to reconstruct an approximation of the useful signal contained in the recorded underwater acoustic data.

In the first part of this paper, we recall the denoised *Hearingogram* principle. Next, in the second section, we describe the way to obtain an approximation of the useful signal from this time-frequency plane. In section three, we present and discuss some results obtained on synthetic and real

978-0-933957-40-4 ©2013 MTS

data. We finish this paper with some concluding remarks.

II. The denoised *Hearingogram*

1. The *Hearingogram* principle

In order to take into account the human physiology, we have proposed in [8], to construct a time-frequency plane, using the Mel's filters [10]. These filters are non-uniformly spaced on the frequency axis, with more filters in the low frequency regions and less filters in high frequency regions. They are asymmetric triangular shaped filters bank in respect to the Mel's scale, defined as follows, for a given frequency ν in Hz:

$$mel(\nu) = 2595 \log_{10}\left(1 + \frac{\nu}{700}\right).$$

If we consider M Mel's filters, $H_m(\nu)$ (with m taking values between 1 and M), each of them is centered on a frequency ν_m and has a width $B(m)$ defined as follows:

$$B(m) = \begin{cases} \nu_2 - mel(\nu_{min}), & m = 1 \\ \nu_{m+1} - \nu_{m-1}, & m = 2, \dots, M-1 \\ mel(\nu_{\max}) - \nu_{M-1}, & m = M \end{cases}$$

where ν_{max} and ν_{min} corresponds to the highest and the lowest components of the studied signal, respectively (generally $F_s/2$ and 0, where F_s denotes the sampling frequency). Moreover, the center frequencies ν_m are determined according to:

$$\nu_m = 700\left(10^{\frac{mel(\nu_m)}{2595}} - 1\right),$$

with:

$$mel(\nu_m) = \frac{m}{M+1}\left(mel(\nu_{max}) - mel(\nu_{min})\right).$$

By the way of the impulse responses $h_m[k]$ (described in [8]), with k taking values between 1 and M_{h_m} (M_{h_m} corresponding to the number of samples, Mel's filter bandwidth dependent, useful to describe $h_m[k]$), one can access to the *Hearingogram* $\Theta_Z[n,m]$, taking the square magnitude of the convolution product between the studied signal $\boldsymbol{Z}$ and the impulse response $\boldsymbol{h_m}$:

$$\Theta_Z[n,m] = \left(\sum_{k=1}^{M_{h_m}} Z[n-k]h_m[k]\right)^2, \qquad (1)$$

with n taking values between 1 and M_Z, M_Z corresponding to the dimension of the sampled recorded signal $\boldsymbol{Z}$.

2. The denoised *Hearingogram*

As the useful signal is deteriorated by some background noises, we have proposed in [9] to add a denoising step in the *Hearingogram* conception scheme. This one is based on the use of wavelet shrinkage [10]. In fact, the *Hearingogram* principle rests on the use of a convolution product between the received signal and the Mel's impulse responses (as seen in relation (1)), it follows that the disturbing terms in the *Hearingogram* are the result of a linear transformation of a random vector. Moreover, assuming that the smaller Mel's impulse response dimension, which corresponds to the number of samples in $h_M[k]$, is large enough, we can invoke the central limit theorem inducing that the noise contribution, denoted $\boldsymbol{N_{h_m}}$, in the *Hearingogram* is Gaussian:

$$\boldsymbol{N_{h_m}} \hookrightarrow \mathcal{N}\left(0, \sigma^2_{N_{h_m}}\right).$$

Thus, for a Gaussian disturbing signal, it is well known that a robust denoising technique is based on the idea of thresholding the wavelet coefficients. Indeed, wavelet coefficients having small absolute value are considered to encode mostly noise and very fin details of the useful signal, while the important information is encoded by coefficients having large absolute value. Removing the small absolute value coefficients and then reconstructing the data allows to obtain an approximation of the useful signal.

