Open access • Journal Article • DOI:10.1109/JOE.2006.875275

Feature Extraction From Underwater Signals Using Time-Frequency Warping Operators — Source link

Cornel Ioana, André Quinquis, Y. Stephan

Published on: 21 Jul 2006 - IEEE Journal of Oceanic Engineering (IEEE)

Topics: Image warping, Instantaneous phase, Signal, Time–frequency analysis and Chirp

Feature Extraction from Underwater Signals using Time-Frequency Warping Operators

Cornel Ioana, André Quinquis, Yann Stephan

Feature Extraction from Underwater Signals using Time-Frequency Warping Operators

Cornel IOANA[*], *Member IEEE,* André QUINQUIS[*], *Senior Member IEEE*
Yann STEPHAN[**], *Member IEEE*

[*] : ENSIETA, E3I2 Research Centre, 2 rue François Verny, 29806 Brest, France,
Phone : +33.298.34.88.86, Fax : +33.298.34.87.50
Emails : ioanaco@ensieta.fr, quinquis@ensieta.fr
[**] : EPSHOM, Military Center of Oceanography, 13 rue de Chatellier, 29603 Brest
Phone : +33.298.22.10.85, Fax : +33.298.22.18.64
Email : yann.stephan@shom.fr

Abstract – *Processing marine-mammal signals for passive oceanic acoustic tomography or species classification and monitoring are problems that have recently attracted attention in scientific literature. For these purposes, it is necessary to use a method which could be able to extract the useful information about the processed data, knowing that the underwater environment is highly non-stationary. In this context, time-frequency or time-scale methods constitute a potential approach.*

Practically, it has been observed that the majority of time-frequency structures of the marine-mammal signals are highly non-linear. This fact affects dramatically the performances achieved by the Cohen's class methods, these methods being efficient in the presence of linear time-frequency structures.

Fortunately, thanks to the warping operator principle, it is possible to generate other class of time-frequency representations (TFRs). The new TFRs may analyze non-linear chirp signals better than Cohen's class does. In spite of its mathematical elegance, this principle is limited in real applications by two major elements. First, as we will see, its implementation leads to a considerable growth of the signal length. Consequently, from operational point of view, this principle is limited to short synthetic signals. Secondly, the design of a single warping operator can be inappropriate if the analyzed signal is multi-component. Furthermore, the choice of "adapted" warping operator becomes a problem when the signal components have different time-frequency behaviors.

In this paper, we propose a processing method of marine-mammal signals, well adapted to a real passive underwater context. The method tries to overcome the two limitations mentioned before. Also, the first step consists in data size reducing by the detection of the time-frequency regions of interests (ROIs). Furthermore, in each ROI, a technique which combines some typical warping operators is used. The result is an analytical characterization of the instantaneous frequency laws of signal components.

The simulations on real underwater data show the performances of this method in comparison with classical ones.

Keywords*: time-frequency analysis, wavelet transform, high-order statistics, underwater environment, warping operators, passive tomography*

1. INTRODUCTION

Motivation for processing marine-mammal signals stems from increasing interest in the behavior of endangered marine mammals, reflected in a number of publications in the scientific literature [1,2]. The ultimate goal of the current research in this field is to develop tools for the simultaneous localization of mammals and analysis of the emitted signal for species identification and monitoring.

On the other hand, the characterization of underwater environment is a challenging topic, due to the richness of the potential information that can be extracted for navigation or communication, for example. One of the major methods is the *oceanic active tomography* [3], which provides an environmental characterization using a man-made transmitted signal. Nevertheless, it is possible to

imagine the passive tomography concept which will benefit by the generated signals by the natural sources (opportunity sources).

These two applications constitute arguments for the developments of some adequate processing tools. In order to detail the purposes of the proposed method, we firstly point out on some considerations about classical passive sonar systems. From signal processing point of view, the receiver of a passive sonar system is organized as in figure 1 [4].

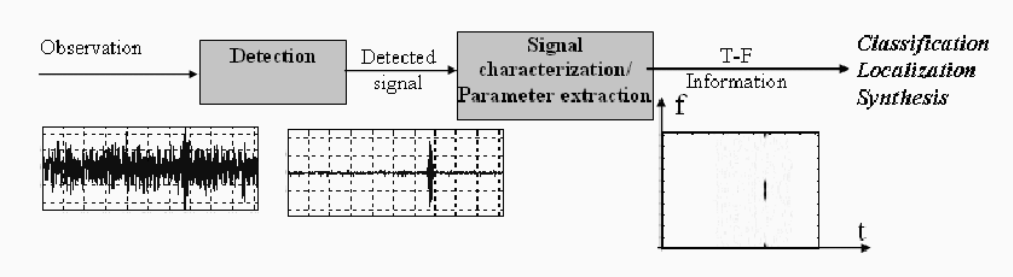


Figure 1. Organization of the passive processing of underwater signals

This figure indicates the two signal processing operations involved in a passive sonar system : *detection* and *characterization*.

It is well known that the sounds produced by marine mammals have generally a non-stationary behavior. It ranges from short transients to long tones, with a huge variety of non-linear chirp-like structures. Analysis of these sounds needs adapted tools in order to accurately characterize the signals in both time and frequency dimensions. In numerous studies [4,5,6] and for well-known reasons the spectrogram [7] has been used as the interface between sounds and the feature extraction stage. Nevertheless, in many applications, such as passive sonar signal processing or speech processing, the spectrogram does not provide enough freedom to the user. When the interest is to obtain, via spectrogram, a "nice picture" of the sound a marine mammals actually transmits, it is indispensable to be able to choose the good parameter set (window type and length, overlapping parameter, etc.). On the other hand, the spectrogram representation is unable to provide information describing the auditory perception of this kind of signal. This information is often required by the post-processing stages [8]. Therefore, with the notion that the auditory system inherently represents acoustic signals in robust manner [9], a promising approach to solve the *detection* of underwater sounds seems to lie in the mimicking of that process. Recent research [9] has indicated an analytically appealing model of human auditory processing in its early stages. This model is based on the Over-Complete Wavelet Transform (OCWT) [10]. Therefore, the proposed detection method is based on the OCWT which leads to signal processing on interest frequency sub-bands. In each of them, an irregular sampling procedure will be used, in order to adaptively detect the signal time-frequency features. On the other hand, the detection stage reduces the size of the data applied to the next phase – *parameter extraction*. Since only the useful parts of the received signal are taken into account, the complexity of the characterization procedure is naturally reduced.

Furthermore, the parameter extraction algorithm will be applied in each detected ROI. In order to deal with the non-stationary behavior of the signal belonging to each ROI, the use of time-frequency representation constitutes a potential approach. Usually [8, 11], the time-frequency (T-F) methods belonging to the Cohen's class [7] are used for the extraction of time-frequency information from non-stationary signals. Nevertheless, these methods perform correctly while the time-frequency components of the analyzed signal have a linear T-F behavior. More precisely, the time-frequency distribution is highly concentrated around the linear instantaneous frequency law (IFL). In the case of non-linear IFLs, some artifacts appear, related to the bi-linearity of this class of TFRs [7]. The attenuation of these artifacts implies the application of some smoothing functions (kernels) in the ambiguity plane. This operation could create some degradation of the quality of time-frequency information [7,12].

Fortunately, it is possible to generate other class of time-frequency representations which may analyze non-linear chirp signals better than Cohen's class does. The technique used to generate the non-linear TFR classes is based on the use of warping operators [12]. Generally, this operator is

interpreted as a unitary axis deforming. In time-frequency analysis context its application produces the "linearization" of the time-frequency behavior of the signal. The application of warping technique imposes the knowledge of the time-frequency behavior of the signal. Moreover, if the T-F components of the signal have different types of T-F structures (example, Hyperbolic Chirps and Linear Chirps) it is necessary to combine several warping operators. In this paper, we propose a new method for time-frequency feature extraction based on the signal-dependent selection of some warping operators. The aim is to obtain a complete time-frequency description of marine-mammal emissions. The examples will illustrate the performances of this method in a real-data context.

In conclusion, the main purpose of this paper is to propose an alternative solution for both detection and parameter extraction stages which are part of a passive sonar system.

The jointly use of the proposed methods will be illustrated on some real marine –mammal emissions.

The paper is organized in two main parts. The first one – "Passive Detection Methods" deals with the problem of detection of the transient parts of the signal. In Section 2 we briefly present the OCWT concept and its relation with the auditory modeling. Furthermore, we propose a new method for underwater signal detection in a passive context. This method is based on an irregular sampling of the time axis, provided by a *split and merge* algorithm [13]. In Section 3, we will study the performances of the new approach using the real underwater mammal emissions. Furthermore, we will compare the obtained results with the ones given by two classical methods.

The second part of the paper is devoted to the "Underwater Signal Characterization". After a short presentation of the warping-based non-linear TFR (Section 4), we introduce, in section 5, a parameter extraction method based on the combination of several time-frequency warping operators. The jointly use of the new detection and parameter extraction methods is illustrated, via some examples, in the Section 6. Section 7 - "Conclusion and Perspectives" - highlights the significance of the results and the realistic perspectives.

PART 1. PASSIVE DETECTION METHODS

2. PASSIVE DETECTION METHOD BASED ON OVER COMPLETE WAVELET TRANSFORM

Human auditory system has a remarkable capacity for acoustic signal detection and interpretation. In the last years, many works tried to design a similar artificial system, based on the mimic of the physical phenomena which occurs in a human auditory system.

When a signal enters the ear it creates a pattern of activity (displacement) along the basilar membrane [9]. This membrane has a length and, hence, a spatial axis (indexed by s) associated with it which extends from its base ($s=0$) towards its apex ($s>0$). The magnitude of activity at any particular spatial location, s, on the membrane may be modeled as the output of a linear (bandpass) filter with an impulse response $h_s(\cdot)$. Because of the physical nature of the membrane, the filters $\{h_s : s > 0\}$ are related in a special way. Physically, this is reflected by the fact that the basilar membrane provides a logarithmic mapping of the acoustic spectrum. One might think of the basilar membrane as acting like a bank of bandpass filters where the center frequency of the filter $H_s(\cdot)$ is related logarithmically to s. As it is shown in [9], this type of relationship can be described nicely in terms of wavelet transformation (see [14], for complete references). In its continuous form, the wavelet transform is expressed as :

$$\left(CWT_{\ g}\, x\right)(t, a_0) \overset{\Delta}{=} \left\langle x, \tau_t D_{a_0} g \right\rangle = |a_0|^{1/2} \int x(u)\, g^*\left(a_0 (t-u)\right) du \tag{1}$$

where g is the analyzing wavelet, x is a given signal, τ_t is the translation operator ($(\tau_t f)(u) = f(u-t)$) and D_s is the scale operator $\left((D_s f)(u) = \frac{1}{a_0} f\left(\frac{u}{a_0}\right)\right)$. Despite the fact that there is an infinity of possible sampling procedure of the CWT, the term *discrete wavelet transform (DWT)*

is commonly used to mean the one associated with the dyadic sampling lattice (a_0=2). For a_0 lower than *2* we get the Over Complete Wavelet Transform (OCWT).

From computational point of view, the OCWT is implemented in two stages :

I. First, we decompose the signal with a linear filter bank structure (an example is given in figure 2.a). The impulse responses of the filter bank are determined by the analyzing wavelet *g* and the scale parameter a_0. The filtering stage result is mathematically stated as :

$$OCWT_g x(t, a_0) = \left(x * D_{a_0} g^*\right)(t) \tag{2}$$

Here a_0 is a scaling index which controls the overlapping between the filter transfer functions (figure 2.b). For a_0=*2,* we obtain the filter bank structure used for the DWT computation. A filter bank example is shown in 2.b., using the Morlet's wavelet as the analyzing function :

$$g_{Morlet}(t) = \frac{1}{\sqrt{\pi\gamma_b}} \cdot e^{j2\pi\gamma_c t - (t^2/\gamma_b)} \tag{3}$$

where γ_b and γ_c are the bandwidth and the frequency center of the Fourier transform of *g*.

