Lossy Vibration Compression through Matching Pursuit

D. Stefanoiu, A. Dumitrascu, J. Culita

"Politehnica" University of Bucharest, Faculty of Automatic Control and Computer Science, Romania
(Tel: 40-21-4029167; e-mails: {dan.stefanoiu, alexandru.dumitrascu, janetta.culita}@acse.pub.ro).

Abstract: This paper introduces a method for lossy compression of mechanical vibrations, based on the Matching Pursuit (a general purpose time-frequency-scale signal transform) and some usual compression algorithms. The compressed vibration can be sent through industrial communication channels. The ultimate goal is to detect possible mechanical faults, while analyzing the vibration at the destination. The algorithm employs a harmonic wavelets dictionary (atoms), from which the best approximates of the initial vibration are extracted. Thus, the samples of the raw vibration are replaced by projection coefficients on the dictionary and their indexes. In other words, instead of compressing the initial vibration samples, one has to compress the most fitted structure of the dictionary atoms, while removing the final residual that cannot encode useful information. Several compression algorithms are applied on both signals (the raw vibration and the harmonic wavelet coefficients), in order to compare their performances. It is shown that the compression ratio significantly decreases when using coefficients and indexes instead of vibration samples.

Keywords: Matching Pursuit, signal compression, time-frequency-scale transform, dictionary waveforms.

1. INTRODUCTION

This paper introduces an approach in which one of the most remarkable signal analysis procedures based on the procedure devised in (Mallat and Zhang, 1993), namely the *Matching Pursuit Algorithm* (MPA), is combined with some compression methods, in order to send mechanical vibrations through industrial communication channels. At the destination, vibration can first be reconstructed (approximately) and then analyzed (especially to detect possible faults of mechanical systems). One of the most exposed mechanical devices to flows during its exploitation is the bearing. (Gears also are very often affected by defects, but this article focuses on bearings.) Many bearing manufacturers have designed and implemented their own fault detectors. Usually, such a detector is some expensive mobile equipment using sensors, acquisition boards and a computational unit, all together being displaced nearby the mechanical system under test. This was and still is (at least for a while) the current practice. Nevertheless, in recent years, new concepts coming from Computer Science and Automation allow changing the classical practice and point of view to new ones. Nowadays, one operates very much with concepts like *cloud computing*, *internet of services* or *internet of things*. The main idea behind them is quite simple and effective: the user of a dynamical system can send tasks or information (*ask for services*) to a resource with very high computing power and unknown location (*the cloud*), via internet, in order to solve specific problems, very often related to parts of that system (*the things*). This allows the user keeping its own resources to a reasonable amount, without excessive expenses. The idea develops rapidly in industry, where one speaks more and more frequently about *virtualization in manufacturing*.

In case of fault detection problem based on mechanical vibration, the cloud computing technology yields sending such vibrations to a virtual fault detector, located in a cloud. So, instead of physically moving the equipment at the mechanical system, the user can acquire the necessary vibrations (by means of affordable accelerometers) and send them to the virtual fault detector. The vibration is one of the most employed signals in industrial monitoring, being characterized by non stationary and fractal nature (especially because of various corrupting noises), which overlaps on the harmonic behavior (due to mechanical rotation). Or, harmonic means redundant. Hopefully, the redundancy stands for an essential feature, which makes the vibration a good candidate for efficient data compression applications. The compression is necessary in order to efficiently exploit the communication channel, when sending data to some destination (e.g. to the virtual fault detector). Redundancy is usually quantified by the compression ratio, coarsely computed as the ratio between the sizes of compressed and original data. Many signal/data compression methods exists nowadays (Dobrescu and Kevorchian, 2002; Stefanoiu, 2003). Two approaches are under consideration when trying to compress vibrations: either the raw signal is directly compressed, or a transformation of vibration is applied on the raw signal (by using *Signal Processing* (SP) techniques), before any compression method. Given the above mentioned vibration characteristics, the second approach might serve better, provided that the transformation is selected wisely.

Quite a large panoply of SP transforms exists. Probably, the most appropriate vibration transforms lie in the field of *time-frequency-scale* (tfs) analysis. This field was devised about four decades ago and, since then, numerous applications traditionally developed in the framework of classical Fourier analysis have been reconsidered. The

scientific community rapidly understood the need to process non stationary signals with different and more accurate methods than the classical ones. The scientific community rapidly embraced the concept of *tfs transformation*. A comprehensive and quite exhaustive description of this field is due to (L. Cohen, 1989, 1995), who has shown that the various specific transformations can be grouped within about eight classes. Some tfs transformations, e.g. the *Wigner-Ville Distribution*, the *Short Fourier Transform* or the *Wavelet Transform*, became very popular during the last two decades, in theory as well as in applications.

The tfs transformations are extremely appealing for use in applications where "difficult" signals such as seismic, underwater acoustic, celestial scintillations, medical images, satellite images, speech or telecommunication waves/images have to be processed. (The difficulty of such signals consists not only of their non stationary behaviour, but also of their fractal nature, part of it being due to various noises.) However, sometimes, the tfs methods lead to NP hard algorithms that cannot efficiently be implemented without the help of Computer Science. Therefore, a symbiosis between some SP methods (especially those involving tfs algorithms, like MPA) and appropriate Evolutionary Computing strategies, in order to significantly reduce the computational complexity, was initiated for example in (Figueras i Ventura et al., 2001 – both articles; Stefanoiu and Ionescu, 2003) or (Qiang et al., 2011). In spite of the computational burden that it can involve, the MPA seemingly is an appropriate procedure to deal with vibrations.

Different types of MPA (classic, orthogonal) have recently been applied, not necessarily in conjunction with Evolutionary Computing techniques, in signals compression (Sun et al., 2016), image reconstruction by compressed sensing (Fang and Yang, 2012; Bi et al., 2014; Rup et al., 2015), image coding, signal classification using sparse representation (Rusu, 2011). Still, these research works do not address the problem of lossy vibration compression.

The article is organized as follows. Section II deals with the construction of a tfs dictionary (different from the dictionary introduced in (Mallat S. and Zhang S., 1993). Within the same section, the MPA is shortly recalled. Section III shortly recalls some basic compression algorithms (such as Shannon-Fanno's, Huffman's) and points out their limitations. Section IV is devoted to a comparative study of the MPA-compression procedure performance when using two types of bearing vibrations: flawless and affected by flaws. Some concluding remarks and a list of references complete the article.

2. THE EXTENDED MATCHING PURSUIT METHOD

Consider that the set of original data denoted by v is like a sentence said in an alien language, for which a translation dictionary is available. Moreover, assume that the sentence is "said" by someone who does not speak very well the language and thus the result is ambiguous or noisy. The main problem is then to decompose the sentence into correct and meaningful "words" extracted from dictionary and, eventually, to compress the meaning of sentence into a smaller number of words.

The approach described in this section relies on dictionaries generated by selecting a *mother waveform* (mw), a priori known, on which three elementary operations are applied: scaling, time shifting (translation) and harmonic modulation. Dictionaries with millions of words (also referred to as *atoms*) can thus be generated. The problem is to extract the appropriate words that express as accurately as possible the meaning of an original "sentence". In other words, the signal has to be represented by using dictionary atoms. A solution to this problem is served by using the MPA. The original version of MPA, as introduced in (Mallat S. and Zhang S., 1993), operates with dictionaries generated by scaling and time shifting only. We noticed that, by adding the harmonic modulation, on one hand, the compression performance improves and, on the other hand, the dictionary covers more classes of signals (for example, fractal harmonic signals can also efficiently be represented with dictionary atoms).

2.1. Building a Time-Frequency-Scale Dictionary

For the sake of clarity, one constructs first a continuous tfs dictionary. Next, one shows how the continuous dictionary can be discretized for practical purposes. The MPA steps are succinctly recalled in the end.

The construction starts from a continuous time and finite energy mw that can be selected from a large class of known signals, according to the nature of the original data. The mw is usually supposed to have a good time-frequency localization, i.e. to exhibit a fast decay (usually of exponential type) both in time and frequency (pulsation) domains. For example, the unit energy Gaussian function below can play the role of mw:

$$g(t) \overset{\text{def}}{=} A\exp\left[-\frac{(t-t_0)^2}{2\sigma^2}\right],\ \forall\, t\in\mathbb{R}, \tag{1}$$

where: $A = 1/\sqrt[4]{\pi\sigma^2}$ is the amplitude, $\sigma > 0$ is the sharpness and $t_0 \in \mathbb{R}$ is the central instant. The *Fourier Transform* (FT) of mw (1) is then:

$$G(\Omega) \overset{\text{def}}{=} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} g(t)\mathrm{e}^{-\mathrm{j}\Omega t}dt = \sigma A\,\mathrm{e}^{-\mathrm{j}\Omega t_0}\mathrm{e}^{-\frac{\Omega^2\sigma^2}{2}},\ \forall\Omega\in\mathbb{R} \tag{2}$$

(a harmonically modulated Gaussian) and its spectrum is $|G|^2$. It is well known that the Gaussians (1) and (2) have a practical support length (*aperture*) of 6σ (in time) and $6/\sigma$ (in pulsation), respectively, subject to the *Uncertainty Principle* (UP). The supports are actually measuring the tf localization. Similarly to the central instant $t_0 \in \mathbb{R}$, a *central pulsation* of spectrum can be defined as follows:

$$\Omega_0 \overset{\text{def}}{=} \frac{\int_{-\infty}^{+\infty}\Omega|G(\Omega)|^2 d\Omega}{\int_{-\infty}^{+\infty}|G(\Omega)|^2 d\Omega}. \tag{3}$$

For mw (1), definition (3) leads to null central pulsation. However, this can be non null for different dictionary atoms.

