PRSA method for transient detection applied to drinking water production surveillance

Philippe Ravier[1], Meryem Jabloun[1], Julien Roussel[2] and Cécile Capdessus[2]

[1]PRISME laboratory - Polytech Orléans
12 rue de Blois, BP 6744, 45067 Orléans cedex 02, France
[2]PRISME laboratory - IUT de Chartres
21 rue de Loigny la Bataille, 28000 Chartres, France
{Philippe.Ravier, Meryem.Jabloun, Julien.Roussel, Cecile.Capdessus}@univ-orleans.fr

Abstract

This paper concerns the monitoring of water ultra-filtration modules by vibration analysis. During the maintenance process, damaged modules can be revealed by the presence of air bubbles that may appear. The bubbles generate transient signals at the surface of the modules. The present study investigates the potential of the TFR-PRSA approach as a new tool for detecting transient signals based on theoretical statistical results of the TFR-PRSA distribution. The performance of the detector is evaluated considering one synthetic transient and five real ones representing the sound classes that were identified during the measurements in the plant. Monte Carlo simulations show that the proposed detector outperforms the TFR-STFT reference detector for four transients in a white Gaussian noise context. In the case of industrial measured noise, the performance of the TFR-STFT method dramatically decreases. Inversely, no real performance changes are observed comparatively to the white noise case for the TFR-PRSA approach: this result reveals a high robustness of the method with the color of the noise.

1 Introduction

Ultrafiltration using hollow fibers is a commonly used method for drinking water production because it allows removing micro particles up to ten nanometers and pathogens without using chemicals. Fibers are organized like spaghettis inside ultrafiltration modules (tubes). High pressure water is injected through the fibers. Part of it runs through the fibers membranes and reaches the inside of the module as pure as source water. However, due to the high water pressure, fibers may be damaged, causing the dirty water to reach the module and be mixed with the pure water.

The diagnosis technique used in practice consists in injecting air into the fibers within a water filled module. The air is normally stopped by the fiber membrane but in the presence of a damaged fiber, the air passes through this damaged fiber creating some bubbles inside the module. This bubble emission exhibits a recognizable acoustic signature, a transient signal, which can be used for detection purpose. Currently skilled technicians listen to these bubble sounds in order to report on the module state but the subjectivity of this test may induce both detection errors and improper estimation of damages.

In this context, an automatic detection of bubbles emission is of particular interest as a diagnosis help during the maintenance phase. In fact, since the technicians stick there ear on the modules, they are more sensitive to the vibrations that are transmitted within their skull than to the sounds they listen to. Moreover, the maintenance operation is achieved in a highly noisy acoustic environment which is typical of water production plants. For all these reasons, we designed an acquisition system with high-sensitivity vibration sensors instead of acoustic ones [1]. To sum up, since bubbles emissions exhibit acoustic signatures, they also will create vibrations on the modules surface, that are characterized as transient oscillatory components.

Many methods can be used for detecting the oscillatory transients, either parametric or not. Parametric methods are based on parametric models that describe the transients. The detection process then consists in deciding whether the estimated parameters of the analyzed data are those of the model or not. However this approach is faced to a potentially high number of different models, which are not necessarily known. Indeed,

a great variability in the transients morphology can be observed because of various bubbles generation and transmission processes depending on the gravity level of the damages and their location within the modules. In a previous study[1], an attempt of classification permitted to identify five transient categories with different time-frequency signatures within the data collected for this problem.

Non parametric methods are based on statistical properties of the data that are expected to be different between the transient's class and the embedded noises class. In [1], the second order or fourth order statistics of the data have been used for detection purpose, either on the raw temporal data or after a transformation of the data, with the aim of increasing the signal-to-noise ratio (transformation in the frequency domain by filtering or transformation in the time-frequency domain by wavelet decomposition). Indeed, it has also experimentally been shown in [1] that the background noise is normally distributed. In the Fourier domain, this background noise remains normally distributed whereas Fourier coefficients associated with the bubble transients deviate from normality. This assertion holds as long as the computation is done on short time windows which reduces the central limit theorem effect. Furthermore, because of nonstationary property of transient signals, time-frequency analyses are naturally required. In this context, a correct detection of transient signals thus needs an efficient measure of deviation from normal distribution in the time-frequency plane.