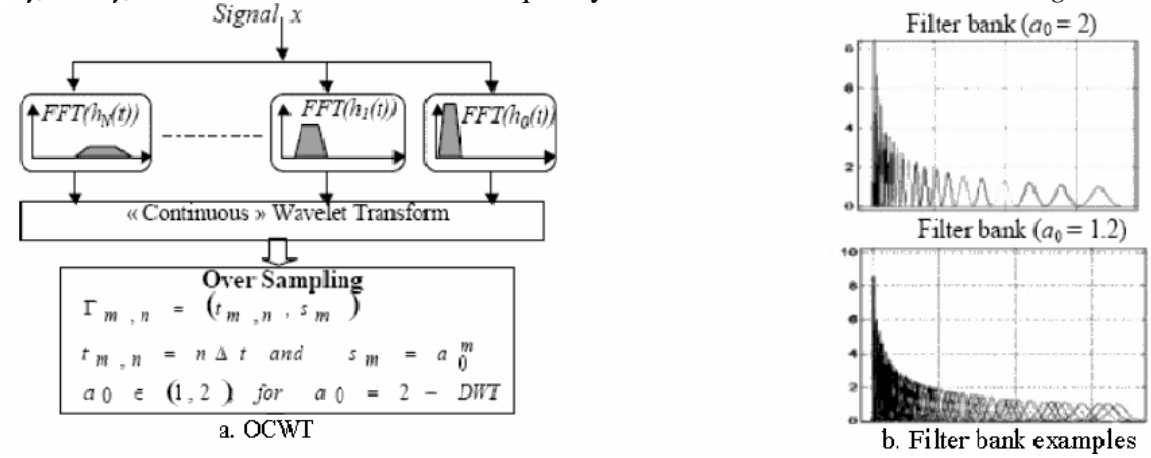


Figure 2. Comparison between filter banks at different scales

II. In the second stage we will sample the signal issued from the filter bank. We take into account the samples at discrete times given by $\{t_{m,n}\}$ (figure 2.a).

Mathematically, OCWT may be interpreted as the CWT sampled version of the signal by a non-dyadic structure. Usually, we use the semi-logarithmic regular sampling, given by [10] :

$$\Gamma(\Delta, a_0) \overset{\Delta}{=} \{n\Delta\} \times \left\{a_0^m\right\}, 2 > a_0 > 1, \Delta > 0 \tag{4}$$

The parameter a_0 controls the filter overlapping, and, implicitly, the redundancy degree. If a_0<2, the wavelet function set will be a frame : the reconstruction is not perfect but we can adapt our distribution to the signal singularities. This can be done by non-uniformly sampling the signal issued from each sub-band, according to its characteristics.

Generally speaking, there are some advantages to adopt a non-uniform sampling strategy in a representation. Many of these advantages are inherited from the ability of a non-uniform sampling to be sensitive to signal time-frequency behaviors [10]. The theoretical framework of the non-uniform sampling strategies is presented in [15] and some applications (for noise suppression, digital communication, compression, etc.) are presented in [10]. In this section we propose a new passive detection method based on a non-uniform sampling technique in a noisy environment. This technique sets different weights to signal samples, according to its statistic properties.

It is well known [13] that the non-gaussian wavelet coefficients provide a large value of the fourth order statistic cumulant (*kurtosis*). On the other hand, the noise coefficients, which have usually a gaussian probability density function, are characterized by a small value of the *kurtosis.* The kurtosis allows us to discriminate between the useful (transitory) and useless (noise) parts of the signal. This principle, illustrated in the next figure, will be applied for signals issued from each OCWT sub-band.

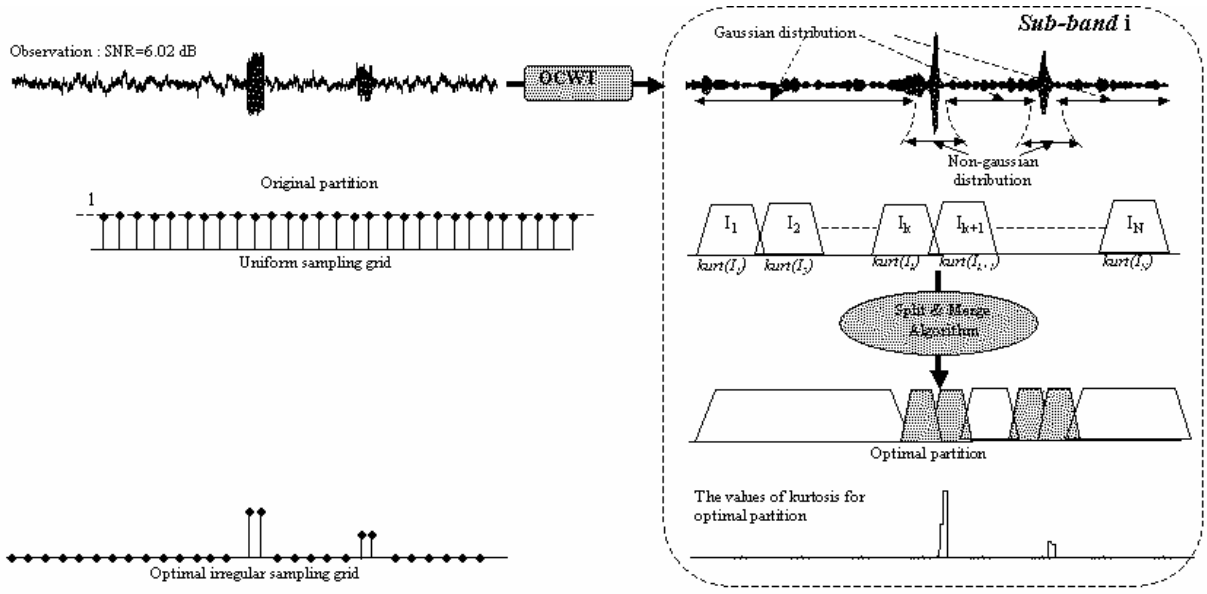


Figure 3. The non-uniform sampling procedure for sub-band i

In the above figure we present the principle of the non-uniform sampling procedure. Initially, the waveform issued from the i^{th}-channel of the OCWT filter bank is uniformly partitioned in equal length intervals I_k composed each one by N samples. For each of them the value of kurtosis is estimated, using the following relation [16] :

$$kurt(I_k) = N\left(\sum_{i=1}^{N} x_i^4\right) / \left(\sum_{i=1}^{N} x_i^2\right)^2 ; \{x_i\} \in I_k \tag{5}$$

Using these values, we apply an iterative Split & Merge algorithm in order to establish the optimal partition. For each two adjacent intervals I_k and I_{k+1}, we test the following condition :

$$\begin{aligned} &H_0 : \text{ if } kurt\left(I_j\right) \le \mu_s \;\&\; kurt\left(I_{j+1}\right) \le \mu_s \Rightarrow I'_j = I_j \cup I_{j+1} \\ &\qquad \Rightarrow \text{kurt}\left(I'_j\right) = \max\left[kurt\left(I_j\right), kurt\left(I_{j+1}\right)\right] \\ &H_1 : \text{ if } kurt(I_j) > \mu_s \; or \; kurt(I_{j+1}) > \mu_s \Rightarrow \text{the intervals will} \\ &\qquad \text{be conserved} \end{aligned} \tag{6}$$

where μ_s is a threshold computed for each sub-band as presented below.

The H_0 hypothesis states that there is no useful part of signal in the considered intervals. More precisely, since the kurtosis values of both intervals are inferior to μ_s we assume that only the noise is presented in the considered intervals. Therefore, these ones will be *merged* (fusion operation). The kurtosis of the new interval will be chosen as the maximum from the kurtosis of the intervals. This over-estimation of the kurtosis minimizes the effect of the estimation errors [13].

Alternatively, the H_1 hypothesis states that one or both intervals contains useful parts of signal and will be *conserved*. The algorithm runs until no fusion is possible.

Since the noise is generally colored, the OCWT coefficients issued from each sub-band have different energies, according to their places in the frequency domain [13]. This justifies the use of kurtosis instead of 4^{th} order cumulant : the kurtosis will normalize the coefficients in each sub-band. Therefore, the estimation being normalized, the variance of kurtosis estimator, in the case of a noise with a variance σ^2 does not depend on σ. The expression of variance is [17] :

$$Var(kurt) = \frac{24}{N_s} \tag{7}$$

where N_s is the number of samples in the sub-band s. It is related to the total samples number via :

$$N_s = \frac{N}{a_0^s} \tag{8}$$

where a_0 is the scale index. Consequently, the threshold μ_s in (6) is expressed, for the sub-band *s,* as

$$\mu_s = \frac{1}{\sqrt{1-\alpha}} \sqrt{a_0^s \frac{24}{N}} \tag{9}$$

where α is the confidence degree, N is the signal length and s the sub-band index (s=1..Number_of_subbands). This threshold is obtained from the maximization of the confidence intervals [13]. It guaranties its availability in the case of short adjacent intervals.

Finally, applying the Split&Merge algorithm, we obtain an optimal partition (figure 3) and and its kurtosis values. The obtained curve weights the samples of the supposed waveform, ensuring an non-uniform sampling of this one : the samples associated to transient parts of signal will be "highlighted", whereas the ones associated to the noise will be shrunk. This effect is illustrated in the figure 4. We consider two chirps atoms (both on 128 samples), mixed with real underwater noise (SNR=6.02 dB). After the OCWT (the number of sub-bands is about 128) we apply the method to the waveform from each channel. The values of kurtosis for the optimal partition provide an optimal sampling grid which improves the representation quality. Repeating the same algorithm for all OCWT channels, we obtain a two-dimensional non-uniform sampling grid which leads to the detection of time-frequency regions-of interest (ROIs) (figure 4). On the other hand, by unifying the kurtosis curves for all sub-bands and for all temporal position, we obtain *the detection curve* in both time and frequency domains. This curve provides an information about temporal localization of the transient parts of the signals.

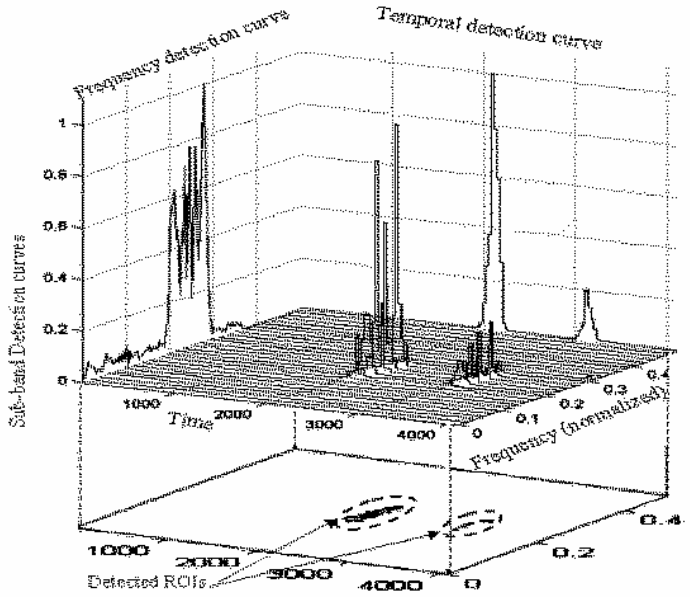


Figure 4. Detection of the time-frequency ROIs, using the OCWT and the non-uniform sampling procedure

With the principle described in this section, the passive detector scheme is organized as shown in the figure 5.

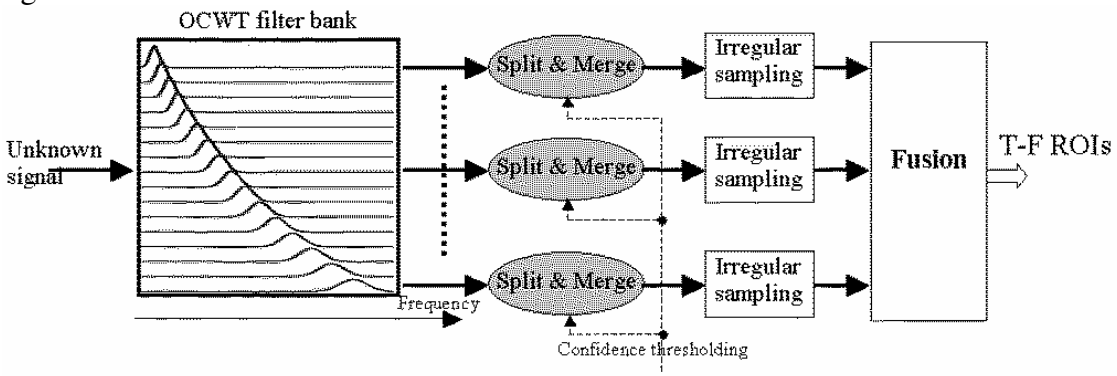


Figure 5. Passive detector based on Irregular Sampling of OCWT (OCWT-IS)

The received underwater signal passes trough a filter bank, designed to cover the interest frequency domain. In our case, since the real signals we used for the experiments were sampled at 44.1 kHz, the useful frequency range is about 100-20 kHz. Nevertheless, the detection method performs in the same way for an arbitrary frequency range. Using an OCWT filter bank, this domain is partitioned in *Ns* sub-bands. The filters of this bank are overlapped, in order to ensure the conservation of global time-frequency information [10].