Another useful concept in construction of dictionary is the *half power bandwidth* defined by: $\Delta\Omega_{1/2}=\Omega_c^+-\Omega_c^-$. Here, $\Omega_c^\pm$ are *half power cut-off pulsations* obtained as solutions of equation:

$$\left|G(\Omega)\right|^2=\frac{1}{2}\left|G(\Omega_0)\right|^2 \quad \text{(half power)}. \tag{4}$$

For mw (1), the solutions of (4) are $\Omega_c^+=-\Omega_c^-=\sqrt{\ln 2}/\sigma$, which involves $\Omega_{1/2}=2\sqrt{\ln 2}/\sigma$.

Three linear operators can be applied on mw g, in order to generate a tfs dictionary:

1. *Scaling* (contraction/dilation) at scale $\alpha>0$:

$$\left(\sigma_\alpha g\right)(t)\overset{\text{def}}{=}\frac{1}{\sqrt{\alpha}}g\left(\frac{t}{\alpha}\right),\quad \forall t\in\mathbb{R}. \tag{5}$$

2. *Time shifting* (delay/anticipation) with duration $\theta\in\mathbb{R}$:

$$\left(\tau_\theta g\right)(t)\overset{\text{def}}{=}g(t-\theta),\quad \forall t\in\mathbb{R}. \tag{6}$$

3. Complex *harmonic modulation* by pulsation $\omega\in\mathbb{R}$:

$$\left(\mu_\omega g\right)(t)\overset{\text{def}}{=}\mathrm{e}^{\jmath\omega t}g(t),\quad \forall t\in\mathbb{R}. \tag{7}$$

Obviously, σ_α, τ_θ and μ_ω from definitions (5)–(7) are non-commutative operators, so they have to be applied in the order listed above. The generic *word/atom* of dictionary is:

$$g_{(\alpha,\theta,\omega)}(t)\overset{\text{def}}{=}\left(\mu_\omega\tau_\theta\sigma_\alpha g\right)(t)=\frac{\mathrm{e}^{\jmath\omega t}}{\sqrt{\alpha}}g\left(\frac{t-\theta}{\alpha}\right)=\frac{A\mathrm{e}^{\jmath\omega t}}{\sqrt{\alpha}}\mathrm{e}^{-\frac{(t-t_0-\theta)^2}{2\alpha^2\sigma^2}},$$

$$\forall t\in\mathbb{R}. \tag{8}$$

The index of a word/atom within the dictionary is denoted by $\xi=(\alpha,\theta,\omega)$. Atoms (8) verify some useful properties, as outlined next. Applying the harmonic modulation in the end leads to a smaller computational effort and, moreover, keeps the harmonic pulsation $\omega\in\mathbb{R}$ independent on the time shifting step θ, which involves a natural localization of atoms on tfs plane (as it will be shown later). Also, definition (8) shows that scaling affects both apertures in time and frequency. However, the unit energy of mw is inherited by all its children (as it can straightforwardly be proven). This property yields the energy conservation of original data when represented within dictionary. According to definitions (2) and (8), the FT of any atom g_ξ is:

$$G_\xi(\Omega)\overset{\text{def}}{=}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}g_\xi(t)\mathrm{e}^{-\jmath\Omega t}dt=\frac{\mathrm{e}^{\jmath\theta\omega}}{\sqrt{2\pi}}\left(\mu_\theta\tau_\omega\sigma_{1/\alpha}G\right)(\Omega),$$

$$\forall\Omega\in\mathbb{R}. \tag{9}$$

Equation (9) shows that a dual dictionary can be constructed in terms of FT. The three operators are applied in the same order, but their parameters are different: $1/\alpha$ for scaling, ω for frequency shift, θ for harmonic modulation. (A constant harmonic modulation by $\theta\omega$ has to be applied in the end as well.) The time aperture of atom g_ξ is $\Delta t_\xi=6\alpha\sigma$, while the spectral aperture is $\Delta\Omega_\xi=6/(\alpha\sigma)$. It can be shown by simple manipulations that the central pulsation (3), denoted here by $\Omega_{0,\xi}$, equals ω (for atom g_ξ) and generally is non null. This shows that the atoms can move along the frequency axis in a manner controlled by harmonic pulsation parameter ω. By solving (4) for atom g_ξ, one obtains the half power cut-off pulsations $\Omega_{c,\xi}^\pm=\omega+\Omega_c^\pm/\alpha$, which involves a half power bandwidth of $\Delta\Omega_{1/2,\xi}=\Delta\Omega_{1/2}/\alpha$. These results are direct and expected consequences of UP.

The tfs dictionary is denoted by $\mathcal{D}(g)\overset{\text{def}}{=}\left\{g_\xi\,\middle|\,\xi\in\mathbb{R}_+^*\times\mathbb{R}\times\mathbb{R}\right\}$ (where $g\equiv g_{1,0,0}$) and includes all atoms defined in (8). It can be proven that if g fulfils some requirements (such as good tf localization), the dictionary $\mathcal{D}(g)$ spans the Hilbert space of continuous time and finite energy (complex valued) signals, $\mathcal{L}^2$ (Daubechies, 1993; Mallat and Zhang, 1993). In this case, any signal of $\mathcal{L}^2$ can be represented as a linear combination of atoms taken from $\mathcal{D}(g)$. Another property results from the *separability* of $\mathcal{L}^2$, i.e. the capacity to support countable bases. A countable system of generators can be extracted from $\mathcal{D}(g)$, by discretizing the three main parameters, for example, as follows: $\alpha=\alpha_0^m$, $\theta=n\theta_0$, $\omega=k\omega_{0,m}$, where m (the *scale index*), n (the *time-shift index*) and k (the *harmonic index*) are integers, whereas $\alpha_0>0$, $\theta_0>0$ and $\omega_{0,m}>0$ are three preset *sampling steps*. Note that the harmonic sampling step has to be adapted to every scale representation, due to UP and atoms properties mentioned above. It follows that the generic discretized atom is (according to (8)):

$$g_{[m,n,k]}(t)\overset{\text{def}}{=}\alpha_0^{-m/2}\mathrm{e}^{-\jmath k\omega_{0,m}t}g\left(\alpha_0^{-m}\left(t-n\theta_0\right)\right),\quad \forall t\in\mathbb{R}. \tag{10}$$

The discretized tfs dictionary is naturally denoted by $\mathcal{D}_{\xi_0}(g)\overset{\text{def}}{=}\left\{g_{[m,n,k]}\,\middle|\,m,n,k\in\mathbb{Z}\right\}$ (where $\xi_0=(\alpha_0,\theta_0,\omega_0)$ and $g\equiv g_{[0,0,0]}$). Atoms (10) still are continuous time functions, so the dictionary is not yet discrete. Unfortunately, $\mathcal{D}_{\xi_0}(g)$ is not necessarily a generators system of $\mathcal{L}^2$ for any $\xi_0\in\mathbb{R}_+^*\times\mathbb{R}\times\mathbb{R}$, although $\mathcal{D}_{\xi_0}(g)\subset\mathcal{D}(g)$. To make $\mathcal{D}_{\xi_0}(g)$ span $\mathcal{L}^2$ (more specifically, a *frame* of $\mathcal{L}^2$), some quite technical results were proven in (Daubechies, 1993). As a practical consequence of Daubechies' theorems, if the mw $g\in\mathcal{L}^2$ has a good tf localization and at least one of the strings $\left\{\omega_{0,m}\right\}_{m\in\mathbb{N}}$ or $\left\{\omega_{0,-m}\right\}_{m\in\mathbb{N}}$ is upper bounded by $2\pi/\theta_0$, then $\mathcal{D}_{\xi_0}(g)$ becomes a frame of $\mathcal{L}^2$, independently on α_0. (The upper bound $2\pi/\theta_0$ results from a version of UP also known as *Gabor Uncertainty Principle* (Gabor, 1946). This result is known as *Daubechies' consequence*.