Let $\boldsymbol{Z_{h_m}}$ be the response of the m^{th} Mel's filter to excitation $\boldsymbol{Z}$. For each m value, in the range 1 to M, we apply the Discrete Wavelet Transform (DWT) to $\boldsymbol{Z_{h_m}}$, following the Mallat multiresolution analysis algorithm [11]. Next, we use the hard-thresholding rule to remove the coefficients that are smaller than a threshold and leave the other ones unchanged. Finally, an approximation of the useful signal $\boldsymbol{\tilde{S}_{h_m}}$ in $\boldsymbol{Z_{h_m}}$ is restored applying the inverse DWT to the retained wavelet coefficients. Concerning the threshold value, this one has been calibrated using the *universal* threshold. It is dependent on the scale level p of the multiresolution analysis and is linked to the dimension M_Z of the noise-corrupted data $\boldsymbol{Z}$:

$$\lambda_m[p] = \alpha\sigma_{N_{h_m}}\sigma_\eta^m[p]\sqrt{2\ln M_Z},$$

where:

- α is a parameter used to reinforce the noise reduction;
- $\sigma_{N_{h_m}}$ is the noise standard deviation estimated by the way of the *scaled Median Absolute Deviation* (*MAD*):

$$\sigma_{N_{h_m}} = MAD(\boldsymbol{Z_{h_m}})/0.6745;$$

- $\sigma_\eta^m[p]$ is the standard deviation, at scale level p, of the response of the m^{th} Mel's filter to a white Gaussian noise excitation.

Finally, the denoised *Hearingogram* simply corresponds to the square magnitude of $\tilde{\boldsymbol{S}}_{\boldsymbol{h_m}}$, making vary m from 1 to M.

III. Denoised useful signal reconstruction

If we consider the filter $H^{Mel}(\nu)$ associated to the whole Mel's filter bank, it corresponds to a band-pass filter:

$$H^{Mel}(\nu) = \sum_{m=1}^{M} H_m(\nu) = 1, \forall \nu \in [\nu_1; \nu_M]$$

with $[mel(\nu_{min}); \nu_1[$ and $]\nu_M; mel(\nu_{max})]$ as transition widths.

In order to ensure the energy conservation, it is necessary to add to this filter, two filters, $H_0(\nu)$ and $H_{M+1}(\nu)$, so that:

$$H(\nu) = H_0(\nu) + H_{M+1}(\nu) + H^{Mel}(\nu) = 1,$$

$\forall \nu \in [0; F_s/2]$.

The impulse responses, $h_0[k]$ and $h_{M+1}[k]$, associated to these filters are defined by:

$$h_0[k] = \nu_1 sinc^2(\pi\nu_1 k)/F_s$$

and

$$h_{M+1}[k] = -\frac{1}{F_s} sinc(\pi f_0 t)\big(2\nu_M \cos(\pi f_1 t) + f_1 sinc(\pi f_1 t)\big) - \ldots \frac{1}{2 f_0 F_s}\left(4\nu_M^2 sinc(2\pi\nu_M t) - F_s^2 sinc(\pi F_s t)\right),$$

where:

$$\begin{cases} f_0 = \dfrac{(F_s - 2\nu_M)}{2} \\ f_1 = \dfrac{(F_s + 2\nu_M)}{2} \end{cases}$$

We present, on figure 1, the frequency responses associated to this filter bank, with in black lines the Mel's filters and in red lines $H_0(\nu)$ and $H_{M+1}(\nu)$.

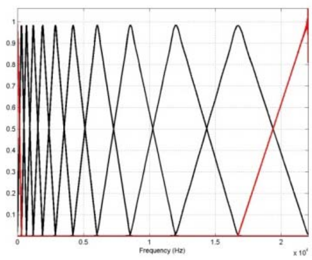

Fig. 1: Filter bank (black lines: Mel's filters, red lines: additional filters to ensure the energy conservation)

In these conditions, the impulse response $\boldsymbol{h}$ associated to this band-pass filter could be approximated by:

$$\boldsymbol{h} = TF^{-1}[H(\nu)] \cong \delta,$$

where δ denotes the Kronecker symbol.