The signal issued from each sub-band filter is analyzed with the procedure described in (6). The purpose is to distinguish, in the considered band, the useful parts of the received signal. Furthermore, using the "Split & Merge" procedure (relation 6), we obtain a non-uniform sampling grid adapted to the sub-band representation of the signal. The sets of the sampling grids associated to the sub-bands form the time-frequency regions of interest.

3. PERFORMANCES OF PROPOSED DETECTOR

In this section we compare the proposed passive detector with some other detectors. We will present comparatively the results of the detection obtained by the method introduced in the previous sections (OCWT-IS – OCWT and irregular sampling) and two classical passive detectors based on the following methods : Short-Time Fourier Transform (STFT) [5] and the signal decomposition on the Local Cosine Basis (LCB) [13]. This last method decomposes the signal in a local cosine basis. The choice of the best basis involves a split and merge algorithm, similar to the one presented previously. This concept has been generalized in the section 2 by the "Split and merge" algorithm applied to each sub-band response.

As test data set, we used the signal emitted by a *Tursiops Truncatus* and recorded in the Marineland site [29] located in South of France. The sampling frequency is 44.1 kHz and we have taken into account an observation of 5.92 seconds. The receiver frequency range was .025-23 kHz.

As a performance measure, the receiver operating characteristics (ROC) are computed. Recall that the ROC is a collection of curves describing the probability of detection as a function of the probability of false alarm for a set of different signal to noise ratio (SNR). Figure 6 displays the ROCs generated for three SNR values : 5, 3, 1 dB using the methods previously enumerated.

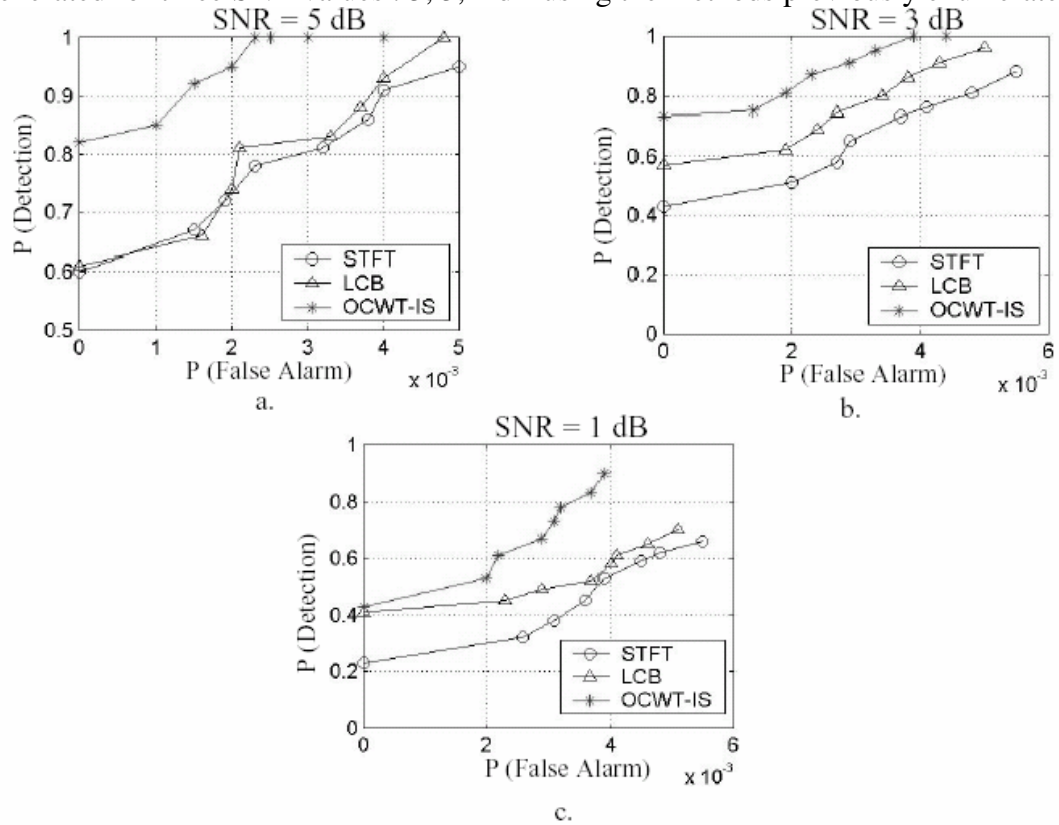


Figure 6. ROCs for detection schemes : a. SNR = 5 dB; b. SNR = 3 dB; c. SNR = 1 dB

Obviously, the results indicate that the OCWT based method (for $a_0=1.12$) outperforms a good detection performance in all the cases. Firstly, we obtain an acceptable value for the detection probability, even for a SNR close to 0 dB. Secondly, the false alarm level is lower than the values generated by the first two methods. Both of these facts demonstrate the superiority of the OCWT-IS based detector.

The good performances of the OCWT-IS detector relatively to the two other methods are explained as following. In the case of the STFT based method, the parameters and the detection threshold are selected before the simulations and maintained constant during the experiments. Since the experiment conditions were changing (for example, the noise parameters were different for each trial), the performances of the STFT based detector decreased much faster than the ones of the OCWT-IS. On the other hand, the use of a regular time-frequency partition does not allow efficient extracting of the signal characteristics.

The second detector, based on the LCB decomposition, performs better than the STFT based detector. This can be explained by the Split&Merge-based algorithm for best basis extraction who adapts the time-frequency partition with respect to signal characteristics. Nevertheless, the dyadic sampling grid is inherently less robust in a noised environment than a non-dyadic one [10]. It explains the superiority of the OCWT-IS based detector illustrated also by the statistical detection performances.

Another comparison criterion is the quality of the ROI's time-frequency information. This criterion is very important while the quality of ROI information has a critical implication for the next stage – *the feature extraction*. As we will see in the next sections, if the ROIs are correctly extracted the feature extraction will be successfully performed from both operational and efficiency points of view.

To study comparatively the three methods, let consider a sequence of a real underwater sound. We start with the results provided by the spectrogram and the LCB-based detectors.

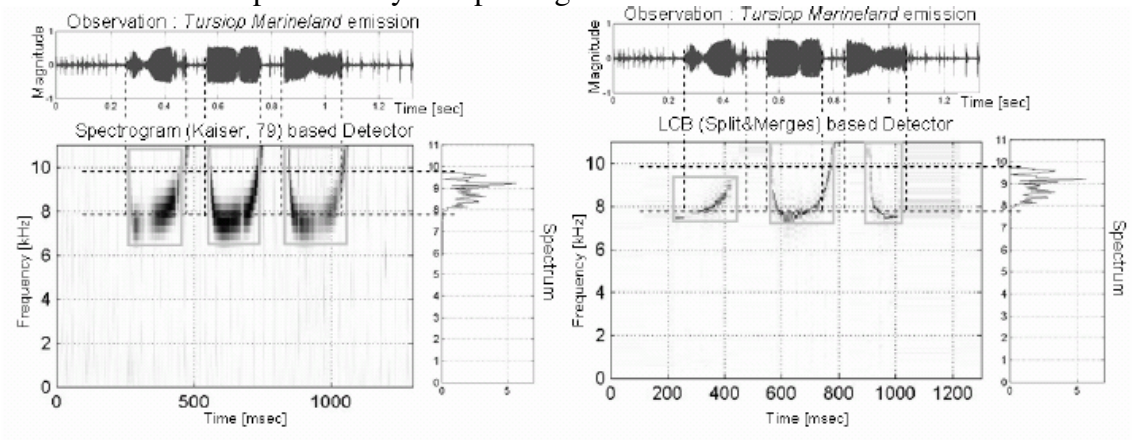


Figure 7. ROIs detection using the spectrogram, LCB and OCWT-IS –based methods

As a reference measure we consider the "ideal" ROIs, defined in the figure 7 as the time-frequency rectangles bounded by the dashed lines who represent the start/stop coordinates of the useful parts of the signal. The horizontal lines represents the spectrum limits whereas the vertical ones are associated to the time supports of the analyzed signal. The detected ROIs are represented by solid gray rectangles.

In the case of spectrogram-based detector (figure 7.a), we have used a Kaiser-type analysis window of size 79. We observe that the detected ROIs are larger, in the frequency direction, than the ideal ones. This is due to the Heisenberg's principle and, consqently, to the trade-off between time and frequency resolution. The best trade-off is difficult to be done and it depends on the optimal choice of spectrogram parameters.

In the case of LCB (Split&Merge)-based detector (figure 7.b) the time-frequency resolution is improved. Neverthless, due to the Split & Merges algorithm, the time supports are not well matched in the detected ROIs.

Finally, we present the results obtained in the case of OCWT-IS-based detector (figure 7.c). It can be observed that the detected ROIs are close to the ideal time-frequency partitions associated to useful parts of the signals. The used filter bank is illustrated in figure 7.d : we considered a 128 Morlet-type filters and an overlapping degree a_0=1.02. In these conditions the ROIs are well detected. Therefore, the signal components are well conserved and it will ensure the good performances for feature extraction stage as shown in the next section.

For the signal considered in the previous example, the OCWT-IS performs also the separation of the three component of the signal. Nevertheless, if the time-frequency components are close, having the same supports in time and frequency domains, the complete separation is impossible. This phenomenon is illustrated in the next figure, using a sound generated by a long-finned pilot whale (*Globicephala melas*). The sampling frequency is 44.1 kHz and we have taken into account an observation of 5.92 seconds. The results are also presented for the three detectors with the same parameters as in the previous example.

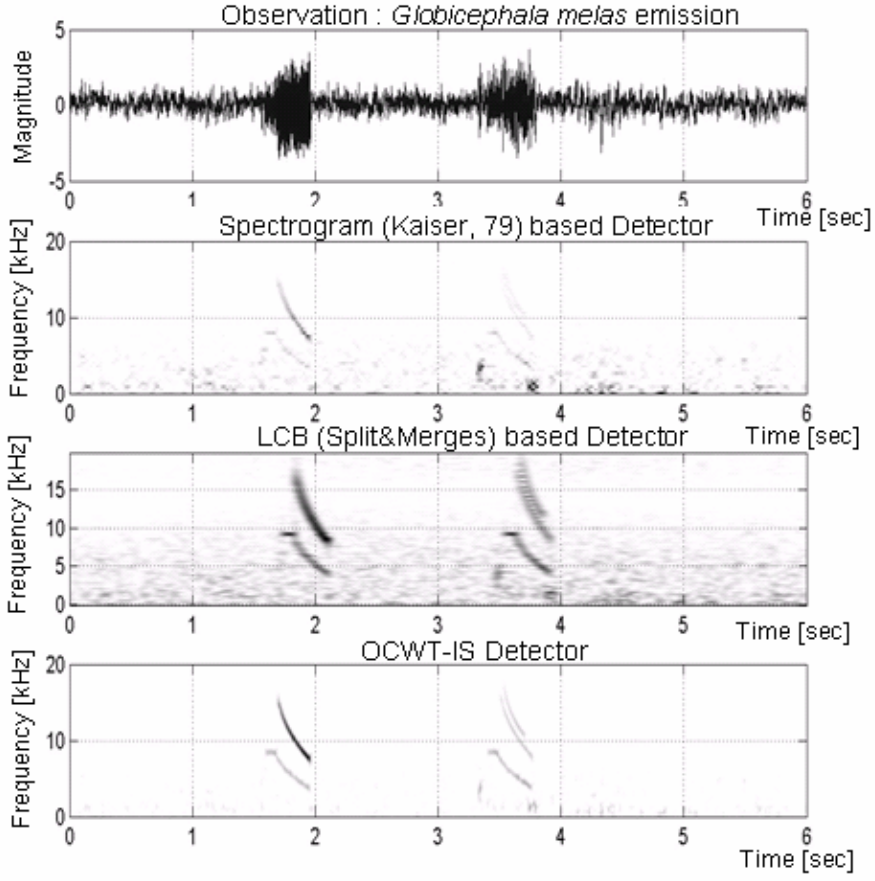


Figure 8. Detection results for a Globicephala melas emission

In the case of this test signal, the detected ROIs contain many time-frequency components enclosed in the same T-F ROI.