For practical purposes, it is not suitable to work within a countable dictionary containing continuous time atoms, but

rather within a finite one, consisting of discrete time atoms. A finite subset should be extracted from $\mathcal{D}_{\xi_g}(g)$ and the waveforms within the subset should be discretized. Denote by $\mathcal{D}_{\xi_g}[g]$ the set of all discretized atoms of $\mathcal{D}_{\xi_g}(g)$. The framework of continuous time finite energy signals $\mathcal{L}^2$ is then replaced by the discrete time, finite energy (complex valued) signals, denoted by ℓ^2. In general, if the discretization has been realized with respect to Shannon-Nyquist Sampling Theorems (Oppenheim and Schafer, 1995; Proakis and Manolakis, 1996), the subspaces of limited bandwidth signals from $\mathcal{L}^2$ and ℓ^2 have similar properties. Moreover, one can prove that $\mathcal{D}_{\xi_g}[g]$ spans ℓ^2, if $\mathcal{D}_{\xi_g}(g)$ spans $\mathcal{L}^2$. Let T_s be the sampling period employed to acquire the data v such that aliasing is attenuated. Then the generic discrete atoms of dictionary $\mathcal{D}_{\xi_g}[g]$ result by sampling atoms (10) with period T_s:

$$g_{[m,n,k]}[l] \overset{\text{def}}{=} \alpha_0^{-m/2} \mathrm{e}^{-\mathrm{j}\,\omega_{0,m} k l T_s} g\left(\alpha_0^{-m}\left(l T_s - n\theta_0\right)\right), \quad \forall\, l \in \mathbb{Z}, \quad (11)$$

where $l \in \mathbb{Z}$ denotes the normalized time instant. Hence, the dictionary $\mathcal{D}_{\xi_g}[g]$ includes all atoms (11). The full discretization (of parameters and time) involves two restrictions. Firstly, note that spectra of atoms (10) are centered around pulsations $\left\{k\omega_{0,m}\right\}_{k,m\in\mathbb{Z}}$. Thus, even if one of the strings $\left\{\omega_{0,m}\right\}_{m\in\mathbb{N}}$ or $\left\{\omega_{0,-m}\right\}_{m\in\mathbb{N}}$ was bounded, the string $\left\{k\omega_{0,m}\right\}_{k,m\in\mathbb{Z}}$ would diverge towards both $\pm\infty$. In other words, only a subset of $\mathcal{D}_{\xi_g}[g]$ consists of correctly sampled atoms (with period T_s). More specifically, only the atoms with spectral apertures included into the pulsations band $\left[0, \pi/T_s\right]$ are sampled with minimal aliasing. The other atoms should be removed from $\mathcal{D}_{\xi_g}[g]$, because they can distort the original data. Secondly, one cannot be sure whenever a finite subset of $\mathcal{D}_{\xi_g}[g]$ (minimal aliasing sampled atoms) still is a generators system of ℓ^2, even in case $\mathcal{D}_{\xi_g}(g)$ spans $\mathcal{L}^2$. Very likely, it is not. For simplicity, let the same notation $\mathcal{D}_{\xi_g}[g]$ indicate a finite subset extracted from $\mathcal{D}_{\xi_g}[g]$ and let $\left\langle \mathcal{D}_{\xi_g}[g] \right\rangle$ be the subspace spanned inside ℓ^2 by $\mathcal{D}_{\xi_g}[g]$ (as finite set of discrete atoms). In general, the original signal v does not belong to the subspace $\left\langle \mathcal{D}_{\xi_g}[g] \right\rangle$. But the distance from v to $\left\langle \mathcal{D}_{\xi_g}[g] \right\rangle$, can be minimized by *tuning* the mw and the dictionary on the original data. (The tuning operation is explained later.)

Evaluate now the frequency characteristics of atoms (11), when generated by mw (1). Obviously, for any atom $g_{[m,n,k]}$, the central pulsation is $\Omega_{0,[m,n,k]} = k\omega_{0,m}$, the half power cut-off pulsations are $\Omega^{\pm}_{c,[m,n,k]} = k\omega_{0,m} \pm \alpha_0^{-m}\sqrt{\ln 2}/\sigma$ and the half power bandwidth is $\Delta\Omega_{1/2,[m,n,k]} = 2\alpha_0^{-m}\sqrt{\ln 2}/\sigma$. Also, the time aperture is $\Delta t_{[m,n,k]} = 6\alpha_0^{m}\sigma$, while the spectral aperture is $\Delta\Omega_{[m,n,k]} = 6\alpha_0^{-m}/\sigma$.

Refer now to *tuning*. This important operation requires to specify from the beginning how the atoms are located over the tfs plane. The plane is defined by embedding tf tiles into local lattices, each of which corresponding to a constant scale. The plane will then be defined such that any relationship between scale and frequency is avoided. In this way, one keeps more freedom in representation and interpretation. Assume that the spectrum of original data v is essentially located into the frequency band $[\omega_l, \omega_r]/2\pi$ (with left and right bounds determined by pre-filtering). Also, consider that v consists of N sampled data, with period T_s. Fix arbitrarily the scale index $m \in \mathbb{Z}$. Then the tf lattice corresponding to scale α_0^m consists of maximum $N \times K_m$ rectangular tiles associated to couples of tf indices (n,k), where K_m is defined by:

$$K_m \overset{\text{def}}{=} \left\lfloor \frac{\omega_r - \omega_l}{\omega_{0,m}} \right\rfloor. \quad (12)$$

Every cell-box of lattice is uniquely associated to an atom $g_{[m,n,k]}$, as illustrated in Figure 1. All lattices together (i.e. for all scales) define the tfs plane.

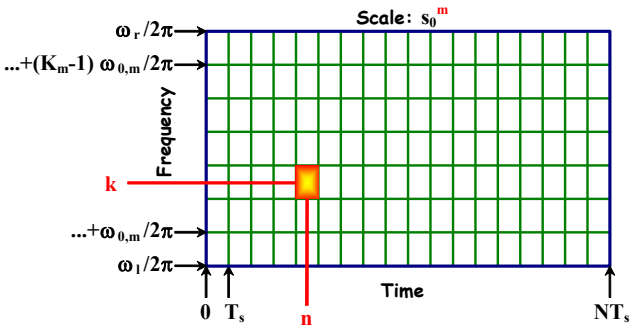

Fig. 1. A tf lattice with non negative axes, at constant scale.

The representation of tf plane intimately depends on the values of harmonic step $\omega_{0,m}$ as function of scale index m (due to UP). In order to set the values of $\omega_{0,m}$, the half power cut-off pulsations are used. Fix $m,n,k \in \mathbb{Z}$ arbitrarily and consider the following couple of discrete time atoms: $g_{[m,n,k]}$ and $g_{[m,n,k+1]}$. Their spectra are centered on $k\omega_{0,m}$ and $(k+1)\omega_{0,m}$, respectively. They are adjacent, as no other atom could have the spectrum centered in sub-band $\left[k\omega_{0,m}, (k+1)\omega_{0,m}\right]$. It is then suitable that the two adjacent spectra only overlap starting with their half band cut-off pulsations, i.e. $\Omega^{+}_{c,[m,n,k]} = \Omega^{-}_{c,[m,n,k+1]}$. This requirement relies on the following interpretation. The spectrum of any atom focuses not only on the spectral line pointed by the central pulsation, but also on a number of spectral lines located around the central pulsation. Whenever the spectral power is superior to the half of maximum, the corresponding spectral

power should be accounted. A sub-band is thus associated to every atom. Its bandwidth equals the half power bandwidth. The half power condition is equivalent to:

$$\begin{cases} k\omega_{0,m} + \dfrac{\sqrt{\ln 2}}{\alpha_0^m \sigma} = (k+1)\omega_{0,m} - \dfrac{\sqrt{\ln 2}}{\alpha_0^m \sigma} \\ \omega_{0,m} = \dfrac{2\sqrt{\ln 2}}{\alpha_0^m \sigma} = \alpha_0^{-m}\omega_0 \end{cases}, \qquad (13)$$

where $\omega_0 = 2\sqrt{\ln 2}/\sigma$ is a constant harmonic sampling step (also employed for scale 0). Equations (13) show that, unlike the time sampling step, the harmonic step has to be adapted to every scale by the inverse scale factor. The UP is naturally hidden behind this argument and causes an adaptation of atoms spectral aperture to the scale. Also, note that one of the strings $\{\omega_{0,m}\}_{m\in\mathbb{N}}$ or $\{\omega_{0,-m}\}_{m\in\mathbb{N}}$ from (13) is upper bounded by ω_0, regardless the value of α_0. If $\mathcal{D}_{\xi_\sigma}(g)$ spans $\mathcal{L}^2$, then it suffices to set the time sampling step such that $\theta_0 \le \pi\sigma/\sqrt{\ln 2}$ (according to Daubechies' consequence).

The tuning operation is completed by an argument aiming to set the remaining parameters with respect to original data characteristics. Consider that $\mathcal{D}_{\xi_\sigma}[g]$ consists of all atoms (11) generated by mw (1), for $m \in \overline{1-M_l, M_r-1}$, $n \in \overline{1-N_l, N_r-1}$, $k \in \overline{0, K_m-1}$, where the left and right bounds M_l, M_r, N_l and N_r are non negative integers that have to be specified next.