It follows that one can access to data $\boldsymbol{Z}$ from the knowledge of its associated $\boldsymbol{Z}_{\boldsymbol{h_m}}$, by the way of a simple summation. Following the same idea, an approximation of the useful signal $\tilde{\boldsymbol{S}}$ can be obtained from $\tilde{\boldsymbol{S}}_{\boldsymbol{h_m}}$ by:

$$\tilde{\boldsymbol{S}} = \sum_{m=0}^{M+1} \tilde{\boldsymbol{S}}_{\boldsymbol{h_m}}$$

This denoising process can be summarized as follows:

- Initialize a M_Z-dimensional vector $\tilde{\boldsymbol{S}}$ to zero;
- For m equal to 0 to $M+1$ do:
  - Computation of impulse response $\boldsymbol{h_m}$;
  - Determination of $\boldsymbol{Z}_{\boldsymbol{h_m}}$ making the convolution product between $\boldsymbol{Z}$ and $\boldsymbol{h_m}$;
  - Multiresolution of $\boldsymbol{Z}_{\boldsymbol{h_m}}$ by the way of a DWT;
  - Wavelet coefficients thresholding;
  - Construction of $\tilde{\boldsymbol{S}}_{\boldsymbol{h_m}}$ applying an inverse DWT;
  - Iterative construction of $\tilde{\boldsymbol{S}}$: $\tilde{\boldsymbol{S}} = \tilde{\boldsymbol{S}} + \tilde{\boldsymbol{S}}_{\boldsymbol{h_m}}$

The flowgraph of this algorithm is proposed in figure 2.

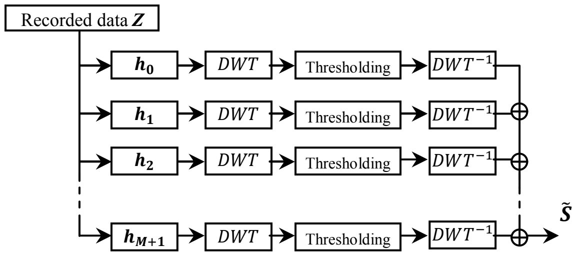

Fig. 2: Flowgraph of the denoising process

IV. Experiments

1. Simulated signal

To illustrate the proposed processing, we have constructed a test signal. This one is proposed on figure 3 (upper part: discrete-time signal, lower part: spectrogram), it has been sampled at a frequency equal to 44100 Hz.

This test signal has been designed in order to illustrate the different classes of transient signals one can encounter in an underwater medium, with from left to right: one vocalization, three impulsive signals, one shock, one broad band signal, four signals corresponding to a mix between shock and impulsive signals, and one discontinuous vocalization. This test signal has been disturbed by a Gaussian-Gaussian mixture (the noise PDF can be viewed on figure 4); the resulted observation is presented on figure 5.

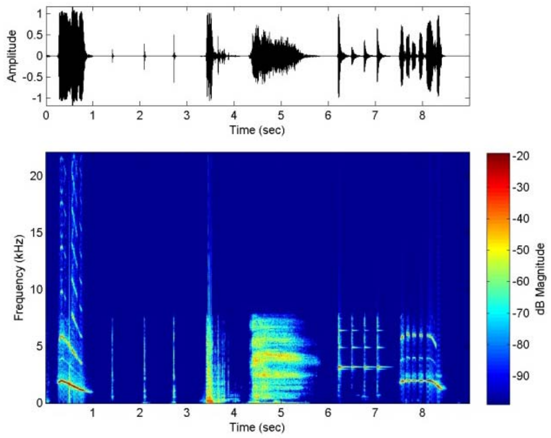

Fig. 3: Test signal (discrete-time signal and spectrogram)