We close this section with some remarks about the operational aspects. The proposed detector OCWT-IS is well adapted to a real passive context : the detection is performed without any information about the received signal. On the other hand, the detection is automatically done : there is only one parameter to set – the confidence interval. Practically, the performance tests have been done with a pre-defined value 85%. Therefore, in practical context, the method can be considered as completely adaptive. The examples provided for two real marine-mammal observations show that the OCWT-IS successfully performs without any changes of the filter bank structure.

PART 2. UNDERWATER SIGNAL CHARACTERIZATION

4. THE PRINCIPLE OF WARPING-BASED TIME-FREQUENCY REPRESENTATIONS

The examples illustrated in the previous section show that, in real cases, the underwater signals are characterized by three items (figure 7, 8). Firstly, signals are particularly non-stationary. Therefore, it explains the interest for time-frequency methods. Secondly, the time-frequency features of the signals are generally non-linear. Thirdly, a ROI can contain many time-frequency structures with a different T-F structure.

In this section we will study a class of methods to deal with the non-linear behavior of time-frequency structures of the analyzed signals. We will see that, in spite of its theoretical potential, this method is limited when a given ROI contains many T-F structures. A possible approach will be proposed in the section 5.

At this stage we consider the problem of time-frequency representation of a signal having non-linear T-F structures.

Typical time-frequency analysis, imposed by the non-stationarity of the sounds generated by the marine-mammals, involves the use of Cohen's class members such as the spectrogram or the Wigner-Ville Distribution - WVD [11] or wavelet-based methods [8]. While these methods are natural for signals containing pulse, sinusoidal, linear chirps, there are many other signal classes that are not well described in terms of time, frequency or scale. It is also the case of the majority of underwater signal classes. Matching these types of signals requires new joint distribution with different instantaneous frequency and group delay localization properties. One of the most known techniques is the unitary similarity transformations. This type of transformations provides a simple powerful tool for generating new classes of joint distributions based on concepts different from time, frequency and scale [12]. These new signal representations focus on the critical characteristics of large classes of signals, and, hence, prove to be useful for representing and processing signals that are not well matched by current techniques. Actually, it is possible to construct (via unitary transformation) distributions that match almost any one-to-one group delay or instantaneous frequency characteristics. One of the most known unitary transforms is the *axis transformation* [12], defined for a signal $s(t)$ as an operator $\mathbf{U}$ on $L^2(\Re)$, whose effect is given by

$$\left(\mathbf{U}s\right)\left(t\right) = \left|\dot{w}\left(t\right)\right|^{1/2} s\left[w\left(t\right)\right] \tag{10}$$

where w is a smooth, one-to-one function comprising a large subclass of unitary transformations ([12]). The term $\dot{w}$ stands for the first derivative of the function w. The functions $w\left(t\right) = e^t$ and $w\left(t\right) = \left|t\right|^k \operatorname{sgn}\left(t\right), k \neq 0$ provide examples of useful warping functions ([18], [19]). Generally, these functions are chosen to ensure the "linearization" of a signal time-frequency behavior. Consequently, for a signal expressed as

$$s\left(t\right) = e^{j2\pi\left(f_0 t - \beta m\left(t\right)\right)} \tag{11}$$

(where $m(t)$ is the frequency modulation law and β is the rate of modulation), the associated warping function is given by [20] :

$$w(t) = m^{-1}(t) \tag{12}$$

The application of this operator produces the linearization of the time-frequency content. The following example illustrates this property. For the signal given by

$$s(t) = e^{j2\pi(0.38t - 0.02t^{1.3})} \tag{13}$$

the associated warping function can be defined as [19]

$$w(t) = t^{\frac{1}{1.3}} \tag{14}$$

In the next figure the effect of the corresponding warping operator in the time frequency plane is plotted. The WVD of the original signal (figure 9, middle part) illustrates the presence of the cross terms which is caused by the non-linearity of its IFL (figure 9). Application of the warping operator (14) linearizes the time-frequency content of the original signals as indicated in figure 9 (left subplot). Consequently, the time warping operator effect consists in transforming a non-linear frequency modulation in a tone. In the case of the frequency warping (figure 9, right side), linearization is done in a dual manner : a non-linear frequency modulation becomes a pulse. According to the concepts introduced in [7] and [12] the linearization is achieved by computing the WVD for the new time and frequency coordinates, given by the general expression (15).

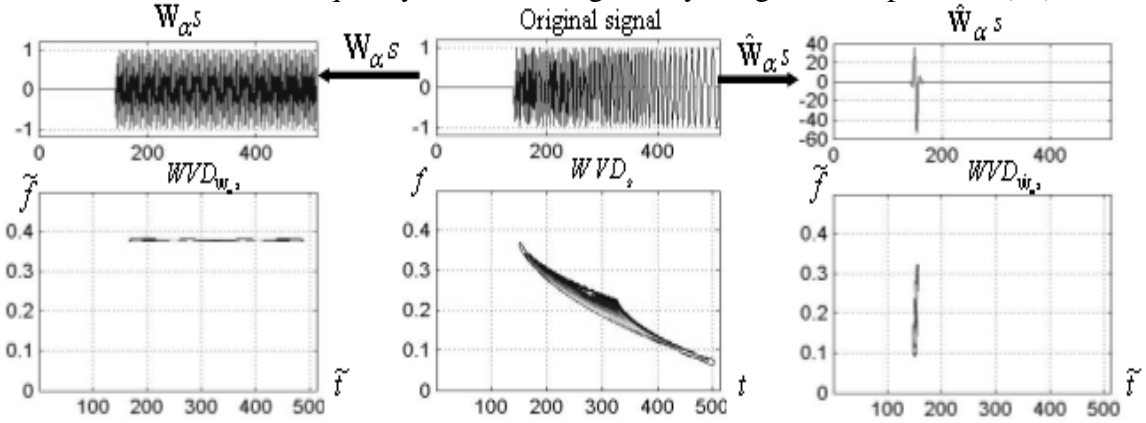


Figure 9. The time-frequency effect of the signal warping

Linearization of the time-frequency content, provided by signal warping, is materialized by taking the WVD of the transformed signal (figure 9, left side).

The new time-frequency coordinates are related to the standard ones via [12]

$$\begin{aligned} &\tilde{t} = w(t) \;\; ; \;\; \tilde{f} = f\dot{w}^{-1}\left(w(t)\right) - \text{for time warping operator} \\ &\tilde{t} = t\dot{w}^{-1}\left(w(f)\right) \;\; ; \;\; \tilde{f} = w(f) - \text{for frequency warping operator} \end{aligned} \tag{15}$$

where w^{-1} is the inverse function of w and $\dot{w}^{-1}$ represents its first derivative. In the case of frequency warping operators, the deformation function w is applied in the frequency domain. To specify its application in the spectral domain, this operator is often denoted by $\hat{w}$.

As exemplified in figure 9, the effect of warping operator is the linearization of time-frequency structure of the analyzed signal. This linearization is materialized in the time-frequency plan by computing the TFR of the warped signal. Usually, the TFRs of the Cohen's class (CTFR) are used. The general form of the CTFR is :

$$CTFR_x(t,f) = \iint \left[\int x\left(t - \frac{\tau}{2}\right) x^*\left(t + \frac{\tau}{2}\right) e^{j2\pi\xi t} dt \right] \phi(\tau,\xi) e^{-j2\pi(\tau f + \xi t)} d\tau d\xi \tag{16}$$

where $\int x\left(t-\frac{\tau}{2}\right)x^*\left(t+\frac{\tau}{2}\right)e^{j2\pi\xi t}dt$ is the ambiguity function of the signal *x*. The weighting function $\phi(\tau,\xi)$, called **kernel** of the distribution, determines completely the properties of the CTFR. Different choices for the kernel function ϕ yield widely different time-frequency representations.

Finally, the warped-based TFR (WTFR) is obtained by unwarping the coordinates of the CTFR [12] using the following formula :

$$WTFR_s^w(t,f) = CTFR_{Ws}\left(w^{-1}(t), f\dot{w}\left(w^{-1}(t)\right)\right) \tag{17}$$

The unwarping procedure consists in transforming the coordinates of the WVD image according to the real time and frequency coordinates (for example figure 9, left side). The result is a cross-terms free time-frequency distribution, providing a correct concentration of the energy across the theoretical instantaneous frequency law. The lack of cross-terms is explained by the linearity of the unwarping transform applied on a cross-term-free 2D data (i.e., the WVD of warped signal). This main property of the warping-based TFR is illustrated in the next figure for the signal given in (13).

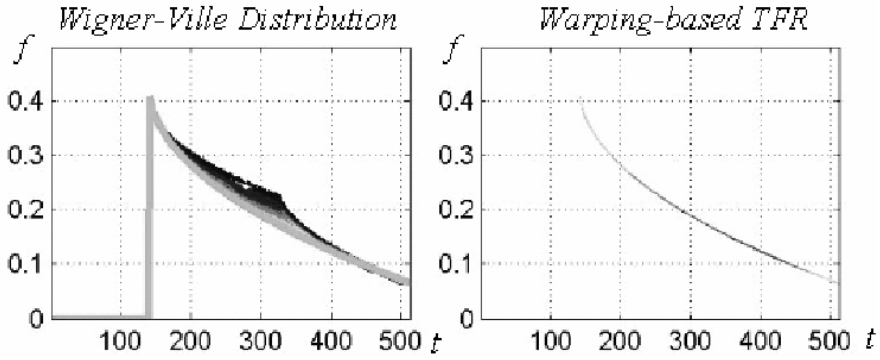


Figure 10. Warping-based TFR vs Wigner-Ville distribution

In spite of its elegance, the principle of the warping-based time-frequency analysis is limited for real data context. Actually, there are two major limitations.

Firstly, the implementation scheme [21] of the warping operators involves a drastic increasing of the signal length. The increasing of sample number, performed by a zero padding operation, is imposed by the interpolation procedure. This one produces new signal values associated to the new warped coordinates. In [21], it is shown that a ten-time increasing of the signal size is necessary to obtain a sufficient accurate estimation of the new warped signal values. Since the warping-based TFR principle supposes the computation of a Cohen's class distribution of the warped signal, the increasing of the signal size leads to serious problems in the case of real data, represented by a large number of samples. In this context, a potential solution is to analyze **only** the signal components bounded by the time-frequency ROIs. Therefore, the solution we adopted to overcome this limitation is to analyze the signal of the detected ROIs by the method previously presented.

Secondly, as it was observed in section 3, the marine-mammal sounds generally contain many non-linear time-frequency structures. The warping-based time-frequency analysis of these types of structures imposes firstly the separation and the characterization of each time-frequency component. Nevertheless, this operation constitutes a very difficult one because it must to be done in a passive context (i.e., there is no any a priori information about the time-frequency structures).

One possible method was proposed in [22] for marine-mammal sounds. The involved technique is based on the signal decomposition on an *extended basis function dictionary* that can accurately represent a multiple TF structure signal, even for the non-linear case [23]. This dictionary contains different frequency modulation (FM) chirps

$$\psi\left(t,\overline{\theta}\right) = e^{j2\pi c\xi(t)} \tag{18}$$

which is uniquely specified by its FM rate c, its monotonic phase function $\xi(t)$ and the parameter set $\overline{\theta} = [t_0, f_0, c, d]$ (which contains the time and frequency centres, chirp rate and signal duration, respectively). The dictionary may contain the following FM chirps : sinusoids ($\xi(t) = t$), linear chirps ($\xi(t) = t^2$) or k-power FM (PFM) chirps ($\xi(t) = t^k$).

Using this extended dictionary and a projection algorithm ("Matching Pursuit" [24], for example) we can obtain an optimal approximation for a given signal as a sum of the corresponding basis function, weighted by the decomposition coefficients. These coefficients are obtained by the inner product between the signal and the associated basis functions. This approximation leads to an adaptive time-frequency representation of the analyzed signal [23]. On the other hand, the signal can be "parameterized" by the basis function features (c and $\xi(t)$), achieving also a data set for signal characterization.