For each scale $m \in \overline{1-M_l, M_r-1}$, the number of frequency sub-bands (of half power bandwidths) is set in definition (12) above, by accounting the original signal bandwidth. As the time support of this signal is of length $(N-1)T_s$, setting the mw aperture to the same duration seems natural. This involves that the central point t_0 and sharpness σ are:

$$t_0 = \frac{N-1}{2}T_s \text{ and } \sigma = \frac{N-1}{6}T_s. \qquad (14)$$

Since the mw is the most dilated atom of dictionary, only contractions are allowed, in order to generate another atoms, which involves $\alpha_0^m < 1$. This requirement can be met either by setting $\alpha_0 < 1$ and $m \ge 0$ or $\alpha_0 > 1$ and $m \le 0$. It is however more practical to work with non negative scale indexes. Therefore, $\alpha_0 = 1/2$ is a choice similar to traditional setting within the framework of wavelets. Thus, any contraction halves the number of samples to represent the atoms and sets M_l to unit. This choice also suggests to set the total number of samples as a power of 2 ($N = 2^P$, with $P \in \mathbb{N}$, large enough). The time shifting sampling step θ_0 can be set equal to sampling period T_s. This setting keeps a relatively high degree of redundancy into the dictionary, with an affordable computational effort and simple atoms interpretation. A bigger θ_0 might result in a loss of information carried by the original data. It turns out that $N_l = N_r = N$. Moreover, due to the value of sharpness σ in (14), the condition $\theta_0 \le \pi\sigma/\sqrt{\ln 2}$ is automatically met. It follows that $\mathcal{D}_{\xi_\sigma}(g)$ (from which $\mathcal{D}_{\xi_\sigma}[g]$ is built) spans $\mathcal{L}^2$. Now, $m \in \overline{0, M_r-1}$, where M_r is the maximum number of analyzing scales allowed by the tfs dictionary. The most dilated atoms lie on scale #0 (their spectra being the sharpest among the other spectra, due to UP). Atoms are maximally contracted on scale $\#(M_r-1)$, while their spectra are maximally dilated. The Daubechies' consequence offers a first limitation (by using (13) for $m = M_r - 1$ and the settings $\alpha_0 = 1/2$, $\theta_0 = T_s$):

$$M_r \le \log_2\left(\frac{N-1}{3\sqrt{\ln 2}}\pi\right). \qquad (15)$$

Another restriction can be imposed as well: the spectrum aperture of the most contracted atoms has to be upper bounded by the original data bandwidth. More specifically, $\Delta\Omega_{[M_r-1,n,k]} = 6\alpha_0^{1-M_r}/\sigma \le \omega_r - \omega_l$. With $\alpha_0 = 1/2$, and (14), this inequality involves:

$$M_r \le \log_2\left[\frac{(N-1)(\omega_r-\omega_l)T_s}{9}\right] - 1. \qquad (16)$$

Clearly, the maximum number of scales M_r allowed by the dictionary $\mathcal{D}_{\xi_\sigma}[g]$ results now from inequalities (15) and (16). Note that, on the scale $\#(M_r-1)$, the frequency axis is split into three sub-bands (if $[\omega_l, \omega_r] \subset [0, \pi/T_s]$) or four sub-bands (if $[\omega_l, \omega_r] = [0, \pi/T_s]$).

The tuning argument above can be employed for any other specific mw, not necessarily of Gaussian type.

The necessary parameters to build the dictionary $\mathcal{D}_{\xi_\sigma}[g]$ are listed in the end of section. An example of dictionary atoms and their corresponding spectra is illustrated in Figure 2. In this case, the dictionary includes more than 4.2 million atoms, ranged on 9 scales.

2.2. Performing Matching Pursuit into the Dictionary

In general, the tfs dictionary $\mathcal{D}_{\xi_\sigma}[g]$ is redundant. Its atoms are not necessarily orthogonal to each other and the spectra of any two adjacent atoms overlap.

As already mentioned, the original signal v may not belong to the subspace $\langle \mathcal{D}_{\xi_\sigma}[g] \rangle$, but the tuning operation can diminish the distance from v to $\langle \mathcal{D}_{\xi_\sigma}[g] \rangle$. Actually, one can exploit this property to give conventional definitions of noise and useful signal. The *useful signal*, denoted by v_u, can be obtained by projecting the original signal v onto $\langle \mathcal{D}_{\xi_\sigma}[g] \rangle$. The *residual signal*, Δv, is orthogonal on $\langle \mathcal{D}_{\xi_\sigma}[g] \rangle$ and corresponds in this case to *useless noises*.

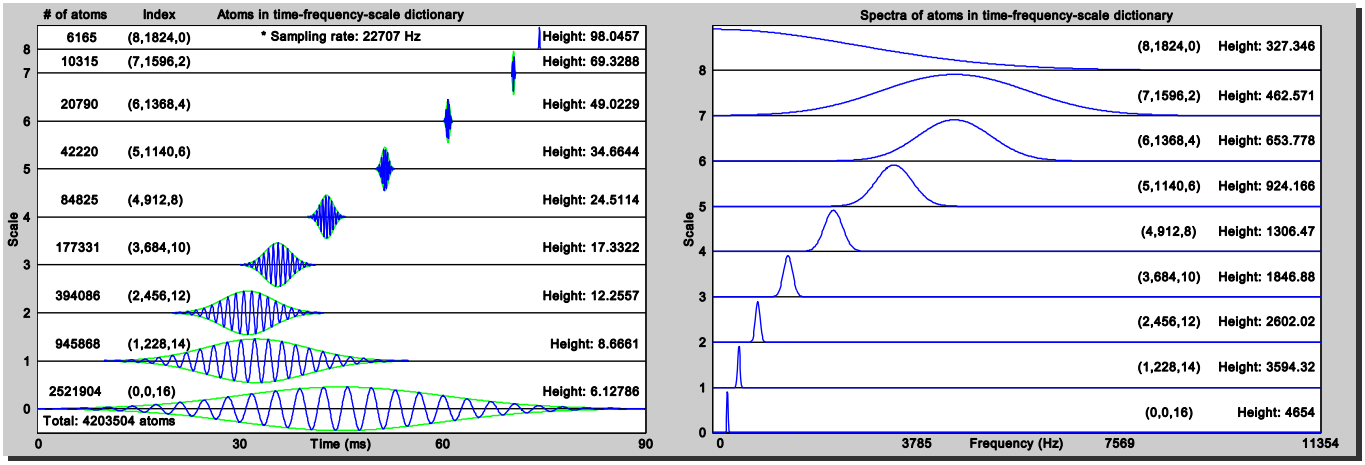

Fig. 2. Some tfs dictionary atoms (left) and their corresponding spectra (right).

Some other noises can belong to the useful signal though. These definitions are purely conventional and their quality is a consequence of how the dictionary has been built. (Tuning also plays an important role in this respect.)

Because of dictionary redundancy, v_u cannot easily be expressed, although it is a linear combination of its atoms. Usually, one estimates v_u with controlled accuracy, for example with the help of MPA, which relies on the concept of *best matching atom* (bma). When projecting the signal onto the dictionary (by means of the scalar product), the bma has *maximum magnitude of resulted projection coefficient*. According to MPA, the useful signal is recursively estimated by approximating the residual through the following process:

$$\Delta^{q+1}x \equiv \Delta^q x - \left\langle \Delta^q x, g_{[m_q,n_q,k_q]} \right\rangle g_{[m_q,n_q,k_q]}, \quad \forall\, q \geq 0\,. \quad (17)$$

The approximation process (17) starts with signal $\Delta^0 x \equiv v$ as the first estimation of residual (the coarser one). The corresponding bma $g_{[m_0,n_0,k_0]}$ is found and the projected signal is subtracted from the current residual. The new resulted residual $\Delta^1 x$ (that refines the estimation of noisy part) is looking now for its bma in $\mathcal{D}_{\xi_g}[g]$. Once the bma has been found, a finer residual estimation $\Delta^2 x$ is produced by subtraction (see equation (17)) etc. Iterations stop when the relative energy of current residual falls below a threshold $\varepsilon > 0$ a priori set, i.e. when the following inequality holds:

$$\frac{\left\|\Delta^{Q_\varepsilon} x\right\|^2 - \left\|\Delta^{Q_\varepsilon+1} x\right\|^2}{\left\|\Delta^{Q_\varepsilon} x\right\|^2} < \varepsilon \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \frac{\left\|\Delta^{Q_\varepsilon} x\right\|^2 - \left\|\Delta^{Q_\varepsilon+1} x\right\|^2}{\|v\|^2} < \varepsilon \frac{\left\|\Delta^{Q_\varepsilon} x\right\|^2}{\|v\|^2}, \quad (18)$$

where Q_ε is the number of bmas that have currently been found. In (18), any reference to the ideal (unknown) residual $\Delta v \equiv \Delta^\infty x$ is avoided. The second expression of (18) is especially interesting because it shows very clearly when the recursive process (17) should be stopped: whenever the relative energy gain obtained by a new projection on the dictionary subspace is smaller than the relative energy of the last projection, any further projection is useless. This means in fact that $\Delta^{Q_\varepsilon} x$ and $\Delta^{Q_\varepsilon+1} x$ are almost identical, i.e. nearly orthogonal on $\mathcal{D}_{\xi_g}[g]$.

What makes possible to consider a stop test like in (18)? The answer is given in (Mallat and Zhang, 1993). The authors have proven that the residual energy $\left\|\Delta^q x\right\|^2$ decreases as q increases. (If the infinite dictionary $\mathcal{D}_{\xi_g}(g)$ spans $\mathcal{L}^2$, then the energy $\left\|\Delta^q x\right\|^2$ approaches 0 as q increases.) The proof is based on a very elegant feature, related to the ancient Pythagorean relation between sides of a rectangular triangle:

$$\left\|\Delta^q x\right\|^2 = \left\|\Delta^{q+1} x\right\|^2 + \left|\left\langle \Delta^q x, g_{\xi_{q+1}} \right\rangle\right|^2, \quad \forall\, q \geq 0\,. \quad (19)$$

which shows that the initial energy is gradually spread along the projection coefficients, *independently on atoms orthogonality*. This property is exclusively due to employment of projections (which are orthogonal).