In this paper, we investigate the potential of the new time frequency - phase rectified signal averaging (TF-PRSA) method [4] to discriminate the transients from the noise. The main reason of applying this method comes from the ability of the method to highlight dominant frequencies which are inherent to the oscillatory behaviour of the transients. The corresponding Fourier components will emerge with respect to the noise ones which enhances the distance from normality.

2 The TF-PRSA detection method

Recently, the Phase Rectified Signal Averaging (PRSA) method and its Time-Frequency Representation (TFR-PRSA) extension were proposed as new tools for highlighting dominant frequencies in noisy nonstationary signals [2]. We recall hereafter the principle of PRSA computation with illustration on simulated data.

2.1 The PRSA principle

The basic idea of PRSA is the averaging of segments extracted from any discrete time signal y. Segments are symmetrically extracted according to anchor points that satisfy a desired condition in the data [5]. In this process, the averaging operation removes correlated or nonperiodic components of the signal (such as noise and artifacts) while the quasi-periodic components are enhanced.

The steps of the PRSA method are illustrated in Fig.1. In this example, the anchor points correspond to the increases in the signal y (Fig.1(a)), *i.e.* instants n such that

$$y_n > y_{n-1}. \tag{1}$$

Assuming a total of M anchor points indexed by n_m, $m = 1,\ldots,M$, segments of length $2L+1$ are centered on these anchor points y_{n_m} (Fig.1(b-d)) such that a samples vector for each segment is constituted as

$$\left[y_{n_m-L}, y_{n_m-L+1}, \ldots, y_{n_m}, \ldots, y_{n_m+L-1}, y_{n_m+L}\right]. \tag{2}$$

All these segments are averaged, which leads to the PRSA signal $\tilde{y}_\ell$

$$\tilde{y}_\ell = \frac{1}{M}\sum_{m=1}^{M} y_{n_m+\ell}, \quad \text{for } \ell = -L, -L+1, \ldots, L. \tag{3}$$

2.2 The PRSA Time-Frequency representation

By locally applying the PRSA technique over a K samples sliding window and taking the Fourier transform of each PRSA signal, a Time Frequency representation can be obtained. More precisely, the entire sequence y_n with $n = 0,\ldots,N-1$ now becomes $y_{k,n} = y_{k-n}$ with $n = 0,\ldots,K-1$ now being limited to the window length and $k = K-1,\ldots,N$ denoting the instant location of each sliding window. In this procedure, if the PRSA

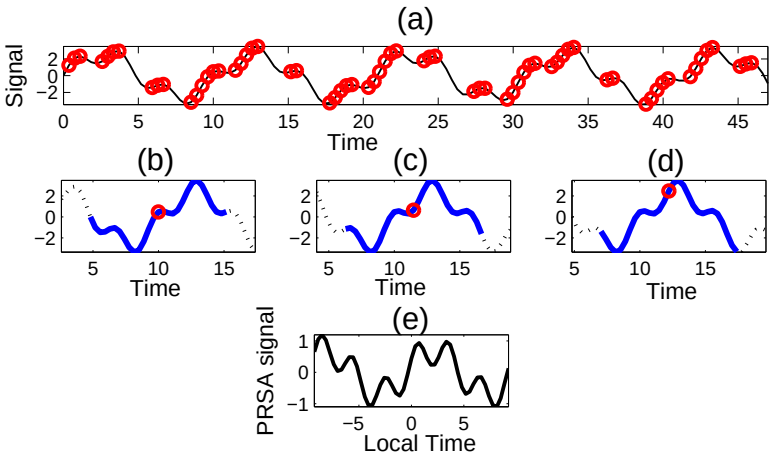

Figure 1: Principle of PRSA: (a) raw signal with red anchor points satisfying $y_n > y_{n-1}$, (b)-(d) three instances of segments of equal length $2L+1$ centered on anchor points, (e) PRSA signal $\tilde{y}_\ell$ (3).