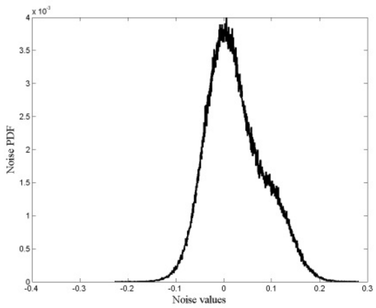

Fig. 4: Noise PDF

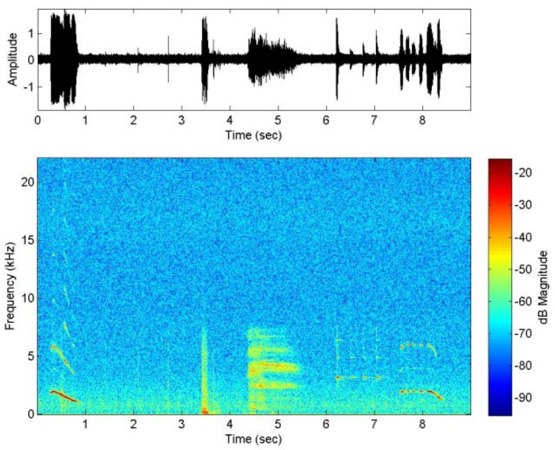

Fig. 5: Simulated noisy received signal and its spectrogram

Applying the proposed algorithm, with 200 Mel's filters, a 4th order Daubechies wavelet and with α equal to 1, allows obtaining the denoised signal presented on figure 6. This result shows without any ambiguity the efficiency of this denoising method: whatever the kind of transient, after process they are well preserved keeping all the useful information. The only exception is for the first impulsive signal, this one having a too low magnitude, compared to the noise level, to be restored.

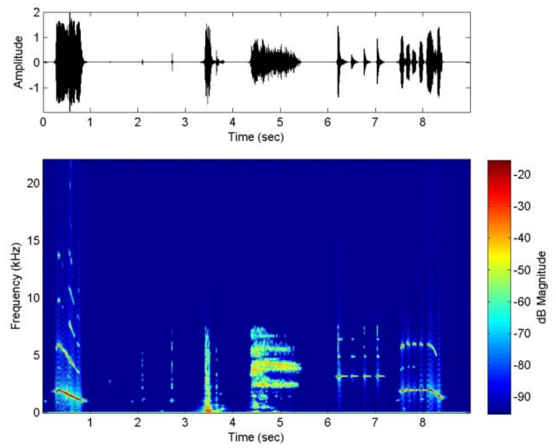

Fig. 6: Denoised data and its spectrogram

In order to benchmark our method, we have compared it with two well-known algorithms. The first one is built from the Ephraim and Malah's MMSE-LSA attenuation rule [12] associated with the decision-directed (DD) *a priori* SNR estimator described in [13]. The second one is based on the Wiener attenuation rule [12] associated with the non-causal *a priori* SNR estimator proposed by Cohen in [14]. The algorithms are tuned as follows: $\alpha = 0.98$ for the DD estimator, while $\alpha = 0.9, \alpha_{bis} = 0.9, \beta = 2$ and $\xi_{min} = -25\,dB$ for the second algorithm.
We recall that the signal has been sampled with a 44100 Hz sampling frequency. Its short-time Fourier transform (STFT) is computed with a 512 points (~11 ms) Hann window with 50% overlapping. Besides, the noise standard deviation is estimated by the way of the MAD. Generally for such denoising problems, it is used to establish the performance assessment:

- the Global SNR (*GSNR*):

$$GSNR = 10\log_{10}\frac{\sigma_S^2}{\sigma_{\tilde{N}}^2},$$

  where $\tilde{\boldsymbol{N}}$ corresponds to the summation of the residual noise and the distortion, which has affected the useful signal after the denoising step: $\tilde{\boldsymbol{N}} = \boldsymbol{S} - \tilde{\boldsymbol{S}}$;