Nevertheless, in order to achieve good analysis performances, the dictionary size must be very large. More precisely, the method proposed in [23] is based on an extended dictionary composed by a set of individual dictionaries associated to the modulation types. Obviously, the dictionary grows with the number of the considered modulations.

For example, the dictionary involved in the example illustrated in [22] contains 20,971,520 elementary functions (knowing that 10 modulation types have been considered). Consequently, some computational difficulties appear when such number of elementary functions must be managed in a sequential manner (as proposed in [23]).

On the other hand, the size of the required dictionary can be reduced considering only the signal within the detected ROIs. For example, in the case of the second detected ROI in figure 8, the parameter t_0 will be searched only in the interval [3.3,3.8]sec instead of [0,6]sec if all the signal samples are considered. For the same precision P, the number of parameters t_0 we are looking for are 0.5/P and 6/P, respectively. We remark that in the case of the ROI signal processing, the number of the t_0 is 30 times smaller than in the case of processing of all signal samples.

In the next section we propose an alternative method which is based on signal component separation using different warping operators. This method will be exemplified in the case of the signal emitted by a long-finned pilot whale.

5. WARPING-BASED SIGNAL DECOMPOSITION

In the previous section we have presented the warping technique principle and its application for warping-based time-frequency representation. We have seen that the implementation of warping operators involves some assumptions concerning the nature of the analyzed time-frequency structures (relation 13). Moreover, the warping-based TFR is obtained by processing the time-frequency distribution of the warped signal (relation 17), and, consequently, the size of the analyzed signal cannot be arbitrarily large. Since the size problem is partially solved by the ROIs detection procedure, the knowledge of the signal time-frequency nature remains a real difficulty in the context of signal analysis via non-linear TFR.

In this section we propose a new method for feature extraction which is applied to the signal of previously detected ROIs. The method is based on the application, in a parallel manner, of many warping operators. Hence, for each issued signal representation; we look for linear time-frequency structures. In fact, applying the warping operation, the T-F components having a T-F behavior associated to the warping law become linear chirp-like components. These new structures can be estimated by chirplet transform principle [25], implemented, for example, by Matching Pursuit method [24].

To illustrate the features of the new method, Warping–based Signal Decomposition (WSD), let consider the following expression for a signal belonging to the ROI k :

$$\text{ROI } k : s(t) = \sum_{C_i} \sum_{j=1}^{N_{C_i}} s_j^{(C_i)} \tag{19}$$

where $\{C_i\}$ is the set of the time-frequency structure types (classes) and $\left\{s_j^{(C_i)}\right\}$ is the set of signal component having a C_i-type time-frequency behavior. C_i can denote the hyperbolic frequency modulation (HFM) class [18], k^{th} order Power Frequency Modulation (PFM) [19], linear chirp class, etc. N_{C_i} is the number of C_i-type time-frequency structures which compose $s(t)$.

We consider $\{\mathbf{W_i}\}$ – the family of warping operators associated to the classes $\{C_i\}$. The effect of the application of an element of this family, $\mathbf{W_i}$, on $s(t)$ is the linearization of time-frequency structures corresponding to $\left\{s_j^{(C_i)}\right\}$:

$$WVD_{\mathbf{W_i}s_j^{C_i}}(t,f) \to \delta\left(f - \underbrace{(ct+f_0)}_{chirp}\right) \qquad (20)$$

$$\mathbf{W_i} \to \text{warping operator associated to } C_i$$

Therefore, the components which become chirp signals in the new representation space form the C_i-type signal subset (denoted by $SbC_i = \left\{\tilde{s}_j^{C_i}\right\}$;). They can be obtained by estimating the N_{C_i} chirps $\left\{\psi_k\left(t,\bar{\theta}\right)\right\}_{k=\overline{1,N}_{C_i}}$ and modulating these functions with the associated modulation operator :

$$\tilde{s}_j^{C_i}(t) = \left(\mathbf{M_i}\psi\right)\left(t,\bar{\theta}\right);\ j=1,...,N_{C_i} \qquad (21)$$

The estimation of the chirplet functions can be performed by a Matching Pursuit-based method [24]. In fact, the idea is to search iteratively, for the $\mathbf{W_i}$ warped version of the signal, the local best matched functions, selected from a dictionary.

$$\begin{aligned} \hat{\bar{\theta}} &= \max_{\bar{\theta}} \left| \left\langle \mathbf{W_i} s(t), \psi\left(t,\bar{\theta}\right) \right\rangle \right| \\ \hat{\bar{\theta}} &= \left[\hat{t}_0, \hat{f}_0, \hat{c}, \hat{d}\right] \end{aligned} \qquad (22)$$

The parameter set, $\hat{\bar{\theta}} = \left[\hat{t}_0, \hat{f}_0, \hat{c}, \hat{d}\right]$, contains the estimation of time and frequency centres chirp rate and signal duration, respectively. Generally, in order to estimate correctly the chirp-like components of a signal, the dictionary must be large enough. Nevertheless, using as *a priori* information the detected ROIs parameters, the dictionary size can be considerably reduced [22].

Using a battery of warping operators and a Matching Pursuit-based Chirplet Estimation (MPCE) procedure, we can estimate all parameters of signal components. The WSD method is illustrated in the figure 11.

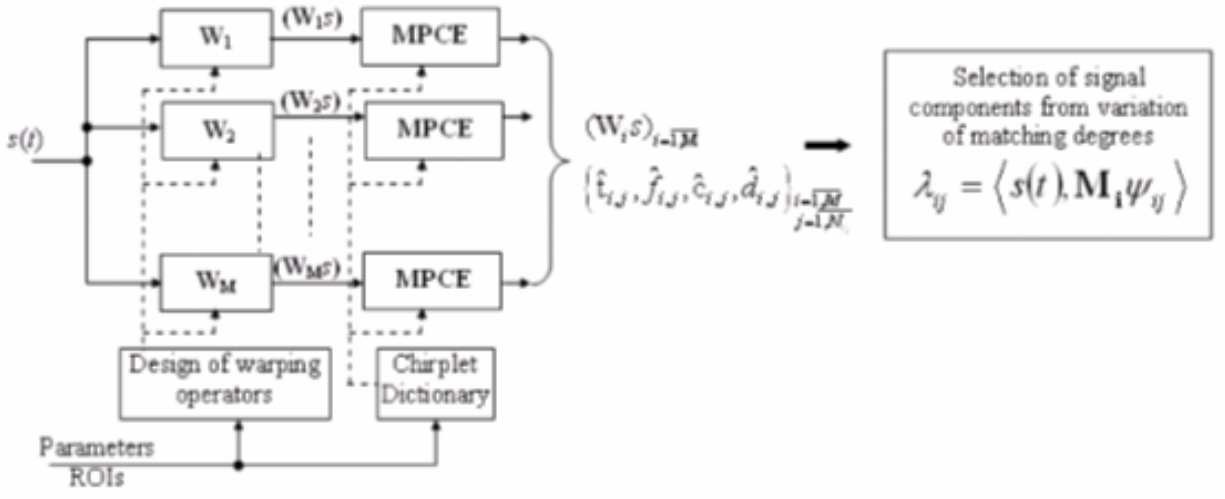


Figure 11. Warping-based Signal Decomposition diagram

The warping operators are implemented according to the desired warping law [21] and the parameters of ROIs (the time and frequency centres, time and frequency supports). These parameters are also used to pre-define the chirplet dictionary. Furthermore, the algorithm is parallelized in M braches, where M is the number of T-F behaviour types of interest. After the application of $\mathbf{W_i}$, all the C_i time-frequency components of the signal will be linearized and the others will have an arbitrary shapes. Consequently, in each branch, looking (via MPCE algorithm) for linear T-F structures, the components $\mathbf{W_i}s(t)$ will be well matched, thanks to their linear T-F structures. For the other structures, even if they are more energetic, the chirplet decomposition coefficients will be smaller than the ones corresponding to the chirplet structures.

The parallel organization of this algorithm allows reduction of the required dictionary size. It avoids the sequential exploration of the signal using several dictionary types as proposed in [23]. By working with a parallel scheme, the warped signal corresponding to each branch is projected on the **same** chirplet dictionary. In other words, only one dictionary is required. Comparing with the method proposed in [23] it is equivalent with the dictionary size reducing

In practical applications [24] the choice of the number of searched chirplet functions constitutes a complex problem. If a small dictionary is considered, some of signal time-frequency structures could not be modelled. This is due to the fact that the corresponding basis functions (that would match these structures) don't belong to the dictionary. At opposite, in case of a large dictionary, the computation time will increase since a large number of functions will be taken into account. Since we have no idea about how the processed signal looks like, the choice of dictionary could be set up according to the application. In our case, since we are interested in signal feature extraction, it is appropriate to oversize this parameter.

Finally, the estimations obtained from each branch are used to generate, by the inverse warping operation, the corresponding T-F components. Naturally, the component number being oversized, the number of estimated components will be lager than the real one. Consequently, the WSD algorithm finishes with a post-operation analysis in order to identify the real T-F components of the analyzed signal. As depicted in the figure 11, we propose the computing of the matching degrees between the signal and all the estimated components, given by the following function :

$$\lambda_{ij} = \left\langle s(t), \mathbf{M_i}\psi_{ij} \right\rangle \tag{23}$$

Analyzing the variation of the values λ_{ij}, by a conjugate gradient method [26], we can estimate the number of components of each time-frequency non-linearity types. However, if some inherent errors occur, the artifacts (i.e. the retained components which do not belong to the analyzed signal) will have relative small contributions on the signal modelling.

Therefore, after the WSD performing, the signal is modelled as

$$\tilde{s}(t) = \sum_{C_i} \sum_{j=1}^{\tilde{N}_{C_i}} \lambda_{ij} \left(\mathbf{M_i}\psi_{ij}\right)(t) \tag{24}$$

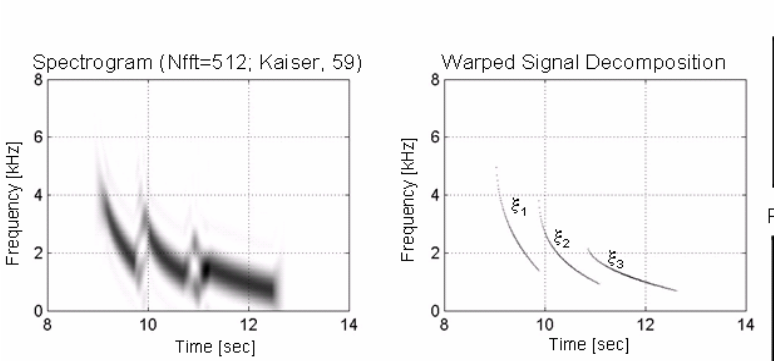
where $\left(\mathbf{M_i}\psi_{ij}\right)(t)$ is the signal estimated component, completely characterized by the parameters of associated chirps signal (time and frequency centres, bandwidth and duration) and the instantaneous frequency law provided by the $\mathbf{M_i}$.

To estimate the frequency tones a spectral estimation method [27] (MUSIC, for example) can be applied on the residual signal. The complete signal characterization is also expressed as :

$$\tilde{s}(t) = \sum_{C_i} \sum_{j=1}^{\tilde{N}_{C_i}} \lambda_{ij} \left(\mathrm{M_i}\psi_{ij}\right)(t) + \sum \lambda_k e^{j2\pi f_k t} \tag{25}$$

where $\{f_k\}$ are the frequency estimations of the tones existing in the signal T-F structure.

In order to show the performing of the WSD algorithm we consider a synthetic signal composed by three frequency modulations.



Real parameters of T-F structures

ξ	t_{0k} [sec]	f_{0k} [kHz]	D_k [sec]	β_k	$m(t)$
ξ_1	9	4.1	0.9	-0.7	$t^{0.4}$
ξ_2	9.85	3.2	1.25	-0.8	$t^{0.3}$
ξ_3	10.8	1.9	1.8	-0.11	$t^{0.5}$

Parameters of T-F components estimated by WSD

ξ	t_{0k} [sec]	f_{0k} [kHz]	D_k [sec]	β_k	$m(t)$
ξ_1	9.05	4	0.83	-0.68	$t^{0.4}$
ξ_2	9.94	3.15	1.04	-0.78	$t^{0.3}$
ξ_3	10.87	1.89	1.75	-0.108	$t^{0.5}$

Figure 12. Warping-based Signal Decomposition for a synthetic signal

The battery of the warping operators used for this example contains the power warping operators of orders *k=0.1÷0.9*. The parameters of the three time-frequency components are accurately estimated as indicated in the figure 12. In the same time, the time-frequency representation provided by the WSD avoids the resolution problems of the spectrogram conserving also the information about the true time-frequency content of the signal.