signal is $\tilde{y}_\ell$ with $\ell = -L,\ldots,L$ and its Fourier transform $\tilde{Y}_q$, the set of sequences $y_{k,n}$ will become $\tilde{y}_{k,\ell}$ and the Fourier transforms $\tilde{Y}_{k,q}$. The Fourier transforms are computed on $2L+1$ length windows and produce Q frequency samples. In the following, we fix $Q = 2L+1$ which is the general case. Finally, the PRSA time-frequency representation which describes the time-evolution of the PRSA spectrum is obtained in analogy with the short-time Fourier transform (STFT) and the spectrogram:

$$\begin{array}{ccccc} \text{Time} & \mathbf{x} & \text{Frequency} & \longrightarrow & \text{PRSA-TFR} \\ (k & , & q) & \longmapsto & \left|\tilde{Y}_{k,q}\right|^2. \end{array} \tag{4}$$

Figure 2 displays an example of temporal data of a real vibration signal measured on the modules surface constituted of ambient noise with transients to be detected. The question of the correct number of transients arises. Figures 3 and 4 show the TF-STFT representation and the TF-PRSA representation of the same signal, respectively. The integrated curves of the TF representations along time and frequency are also shown. One can see that the integration over the time produces a lower level of the ambient noise components with respect to the peak frequency using TF-PRSA method comparatively to TF-STFT method. This illustrates the dominant frequencies highlightening effect of the PRSA method. Integrating over the frequency, transients clearly appear using TF-PRSA method comparatively to the same curve obtained with the TF-STFT method of Fig.3.

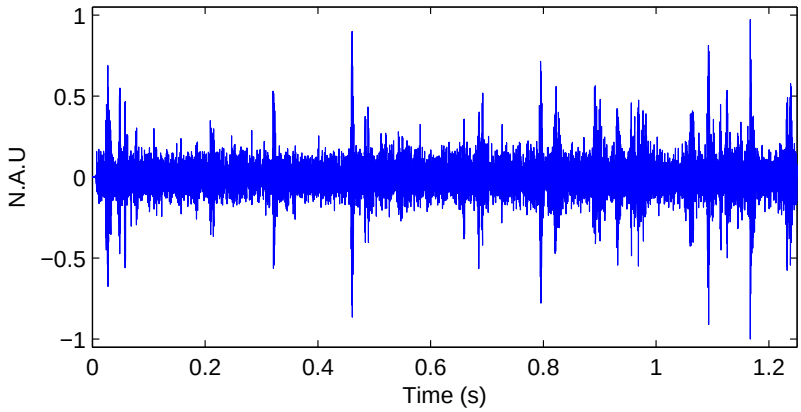

Figure 2: Example of temporal data with transients to be detected. How many transients? Were are they localized?

2.3 The PRSA-TF detection process

In the present paper, we investigate the potential of the TFR-PRSA method as a new tool for detecting oscillatory transient signals. The bubble detection is achieved by measuring the deviation of their empirical