- the Segmental SNR (*SSNR*), which is defined as the average of the SNR_i values computed over segments with useful signal activity (SNR_i is commonly called the Local SNR (*LSNR*) computed for the *i*-th segment of the signal):

$$SNR_i = 10\log_{10}\frac{\sigma_{S,i}^2}{\sigma_{\tilde{N},i}^2},$$

where $\sigma_{S,i}^2$ and $\sigma_{\tilde{N},i}^2$ are powers in the i-th frame respectively;

- the mean square error (*MSE*):

$$MSE = E\left\{\left(\boldsymbol{S} - \tilde{\boldsymbol{S}}\right)^2\right\}.$$

We present, on table 1, the results obtained for the studied denoising methods. We have computed these criteria for the different classes of transient signals present in the studied synthetic signal and for a part of the observation, where only noise occurred.

Table 1

Objective comparison between Denoised Hearingogram (DH), Ephraim and Malah (LSA-DD), Wiener and Cohen (Wien-NC)

	Noisy	LSA-DD	Wien-NC	DH
GSNR	16.02	28.20	29.10	20.30
SSNR	- 3.90	6.94	13.05	8.50
MSE	44×10^{-4}	7×10^{-4}	6×10^{-4}	20×10^{-4}

– Vocalization –

	Noisy	LSA-DD	Wien-NC	DH
GSNR	- 10.47	2.09	3.20	3.30
SSNR	- 43.41	- 26.70	- 10.14	- 8.40
MSE	43×10^{-4}	2×10^{-4}	2×10^{-4}	2×10^{-4}

– Impulsive signals –

	Noisy	LSA-DD	Wien-NC	DH
GSNR	9.40	15.70	16.40	14.50
SSNR	-12.90	- 1.59	3.15	2.30
MSE	41×10^{-4}	10×10^{-4}	8×10^{-4}	15×10^{-4}

– Shock –

	Noisy	LSA-DD	Wien-NC	DH
GSNR	7.94	12.17	12.24	9.16
SSNR	- 1.43	5.78	7.45	5.69
MSE	41×10^{-4}	16×10^{-4}	15×10^{-4}	28×10^{-4}

– Broadband signals –

	Noisy	LSA-DD	Wien-NC	DH
GSNR	7.69	20.35	22.04	18.13
SSNR	- 11.68	3.32	8.62	6.02
MSE	43×10^{-4}	2×10^{-4}	2×10^{-4}	6×10^{-4}

– Mix shocks / impulsive signals –

	Noisy	LSA-DD	Wien-NC	DH
GSNR	13.07	24.77	25.85	19.28
SSNR	- 9.50	4.07	11.28	10.14
MSE	42×10^{-4}	3×10^{-4}	2×10^{-4}	19×10^{-4}

– Discontinuous vocalization –

	Noisy	LSA-DD	Wien-NC	DH
GSNR	- 42.07	- 24.61	- 11.12	2.4
SSNR	- 43.58	- 25.48	- 5.96	6.9
MSE	42×10^{-4}	7.5×10^{-5}	3.4×10^{-6}	1.5×10^{-7}

– Noise –

An analysis of these results reveals that the average performance of our process is better than the LSA-DD method, but not as good as the Wien-NC method, with the exception of impulsive signals and noise cases. Let us remark, that our process does not required some input parameters (the α value can be set by default to 1), unlike the Wien-NC method where four parameters are adjusted according to the nature of the observation. This is a key point for such a denoising step in fully automatic and operational systems.

2. *Real underwater records*

To illustrate our processing on real data, we have chosen to apply it on a record of killer whale echolocations. This signal is presented on figure 7, with its spectrogram. It has been recorded with a sampling frequency equal to 44100 Hz.