Furthermore, we exemplify the proposed WSD algorithm in the case of a real underwater data extracted from a ROI. As a test signal we consider the second arrival of the Globicephala Melas emission (figure 9). The battery of the warping operators contains the following elements :

- the power frequency warping operators [19]

$$\mathbf{W}_{\mathbf{t^k}} = t^{1/k} \tag{26}$$

of orders k=0.1, 0.3, 0.4, 0.5, 0.6, 0.8. These warping operators are associated to the instantaneous frequency law of type t^k.

- the logarithmic frequency operator [18]

$$\mathbf{W}_{\mathbf{log}} = \exp(t) \tag{27}$$

which corresponds to logarithmic frequency modulations $log(t)$.

The processing algorithm is depicted in the next figure.

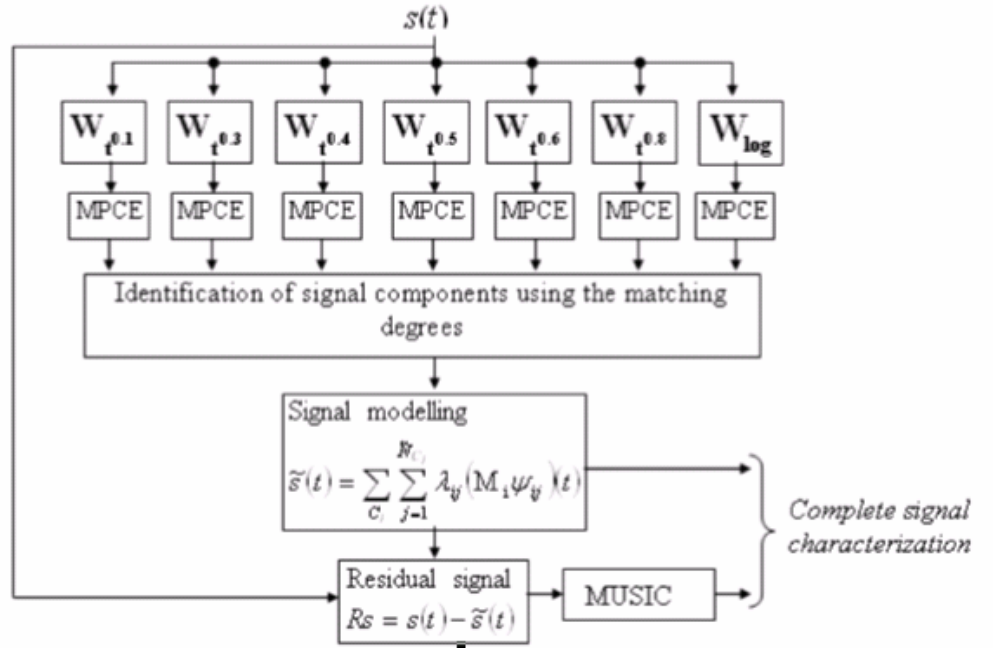


Figure 13. WSD algorithm for characterization of a marine-mammal emission

After the signal transformations according to each warping operator, the chirplet estimation is performed. For each signal transformation, we estimate *8* chirplet functions. This number has been experimentally chosen since for the tests that have been done it provides an accurate estimation.

The matching degrees, estimated with (23), provide information about the components that are parts of signal. Using the residual signal, we estimate the tone components of the signal.

In the next figure we show the time-frequency representation of one of the ROI detected by the OCWT-IS algorithm (figure 8).

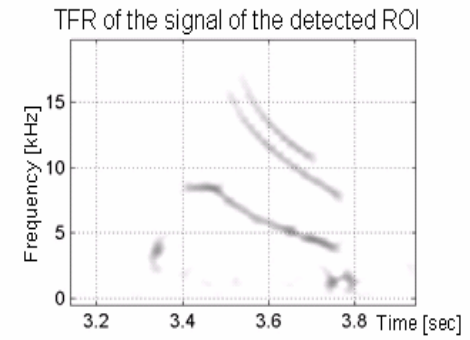


Figure 14. TFR of the considered ROI

Obviously, the time-frequency representation of the detected signal provides only general information about its time-frequency behavior. For this signal, the results of warping operators and the estimation of the chirplet components are illustrated in the figure 15.

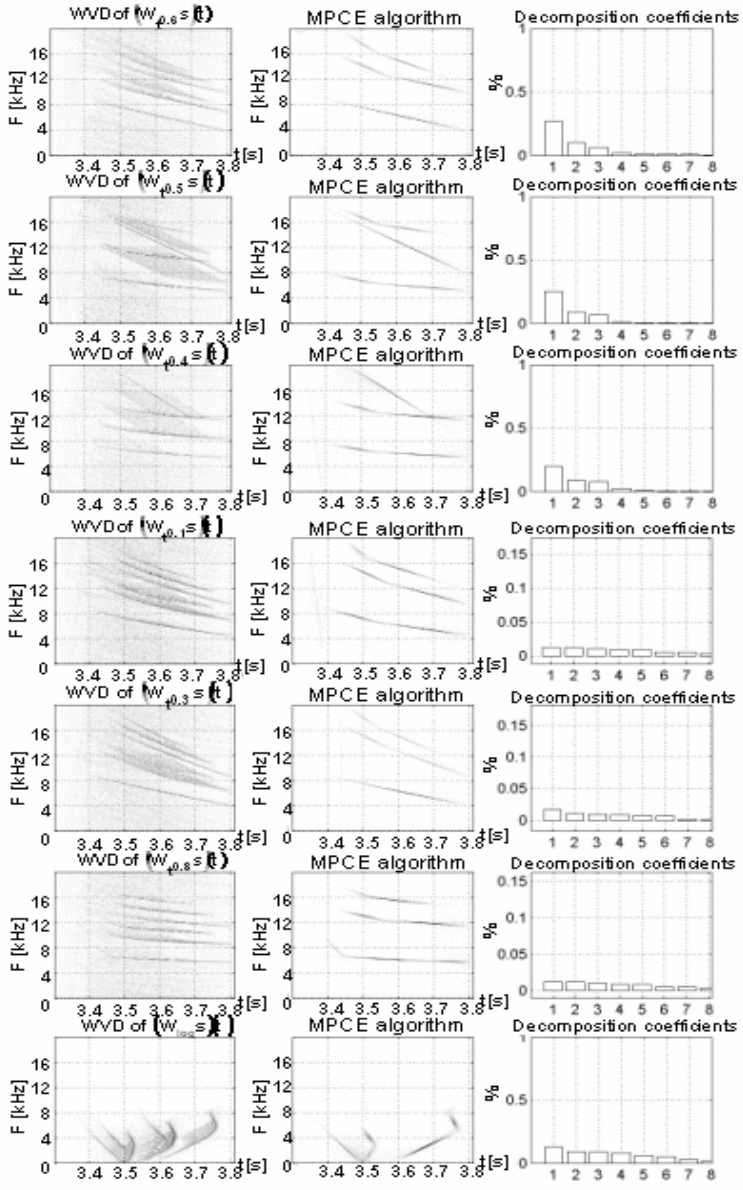


Figure 15. Globicephala Melas emission processing using WSD algorithm

The left subplots illustrate the WVDs of the signals obtained after the warping operators composing the WSD scheme (figure 11). The time-frequency complexity and the noise existing in the analyzed ROI do not permit to distinguish, using the information provided by the WVD, the signal T-F components. For each branch, the signal approximation by chirplet function (MPCE) is represented in the middle subplots. Each warped signal will be approximated by eight chirplet functions using the algorithm proposed in [28].

In the case of power warping of orders 0.6, 0.5, 0.4 we observe that one dominant linear T-F component appears. It is also traduced in the distribution of the decomposition coefficients, depicted, for each branch, in the right side of figure 15. The largest coefficient corresponds to the major chirplet function. The existence of this dominant chirplet coefficient indicates that it corresponds to the power frequency modulations of order 0.6, 0.5 and 0.4, respectively. The other chirplets functions, which are less energetic, correspond to different PFM types. This is given by the amplitude of the associated coefficients and by analyzing comparatively the results for the orders 0.6, 0.5, 0.4. The linear chirplets (approximating a PFM), which are connected, do not appear in the decomposition associated to the correct warping operator. Therefore, three components are estimated : a 0.6^{th} order PFM, a 0.5^{th} order PFM and 0.4^{th} order PFM. The other parameters are given by the elements of estimated chirplet functions. For example, in the case of the PFM $t^{0.6}$, the chirplet function obtained by the corresponding warping is characterized by the following parameters : duration D=0.35 sec, a bandwidth B=4.82 kHz, a time centre t_0=3.62 sec and the frequency centre f_0=2.41 kHz. The relationship between the parameters of the chirplet function and the associated non-linear T-F structure is a consequence of the unitary property which characterizes the warping operators.

For the other warped signals, the decomposition coefficients are energetically reduced. On the other hand, the T-F representations of the estimated chiplets do not contain a dominant chirplet function (all 8 chirplets are represented and the decomposition coefficients have a reduced dynamic – figure 15). Consequently, the analyzed signal does not contain the assumed non-linear T-F natures.

Finally, the residual signal is estimated. Applying the MUSIC method we estimate the tone components existing in the signal structure. In the figure 16, the MUSIC pseudo spectrum and the time-frequency distribution of the remained tone are represented.

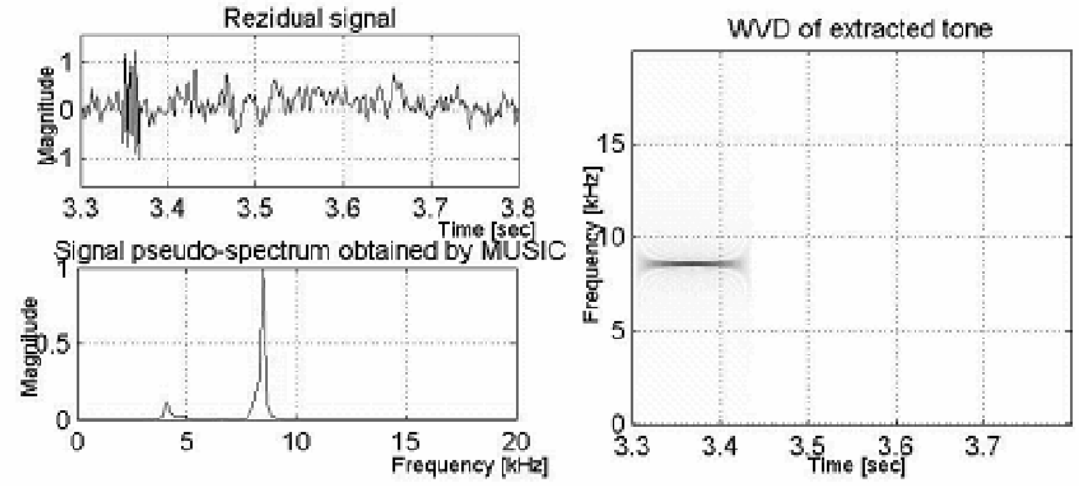


Figure 16. Processing of residual signal

The pseudo-spectrum, evaluated by MUSIC method, indicates the existence of a single tone component, located to the frequency 8.5 kHz.

After the estimation of all searched components, the combination of the adapted TFR of each extracted components provides the global time-frequency representation of the analyzed signal.

$$GTFR_s(t,f) \approx \sum_{C_i} \sum_{j=1}^{N_{C_i}} \left|\lambda_{ij}\right|^2 WTFR^{C_i}_{\mathbf{M}_{ij}\psi_{ij}}(t,f) \tag{28}$$

The issued T-F image constitutes the *non-parametric information* provided by the WSD algorithm. For the signal extracted from the detected ROI illustrated in figure 14, the global TFR is depicted in the figure 17.

