Compression methods for mechanical vibration signals: Application to the plane engines

Marius Oltean, José Picheral, Elisabeth Lahalle, Hani Hamdan, Julien Griffaton

Compression methods for mechanical vibration signals: application to the plane engines

Marius Oltean[a,b], José Picheral[a,*], Elisabeth Lahalle[a], Hani Hamdan[a], Julien Griffaton[c]

[a] *E3S-SUPELEC System Sciences, 3 Rue Joliot Curie, 91192 Gif sur Yvette CEDEX, France*
[b] *POLITEHNICA University of Timisoara, Romania*
[c] *SAFRAN Snecma*

Abstract

A novel approach for the compression of mechanical vibration signals is presented in this paper. The method relies on a simple and flexible decomposition into a large number of subbands, implemented by an orthogonal transform. Compression is achieved by a uniform adaptive quantization of each subband. The method is tested on a large number of real vibration signals issued by plane engines. High compression ratios can be achieved, while keeping a good quality of the reconstructed signal. It is also shown that compression has little impact on the detection of some commonly encountered defects of the plane engine.

Keywords: fault detection ; vibration signals ; uniform adaptive quantization ; subband coding ; DCT ; compression

1. Introduction

Vibration signals provide useful information for the diagnosis of rotating machineries and they represent a key prerequisite of mechanical health monitoring systems. The goal of health monitoring systems is to perform off-line

*Corresponding author

Email addresses: marius.11.oltean@gmail.com (Marius Oltean), jose.picheral@supelec.fr (José Picheral), elisabeth.lahalle@supelec.fr (Elisabeth Lahalle), hani.hamdan@supelec.fr (Hani Hamdan), julien.griffaton@snecma.fr (Julien Griffaton)

default detection in order to anticipate damage and maintenance. Defaults that should be detected are mechanical defaults such as damaged bearings (spalling or brinelling of the races and the rolling elements), damaged gears (scuffing of the teeth), coupling between aero and mechanical abnormal behavior like fan flutter, fleeting events like a shock caused by a small FOD (Foreign Object Damage, small birds ingestion for instance) or rubbing between a rotor and a stator. Impending failures (few minutes) are not in the scope of health monitoring. The scope of health monitoring is to detect damages few flights before failure, but it happens that the early symptom of a future failure shall be detected during a single flight (this is the case of fleeting and non-repetitive events). Such systems require, in general, a significant amount of data to learn the features that are associated with faulty behavior of the engine, as well as with its normal functioning. Furthermore, the high sampling frequency of the vibration signals (for most engines applications, 2 accelerometers with a bandwidth up to 20kHz), as well as the long time intervals required to perform a full acquisition (*i.e.* corresponding to a particular stage of the flight, for instance an engine shut-down that lasts approximately 2 minutes, at the end of the flight, or cruise that lasts few hours), lead to a huge amount of data needing to be stored and transmitted. Thus, compression becomes a mandatory component of a health monitoring system and required compression rates into the range 10 to 100.

Within this framework, the main challenge we deal with is to define a compression algorithm which is capable of achieving a high Compression Ratio (CR), while keeping a reasonable reconstruction quality, in terms of Reconstruction Signal to Noise Ratio (RSNR) or Mean Square Error (MSE). Besides the aforementioned performance criteria, we aim to limit the impact of compression on the signature of some particular potential faults of the engine. Nevertheless, since in such an application we often deal with complexity constraints (*e.g.* embedded systems, real-time monitoring), one important objective is to keep compression algorithms at a reasonable level of complexity.

All these specific constraints motivated our research efforts, which aimed to define a new compression approach for mechanical vibration signals. The proposed algorithm meets the aforementioned objectives, by exploiting the particular characteristics of the mechanical vibration signal.

Whereas the generic field of data compression is extensively treated in the literature, the number of works dedicated to the specific topic of vibration signals compression is rather limited. The most popular approach relies

on transform based methods, mostly using various wavelet decompositions such as in [1–3]. The affinity for wavelets comes from two directions: their suitability for the compression of signals with non-stationary characteristics and their extensive use in feature extraction for fault diagnosis purposes [3–6]. However, in the reference [3], the author recalls that Wavelet Transform (WT) shows a significant gain over the harmonic approaches - Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT) - only for vibration signals with strong non-stationary characteristics.

The second category of methods is represented by predictive coding [7, 8]. In general, various Adaptive Differential Pulse Coded Modulation (ADPCM) schemes are adopted. However, these papers mostly propose basic approaches, and the reported results rely on a limited set of experiments.

In the category of more specific approaches, we can mention [9], where the signal is transformed into an image and then compressed in the domain of a lifting-scheme WT, or the use of Empirical Mode Decomposition (EMD) [10], which, due to its redundancy, is better suited for analysis than for compression.

It should be noted that most of the aforementioned papers focus mainly on signal analysis and feature extraction, compression being only marginally considered. As it has already been noticed, vibration signal processing relies usually on the use of some transformations/decompositions (WT, EMD). In this way, the signal is decomposed into several subbands, where the features of interest can be easily identified. It is well known, on the other hand, that Subband Coding (SBC) is one of the classical approaches in the data compression field [11]. In the case of mechanical vibration signals, a widely considered model is based on a sum of harmonics [12], which are related to the speed of the rotating machinery. These frequency components can be easily highlighted by a subband decomposition and effective compression can be achieved by an adaptive quantization of the subbands. Surprisingly, although it seems particularly suitable to vibration signal compression and analysis, classical subband decomposition was never considered for this subject, to the best of the authors' knowledge.

In this paper, we propose a new compression method for mechanical vibration signals. This method is referred to as Subband Adaptive Quantization (SAQ). SAQ can be viewed as an alternative approach to the classical SBC, based on filter banks [13]. It is well adapted to vibration signals properties of an aircraft engine environment: the frequency content is the superposition of a non-uniform colored noise (due to the combustion chamber, and structural

frames and mounts eigenmodes) and many integer harmonics of the fluctuating rotating speeds of the shafts (unbalances, misalignments, blade passing frequencies and gearmesh frequencies). During a flight, the rotating speed of each shaft is almost constant around few hundreds of Hertz during cruise with small fluctuations of few Hertz. The main shafts are not mechanically coupled; their rotating speeds are the result of the aerodynamic balance between compressors and turbines for a throttle set by the pilot. From flight to flight, those rotating speeds may vary by more than a dozen Hertz because of the contextual environment changes between two different flights (external ambient temperature change, for instance). In order to take into account those variations of the rotating speeds, a sampling rate of a dozen Hertz is sufficient for the acquisition of those parameters. SAQ of the vibrations for this application is therefore well adapted, because it allows simple and flexible subband decomposition on a large number of narrow-band signals, using suitable orthogonal transforms. Each subband is uniformly quantized on a number of bits which are adaptively allocated using an energetic criterion. The use of a uniform quantization, combined with an adaptive bit distribution, preserves system simplicity and allows high compression ratios to be obtained with a reasonable amount of distortion. This method is applied to a large number of real vibration signals, issued by plane engines in various working regimes and health statuses. Compression performance, evaluated from the reconstruction quality perspective, is compared with several existing approaches. It shows that SAQ is better suited to the vibration signals than the other methods. We also study the impact of SAQ-based compression on the detection of two of the most common defects of the engine: rolling elements fatigue and rotor-stator rubbing. We show that these defects remain detectable even when a strong compression has been applied to the signal.

In section 2, we explain how subband decomposition can be implemented using an orthogonal transform. The choice of the appropriate transform is investigated in section 3. In section 4, we introduce the principles of SAQ. Compression is tested on a large dataset of vibration signals, the results being presented in section 5. Compression impact on the fault detection of two particular defects of the engine is investigated in section 6, whereas the last section is dedicated to concluding remarks.

2. Transform-based subband coding

Since the compression method proposed in this paper relies on the SBC principle, we will focus in this section on SBC implementation using an orthogonal transform. In [14, 15] Malvar exhibits the link between transform coding and SBC. As Malvar emphasizes, SBC systems, in general, can only work with a reduced number of subbands, because of the computational complexity of the Quadrature Mirror Filters (QMF) that are usually employed. The use of a transform in an SBC implementation allows these issues to be bypassed. In this way, the SBC system can be implemented in a simple and flexible manner, making signal decomposition possible in a large number of subbands.