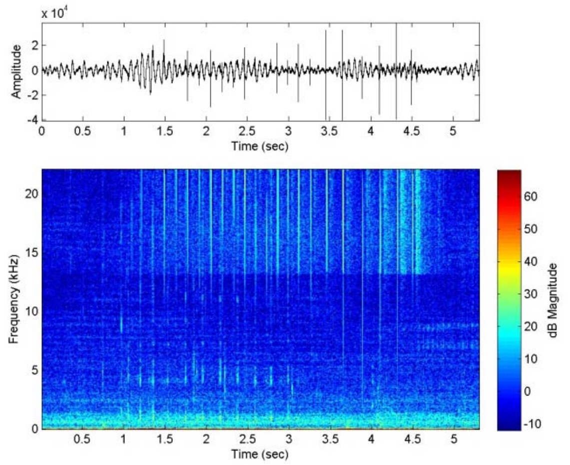

Fig. 7: Killer whale echolocations and associated spectrogram

Although these echolocations correspond to very impulsive signals, they are well preserved after denoising, making easier their detection and interpretation (see figure 8). For this experiment, we have used the same parameters than for the simulated data, presented in the previous section.

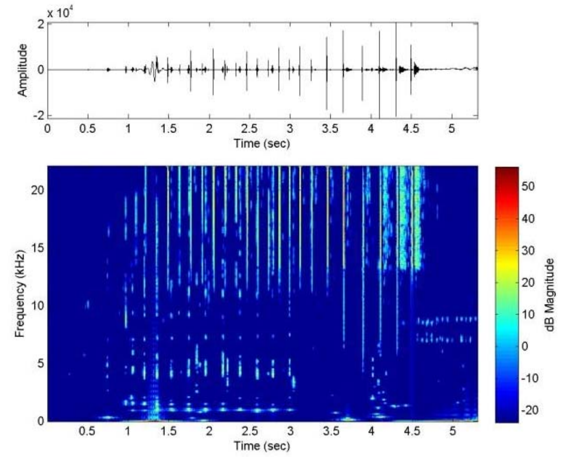

Fig. 8: Denoised signal presented on figure 7, with its associated spectrogram

The last example corresponds to killer whale vocalization sampled at a frequency equal to 22050

Hz. This signal is presented on figure 9. It has a 30 seconds duration and presents at its end (between the 26th second and the 29th second) some echolocation signals not clearly visible on the spectrogram.

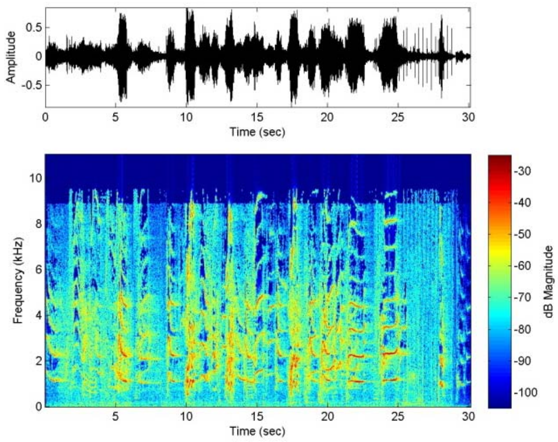

Fig. 9: Killer whale vocalizations and associated spectrogram

Applying the proposed processing allows to obtain the denoised data presented on figure 10. One more time, this result reveals the efficiency of this denoising approach, keeping the whole richness of these harmonic narrow-band signals. Moreover, the echolocation signals are now clearly visible on the spectrogram.