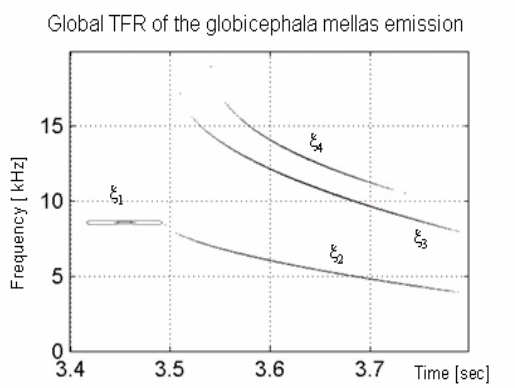


Figure 17. Global TFR of the considered test signal

The issued time-frequency image provides a correct representation of the T-F components of the analyzed signal. Moreover, the obtained representation is cross-terms free and uncorrupted by any artefacts (noise, false components, etc).

On the other hand, this algorithm furnishes a *parametric description* of T-F components of signal. Therefore, these components can be expressed by the corresponding instantaneous frequency laws (IFL). For the considered test signal, the estimated IFLs (represented in figure 17) of components are expressed as :

$$\xi_1(t)[kHz] = 8.55;\ \ t \in [3.3, 3.42]\text{sec}\ . \tag{29.a}$$

$$\xi_2(t)[kHz] = 8.57 - 0.0054 \cdot t^{0.6};\ \ t \in [3.45, 3.8]\text{sec}. \tag{29.b}$$

$$\xi_3(t)[kHz] = 17.3 - 0.018 \cdot t^{0.5};\ \ t \in [3.43, 3.8]\text{sec}. \tag{29.c}$$

$$\xi_4(t)[kHz] = 19.2 - 0.035 \cdot t^{0.4};\ \ t \in [3.4, 3.75]\text{sec}. \tag{29.d}$$

The analytical estimation of the IFLs provides useful information for the further processing stage. The parameters of these IFLs constitute the extracted feature of the analyzed signal and they could be used for signal identification, classification or signal characteristic tracking.

In conclusion, the method proposed in this section, based on the exploration of the warped versions of the analyzed signal (Warping based Signal Decomposition), provides a *correct* and *rich* information about the studied process. The *correct* means that the most important T-F components of the analyzed signal are characterized. Of course, some components may be lost, but, as it was illustrated by the previous example, the most representative T-F structures of the signal are extracted. On the other hand, the information is *rich*, because we obtain a lot of parameters, such as, modulation type, bandwidth, duration, location in time, etc.

In the next section we will illustrate more results obtained for real marine-mammal signals.

6. RESULTS FOR REAL DATA

In this section we present the results provided by the algorithm proposed previously. Three marine-mammal signals are the subjects of our test. The first one is an emission of a *Tursiops Truncatus* whose ROI detection is illustrated in the figure 7. The second is the signal bounded by the detected ROIs for a *Globicephala melas* emission, illustrated in figure 8. The both signals are sequences of underwater recordings provided by Technical Centre of Naval Systems from French Department of Weapons.

The last tested signal corresponds to a common dolphin whistle, downloaded from the web site www.dolphlinear.com/de_sound.

For these test signals we apply the proposed algorithm. In the first stage, the signal regions of interest are detected, using the OCWT-IS based detector. In our work context – feature extraction – this first operation has two main purposes. Firstly, the reduction of signal length offers the conditions for the implementation of warping operators, used in the further stage. Secondly, processing signal only in a T-F region minimizes the number of artifacts caused to the noise.

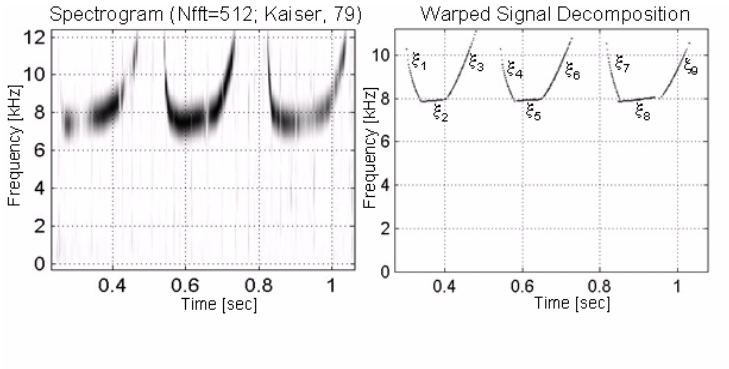
After the ROIs detection, for each corresponding signals, we apply the feature extraction procedure obtaining also the parameters of T-F components. The algorithm proposed in this paper has been designed to deal with multi-component and non-linear time-frequency behaviors. Therefore, the combination between the warping operator concept and the chirplet decomposition has been used. More precisely, the warping operators transform the non-linear T-F structures in linear ones. Furthermore, the new structures are estimated using a chirplet decomposition method. The final results are the estimation of the instantaneous frequency law of the T-F components. Generally, the estimated IFL of the k^{th} component can be expressed as :

$$\xi_k(t) = f_{0k} + \beta_k m_k(t); t \in \left[t_{0k} - \frac{D_k}{2}; t_{0k} + \frac{D_k}{2} \right] \tag{30}$$

where t_{0k}, f_{0k} are the centers in time and frequency coordinates, D_k the duration, β_k the modulation rate. These parameters are provided by the MPCE algorithm when applied on the linear T-F structure obtained after the warping operation $\mathbf{W_k}$, directly related to the modulation law, $m_k(t)$.

The virtues of this algorithm are illustrated by the following examples. We compare the obtained T-F image with the one provided by the spectrogram, which is the main tool for underwater signal analysis.

Firstly, we consider a real signal corresponding on the detected ROIs of a *Tursiops Truncatus* emission (figure 7). Applying the WSD algorithm (similar to the one illustrated in figure 12) for the signals existing in the detected ROIs, we estimate nine T-F components, given by the figure 18.



Parameters of T-F components estimated by WSD

ξ	t_{0k} [sec]	f_{0k} [kHz]	D_k [sec]	β_k/F_s	$m(t)$
ξ_1	0.32	8.75	0.042	-0.028	$t^{0.4}$
ξ_2	0.372	7.97	0.045	$2.85 \cdot 10^{-5}$	$t^{1.1}$
ξ_3	0.426	9.62	0.04	$8.6 \cdot 10^{-4}$	$t^{1.4}$
ξ_4	0.612	9.12	0.045	-0.028	$t^{0.4}$
ξ_5	0.664	7.97	0.05	$1.71 \cdot 10^{-5}$	$t^{1.1}$
ξ_6	0.719	9.58	0.044	$7.73 \cdot 10^{-4}$	$t^{1.4}$
ξ_7	0.89	9.2	0.037	-0.02	$t^{0.4}$
ξ_8	0.93	7.95	0.043	$2.85 \cdot 10^{-5}$	$t^{1.1}$
ξ_9	1	9.61	0.045	$7.71 \cdot 10^{-4}$	$t^{1.4}$

Figure 18. Feature extraction from a Tursiops Marineland emission

In the case of spectrogram, the representation of T-F contents is affected by the trade-off between time and frequency resolution. A better representation may be achieved by testing many spectrogram parameters (window type and length).

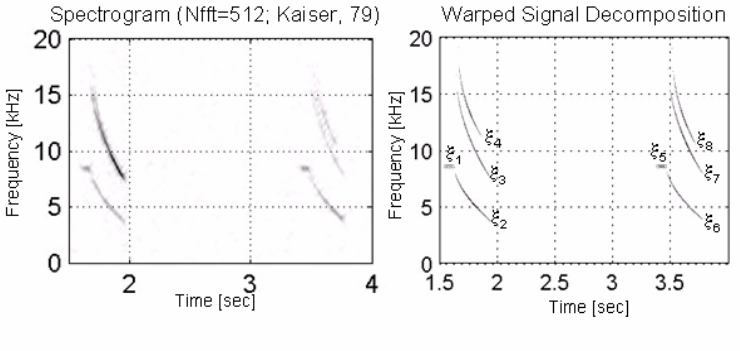
Using the Warped Based Decomposition, for each detected ROI, three PFM components are estimated. The parameters of all PFM components are given in figure 18. The correspondence between these parameters and the associated T-F components are represented in the same figure. We observe that the T-F components estimated by WSD algorithm are visually close to the T-F behavior illustrated by the spectrogram. In addition, the WSD provides an analytical description of the IFLs. Based on this description, the signal emitted by a Tursiops Truncatus is composed by three time-frequency regions of interest. Each region of interest contains three power frequency modulations : the first one frequency decreasing, the second and third ones – frequency increasing.

The short durations of the T-F components prove the capability of the WSD to deal with transient signals. Since the visual T-F information provided by both spectrogram and WSD

algorithm indicates three similar component sets, the values of the parameters of each ROIs justify objectively this similarity.

Consequently, the both information types – visual and parametric - illustrated in figure 18 provides a complete description of analyzed signal.

We continue our analysis with a *Globicephala Melas* emission which has been the subject of the example given in previous section. Now, we consider both detected ROIs illustrated in figure 9. The analysis, via WSD, is illustrated in figure 19. As in the previous example, the spectrogram TFR is also pictured.



Parameters of T-F components estimated by WSD

ξ	t_{0k} [sec]	f_{0k} [kHz]	D_k [sec]	β_k/F_s	$m(t)$
ξ_1	1.61	8.5	0.12	0	-
ξ_2	1.97	4.7	0.35	-0.0054	t^{06}
ξ_3	1.95	10.92	0.37	-0.018	t^{05}
ξ_4	1.9	13.2	0.35	-0.035	t^{04}
ξ_5	3.36	8.55	0.12	0	-
ξ_6	3.62	4.82	0.35	-0.0054	t^{06}
ξ_7	3.61	11.02	0.37	-0.018	t^{05}
ξ_8	3.57	13.81	0.35	-0.035	t^{04}

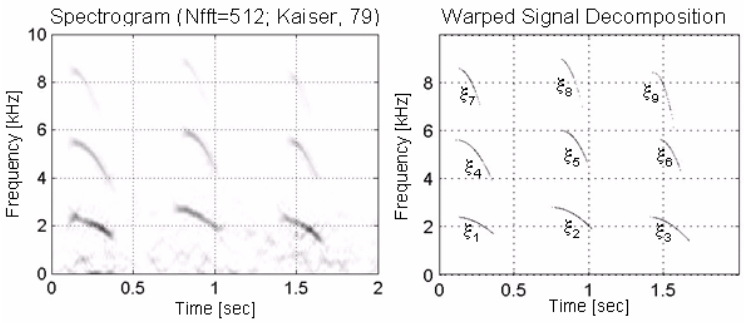
Figure 19. Feature extraction from a Globicephala melas emission

Obviously, as it is visually indicated by both spectrogram and WSD, the analyzed sequence is composed by two similar signal packets (figure 19). Each packet contains one tone signal (ξ_1 and ξ_5, respectively) and three power frequency modulations. In addition, the feature extraction provided by WSD algorithm objectively indicates the similarity degree between the both packets. Examining the table of figure 19, we observe that similar modulation types act in both packets : the orders of the PFMs and the modulation rates have the same values. Nevertheless, we can observe that the frequency centers of the T-F components have not the same values : the frequency center values of the second ROI are superior than the analogue values of the first packet. It could be explained by the Doppler effect induced by the source-receiver relative motion.

We can also observe that the differences between this parameter are not the same. Consequently, the propagation environment has frequency dependent characteristics. The evaluation of the parameter variation, for many received characteristics, may lead to the quantitative analysis of propagation phenomenon of the environment where the source evaluates.

With the remark that the spectrogram is unable to quantify the signal parameters variation, this example illustrates the capability of the WSD algorithm to analytically express the time-frequency features.

The last example, consisting in a Dolphin whistle analysis, is illustrated in the figure 20.



Parameters of T-F components estimated by WSD

ξ	t_{0k} [sec]	f_{0k} [kHz]	D_k [sec]	β_k/F_s	$m(t)$
ξ_1	0.25	2.3	0.27	$-9.3\cdot10^{-7}$	t^{19}
ξ_2	0.89	2.37	0.28	-10^{-6}	t^{19}
ξ_3	1.58	2.25	0.28	$-2.12\cdot10^{-6}$	t^{18}
ξ_4	0.24	4.69	0.26	$-8\cdot10^{-7}$	t^{21}
ξ_5	0.91	5.21	0.21	$-6.6\cdot10^{-7}$	t^{22}
ξ_6	1.54	4.91	0.17	$-1.61\cdot10^{-6}$	t^{21}
ξ_7	0.2	7.66	0.18	$-1.75\cdot10^{-6}$	t^{21}
ξ_8	0.87	8.11	0.16	$-3.71\cdot10^{-6}$	t^{2}
ξ_8	1.52	7.73	0.13	$-2.05\cdot10^{-6}$	t^{22}

Figure 20. Feature extraction from a Dolphin whistle

In the case of Dolphin whistle, the visual and parametric (figure 20) information provides an analytical characterization of signal features. In the figure 20, we observe that the WSD can eliminate the noise from the T-F image. Observing also the noise existing in the spectrogram

representation, the noise robustness of the WSD algorithm is illustrated. This property is explained by the signal processing only in the detected ROI and by the performances of the WSD algorithm.