In order to illustrate the link between SBC and transform coding, we show in figure 1 the block scheme of a classical SBC system and in figure 2 its equivalent using transform-based implementation. The successive-in-time samples of each parallel output of the transform from figure 2 represent different subband components of the input signal. The link between the classical SBC system and its equivalent transform-based implementation is described by the following rules [14]: 1. The impulse responses of the synthesis filters are the transform basis functions; 2. The analysis filters are time-mirrored versions of the synthesis filters; 3. The decimation factor from SBC simply equals the transform length; 4. Perfect reconstruction is possible due to the orthogonality of the decomposition basis used in the transform-based approach, which is equivalent to the "lossless" property, as defined in [16].

In the following, we will mathematically formalize these properties. Let us consider the decomposition of the input signal into M subbands. In order to apply an orthogonal transform, the input signal must be split in blocks of L samples, represented by the vector : $\mathbf{x} = [x(0)\ x(1)\ ...\ x(L-1)]^T$. The orthogonal transform can be described in a matrix form as follows:

$$\mathbf{y} = \mathbf{W} \cdot \mathbf{x} \tag{1}$$

where $\mathbf{y} = [y(0)\ y(1)\ ...\ y(M-1)]^T$ is the vector of the transform domain coefficients and $\mathbf{W}$ is the transform matrix. For the usual discrete transforms (DCT, DFT), $L = M$, and $\mathbf{W}$ is a square matrix. These transforms meet the conditions for being part of a SBC system as shown in figure 2. Unfortunately, they exhibit an important drawback: the length of the basis vectors (and consequently the length of the filters from the SBC approach) equals the number of subbands. This leads to some undesirable errors around the block

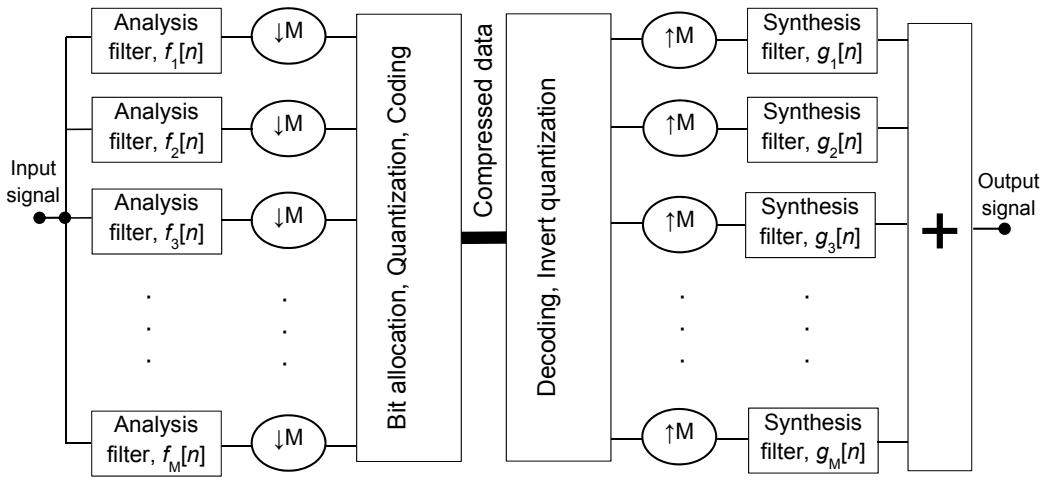

Figure 1: The block-scheme of a classical SBC system.

edges when quantization is performed in the transform-domain samples. This can only be avoided using a vector basis that spans over the data block to be processed.

Queiroz [17] has proven that the equivalence between SBC and transform coding holds for a special class of orthogonal transforms, called lapped transforms. For these transforms, the input signal vector is composed of samples spanning more than one data block ($L = K \cdot M$, K-integer). In this case, the buffer in figure 2 can be modeled as a L-size sliding window, which, in order to build up a new input block for the orthogonal transform is shifted by M samples only, thus creating superposition (lapping) between the adjacent blocks. Obviously, by choosing $L = M$, there will be no overlapping between the adjacent data blocks, which matches the case of the "regular" (non-overlapped) orthogonal transforms.

If we consider the most general case, the $M \times L$-size matrix $\mathbf{W}$ of the orthogonal transform ($L \geq M$), can be written as:

$$\mathbf{W} = [\mathbf{w_0}\ \mathbf{w_1}\ ...\ \mathbf{w_{M-1}}]^T \tag{2}$$

where $\mathbf{w_k}$ are L-size vectors of the orthonormal basis. The link between the classical SBC and its transform-based version is given by:

$$f_k(L-1-n) = g_k(n) = w_k(n) \tag{3}$$

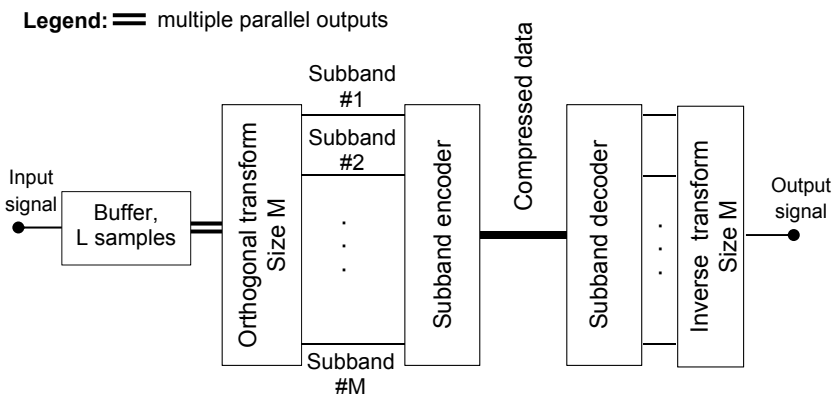

Figure 2: SBC implementation using an orthogonal transform.

where $f_k(n)$ and $g_k(n)$ are the impulse responses of the analysis and synthesis filters respectively, for the k-th subband.

3. Orthogonal transforms for the SBC of mechanical vibration signals

There are two key components of the SBC scheme shown in figure 2: the transform block and the subband encoder. In this section, we will focus on the first one, whereas the subband encoder will be examined in the next section.

The main quality of an orthogonal transform from a data compression perspective is its ability to concentrate signal energy into a small number of coefficients (or subbands), which allows effective adaptive quantization to be performed. Two transforms are widely used in data compression: the DCT and the DWT. Among various transforms, the DCT has been shown to be a good approximation of the statistically optimal Karhunen-Loeve transform, for a wide class of signals. The DWT is also successfully employed in various compression applications, due to its energy concentration properties as well as its ability to analyze a signal at variable time and frequency resolutions. On the other hand, the Lapped Orthogonal Transform (LOT), although less used in compression standards, has been shown to have similar or even better energy compaction properties compared to the DCT in some image compression applications [18].