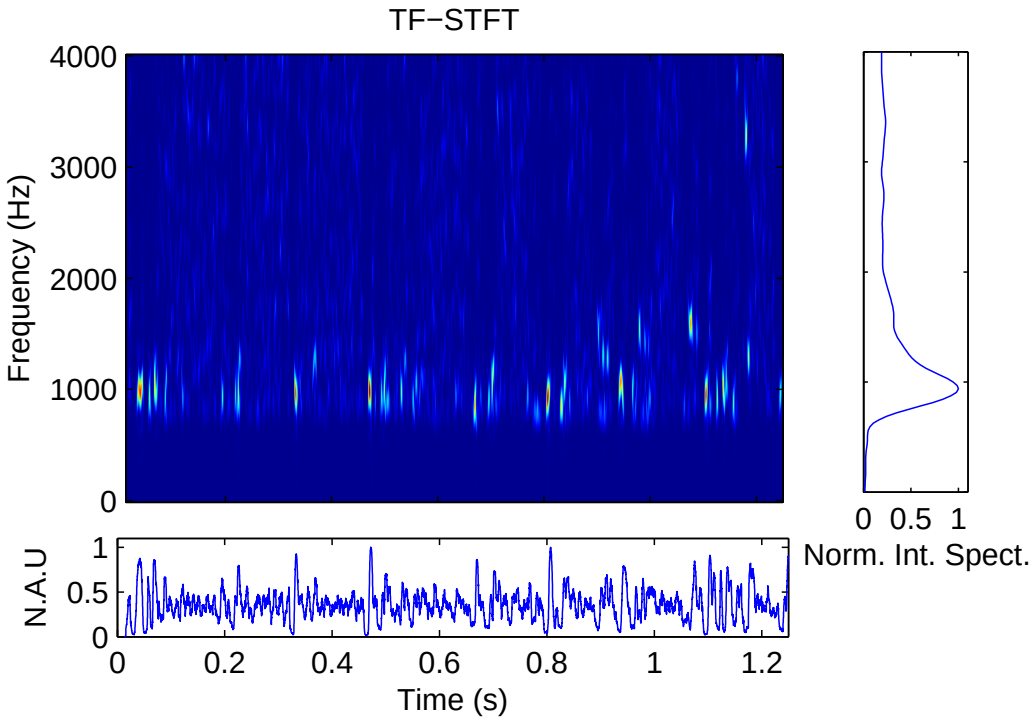

Figure 3: Time Frequency representation with the Short Time Fourier Transform technique applied on the signal of Fig.2.

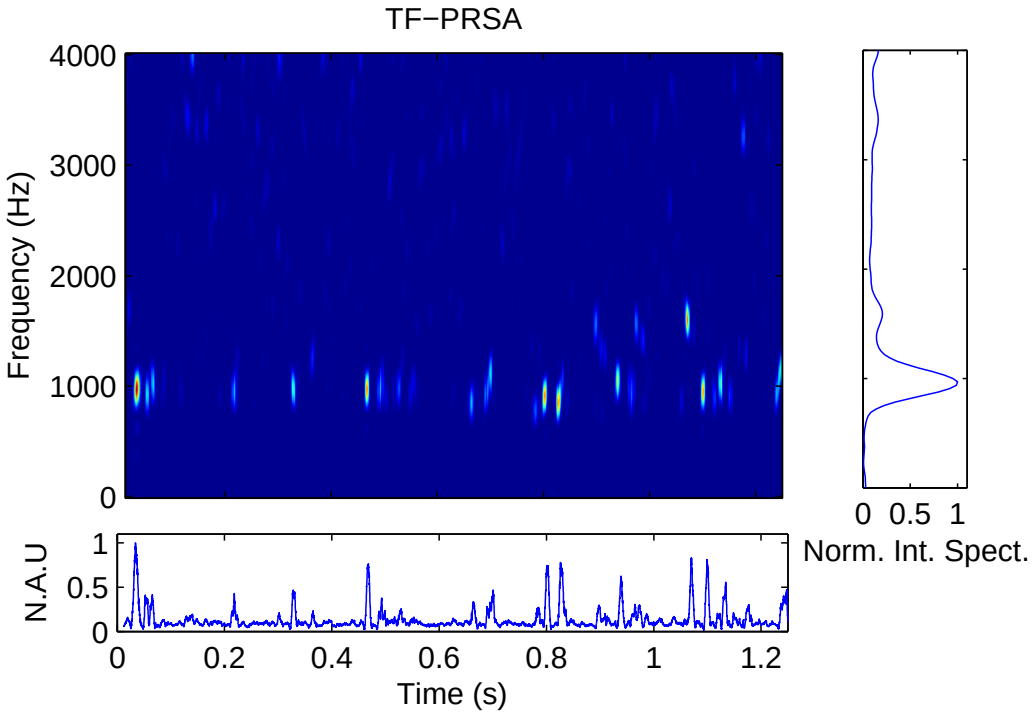

Figure 4: Time Frequency representation with the sliding window PRSA technique applied on the signal of Fig.2. Transients clearly appear on the temporal TF-PRSA integrated curve comparatively to the same curve obtained with the TF-STFT method of Fig.3

TFR-PRSA coefficients distribution from the theoretical distribution of an i.i.d Gaussian noise. We recall hereafter the theoretical results derived in [4] that give the probability density function of the energy distributed in the TFR-PRSA for a white Gaussian noise.