The examples of this section, provided on real data, illustrate the performances of the proposed approach. We have seen that the WSD algorithm furnishes two information types : visual and parametric. The visual information consists in the time-frequency distribution of the signal. But, unlike the spectrogram, the information is clearly represented because only the estimated T-F components are plotted.

The parametric information of the WSD provides the analytical characterization of the IFLs of the T-F components. This information, less intuitive than the visual one, could have a lot of applications, such as signal identification and classification, underwater environment characterization, underwater source separation and description, etc.

The features and the possible applications of the proposed algorithm will be briefly commented in the next section.

7. CONCLUSION AND PERSPECTIVES

In this paper, we have presented a new method for underwater signal characterization in a passive context. It is well known that the underwater environment is a very difficult one, from signal processing point of view. The noise, the non-stationnarity and the diversity of the signals, are only few characteristics of underwater signal processing field. To deal with these, the time-frequency methods constitute an attractive solution.

In this context, we proposed a new time-frequency based technique for feature extraction of underwater signal. We focused the development of this method on the marine-mammal emissions, which constitute richness sources of information about the underwater environment. Generally, the characterization of these types of signals implies the *signal detection* and *feature extraction*. In this paper, these signal processing stages have been the subjects of two proposed approaches.

Firstly, we have proposed a filter bank-based method for the *detection* of transient parts of the signal. We have used the OCWT filter bank which provides a robust tool to detect the transient signals. Furthermore, in order to adapt the time-frequency partition according to the signal particularities, we proposed the use of an irregular sampling procedure. The construction of the irregular sampling grid is based on the discrimination between signal and environmental noise. This is provided by the 4^{th} order statistical moments. The application of the irregular sampling procedure guaranties the satisfaction of theoretical conditions related to the noise whiteness. The proposed method, OCWT-IS, provides an optimal time-frequency partition and a set of detection curves. The set of these curves furnishes the signal regions of interest.

The proposed OCWT-IS detector has been compared with two classical ones, based on spectrogram and Local Cosine Basis. The tests effectuated on synthetic and real data have proved the quality of the proposed approach from both detection performances and feature conservation points of view.

In the second stage, we studied the problem of the *characterization* of signals contained by the detected ROIs. Since the detected signals are non-stationary, the time-frequency methods constitute the potential solution. For time-frequency representation of signal features, thanks to its operational capabilities and to the intuitive T-F information, the spectrogram is the most popular tool. Nevertheless, since the spectrogram analysis is a windowing-based method, the induced smoothing leads to T-F structure localization at different coordinates than the correct ones. Consequently, the accurate extraction of T-F parameters needs some additional procedures, difficult to design for a passive context.

An alternative solution to deal with these types of T-F components consists in generating new time-frequency distributions based on warping operator concept. In this paper, the warping principle is used to furnish both signal modeling and representation.

The first purpose-signal modeling, is based on the estimation of signal time-frequency non-linearities. To do that, the analyzed signal is warped in some fashions and we estimate, for each

deformation, the virtual linear time-frequency structures (chirplets). These chirp-like components corresponds to the associated non-linear T-F components existing in the signal structure. The analysis of extracted chirplets and corresponding decomposition coefficients from all branches leads to the identification of T-F components of the signal. Therefore, the signal modeling provides the parametric information about the signal. The extracted parameters can be associated to the signal features.

The second purpose - signal representation is aimed to furnish visual information about analyzed phenomenon. This is done by combining the warping-based TFRs of extracted component.

The performances of this algorithm, called Warping-based Signal Decomposition, have been evaluated in a real context, using some marine-mammal emissions. The main features of this algorithm have been illustrated.

Firstly, the analytical description of the IFLs, directly related to the visual time-frequency image, provides rich information about the analyzed signal. These two information types may be very useful in practical applications as underwater signal identification or classification. It is possible to synthesize the signal, since it is fully characterized by its parameters.

Secondly, we have seen that the proposed method furnishes an adaptive estimation over the time. Therefore, the parameter tracking could provide interesting information about the source evolution, environments parameters, etc.

The good performances in noise environment, illustrated on real data, guarantee the correct processing in real environment.

Nevertheless, the performances of the WSD algorithm depend on the number of warping operators that are used. Practically, this number is limited, and the consequences may be the apparition of false component or the looseness of some useful ones. On the other hand, since in a passive context the information about the underwater mammals emissions cannot be assumed, the complete analysis supposes a multi-disciplinary studies : signal processing, biologic, underwater engineering, etc. In this context, the main task of the proposed algorithm is to help in the processing of a huge volume of data. Furthermore, WSD can offer an alternative characterization of the underwater signals in terms of time-frequency component estimation.

In future works, we will use this alternative description in underwater signal classification and oceanic passive tomography.

For the first topic, the analysis of many underwater signal classes will be done. The study of discrimination capabilities provided by the extracted parameters must be also done.

In theoretical field, two directions will be explored. Firstly, new implementation structures will be analyzed in order to improve the operational characteristics. Secondly, new warping techniques will be studied, to achieve a more accurate description of real signals.

Acknowledgments. This work was supported by the French Military Center of Oceanography under the research contract CA/2003/06/CMO.

REFERENCES

[1] W.A. Watkins, M.A. Daher, K.M. Fristrup, T.J. Howald, *Sperm whales tagged with transponders and tracked underwater by sonar,* Deep Sea Research, vol.9, pp. 55-67, 1993.

[2] W.J. Richardson, C.R. Greene, C.I. Malme, D.H. Thomson, *Marine Mammals and Noise,* Academic Press, California, 1995.

[3] W. Munk, P. Worcester, C. Wunsch, *Oceanic acoustic tomography,* Cambridge University Press, 1995.

[4] W.S. Burdic, *Underwater acoustic system analysis,* Pretince-Hall, 1984.

[5] R.A. Altes, *Detection, estimation and classification with spectrograms,* Journal of the Acoustic Society of America, vol. 67, no. 4, pp. 1232-1246, Apr. 1980.

[6] G. Roberts, A.M. Zoubir, B. Boashash, *Time-Frequency Classification using a Multiple Hypotheses Test: An Application to the Classification of Humpback Whale Signals,*

[7] L. Cohen , *Time-Frequency Analysis*, Pretince Hall, New Jersey, 1993.
[8] E. Delory, J.R. Potter, *Objectivity in the study of marine mammal vocalisations: a wavelet approach*, European Cetacean Society Proceeding, Valencia, April, 1999.
[9] A. Theolis, S. Shamma, *Classification of Transient Signals via Auditory Representation*, Technical Research Report TR 91-99, University of Mayland, 1999.
[10] A. Theolis, *Computational Signal Processing with Wavelet*, Birkhauser Press, Boston, 1998.
[11] B. Boashash, P. O'Shea, *A methodology for detection and classification of some acoustic signals using time-frequency analysis techniques,* IEEE Trans. On Acoustics, Speech and Signal Processing, 38 (11):1829-1841, Nov 1990.
[12] R. Baraniuk, *Unitary Equivalence : a new twist on signal processing*, IEEE Trans. On Signal Processing, vol. 43, no. 10, October, 1995.
[13] P. Ravier, P.-O. Amblard, *Wavelets Packets and De-noising Based on Higher-Order-Statistics for Transient Detection*, IEEE Trans. on Signal Processing, Vol. 81/9, pp. 1909-1926, Aug, 2001.
[14] S. Mallat, *A Wavelet Tour of signal processing*, Academic Press, 1998.
[15] K. Gröchenig, *Irregular sampling of wavelet and short time Fourier Transform*, Constr. Approx., No 9, 282-297, 1993.
[16] J.M. Mendel, *Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and some applications",* Proceedings of the IEEE 79(3):278-305, March 1991.
[17] M.G. Kendall, A. Stuart, *The advanced theory of statistics*, Charles Griffin & Company Limited, 1963.
[18] A. Papandreou, S.M. Kay, G.F. Boudreaux-Bartels, *The use of hyperbolic time-frequency representation for optimum detection and parameter estimation of hyperbolic chirps,* IEEE-SP Symposium on Time-Frequency/Time-Scale, Philadelphia, , pp. 512-518, Oct. 1994.
[19] F. Hlawatsch, A. Papandreou, G.F. Boudreaux-Bartels, *The power classes- quadratic time-frequency representations*, IEEE Trans. Signal Processing, vol. 47, pp. 3067-3083, Nov. 1999.
[20] T. Twaroch, F. Hlawatsch, *Modulation and warping operators in joint signal analysis*, in Proc. IEEE-SP Int. Sympos. Time-Frequency Time-Scale Analysis, Pittsburgh, PA, pp. 9-12, Oct. 1998.
[21] K. Canfield, D.L. Jones, *Implementing Time-Frequency Representations for non-Cohen Class*, Proceedings of the 27th ASILOMAR Conference, Pacific Grove, 1993, pp. 1464-1468
[22] C. Ioana, A. Quinquis, *On the use of time-frequency warping operators for analysis of marine-mammal signals,* Proceedings of ICASSP 2004, Montreal, pp. 605-608, May 2004.
[23] A. Papandreou-Suppappola, S. Suppappola, *Adaptive time-frequency representations for multiple structures*, Proceedings of the 10th IEEE Workshop on Statistical Signal and Array Processing, Pocono Manor, , pp. 579-583, 2000.
[24] S. Mallat, Z. Zhang, *Matching pursuit with time-frequency dictionaries*, IEEE Trans. Signal Processing, vol.41, no.12, PP. 3397-3415, Dec., 1993.
[25] S. Mann, S. Haykin, *The Chirplet Transform : A Generalisation of Gabor's logon*, Canadian Image Processing and Patern Recognition Society, Oct. 1991.
[26] E. Polak, *Computational Methods in Optimization*, Academic Press; Oct. 1997.
[27] P. Stoica, R. Moses, *Intoduction to spectral analysis*, Pretince Hall, 1997.
[28] J.C. O'Neill, P. Flandrin, *Chirp Hunting,* IEEE Int. Symp on Time-Frequency and Time-Scale Analysis, pp. 425-428, Pittsburg, Oct. 1998.
[29] www.marineland.fr

Cornel Ioana received the Dipl.-Eng degree in electrical engineering and informatics in 1999 from the Romanian Military Technical Academy of Bucharest. Between 1999 and 2001 he activated as a military researcher in a research institute of Romanian Ministry of Defense (METRA, Bucharest). He received the M.S. degree in telecommunication science in 2001 and the PhD in electrical engineering field in 2003, both from University of Brest-France. Since 2003 he works as researcher and development engineer in ENSIETA, Brest-France. His current research activity deals with the implementation of signal processing methods adapted to the passive underwater tomography. His technical interests are non-stationary signal processing, electronic warfare, sonar and real-time systems.

André Quinquis received the M.S. degree in 1986 and the Ph.D. degree in 1989 in signal processing, both from the University of Brest. Between 1989 and 1992 he taught and developed research activities in signal and image processing at the Naval Academy in Brest. In 1992 he joined the Engineering School ENSIETA of Brest, where he held the positions of Senior Researcher and Head of the Electronics and Informatics Department. Since 2001 he has been Scientific Director of ENSIETA. He is mainly interested in signal processing, time-frequency methods, and statistical estimation and decision theory. Dr. Quinquis is author of 8 books and of more than 100 papers (international journals and conferences) in the area of signal processing.

Yann STEPHAN was born in Lannion, France, in 1966. He received his engineering diploma in Electrical Engineering from ENSIEG in Grenoble. He received his Ph.D. degree in Computer Sciences in 1996 from CNAM in Paris. Since 1992, he has worked with the *Service Hydrographique et Océanographique de la Marine (SHOM)* within the Center for Military Oceanography in Brest. His current topics include inverse methods, acoustic environmental assessment and tactical use of the environment.