Some principles related to the suitability of various orthogonal transforms for compression are recalled in [3], in the framework of mechanical vibration

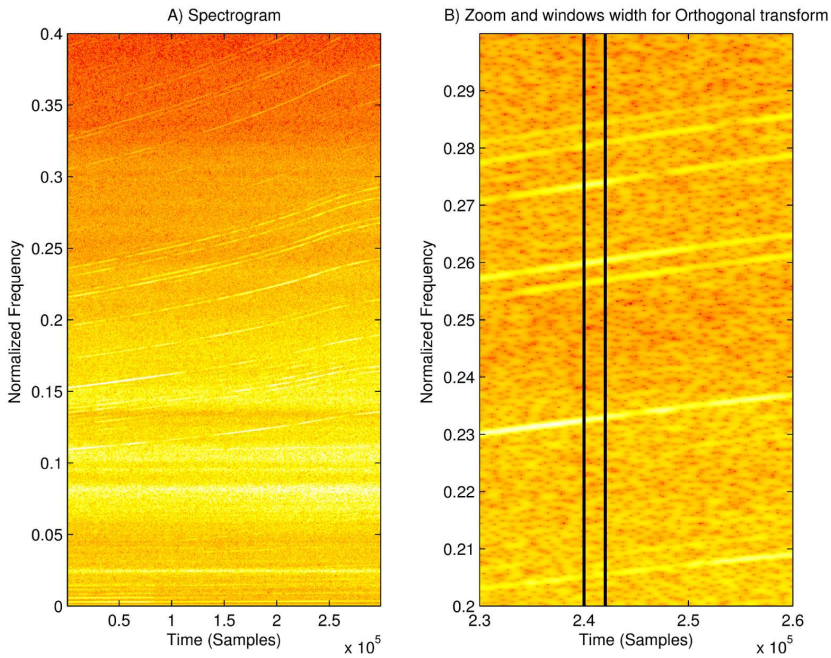

Figure 3: Spectrogram of a typical vibration signal.

signals. The choice of the transform must be made in accordance with the nature of the signal, and it is therefore highly dependent on the application [12]. Every vibration component may be described by its envelope and its instantaneous frequency. An important property of vibration signals is the speed of variation of these parameters. In the case of slow-varying characteristics, an harmonic basis (such as the one provided by DCT or LOT) should be preferred for decomposition, whereas the DWT is better suited for signals with strong non-stationary characteristics.

The signals measured by the accelerometers mounted on the plane engine are noisy non-stationary harmonic signals. A spectrogram of a typical signal is presented in figure 3. It may be seen in figure 3 a quite rich spectral content of noisy time varying spectral lines. A zoom is performed on the spectrogram to highlight one of the fastest variations of the signal parameters compared to the size of the processing window. It may be noticed that even in this worst case, the variation of the signal parameters is small with respect to the frequency resolution. Therefore, one can expect that the DCT and LOT perform better than the DWT, an empirical validation of the transform that is better suited for this kind of vibration signals is performed in the following.

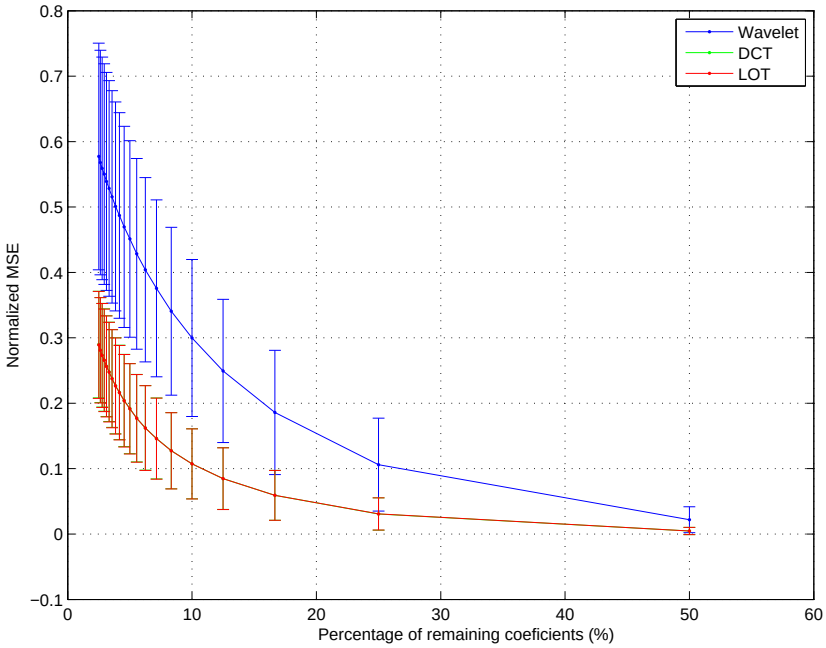

Figure 4: Normalized Means Square Error after compression using DWT($db20$), DCT and LOT.

A simple hard-thresholding procedure is used in order to evaluate the ability of various transforms (DWT, DCT, LOT) to concentrate energy on few coefficients. Thus, the lowest coefficients in the transform-domain are discarded and the signal is then reconstructed using only the remaining coefficients (no further quantization of the remaining coefficients is performed). The transforms are applied to blocks of $M = 2048$ samples.

The reconstruction quality is evaluated in function of the percentage of the remaining coefficients after thresholding using the Normalized Mean Square Error (NMSE):

$$\text{NMSE} = \frac{\sum_{n=0}^{M-1}(x[n] - \hat{x}[n])^2}{M\sigma_x^2} \tag{4}$$

where $x[n]$ is the original signal, and $\hat{x}[n]$ is the signal reconstructed after compression.

For the DCT implementation, several forms are known, issued from different boundary conditions. We choose the most used in signal and image processing which is the DCT of type II. In this case, the transform vectors that compose the matrix $\mathbf{W}$ are:

$$w_k(n) = \cos\left[\frac{\pi}{M}\left(n + \frac{1}{2}\right)k\right],\ k, n = 0, ..., M-1 \tag{5}$$

Various implementations are available for the LOT too. In this paper, we use a fast implementation proposed by Malvar [14], which is efficient for large values of the transform length, M. This approach uses a linear combination of DCT matrices of type I, II and IV in order to generate the basis vectors of the LOT. LOT basis vectors can be seen as smooth trigonometric waveforms. We recall that these vectors have the size $L = K \cdot M$, with K integer ($K = 2$ for the LOT used in this paper).

All the Daubechies' wavelet functions have been evaluated. The best performance for the DWT is provided by the twentieth order Daubechies' wavelet function ($db20$) more suitable for stationary regular data [3]. It has been used here.

In figure 4, the mean value as well as the standard deviation of the NMSE (evaluated using a dataset of 54 signals) are plotted versus the percentage of the remaining coefficients after thresholding in the transform-domain (DWT ($db20$), DCT, LOT). It is obvious that, while the DCT and the LOT lead to similar reconstruction performance, the DWT is not able to concentrate

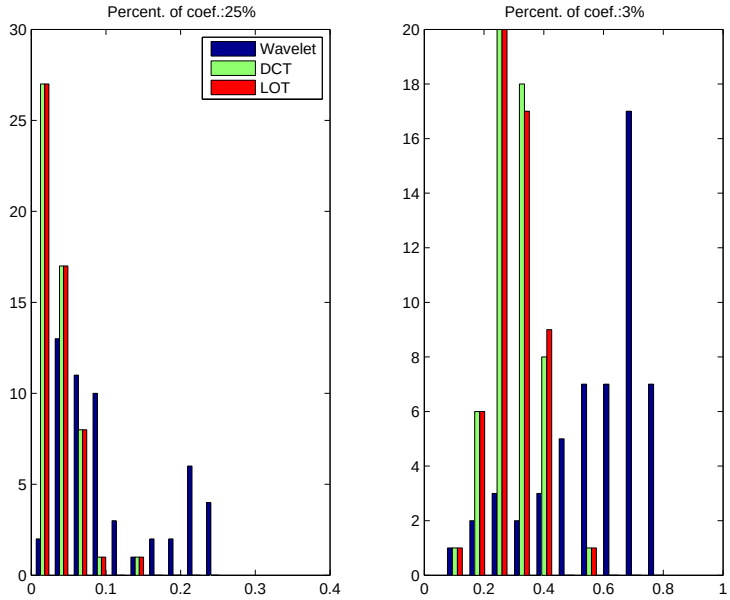

Figure 5: Histogram of the NMSE using DWT($db20$), DCT and LOT.

signal energy and, for a fixed percentage of coefficients, the reconstruction quality will be much lower. The figure 5 shows the histograms of the NMSE using DWT($db20$), DCT and LOT for 25% (left) and 3% (right) of remaining coefficients. It may be seen that the dispersion of NMSE for DWT is larger and that DWT can achieve mostly the same performance on some specific signals but for a large part of signals, the DWT provides worst performance especially for high compression rate with 3% of remaining coefficients.