Let us consider unit and zero-mean white Gaussian noise y, $y_n \sim \mathcal{N}(0,1)$. The probability distribution of the PRSA $\tilde{y}$ of y has been shown to be approximated by:

$$\begin{array}{rcl} \tilde{y}_\ell & \sim & \mathcal{N}\left(0, \frac{1}{\sqrt{M}}\right) \ \forall\, \ell \in \{-L, -L+1, \ldots, L\} \setminus \{-1, 0\}, \\ \tilde{y}_0 & \sim & \mathcal{N}\left(\frac{1}{\sqrt{\pi}}, \frac{\sqrt{\pi-1}}{\sqrt{M\pi}}\right), \\ \tilde{y}_{-1} & \sim & \mathcal{N}\left(\frac{-1}{\sqrt{\pi}}, \frac{\sqrt{\pi-1}}{\sqrt{M\pi}}\right), \end{array} \tag{5}$$

where M is the number of anchor points.

The pdf of the squared modulus of the TF-PRSA transform $S[q] = \left|\tilde{Y}_{k,q}\right|^2$ introduced in (4) can be shown to be approximated as a $\chi^2(2)$ distribution, for high frequencies (large q):

$$pdf_S\left(\left|\tilde{Y}_{k,q}\right|^2 = S[q]\right) = Me^{-MS[q]}. \tag{6}$$

For $q = 0$, the $S[q]$ follows a $\chi^2(1)$ distribution expressed as:

$$pdf_S\left(\left|\tilde{Y}_{k,0}\right|^2 = S[0]\right) = M\frac{S[0]^{-\frac{1}{2}}}{\sqrt{2\pi}}e^{-M\frac{S[0]}{2}}. \tag{7}$$

The deviation between the theoretical distribution pdf_S and the experimental one $\widehat{pdf_S}$ is measured by means of the Kullback-Leibler distance expressed as:

$$\mathscr{D}(pdf_S, \widehat{pdf_S}) = \int_0^{+\infty} \widehat{pdf_S} \log\left(\frac{\widehat{pdf_S}}{pdf_S}\right) ds \tag{8}$$

3 Probability detection curves

3.1 Setup procedure

In order to demonstrate the performance of the proposed approach, we evaluate the quality of detection with an additive noise model. We compute the probability of detection as a function of the signal to noise ratio (SNR) for a fixed false alarm rate by means of 500 Monte Carlo simulations. The two assumptions noise only and transient plus noise are created for the computation of the probability of false alarm and probability of detection, respectively. The false alarm rate is fixed to 0.1% for all the experiments.

For each hypothesis testing case, we calculate the TF-PRSA representation with K=200 samples window lengths and L=63. In the white Gaussian case, the number of anchor points is expected to be half the number of L-segments in the window, that is $M = (K-L)/2 = 68$. A sliding operation of the windows by a value of L is considered. Finally the sliding Kullback-Leibler distance between the empirical TF-PRSA representation and the Gaussian theoretical one is computed with the PRSA-TFR of two consecutive K-windows. This leads to an online detection curve that is used for the probability estimations.

We also compare this TF-PRSA method with the TF-STFT approach for which the distribution is known to be $\chi^2(2)$ if non overlapping sliding windows are used. In this case, we consider the sliding operation with 64 samples windows and no overlapping.

In a first study, we consider additive white Gaussian noise with SNR values ranging from -15dB to +5dB. The SNR is considered as the power ratio between the transient and the noise evaluated on the transient support, *i.e.* transient duration.

In a second study, we consider the realistic ambient noise that were recorded during the experiments in the water production plant, at the 8kHz sampling frequency rate. As described in [1], the background noise has experimentally been shown to be normally distributed and colored.