Actually, the identity (19) reveals that the principle of energy conservation holds (also due to the unit energy of dictionary atoms), as illustrated in Figure 3, where each arrow thickness is proportional to the residual energy.

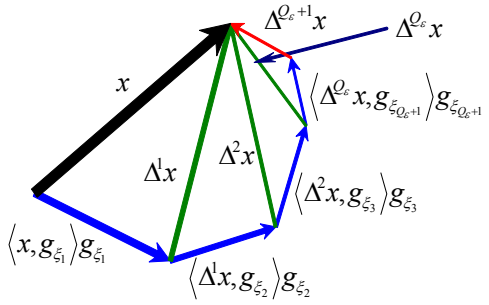

Fig. 3. The matching pursuit principle.

Beside this property, MPA also provides an exact synthesis equation:

$$x \equiv \sum_{q=0}^{Q_\varepsilon-1} \left\langle \Delta^q x, g_{\left[m_q,n_q,k_q\right]} \right\rangle g_{\left[m_q,n_q,k_q\right]} + \Delta^{Q_\varepsilon} x \,. \qquad (20)$$

Obviously, in (20), the residual is additive to the useful signal, x_u, which is estimated by accounting the bmas only:

$$x_u \cong \sum_{q=0}^{Q_\varepsilon-1} \left\langle \Delta^q x, g_{\left[m_q,n_q,k_q\right]} \right\rangle g_{\left[m_q,n_q,k_q\right]} \,. \qquad (21)$$

Note that, in (20) and (21), the projection coefficients (the scalar product) become from successive residuals and not only from the original data.

The main problem of MPA is to find the bmas within the discrete dictionary $\mathcal{D}_{\xi_q}[g]$, up to a number Q_ε, for a given signal. One copes thus with the following non linear maximization problem, with respect to indices $m \in \overline{0, M_r - 1}$, $n \in \overline{1-N, 1+N}$ and $k \in \overline{0, K_m - 1}$, at each stage $q \geq 0$:

$$\max_{m,n,k} \left| \sum_{l \in \mathbb{Z}} \Delta^q x[l] g^*_{[m,n,k]}[l] \right| , \qquad (22)$$

where a^* is the complex conjugate of a. The sum in (22) is actually finite, because the residual and atom supports are practically finite. Though the searching space $\mathcal{D}_{\xi_q}[g]$ is also finite, it usually includes a huge number of atoms to test (e.g. more than 4.2 million atoms – see Figure 2). The exhaustive search is definitely inefficient (actually NP hard). The Gradient-based Optimization Methods are also impractical, because the cost function (the scalar product) is extremely irregular and changes in every step of iteration, for every new residual. In order to clearly reduce the computation effort in search of the best matching atom structure, one of the optimization method based on metaheuristics described in (Stefanoiu et al. (2014)) was effectively integrated in MPA.

To conclude this section, the main steps of vibration transformation algorithm are listed below.

1. Input the vibration frame v and the stop threshold ε.
2. Input the dictionary configuration parameters: the type of mother waveform, g (e.g. Gauss); the signal frame length, $N = 2^P$; the sampling rate, $F_s = 1/T_s$; the estimated cut-off pulsations, ω_l (left) and ω_r (right); the mw central point, t_0; the mw sharpness, σ; the scaling factor, α_0; the time-shift step, θ_0; the harmonic modulation step, ω_0.
3. Evaluate: the number of analyzing scales, M_r; the maximum number of time-shift steps, $2N-1$; the number of frequency sub-bands per scale, $\{K_m\}_{m \in \overline{0, M_r - 1}}$.
4. Set the first residual $\Delta^0 x$ equal to the vibration frame v.
5. While the stop threshold is not reached (according to inequality (18)): use some metaheuristic to solve the optimization problem (22) for the current residual $\Delta^q x$ ($q \geq 0$) and store the corresponding bma (indexes and value); compute the next residual according to equation (17); increment the counter of residuals ($q \leftarrow q+1$).

3. THE TESTED COMPRESSION ALGORITHMS

The stop threshold $\varepsilon > 0$ actually controls the energy loss and, consequently, the compression performance. Compression is achieved when the projection coefficients $\left\{\left\langle \Delta^q x, g_{\left[m_q,n_q,k_q\right]} \right\rangle\right\}_{q \in \overline{0, Q_\varepsilon - 1}}$, together with their corresponding indexes $\left\{\xi_q = \left[m_q, n_q, k_q\right]\right\}_{q \in \overline{0, Q_\varepsilon - 1}}$ from (20) or (21), involve smaller transmission bit flow than the initial data. If the residual $\Delta^{Q_\varepsilon} x$ is ignored, the compression is lossy. Usually, the *standard deviation* (std) of $\Delta^{Q_\varepsilon} x$, namely $\sigma_x^{Q_\varepsilon}$, is also sent to destination (instead of the residual itself). Thus, the user can add a Gaussian white noise with the same std $\sigma_x^{Q_\varepsilon}$ to the useful signal x_u, which may lead to a better approximation of the genuine signal x. But, even in this case, the compression is lossy. Beside the atoms indexes and the residual std, another side information that can be compressed and transmitted is the set of dictionary configuring parameters. Nevertheless, very often, this information is a priori known at the destination.

In order to compress the transformed signal (through MPA), three lossless compression methods were considered, such as: *Shannon-Fano* (S-F), *static Huffman* (sH) and *dynamic (adaptive) Huffman* (dH). They can be applied both on the genuine signal and the transformed one, in order to compare the performance. The corresponding compression algorithms are described in detail, for example, in (Dobrescu and Kevorchian, 2002; Stefanoiu D., 2003). They belong to alphabet based procedures class. A short discussion on some generic characteristics is presented in the sequel.

The essential of S-F algorithm consists of defining the re-encryption codes according to a binary tree constructed by using the data set. On the output data flow, one sends two types of data: effective data encoding the useful information (string of new codes) and auxiliary data encoding the side information (alphabet and symbol counters). The algorithm is efficient if the length of auxiliary data is smaller than length of effective data, but the auxiliary data are necessary in decompression phase. Also, the algorithm is more efficient for larger data sets, because the weight of side information is smaller. To decompress the data, it suffices to build the tree starting from the side information and to parse it according to the useful information. More specifically, the new codes are bit-by-bit parsed, in parallel with the tree. Every time a leaf is reached, the tree parsing re-starts from root.

Within the sH algorithm, the compression and decompression phases are the same as within the S-F one, but the construction of binary tree begins from leafs to the root. This means that the codes of symbols are built from the least significant bit to the most significant bit. Nevertheless, the Huffman codes are different from Shannon-Fano codes. Normally, the performance of S-F and sH algorithms is quite similar. Differences can occur in case the data sets are very large, in which case sH may exhibit slightly better performance than S-F.

The dH algorithm is governed by a very simple principle: the re-encryption codes are generated/decoded simultaneously with the compression structure (the binary tree), as soon as the data stream is parsed. Thus, the code of a symbol may vary during the procedure, depending on the value of the symbol counter at the very instant of the code generation/symbol decoding. The codes are thus *adapted to the data stream*. In this case, the volume and the alphabet of data sets are not a priori known. Therefore, the binary tree is constructed such that the so called *fraternity property* is adaptively preserved. This property is defined by three restrictions: (a) the sons of the same parent (referred to as *brothers*) are increasingly indexed, from left to right; (b) each node has a weight that decreases or stays constant when its index increases; (c) when two parents lie on the same hierarchical level in the tree, the indexes of left-parent sons are smaller than the indexes of right-parent sons.

Comparing to static methods, the dynamic method is able to perform an effective compression even for short data sets. For large data sets, the compression performance is even better. But the complexity of dH algorithm is bigger. Another advantage of the dynamic method is revealed by the fact that the original data set is transited only once, whereas within the static methods the data are transited twice.

In the framework of this article, the compression performance has two measures: the *signal-to-noise ratio* (SNR) and the *compression ratio* (CR). They are defined as follows:

$$\mathrm{SNR} \overset{\mathrm{def}}{=} 20\log_{10}\left(\frac{\sigma_x}{\sigma_x^{Q_\varepsilon}}\right), \quad \mathrm{CR} \overset{\mathrm{def}}{=} 100\times\frac{\beta_c}{\beta_d}, \qquad (23)$$

where: σ_x is the std of genuine signal (x), β_c is the compressed data volume (in bytes) and β_d is the initial data volume (in bytes). As the definitions (23) show, SNR is expressed in *decibels* (dB), while CR is expressed in percents (%). The energy loss is measured by the inverse of SNR: the bigger the SNR, the smaller the energy loss. The efficiency of lossless compression is measured by the inverse of CR: the smaller the CR, the better compression.