The results shown in figures 4 and 5 clearly demonstrate that, although the DWT is widely used in compression applications, for the test dataset of 54 signals (see section 5 for more details on the dataset), the choice of a harmonic decomposition basis is by far more suitable than a wavelet decomposition. These conclusions are coherent with the analysis of the vibration signal specificities detailed at the beginning of this section.

For sake of completeness, we also evaluate the performance for few transient signals observed in the case of fleeting events like a shock caused by a small FOD (Foreign Object Damage, small birds ingestion for instance). An example of this kind of signal is shown on Figure 6 (black vertical lines represent the blocks width). Figure 7 presents the performance of the transforms for this class of signal. In this case, the NMSE is evaluated only for the block

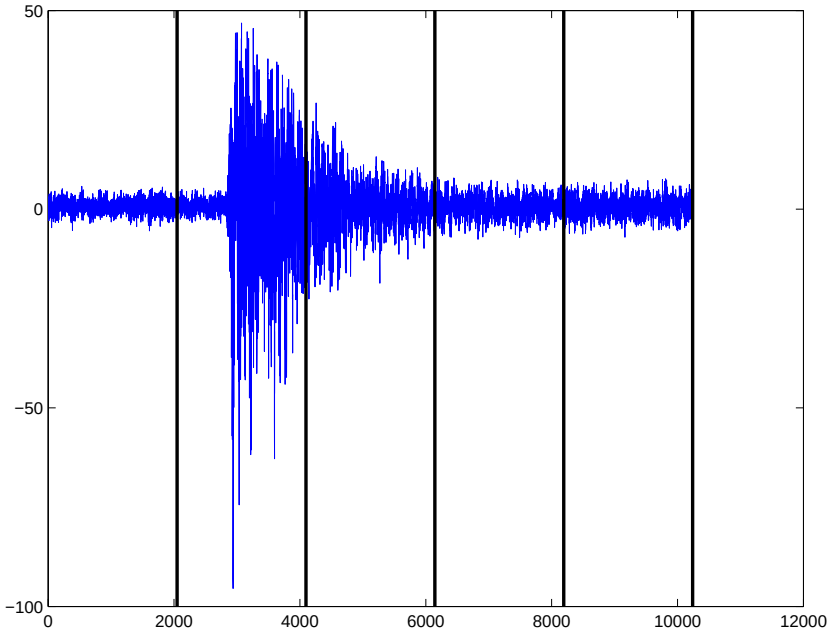

Figure 6: Example of signal vibration with transient phenomena due to small birds ingestion.

centered on the transient phenomena and the results are averaged over the 12 signals with FOD of the dataset. These numerical results show that the wavelet provides better performances: this fact can be easily explained by the ability of the wavelet to deal with highly non stationary signal. Nevertheless, due to the scarcity of this kind of phenomena for the signal we had to compress, we will focus our attention on the DCT and the LOT, which will be used in the implementation of the SBC compression system.

4. Subband Adaptive Quantization

The second key component of the SBC system shown in figure 2 is the subband encoder. This block has several components: a buffer (required to store successive in-time samples on each subband), a quantizer and an encoder of the quantized values. The quantizer works adaptively, based on an algorithm that effectively allocates the available quantization bits within subbands.

For the quantization operation, there is a large variety of choices [11]. The most widely used approach is a scalar quantization, in which each input

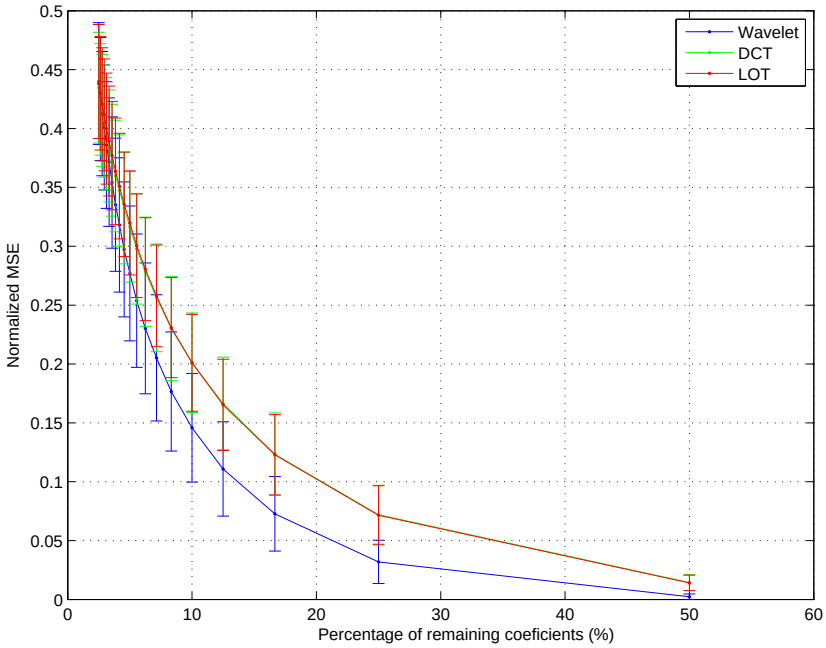

Figure 7: Normalized Means Square Error after compression using DWT($db20$), DCT and LOT for signals with transient phenomena.

sample of the source is individually mapped to a corresponding codeword. The simplest scalar quantization is the uniform one, but this approach is optimal only for cases where the data source to be quantized is uniformly distributed. The opposite approach is the non-uniform quantization, optimized using the probability density function (*pdf*) characteristic of the data source [19]. However, a considerable level of computational complexity is required to obtain a close match between the theoretical *pdf* and the real distribution of the data source. A mismatch of the *pdf* can significantly degrade the performance [11]. An alternative approach is "robust quantization" (or: companded quantization) that allows non-uniform quantization whose distortion is independent of the probability distribution of the data source [11].

The forward adaptive quantization [11], represents a good trade-off between the level of complexity and the performance. This approach requires data to be buffered in blocks prior to quantization. The knowledge of the properties of the data block (dynamic range, histogram) can be used to tune the quantizer accordingly. Furthermore, in the method proposed in this paper, the adaptability of the quantizer is not only defined on a per block basis, but also inside each block. The latter is obtained by making a smart bit allocation within the block, which is to adapt the uniform quantization to the energy of the subband components.

The M components provided by the orthogonal transform are quantized in the following way. The bit allocation is applied to blocks on N coefficients for each of the M subbands. They are represented by the, $M \times N$, $\mathbf{Y}$ matrix:

$$\mathbf{Y} = \begin{bmatrix} y_0(0) & y_1(0) & \dots & y_{N-1}(0) \\ y_0(1) & y_1(1) & \dots & y_{N-1}(1) \\ \vdots & \vdots & \vdots & \vdots \\ y_0(M-1) & y_1(M-1) & \dots & y_{N-1}(M-1) \end{bmatrix} \tag{6}$$

Each row of the matrix $\mathbf{Y}$ can be regarded as a subband component of the input signal. The compression relies on the fact that the energy of the signal is not equally distributed across the subbands and therefore it makes sense to allocate more quantization bits to the most "important" subbands. There are various ways to perform this allocation. In the reference [20], the bits are allocated proportionally to the maximal values of the samples in each subband. In this way, a compression of the dynamic range of the input signal is made, while the quantization step is kept approximately constant in each

subband. However, this method is rather empirical. Saywood [11] gives a criterion which optimizes the bit allocation for transform coding, in order to minimize the distortion introduced by the quantizer. The extension of this method to the SBC case is straightforward, and the bit allocation mechanism will be described in the following:

Step1. Let R denote the average desired number of bits used to code one sample. In this case, $N_b = M \cdot R$ denotes the total number of bits initially available for distribution to subbands. We set the initial number of bits allocated for sample coding in each subband to 0 ($N_k = 0$);

Step2. Compute the variance σ_k^2 of each row k of the matrix $\mathbf{Y}$. This accounts for the energy of each subband (we assume that there is no DC component);

Step3. Sort the variances $\sigma_k^2, k = 1, ..., M$. Suppose that the maximum is obtained for $k = k_M$;

Step4. Increment the number of bits corresponding to the subband k_M ($N_{k_M} = N_{k_M} + 1$). Divide $\sigma_{k_M}^2$ by 2;

Step5. Decrement the total number of available bits, N_b, by 1. If $N_b = 0$, stop, otherwise go back to step 3.