3.2 BiAR synthetic transient

We firstly consider a synthetic transient signal called biAR1. This transient is constituted of a modulated sine wave, shaped by an onset ascending envelope followed by an offset descending envelope. The transient is composed of two consecutive parts: the ascending envelope is described by the impluse response of a first order anticausal autoregressive filter while the consecutive descending part is described by a causal one. The result is a transient signal which durations and central frequencies can can be controlled by the position of the filter poles in the complex z-plane.

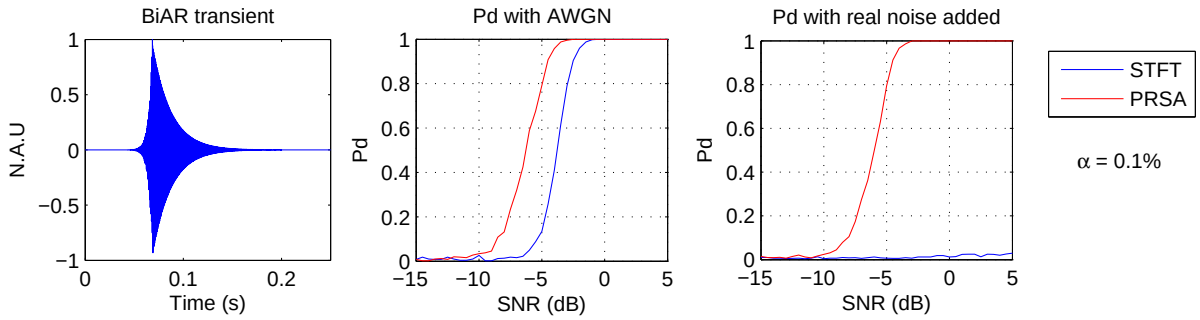

Figure 5: (left) **Synthesized biAR transient** - Probability of detection as a function of the SNR for the TF-STFT (blue) and the TF-PRSA (red) detectors with AWGN (middle) and real noise added (right). The detectors measure the deviation between the theoretical distribution and the experimental one by means of the Kullback-Leibler distance.

The synthetic biAR transient is shown on Fig.5, left. The middle plot shows the probability of detection with additive white Gaussian noise (AWGN) simulations. The right plot shows the probability of detection with real noise added simulations. The best detectors are those displaying high probabilities of detection for low SNR (most left curve on Fig.5). The TFR-PRSA detector shows better performance with respect to the TFR-STFT one: in the AWGN case, a shift of 2dB can be observed between the curves and in the real noise case, the TFR-STFT detector is inefficient while the quality of detection for the TFR-PRSA detector remains the same. This result shows the robustness of the proposed detector with respect to different types of noise, in the synthetic transient case.

3.3 Five real transients

In this section, we consider a representative real transient for each one of the 5 classes that were identified in [1]. Nicknames of the classes are **GlouGlou, Slup, TacTac, TapTap, TicTic**. The results are shown in Fig.6 to 10 respectively, the temporal waveform of each real transient being represented on the left plot of each figure.

Performance results differ between transients because of various transient envelopes making energy and frequency distribution varying within the transients duration. Under synthetic noise conditions, only the Tac-Tac and TicTic transients show better performance with TFR-STFT detector comparatively to the TFR-PRSA detector with a maximum shift of about 2 dB for the TacTac transient. In the other cases, the TFR-PRSA detector outperforms with a maximum shift of about 5 dB for the TapTap transient. This result can be explained by the high-pass effect of the PRSA method that acts in the detection process in favour of the TapTap transient which is composed of high-frequencies. The bandwidth of the TacTac transient has been identified to be in the range 0.977 kHz - 9.131 kHz showing the highest frequency bound among all the transients.

Under real noise conditions, all the TFR-STFT detectors show very bad performance, whatever the transients. This means that this method is very sensitive to the frequency distribution of the noise. More surprisingly, the TF-PRSA detector seems to be insensitive to the power spectral distribution of the noise. This may be explained in this case by the insensitivity of the PRSA transform to very low frequency components [3].