One last problem arises in this context: how to use the lossless compression algorithms (operating on strings of bytes) in case of floating point double precision data (like vibration samples, tfs coefficients, std values)? Normally, one can directly apply the algorithm on the raw data block. But, in our case, the vibration acquisition apparatus (as described in next section) only provides data in ASCII format with a preset number of digits (say 20), given that the vibration samples are expressed in $\mathrm{cm/s^2}$ and vary in range of $[-10,+10]$. Conversion to ASCII is a supplementary source of error and, maybe, of energy loss. But the user can control the conversion accuracy, by setting the appropriate number of digits to be expressed as ASCII codes.

4. SIMULATION RESULTS

The tests were developed through the vibration platform illustrated in the left side of Figure 4. Three main systems are connected: a mechanical stand, a vibration data acquisition and pre-processing apparatus and a PC. All technical characteristics are written on the figure. The mechanical stand includes: a three phase electrical engine; a couple of bearings mounted into detachable mechanical seats (the bearing close to the engine is a standard high quality one, without defects; the other bearing is the tested one and could be flawless or with possible defects; a couple of metallic discs mounted between bearings (to produce a constant load applied on bearings and to create an inertial momentum that keeps the rotation alive for more than 5 s after the engine has been shut off); an elastic coupling between engine axis and load axis (to attenuate the engine self sustained vibrations or shaft wobbling, that could corrupt the data). The vibration is acquired by using 2 light accelerometers in quadrature. Thus, complex valued vibration data are provided, where the horizontal accelerometer gives the real part, while the vertical one gives the imaginary part. The acquisition device is LMS Roadrunner and yields accurate pre-filtering of data simultaneous acquisition on at least 2 channels and selectable accuracy. Data are only provided in ASCII format with a number of digits selected by the user. The maximum allowed sampling frequency is 100 kHz. In this application, the sampling frequency has been set to 20 kHz. Data are represented by 20 digits/value and transferred to a PC.

The tested bearings have characteristics displayed in the right side of Figure 4. Four types of bearings were considered: S3850609 (standard, without defects), I3850609 (with a crack on the inner race), O3850609 (with a chop on the outer race) and M3850609 (with multiple defects on balls and races). The number "3850609" comes from the geometrical main characteristics of bearings (see Figure 4 again: inner diameter of 38.5 mm, ball diameter of 6 mm and 9 balls). Figure 4 also shows the value of sampling rate ($\nu_s = 20\,000$ Hz), the nominal rotation speed of shaft ($\nu_s = 44.3$ Hz – about 2660 rot/min) and the natural frequencies of bearing in decreasing order (BPFI – Ball Pass Frequency on Inner race; BPFO – Ball Pass Frequency on Outer race; BRF – Ball Rotation Frequency; CFI – Cage Frequency with respect to Inner race; CFO – Cage Frequency with respect to Outer race). All these frequencies play a major role in bearing faults detection and identification.

Vibrations have been pre-filtered in range 0.5-11.35 kHz and segmented in frames of 2048 samples. For each bearing, 7 to 8 frames were acquired. The gross estimated SNR is less than 6 dB (i.e. more than 30% of acquired data are noises), while the dictionary includes more than 4.2 million Gaussian atoms generated by mw (1). The MPA is applied on frames and the resulted tfs coefficients are compressed, as previously described. In all cases, the compression performance was superior to the case when the raw vibration is employed instead of transformed vibration.

Hereafter, the discussion focuses on the first frame of each vibration signal. Figures 5-8 display the first frame of acquired vibrations for bearings {S,I,O,M}3850609. Each figure shows the real part (left) and the imaginary part (right) of signals. On top the raw vibration is drawn. Next, in the middle, the removed residual is drawn. Raw vibration versus removed residual are drawn at the bottom. The residual was obtained by setting $\varepsilon = 0.01$ in (18).

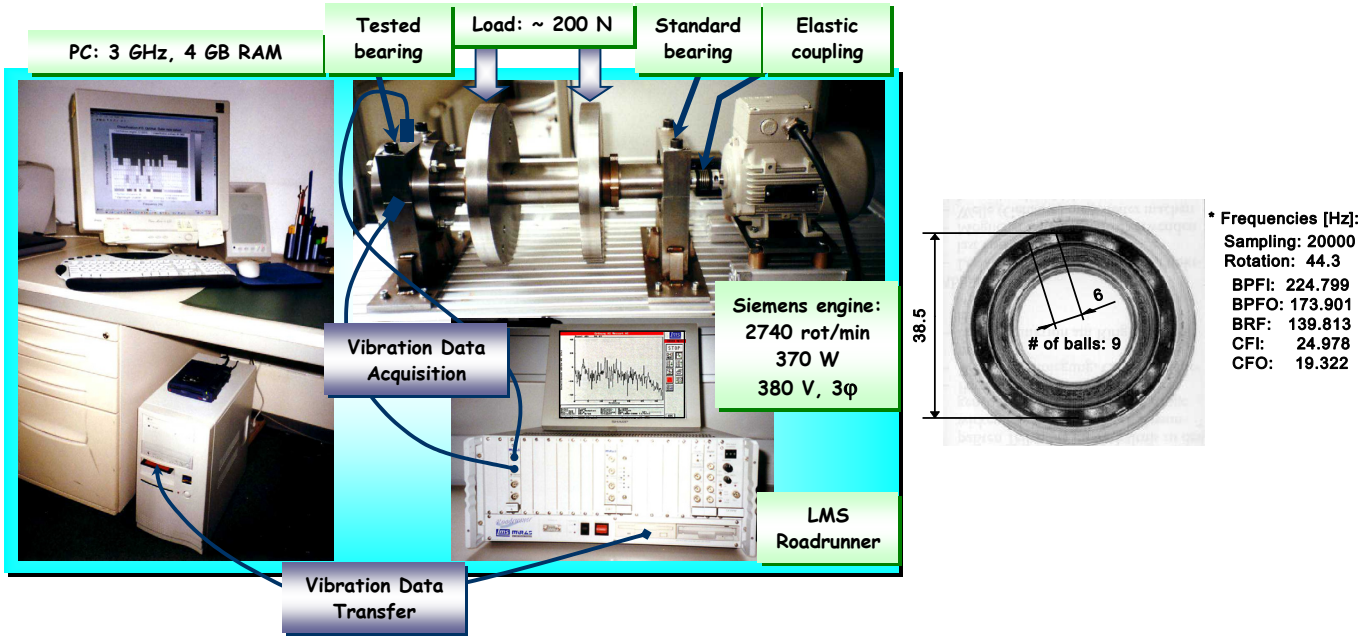

Fig. 4. Bearings testing platform (left) and bearings general characteristics (right).

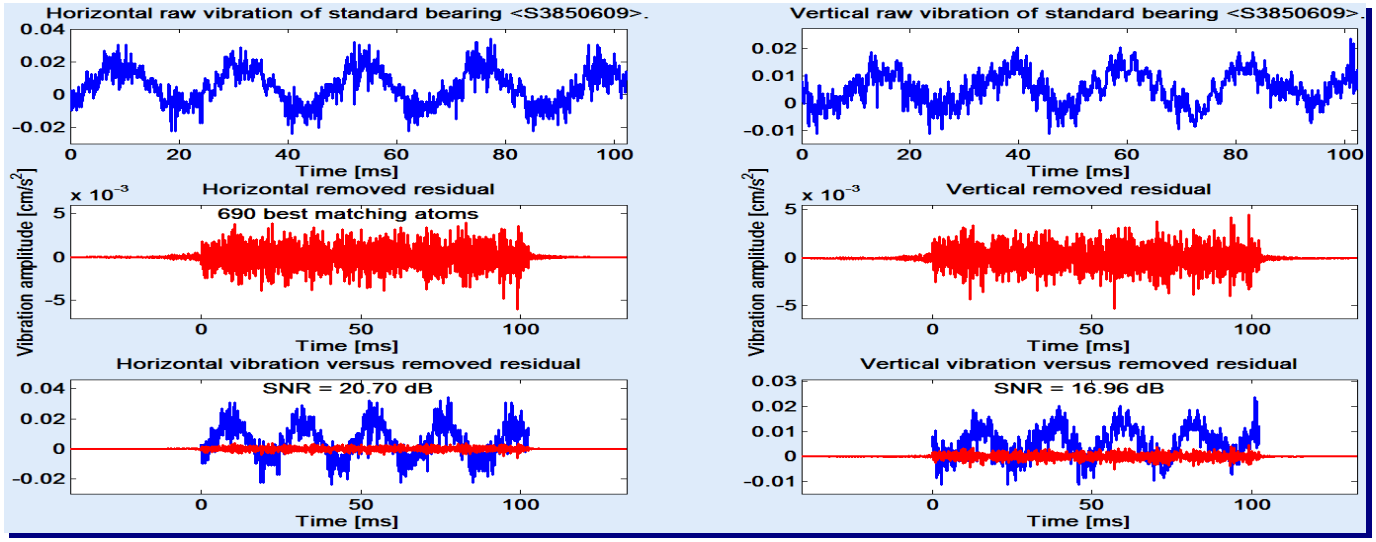

Fig. 5. Real part (left) and imaginary part (right) of S3850609 vibration and removed residuals (first frame only).