Such a procedure ends up by allocating more bits to the bands with higher energy. Its output will be a M-size column vector ("bit allocation table"), containing the values N_k, which needs to be saved as side information. In order to save resources, this table can be further encoded.

Once the number of bits which can be utilized in each subband has been determined, the uniform quantization can be implemented by a simple normalization. Thus, if we denote by $L(k) = 2^{N_k}$ the number of quantization levels in each subband, the codewords that are associated to the samples of the matrix $\mathbf{Y}$ can be determined by:

$$C_m(k) = \left\lfloor \frac{L(k)}{2} \cdot \frac{y_m(k)}{y_{\max}(k)} \right\rfloor \ , k = 0, ..., M-1; \ m = 0, ..., N-1 \qquad (7)$$

where $y_{\max}(k)$ are the maximum absolute values on each row of the matrix $\mathbf{Y}$. The codewords given by equation (7) are a set of integers in the interval $\left[-\frac{L(k)}{2}, ..., \frac{L(k)}{2} - 1\right]$. The quantized values, $\hat{y}_m(k)$, are obtained by inverting equation (7). This inversion leads to a distorted value of the coefficients from the matrix $\mathbf{Y}$, due to the integer part operator used in this equation. Because this operation uses the values $y_{\max}(k),\ k = 0, ..., M-1$, they need to be stored as side information. Finally, reconstruction is carried out by

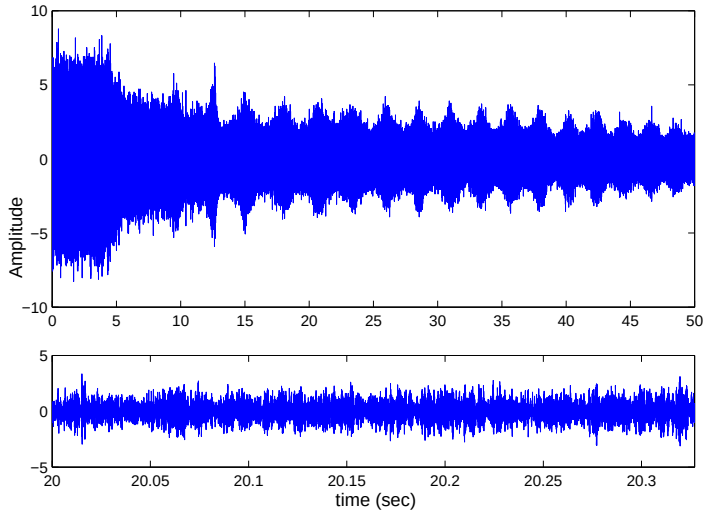

Figure 8: Example of signal measured during the deceleration phase: the full capture (up), zoom in on one processing block (down).

successively applying the inverse orthogonal transform on each column of the matrix of quantized values.

The compression ratio (CR) of a block of $M \cdot N$ samples can be computed as:

$$CR = \frac{M \cdot N \cdot n_b}{N \sum_{k=1}^{M} N_k + n_{\text{SIDE}}} \tag{8}$$

where n_b is the number of quantization bits of the original signal, N_k stands for the number of bits allocated for the quantification of each sample from the k-th band and n_{SIDE} is the number of bits used to encode the side information.

5. Experiments and results

For testing purposes, a large number of vibration signals were acquired using vibration sensors mounted on the engine. One example of signal, collected in the deceleration stage (after landing) is presented in figure 8. We show, in the upper part, the full capture, and in the lower part a zoom in one processing block.

We deal with non stationary signals, whose envelope and instantaneous frequency vary in time. Since we apply a block-based processing to non-

stationary signals, a particular interest must be taken in the choice of the block size. The degree of non-stationarity of the test signals is closely related to the engine's speed variation. The measures collected by two tachometers mounted on the engine allow straightforward computation of the rotation speed of the shaft. Therefore, the size of the data block can be tuned so as to meet a "quasi-stationarity" constraint. Thus, we have used a time window of $S = M \cdot N = 256 \cdot 64 = 16384$ samples, which corresponds to a time duration of 327 ms. An example of waveform for such a window is shown in the lower part of figure 8. By computing the variation of the rotation speed over the time window used, we conclude that the variation speed does not exceed 60 rpm, which corresponds to a "coherence time" of 1 s, during which signal characteristics do not change fundamentally.

Our objective is twofold: while in this section we assess reconstruction quality by using RSNR (9) versus CR plots, in the next section we evaluate compression impact on the fault detection algorithms.

$$RSNR = \frac{\sum_{n=0}^{M-1} x[n]^2}{\sum_{n=0}^{M-1} (x[n] - \hat{x}[n])^2} \tag{9}$$

The SAQ system described in section 4 is implemented using two orthogonal transforms, LOT and DCT. The two methods will be referred to as SAQ-LOT and SAQ-DCT respectively. In order to assess the feasibility of the compression algorithm, we apply this algorithm on 44 vibration signals acquired by three accelerometers, during different stages of the flight (start-up, take-off, landing, stop). All signals were sampled at $f_s = 50$ KHz, with a resolution of $n_b = 16$ $bits/sample$ and their duration depends on the working regime during which they were collected.

In figure 9 (up), we show an example of signal reconstruction, after compression with both SAQ versions described in this paper. Both reconstructed signals were compressed using an average rate of $R = 2$ $bits/sample$, which means that the amount of data needed to represent the signal was reduced nearly 8 times. One can notice the good quality of the reconstruction, despite the strong compression that was performed. However, while the SAQ-LOT approach leads to a reconstruction which is perceptually identical to the original, for the DCT-based method, some small deviations from the original waveform can be observed. In objective terms, the LOT-based method provides a gain of 3 dB over the DCT approach ($RSNR$ of 30.5 dB and 27.3 dB respectively).

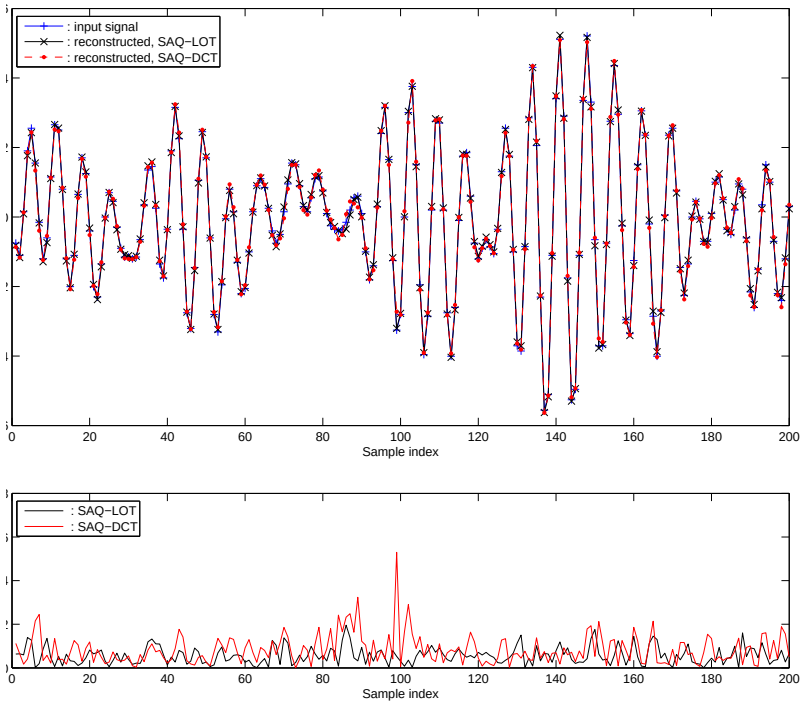

Figure 9: Signal reconstructed after compression (up) and absolute value of the reconstruction error (down).