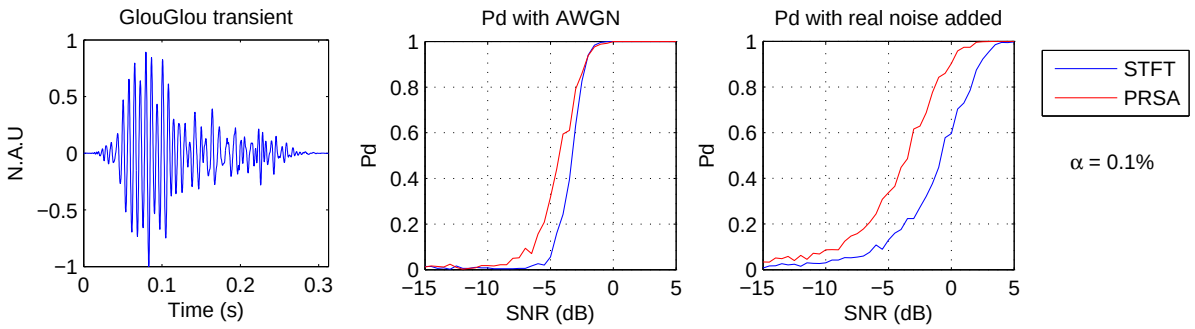

Figure 6: (left) **Real GlouGlou transient** - Pdf as a function of the SNR.

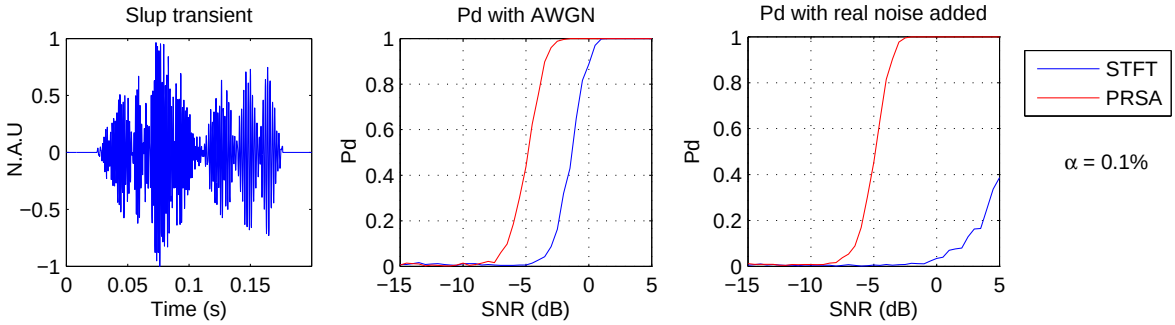

Figure 7: (left) **Real Slup transient** - Pdf as a function of the SNR.

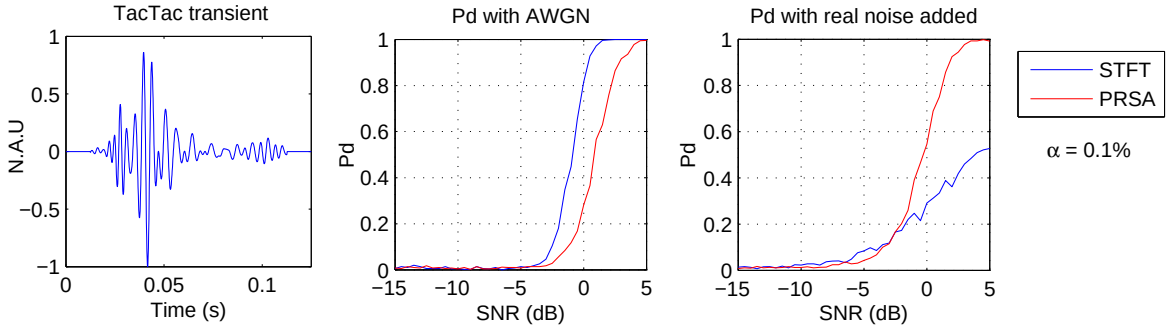

Figure 8: (left) **Real TacTac transient** - Pdf as a function of the SNR.