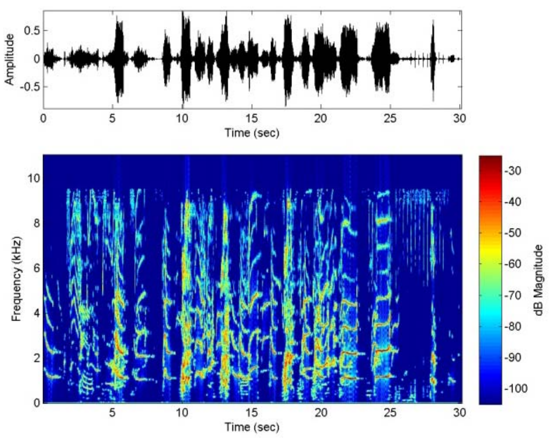

Fig. 10: Denoised signal presented on figure 9, with its associated spectrogram

V. Conclusion

This paper describes an innovative approach to reduce the noise level in noisy underwater acoustic signals. This one is based on the use of the *Hearingogram* principle and requires no information about the useful signal and noise characteristics or statistics. Doing so, this blind process is fully automatic and can be apply whatever the kind of considered signals. Thus, the use of such a process could be of great interest to improve the performances of dedicated systems, where the noisy environment is a source of problems.

References

[1] Ph. Courmontagne, *The Stochastic Matched Filter and Its Applications to Detection and De-Noising*, Stochastic Control, Chris Myers (Ed.), ISBN: 978-953-307-121-3, InTech, DOI: 10.5772/9737, 2010.

[2] Therrien, C.W., Frank, K.L. Jr., and Ruiz Fontes, N., *A Short-Time Wiener Filter for Noise Removal in Underwater Acoustic Data*, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 543-546, Munich, Germany, April 1997.

[3] Tu, K., Jiang, Y.Y., *Development of Noise Reduction Algorithm for Underwater Signals*, IEEE International Symposium on Underwater Technology, p.175–179, 2004.

[4] PL. Brockett, M. Hinich, GR. Wilson, *Nonlinear and non-Gaussian ocean noise*, The Journal of the Acoustical Society of America **82**(4), 1386–1394 (1987).

[5] D. Middleton, *Channel modeling and threshold signal processing in underwater acoustics: an analytical overview*, IEEE Journal of Oceanic Engineering **12**(1), 4–28 (1987).

[6] M. Bouvet and S. C. Schwartz, *Detection in underwater noises modeled as a Gaussian-Gaussian mixture*, *Proc. ICASSP*, pp.2795 -2798, 1986.

[7] Tesei, A., Regazzoni, C.S., Tacconi, G., *HOS-based modelling of low-frequency underwater acoustic noise for signal detection in shallow waters*, OCEANS '96. MTS/IEEE. Prospects for the 21st Century. Conference Proceedings, On page(s): 1421 - 1426 vol.3 Volume: 3, 23-26 Sep 1996.

[8] S. Ouelha, P. Courmontagne, F. Chaillan, *On time-frequency representations for underwater acoustic signal*, Oceans'12 MTS/IEEE Hampton Roads, USA, DOI: 10.1109/OCEANS.2012.6405047, October 2012.

[9] S. Ouelha, P. Courmontagne, F. Chaillan, *A new time-frequency representation for underwater acoustic signals: the denoised hearingogram*, Oceans'13 MTS/IEEE Bergen, Norway, June 2013.

[10] D. Donoho, *De-noising by soft thresholding*, IEEE Trans. on Information Theory, vol. 38(2), pp. 613-627, 1995.

[11] Mallat, S.A., *A theory for multi-resolution signal decompositions: the wavelet representation*, IEEE Trans. on Pattern Analysis and Machine Intelligence, **11**(7):674–693, 1989.

[12] Y. Ephraim and D. Malah, *Speech enhancement using a minimum mean square error log-spectral amplitude estimator*, IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-33, no. 2, pp. 443–445, Apr. 1985.

[13] Y. Ephraim and D. Malah, *Speech enhancement using a minimum mean square error short-time spectral amplitude estimator*, IEEE. Trans. Acoust., Speech, Signal Process, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[14] I. Cohen, *Relaxed statistical model for speech enhancement and a priori SNR estimation*, IEEE Trans. Speech Audio Process., vol. 13, no. 5, pp. 870–881, Sep. 2005.