One important part of this gain is produced by the reduction of the block-edge errors when using the LOT. In order to highlight this effect, we show in the lower part of figure 9 the absolute values of the reconstruction error for both methods. For the DCT-based method, we can clearly distinguish some peaks in the error signal around the block edge (the 100-th sample from figure 9). There are methods to reduce these kinds of errors, such as local filtering around the block edge or the use of overlapped blocks. While the first approach increases the complexity of the compression system, the second one reduces compression efficiency. It should be noted that the LOT already incorporates the later solution, in an efficient manner, with no impact on the CR, since L input samples are transformed into M LOT coefficients.

In figure 10, the compression results of both SBC methods are compared with other classical compression approaches. Several reference methods are considered: DWT thresholding (as applied in [3] for mechanical vibration signals), Robust Non-Uniform Quantization (RNUQ) and Adaptive Quantization of the Adjacent Frequency Components (AQAFC). RNUQ (also referred to as companded quantization) is one of the "classical" quantization approaches, presented, for example by Saywood [11]. The AQAFC is proposed by the authors in [21] for speech compression, but it has been shown to provide good results for vibration compression too [20]. Both methods are applied in the DCT domain.

Compression results, obtained by averaging $RSNR$ values over the whole dataset are shown in figure 10. It can be noted that the best result is reached by using the proposed SAQ-LOT. For $CR = 5$, the reconstruction gain provided by SAQ-LOT, in terms of RSNR, is approximately 4 dB over the alternative DCT-based approach, 8 dB over AQAFC and 14 dB over the the non-uniform quantization. The latter pays tribute to the constant codeword length that is used to encode all the quantization values. The method proposed in [3] has the poorest results, because, as shown in section 3, DWT is not able to concentrate the energy of the signals from the test dataset, and because of the non-adaptive nature of the method (a fixed quantization at 6 bits/sample is applied in all subbands). It can also be noted that, using the SAQ methods, we are able to reach a compression ratio of 20, while keeping, on average, the reconstruction quality over 10 dB.

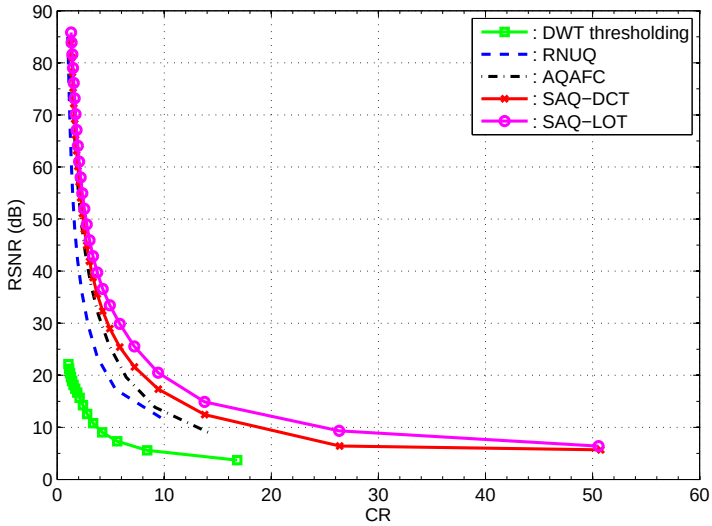

Figure 10: Global compression performance results.

6. Compression and fault detection

Although the $RSNR$ provides an objective criterion to evaluate the compression performance, this measure does not fully prove the feasibility of the compression method, when used as a pre-processing step in mechanical health monitoring systems. Indeed, more meaningful information can be obtained by evaluating the influence of the compression on fault detection. In the following, we will focus our attention on the detection of two common issues that can be encountered in plane engines. Two particular faults are considered in the following: rolling element wear and rotor-stator rubbing. Our main interest is to assess the influence that compression has on the capacity of detecting such faults, commonly encountered in plane engines. With this aim, we will compare detection results of the original signal with those obtained from the signal compressed at various rates. Due to its advantages, previously presented, we will keep SAQ-LOT as the compression method for the results that will be presented in what follows.

6.1. Rolling elements wear

This kind of problem is caused by the degradation over time of the rolling element bearing, which can rapidly turn into phenomena that can severely

damage the engine. This is why the preventive detection of such an issue is critical. The "signature" of this kind of fault consists in a series of energetic peaks grouped around central harmonics that occur at multiples of a fundamental frequency $(f_o - f_i)$, where f_o (respectively f_i) is the rotation frequency of the outer race (respectively the inner race). The frequency of energy peaks can be expressed as [22]:

$$f_{\mathrm{BSR}} = (f_o - f_i) \cdot \frac{D}{2d} \cdot \left[1 - \left(\frac{d}{D} \cdot \cos\alpha\right)^2\right] \tag{10}$$

where the geometrical bearing parameters are: D, the pitch diameter, d, the ball diameter and α, the angle of contact between the ball and the race. According to equation (10), f_{BSR} and all its harmonics are multiples of the difference $f_o - f_i$, referred to as "orders" of this difference.

Detection of rolling elements wear is done by looking for energy peaks around the even orders of $f_o - f_i$. Since the position in frequency of the peaks is known with a good accuracy, detecting this kind of fault is theoretically a straightforward procedure. However, due to signal non-stationarity, a simple Fourier analysis of the signal is not meaningful and time-frequency approaches must be used instead. Furthermore, since there is a direct link between the signature of the damage and the ball spin rotation, it is useful to re-sample the signal at a constant angle, instead of a constant time step. The spectrogram of such a signal then leads to a "time-order" representation. Figure 11, shows a zoom of the spectrogram of a signal corresponding to an engine affected by rolling element bearing damage, as well as the marginal energetic contribution of various frequencies. Note that the signal was measured during the deceleration stage (*e.g.* after landing), to obtain "the cleanest" vibration signal which can be used for fault detection purposes. The energetic signature of the fault can be identified on the spectrogram by some horizontal lines of high energy, occurring near the "expected frequencies", given by equation (10). The marginal energy distribution shown to the right of figure 11 clearly exhibits the faulty behavior of the engine. The impact of the compression is assessed by analyzing the peaks of the marginal energy distribution $(X(k))$.

In figure 12, we the influence of compression on the value of the first peak that occurs near $2f_{BSR}$ is highlighted. This example shows that there is no visible compression impact on the amplitude of the peak until $CR = 25$. Besides, the peak remains clearly distinguishable from the "baseline" even

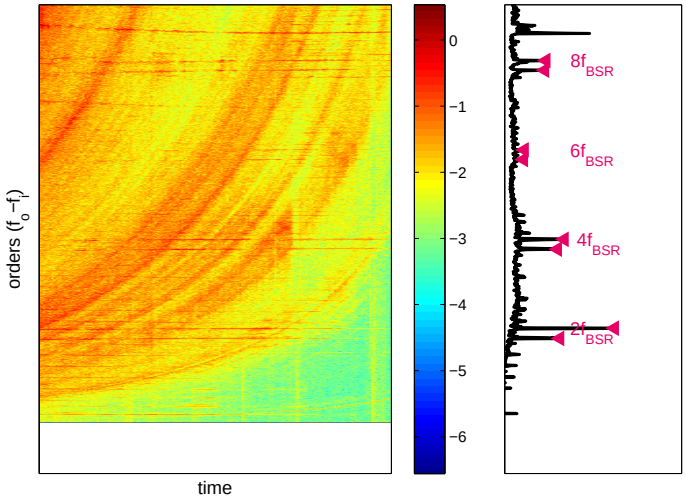

Figure 11: Spectrogram of the vibration signal issued by an engine under the wear of the rolling element bearing and the marginal sum of frequency components.

for very strong compressions ($CR = 50$).