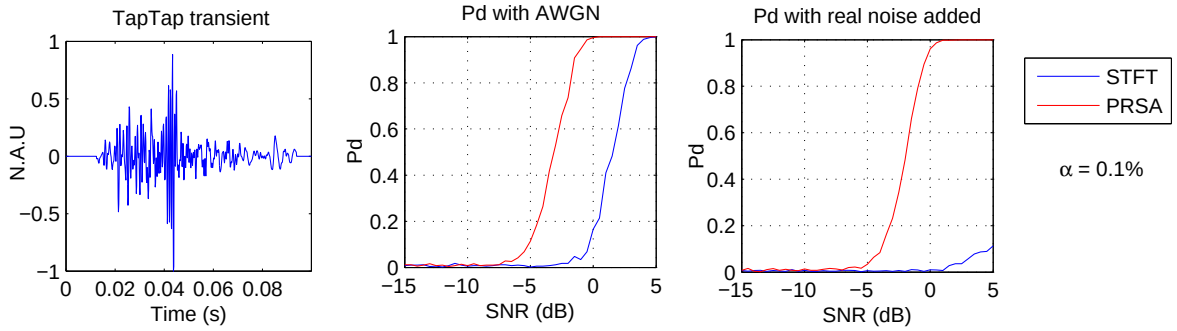

Figure 9: (left) **Real TapTap transient** - Pdf as a function of the SNR.

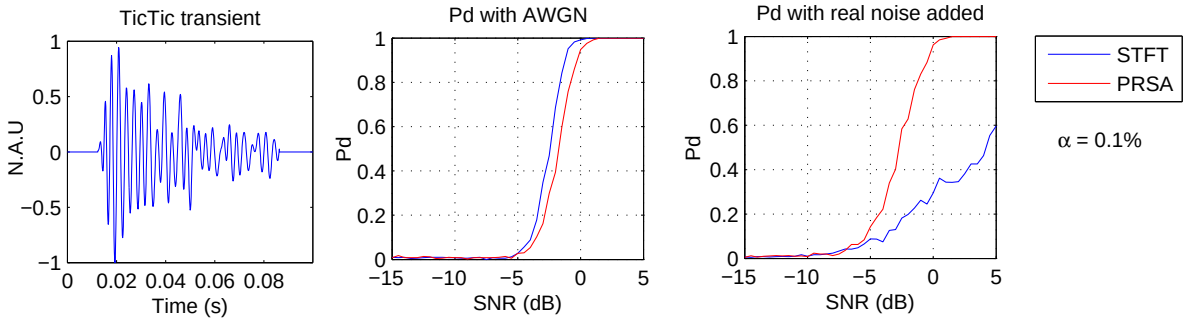

Figure 10: (left) **Real TicTic transient** - Pdf as a function of the SNR.

4 Conclusion

In this paper, we discussed the potential of the new PRSA-TFR to detect transient signals in an industrial context. We proposed a detection process that is based on the Kullback-Leibler distance between the experimental PRSA-TFR and the theoretical one, when white Gaussian noise only is present. The results show that the proposed PRSA-TFR outperforms the STFT-TFR approach, especially when facing the detector to real industrial measured noise.

References

[1] J. Roussel, P. Ravier, C. Capdessus and M. Jabloun, Vibration Analysis for Damage Detection in Ultra filtration Fibers, Surveillance 6, Compiègne, France, 25-26 October 2011.

[2] M. Jabloun, J. Van Zaen and J.-M. Vesin, Time-frequency analysis based on the phase-rectified signal averaging method, EUSIPCO, Glasgow, pp. 2303-2307, August 2009.

[3] M. Jabloun, P. Ravier and O. Buttelli, Phase-Rectified Signal Averaging method applied to heart rate variability signals for assessment of the changes in sympathovagalo balance during rest and tilt, EUSIPCO, Poznan, September 2011.

[4] M. Jabloun, P. Ravier and O. Buttelli, Theoretical insight into the phase rectified averaging method and application to HRV analysis, EUSIPCO, Marrakech, September 2013.

[5] A. Bauer, J. Kantelhardt, A. Bunde, P. Barthel, R. Schneider, M. Malik, G. Shmidt, Phase rectified signal averaging detects quasiperiodicities in non-stationary data, Physica A, Vol. 364, pp. 423-434, 2006.