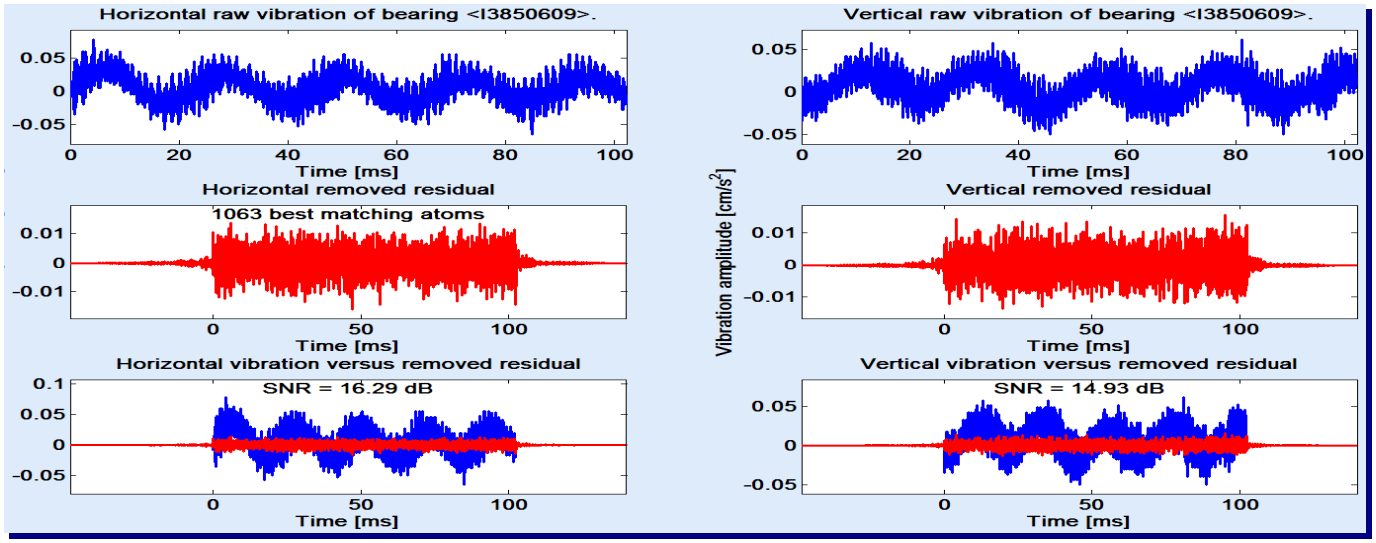

Fig. 6. Real part (left) and imaginary part (right) of I3850609 vibration and removed residuals (first frame only).

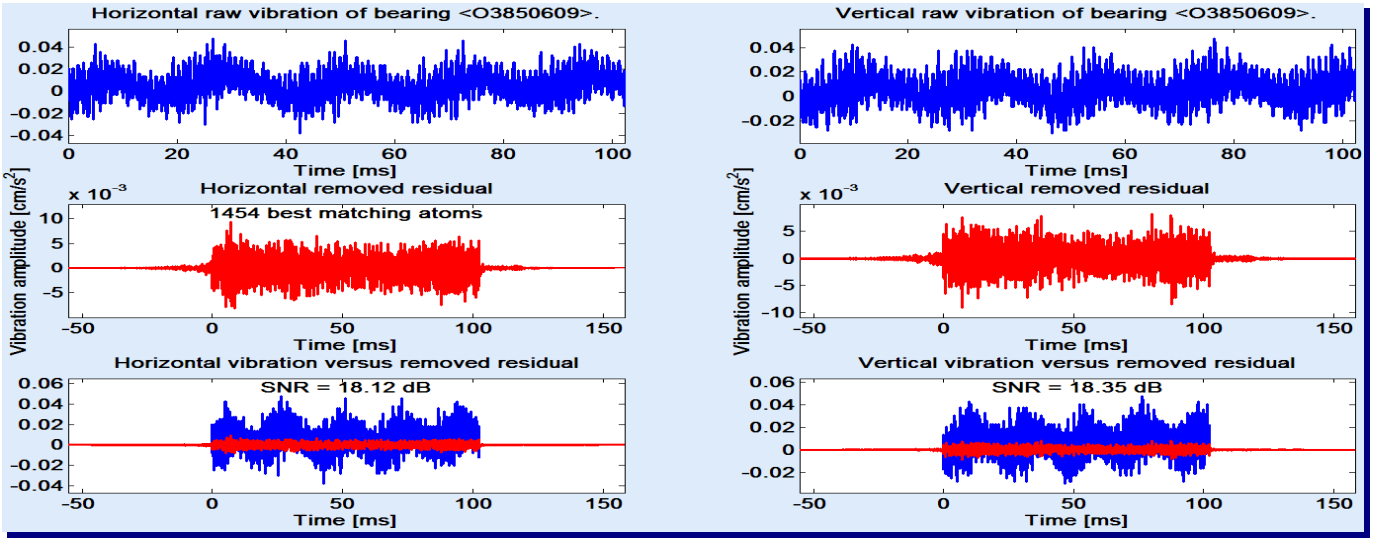

Fig. 7. Real part (left) and imaginary part (right) of O3850609 vibration and removed residuals (first frame only).

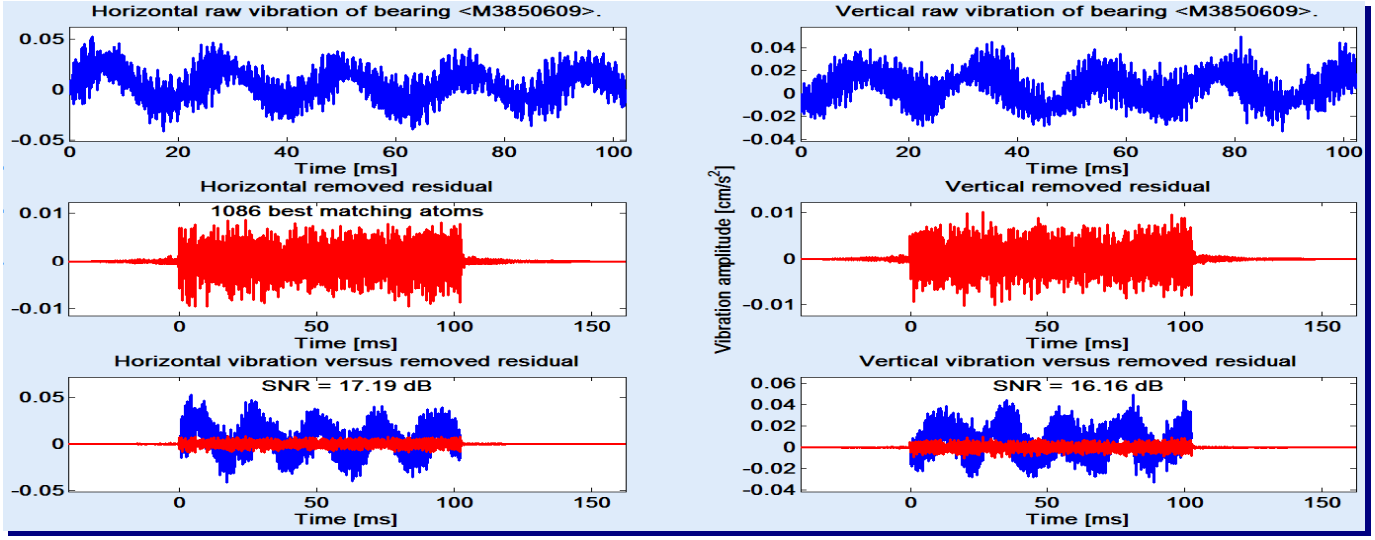

Fig. 8. Real part (left) and imaginary part (right) of M3850609 vibration and removed residuals (first frame only).

This threshold led to extraction on various numbers of bmas taken from dictionary, according to the bearing type, as outlined on the residual real part variation. Also, the SNR of vibration real and imaginary part have been estimated separately – see their values written on the last variations in Figures 5-8. There is more noise in defect vibration than in flawless vibration, as expected.

The removed residuals exhibit larger supports than the genuine vibration frame, namely $\overline{0,2047}T_s$, because the bmas supports are not enforced to be included into $\overline{0,2047}T_s$. They just overlap to this support. Defect encoding vibrations include more auto-correlated noise than the flawless one (as proven by the SNR values and shape of residuals). The standard bearing residual is almost white, while for the other residuals, one can notice some correlation. This is an expected result. Another expected outcome is concerned with the number of bmas found, for the same stop threshold in MPA (ε). While the standard bearing led to 690 bmas, all the other bearings needed more than 1000 bmas, in order to represent the useful signal. This is a direct consequence of defect presence in acquired vibrations. One can thus see how some noise was included into the useful signal, as, actually, coherent information about defects, which was encoded by the noise, is transferred to the bmas. This will result in weaker compression performance for defect encoding vibrations.

Lossless compression has been applied in following conditions. First, one noticed that S-F and sH algorithms led to the same compression ratios. This allowed removing sH from the tested algorithms list and keeping S-F as being simpler to implement. Then, two general purpose (but licensed) compressors were added to S-F and dH, namely WinZIP and WinRAR. Finally, the compression algorithms were applied on the ASCII file of raw vibration (namely `vibra_{S,I,O,M}3850609.txt`), comparing to the ASCII file of transformed vibration (namely `bma_{S,I,O,M}3850609.txt`). Both files also include a text label yielding identification of bearing under test and the value of sampling frequency, F_s. Moreover, the bmas

ASCII file include the value of removed residual std, $\sigma_x^{Q_\varepsilon}$ (which is not included into the raw vibration ASCII file).