In order to highlight the compression impact on the whole harmonics, we introduce a measure of the Relative Peak Attenuation (RPA):

$$RPA(CR, p_k) = \frac{(D_o(p_k) - D_r(CR, p_k))^2}{D_o(p_k)^2} \tag{11}$$

where p_k is the location of the peak, $D(p_k) = X(p_k) - \mu(p_k)$, and $\mu(p_k)$ is the average value of $X(k)$ around the peak. We use the subscript o for the original signal and r for the signal reconstructed after compression. An average measure of this criterion for the N_p detectable peaks is given by:

$$\overline{RPA}(CR) = \frac{\sum_{k=1}^{N_p} RPA(p_k, CR)}{N_p} \tag{12}$$

The dataset analysis has shown that the signature of the rolling element bearing wear is detectable for approximately 60 peaks. Average RPA results, obtained for the test signal compressed at various rates are shown in figure 13. They demonstrate that even for very strong compression ($CR = 50$), the impact on the energetic peaks remains within tolerable boundaries. Thus,

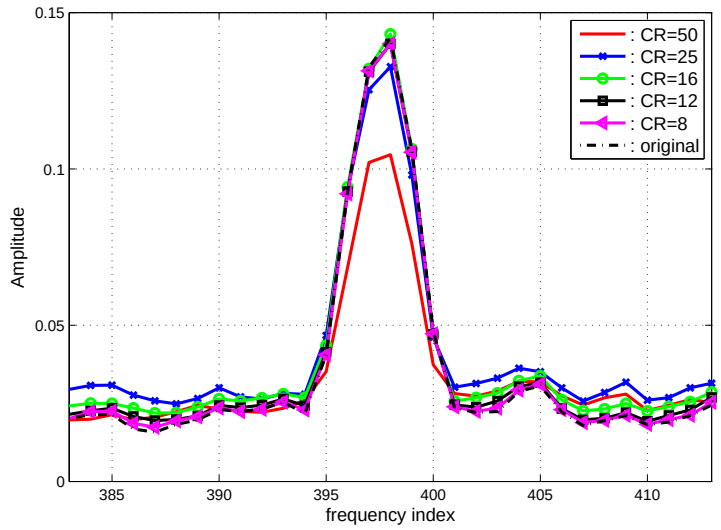

Figure 12: Example of compression influence on the first energetic peak caused by the wear of the rolling element bearing.

the RPA value of 0.04 (the highest in figure 13) corresponds to an average amplitude attenuation of 20%, which means that the peaks remain easily detectable.

As illustrated in figure 11, the energetic signature of the searched fault significantly varies across frequency. Since energetic peaks tend to decrease with the orders of $f_o - f_i$, the average results presented in figure 13 are subject to an important standard deviation around the mean. In order to have a more complete image on the way that the compression impacts the signature of the fault, we show in figure 14 the experimental Cumulative Density Functions (CDF) of the RPA parameter, at several compression ratios. Considering that an attenuation of 1% has little to no impact on the fault detection procedure, we may see in figure 14, that even for very strong compression ($CR = 50$), over 60% of the peaks carrying the energetic signature of the fault can be easily identified. This value exceeds 80% for $CR = 25$ and reaches 100% with a fairly good compression ($CR = 8$).

The results presented in this subsection prove that the SAQ-LOT compression method has little impact on the energetic signature of the particular fault investigated here. Visual detection of the rolling element bearing wear can be successfully performed by the expert, even for signals subjected to

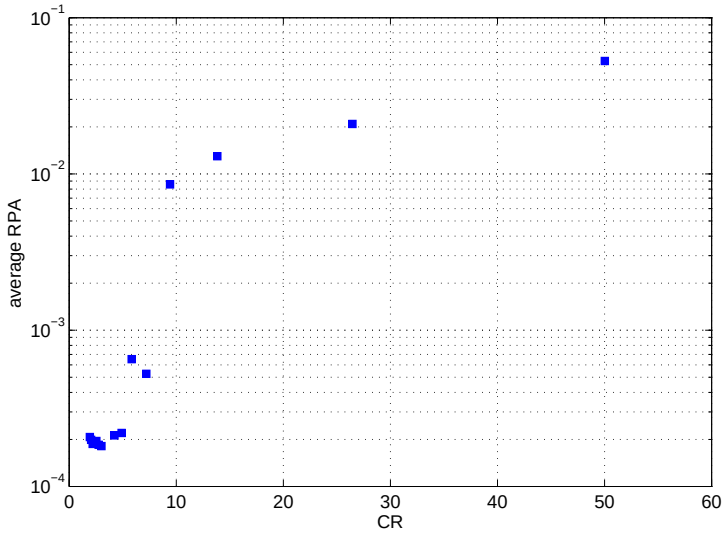

Figure 13: Compression impact on the energetic peaks determined by the rolling element bearing wear.

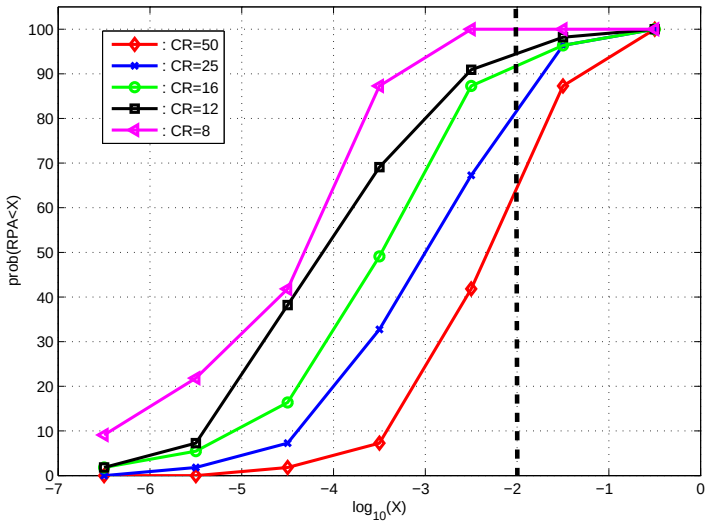

Figure 14: CDF of the RPA parameter at various compression rates.

very strong compression.

6.2. Rotor-stator rubbing

The rubbing between rotor and stator surfaces is a common fault for rotating machineries [22]. This fault can quickly have catastrophic effects for the rotating machinery. Cepstrum computation represents one of the strategies that can be used for defect detection in high speed machineries. The real Cepstrum of a signal, $x(t)$, which is used in our analysis for rubbing detection purposes is defined as:

$$C(\tau) = \mathrm{FT}^{-1}(\ln(\mathrm{FT}(|x(t)|)) \tag{13}$$

where FT stands for the Fourier Transform. The variable τ, is usually referred to as "quefrency", while its magnitude is also anagrammed into "gamnitude". The rotor-stator rubbing determines a particular signature to appear in the vibration signal with the periodicity of one revolution. The Cepstrum is a useful tool to highlight physical phenomena that cause periodic frequency components to occur in the spectrum. In order to analyze the compression effect on the vibration signals affected by rubbing we dispose of 6 signals measured for an engine where the rubbing phenomenon was induced for a short period of time.

In figure 15 we show the Cepstrum of a part of the signal when no rubbing exists, compared to the Cepstrum computed, for the same measured signal, during a rubbing episode. Experience has shown that the contact is well retrieved on the five first "rahmonics" of the cepstrum. Thus, a good criterion of detection would be obtained by comparing the sum of these rahmonics during the contact with that obtained for a healthy engine, or when no contact exists. The ability to distinguish a peak caused by rubbing can be formalized, for example by the parameter Rubbing Detection Criterion (RDC):

$$RDC = \frac{\sum_{n=1}^{5} C(\tau_n)}{5} - \overline{C} \tag{14}$$

where τ_n are the quefrencies corresponding to the five first rahmonics and $\overline{C}$ is the average of the cepstrum samples computed for $\tau \neq \tau_n$. In figure 16, we show the value of the RDC parameter versus the revolution index, for the original signal, as well as for the signal reconstructed after compression at various rates. Although a strong compression significantly impacts the computation of the RDC parameter (whose maximum is attenuated by 25%

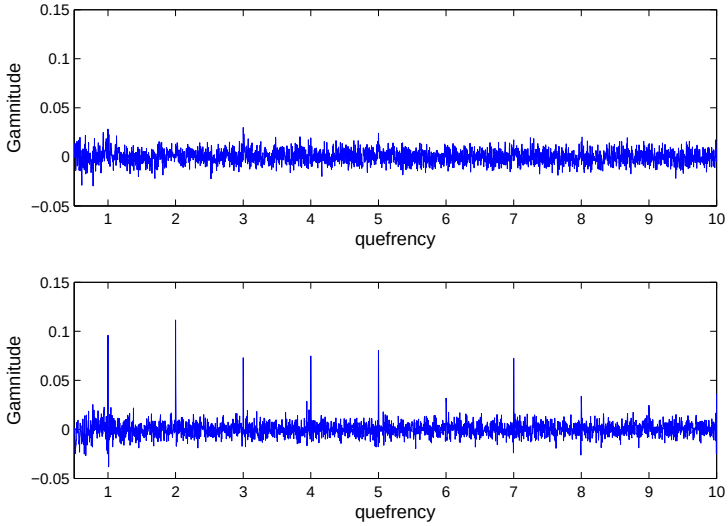

Figure 15: Example of Cepstrum representation during normal functioning (up) and during rotor-stator contact respectively (down).