The **vibra_{S,I,O,M}3850609.txt** files start with label and F_s. The real and imaginary parts of frame are listed next (with a separator). Each numerical value of vibration samples is listed with 20 digits.

The **bma_{S,I,O,M}3850609.txt** files also start with label and F_s, but the std $\sigma_x^{Q_\varepsilon}$ is listed too. In order to list the bmas, the following matrix was created:

$$\mathbf{A} \overset{\text{def}}{=} \left[m_q \quad n_q \quad k_q \quad \left\langle \Delta^q x, g_{[m_q,n_q,k_q]} \right\rangle \right]_{q \in \overline{0,Q_\varepsilon-1}}. \qquad (24)$$

The matrix $\mathbf{A}$ has 4 columns and Q_ε rows.

On the first 3 rows, the bmas indexes are listed. They take integer values. The last column includes all tfs coefficients corresponding to the bmas. Actually, this column is a complex valued vector in floating point double precision representation.

The results of compression algorithms are presented in Table 1. Beside the CR, two new performance parameters have been estimated: the compression duration (τ [ms]) and the compression gain (γ). The gain is defined as the ratio between the compressed vibration data volume (ρ_c^v) and the compressed bmas data volume ($\rho_c^{\mathbf{A}}$):

$$\gamma \overset{\text{def}}{=} \frac{\rho_c^v}{\rho_c^{\mathbf{A}}}. \qquad (25)$$

In order to prove that sending the bmas information is better than sending the raw vibration, it is necessary to have $\gamma > 1$ (the higher the better). The table reveals several insights. Surprisingly, the S-F algorithm seems to perform better than the dH algorithm in terms of CR, duration and gain. Moreover, it is of lower complexity. This can be explained by the fact the dH procedure starts to work efficiently for larger amount of data than the one of ASCII files in this simulation.

Another surprise is provided by the bearing O3850609, with an outer race defect. As one can see, the ASCII file of bmas is weighting more than the file of raw vibration. This large data volume can be explained by the large number of necessary bmas to build the useful signal (1454 – see Figure 7 again). A feature that has not been evoked in context of this article is the *fault severity*. The bearing O3850609 is more damaged than the other bearings. This induces stronger and more auto-correlated noises that corrupt the vibration. Consequently, more bmas are necessary to encode such a fault. Normally, in this case, the raw vibration should be compressed and sent to the cloud. But, if the goal of vibration analysis is to perform fault diagnosis, then sending the bmas information would spare time, as the MPA has already been applied.

Clearly the general purpose compressors WinZIP and WinRAR are by far superior to S-F and dH. (WinRAR perform slightly better than WinZIP.) They also are designed by using more sophisticated compression methods, context based (such as Lempel-Ziv-Welch one), in combination with simpler ones (such as Huffman's or Run Length Encoding one). Moreover, they are licensed trade marks.

Though, using S-F algorithm in conjunction with MPA is an affordable solution for the user, easy to implement with regular resources.

Table 1. Lossless compression performance over the vibration signals.

Bearing		β_d [bytes]	S-F		dH		WinZIP		WinRAR	
			CR [%]	τ [ms]	CR [%]	τ [ms]	CR [%]	τ [ms]	CR [%]	τ [ms]
S3850609	**vibra**	71335	55.16	11	57.78	411	15.51	112	15.12	110
	bma	40754	50.18	10	54.17	150	14.01	103	13.80	99
	γ		1.9242		1.8670		1.9368		1.9175	
I3850609	**vibra**	73110	55.33	13	52.22	330	22.11	160	21.69	158
	bma	62760	50.21	13	53.26	252	17.62	130	17.18	126
	γ		1.2838		1.2733		1.4612		1.4706	
O3850609	**vibra**	73069	56.45	15	59.18	302	8.48	107	8.28	105
	bma	85942	50.06	17	55.41	342	25.25	141	24.72	137
	γ		0.9587		0.9079		0.2855		0.2846	
M3850609	**vibra**	73063	55.9	19	58.50	291	22.47	161	22.07	156
	bma	63832	49.92	10	53.02	265	19.70	137	19.15	142
	γ		1.2817		1.2627		1.3055		1.3191	

5. CONCLUSION

Processing of vibration (noisy/fractal signals) usually requires methods with high complexity degree. Some of such methods lead to NP-hard procedures, similar to MPA, as described into this paper. The implemented algorithm proved that, in general, it is more efficient to send information about vibrations after using the MPA than without using this transform. The MPA performance can be improved by operating with different mw, adapted to the nature of signal to be processed. For example, wavelets with controlled regularity are appropriate for analysis of fractal signals, whereas, in case of harmonic signals (like vibrations), the modulated Gaussian waveforms find better use. The

algorithm can also be employed to perform other operations on the vibration signals such as denoising and fault diagnosis.

ACRONYMS

bma(s) – best matching atoms(s)
CR – Compression Ratio
dB – decibels
dH – dynamic Huffman (algorithm)
FT – Fourier Transform
MPA – Matching Pursuit Algorithm
mw – mother waveform
NP – Non Polynomial
S-F – Shannon-Fano (algorithm)
sH – static Huffman (algorithm)
SNR – Signal-to-Noise Ratio
SP – Signal Processing
std – standard deviation
tf – time-frequency
tfs – time-frequency-scale
UP – Uncertainty Principle

REFERENCES

Bi, X., Chen, X., Li, X., Leng, L. (2014). Energy-based adaptive matching pursuit algorithm for binary sparse signal reconstruction in compressed sensing. *Signal, Image and Video Processing.* Vol. 8, No 6, pp 1039–1048

Cohen L. (1989). Time-Frequency Distributions – A Review. *Proceedings of the IEEE*, Vol. 77, pp. 941–981.

Cohen L. (1995). Time-Frequency Analysis. *Prentice Hall*, New Jersey, USA, 1995.

Daubechies I. (1993). The Wavelet Transform, Time-Frequency Localization and Signal Analysis. *IEEE Transactions on Information Theory*, Vol. 36, No. 9, pp. 961–1005.

Dobrescu, R., Kevorchian, S. (2002). Compresia datelor. *Ed. Academiei.*

Fang, H., Yang, H. (2012). A new compressed sensing-based matching pursuit algorithm for image reconstruction. *5th International Congress on Image and Signal Processing.*

Figueras i Ventura R.M., Vandergheynst P. (2001). Matching Pursuit through Genetic Algorithms. *Technical Report, Signal Processing Laboratory (LTS),* École Polytechnique Fédérale de Lausanne, Switzerland.

Figueras i Ventura R.M., Vandergheynst P., Frossard P. (2001). Evolutionary Multiresolution Matching Pursuit and its relations with the Human Visual System", *Technical Report 01.04., Signal Processing Laboratory* (LTS), École Polytechnique Fédérale de Lausanne, Switzerland.

Gabor D. (1946). Theory of Communication. *Journal of the IEE*, London (UK), No. 93, pp. 429–457.

Mallat S. and Zhang S. (1993). Matching Pursuits with Time-Frequency Dictionaries. *IEEE Transactions on Signal Processing*, Vol. 41, No. 12, pp. 3397–3415.

Oppenheim A.V. and Schafer R. (1985). Digital Signal Processing. *Prentice Hall*, New York, second edition.

Proakis J.G. and Manolakis D.G. (1996). Digital Signal Processing. Principles, Algorithms and Applications. *Prentice Hall*, New Jersey, the third edition.

Qiang, G., D. Chendong, F. Xiangbo and L. Benchao (2011). A Study on Matching Pursuit Based on Genetic Algorithm. *Third International Conference on Measuring Technology and Mechatronics Automation*. pp. 283-286.

Rup, S. , Pati, N., Dash, B. (2015). A Dictionary Based Image Coding using Orthogonal Matching Pursuit. *International Journal of Computer Science and Mobile Applications.* Vol.3, No 10, pp. 06-12.

Rusu, C. (2011). Classification of music genres using sparse representations in overcomplete dictionaries. *Control Engineering and Applied Informatics*, Vol. 13, No 1, pp. 35-42.

Sun, G., Li, Y., Yuan, H., He, J., Geng, T. (2016). The Improvement of Compressive Sampling and Matching Pursuit Algorithm Based on Pre-estimation. *International Journal of Wireless Information Networks*. Vol. 23, No 2, pp 129–134.

Stefanoiu, D., Borne, P., Popescu, D., Filip, F.G., El Kamel, A. (2014). Optimization in Engineering Sciences – Metaheuristics, Stochastic Methods and Decision Support, *John Wiley & Sons*, London, U.K.

Stefanoiu, D. (2003). Data Compression. *Printech Press* Bucharest, Romania.

Stefanoiu, D., Ionescu, F. (2003). A Genetic Matching Pursuit Algorithm, *The seventh IEEE-EURASIP International Symposium on Signal Processing and its Applications*, ISSPA 2003, Paris, France, pp. 577–580.