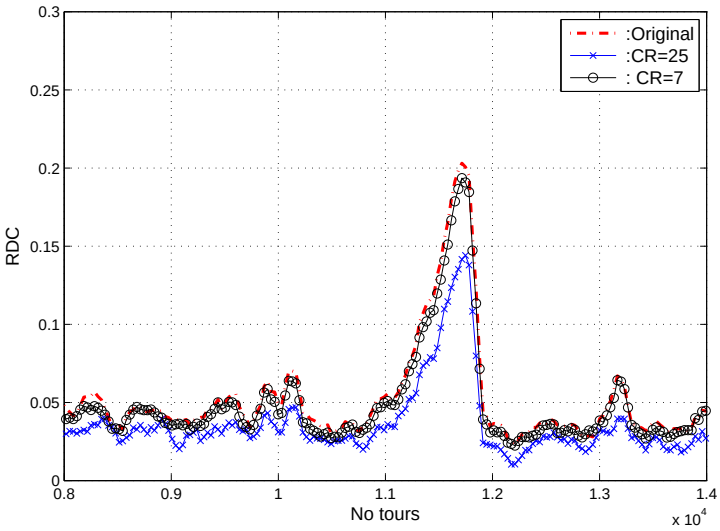

Figure 16: Rubbing detection example for strong and moderate compression.

for $CR = 25$), the impulse caused by the occurrence of rubbing remains clearly distinguishable. Furthermore, there is no meaningful compression influence, in this example, for moderate compression ($CR = 7$). While a binary decision regarding the existence of rubbing is straightforward even for strong compression, a finer analysis (*e.g.* regarding the number of revolutions for which the rubbing persisted) can be influenced by the compression. A solution could be to use an on-line detection of the rubbing phenomenon (*e.g.* when the criterion exceeds a given threshold) in order to adapt the compression rate when such a fault is detected for an accurate analysis after compression. It should, finally,be noted that all test signals we used in our study check the aforementioned conclusions, the rubbing signature remaining detectable even at strong compression.

7. Conclusions

In this paper we introduce a novel compression method for mechanical vibration signals. The method consists in a particular form of SBC implementation, based on the Lapped Orthogonal Transform. The selection of this transform is validated by the experiments carried out on our dataset. While the concept of SBC is well suited to the nature of the vibration signal, its transform-based implementation brings simplicity and flexibility. An adaptive uniform quantization is applied to the subbands, which provides results that are comparable with more sophisticated quantization approaches, while keeping the system simple enough.

The method proposed in this paper is tested on a large number of vibration signals collected by accelerometers mounted on plane engines. We show that, due to the properties of this transform, the LOT-based method provides better results compared to the DCT-based method, as well as to some other classical compression approaches. Strong compression can be performed with a reasonable amount of signal distortion: using LOT-SAQ the signals can be compressed 25 times with an average $RSNR$ of 10 dB.

Furthermore, we investigate the compression effect on the detection of two commonly encountered issues of the plane engine: rolling element bearing wear and rotor-stator rubbing. We show that even for strong compression, the impact on the energetic signatures of these misbehaviors remains moderate and allows straightforward fault detection on the compressed signal. Thus, at a CR of 25, 80% of the peaks carrying the energetic signature of the first defect are attenuated by less than 1%. For the same compression

ratio, the energetic signature of the rubbing phenomenon, while attenuated by approximately 25% in amplitude, remains clearly distinguishable for all the test signals.

[1] Y. Zhang, J. Li, Vibration sensor data compression and its effect on structural system identification, Smart structures and materials (Proceedings of SPIE) 6174 (2006) 61743E1–61473E11.

[2] M. Tanaka, M. Sakawa, K. Kato, Application of wavelet transform to compression of mechanical vibration data, Cybernetics and Systems: an International Journal 28 (1997) 225–244.

[3] W. Staszewski, Wavelet based compression and feature selection for vibration analysis, Journal of Sound and Vibration 211 (1998) 735–760.

[4] Z. Peng, F. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, Mechanical Systems and Signal Processing (18) (2004) 199–221.

[5] B. Liu, Adaptive harmonic wavelet transform with applications in vibration analysis, Journal of Sound and Vibration 262 (2003) 45–64.

[6] W. Wang, P. McFadden, Application of wavelets to gearbox vibration signals for fault detection, Journal of Sound and Vibration 192 (1996) 927–939.

[7] Y. Zhang, J. Li, DPCM-based vibration sensor data compression and its effect on structural system identification, Earthquake Engineering and Engineering Vibration 4 (1) (2005) 154–163.

[8] Z. He, L. Wu, H. Xu, Periodic prediction and switched adaptation in vibration signal compression, in: Proceedings of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Victoria, Canada, 1991.

[9] W. Bao, W. Wang, R. Zhou, N. Li, J. Yang, D. Yu, Application of a two-dimensional lifting wavelet transform to rotating mechanical vibration data compression, Journal of Mechanical Engineering Science, Proceedings of IMechE 223 (2009) 2443–2449.

[10] Z. Peng, P. Tseb, F. Chua, An improved Hilbert-Huang transform and its application in vibration signal analysis, Journal of Sound and Vibration 286 (2005) 186–205.

[11] K. Saywood, Introduction to data compression -third edition, Elsevier, 2006.

[12] M. Feldman, Hilbert transform applications in mechanical vibrations, Wiley, 2011.

[13] P. Vaidyanathan, Quadrature mirror filter banks, M-band extensions and perfect-reconstruction techniques, ASSP Magazine, IEEE 4 (3) (1987) 4 –20. doi:10.1109/MASSP.1987.1165589.

[14] H. Malvar, Lapped transforms for efficient transform/subband coding, IEEE Transactions on Acoustics Speech and Signal Processing 38 (6) (1990) 969–978.

[15] H. Malvar, Extended lapped transforms: properties, applications, and fast algorithms, IEEE Transactions on Signal Processing 40 (11) (1992) 2703–2714.

[16] Z. Doganata, P. Vaidyanathan, T. Q. Nguyen, General synthesis procedures for FIR lossless transfer matrices for perfect-reconstruction multirate filter bank applications, IEEE Transactions on Acoustics Speech and Signal Processing 36 (1987) 1561–1574.

[17] R. de Queiroz, On lapped transforms, Ph.D. thesis, The University of Texas at Arlington (1999).

[18] X. Zou, W. Pearlman, Lapped orthogonal transform coding by amplitude and group partitioning, Applications of Digital Image Processing XXII, Proceedings of SPIE 3808 (1999) 293–304.

[19] J. Max, Quantizing for minimum distortion, IRE Transactions on Information Theory 6 (1960) 7–12.

[20] M. Oltean, J. Picheral, E. Lahalle, H. Hamdan, Vibration signals compression with time-frequency adaptive quantization, in: Proceedings of the International Symposium on Intelligent Signal processing, WISP, Floriana, Malta, 2011.

[21] D. Isar, A. Isar, Speech adaptive compression using cosine packets, in: Proceedings of the 4th International Conference on Communications COMM02, Bucharest, Technical Military Academy, 2002.

[22] I. Howard, A review of rolling element bearing vibration: "Detection, diagnosis and prognosis", Tech. rep., DSTO Aeronautical and Maritime Research Laboratory (1994).