MATEC Web of Conferences **20**, 07005 (2015)
DOI: 10.1051/matecconf/20152007005

Stochastic Resonance algorithms to enhance damage detection in bearing faults

Roberto Castiglione, Luigi Garibaldi and Stefano Marchesiello

Dynamics & Identification Research Group, Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

Abstract. Stochastic Resonance is a phenomenon, studied and mainly exploited in telecommunication, which permits the amplification and detection of weak signals by the assistance of noise. The first papers on this technique are dated early 80 s and were developed to explain the periodically recurrent ice ages. Other applications mainly concern neuroscience, biology, medicine and obviously signal analysis and processing. Recently, some researchers have applied the technique for detecting faults in mechanical systems and bearings. In this paper, we try to better understand the conditions of applicability and which is the best algorithm to be adopted for these purposes. In fact, to get the methodology profitable and efficient to enhance the signal spikes due to fault in rings and balls/rollers of bearings, some parameters have to be properly selected. This is a problem since in system identification this procedure should be as blind as possible. Two algorithms are analysed: the first exploits classical SR with three parameters mutually dependent, while the other uses Woods-Saxon potential, with three parameters yet but holding a different meaning. The comparison of the performances of the two algorithms and the optimal choice of their parameters are the scopes of this paper. Algorithms are tested on simulated and experimental data showing an evident capacity of increasing the signal to noise ratio.

1. Introduction

The phenomenon of stochastic resonance (SR) has applications in a number of different fields and scientific domains. The possibility of resonance in dynamical systems, which behave stochastically, was introduced by R. Benzi et al. [1] in 1981 and originally exploited for studying the evolution of the Earth's climate. Its first applications were in a wide range of problems connected to physical and life sciences. Other observations of this phenomenon concern experiments on electronic circuits, chemical reactions, semiconductor devices, nonlinear optical systems, magnetic systems and superconducting quantum interference devices [2].

The studies of SR for mechanical application, especially for mechanical fault diagnosis, began in the mid-90s and great improvements have been realized mostly during the last years. Several techniques exist and are applied for the detection of defects in rotating machines such as gears or bearings in many industrial applications, but SR is the only one that takes advantage of noise. In fact generally mechanical acquisitions are strongly corrupted by background noise from other elements of the system and this component is usually neglected, on the contrary it is used by SR to enhance the features of faults [3].

2. The theory of SR

Stochastic Resonance is a tool used in signal processing to increase the signal-to-noise ratio (SNR) of the output of a non-linear dynamic system, in order to extract the characteristic features of the system under investigation from background noise. This is obtained by adding, to the measured signal corrupted by noise, a nonlinear dynamic system and, by properly tuning the latter, enhancing the signal of interest (Fig. 1).

Usually noise is considered as a disturbance that makes measured data unusable. The basic idea behind each processing data procedure include the filtering or removal of noise, but it may happen that useful information may be corrupted or destroyed by weakening or removing it, so much attention has to be paid. In SR, instead, noise is a basic element of the process in fact, by adding a well-tuned noise to the full measurement, signal detection is facilitated.

The amplification of weak signal is obtained by varying the noise level, through the addition of a potential, but keeping the input modulation signal. SR mechanism implies that, if a sinusoidal driving frequency mixed with noise is given as input to a nonlinear system, its output contains a high peak corresponding to the driving frequency which varies its amplitude as a function of the noise added through the system [3,4].

The algorithms for SR, especially for weak impulses or aperiodic signals, work in time domain and several types of implementation exist through the use of different kind of potentials. Most of literature makes use of SR with a single bi-stable system for the recognition of weak periodic signals but, recently, a new strategy based on Woods-Saxon potential has been investigated [5].

The dynamic behavior of SR can be described by the Brownian motion equation of particles, where $s(t)$ and $n(t)$ are respectively the input signal and the noise, whose sum gives the measurement signal, and $V(x)$ is the potential

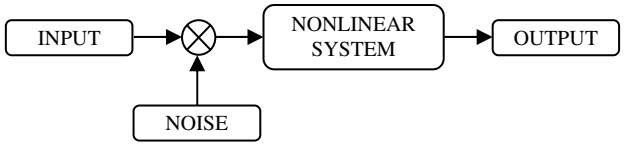

Figure 1. SR mechanism.

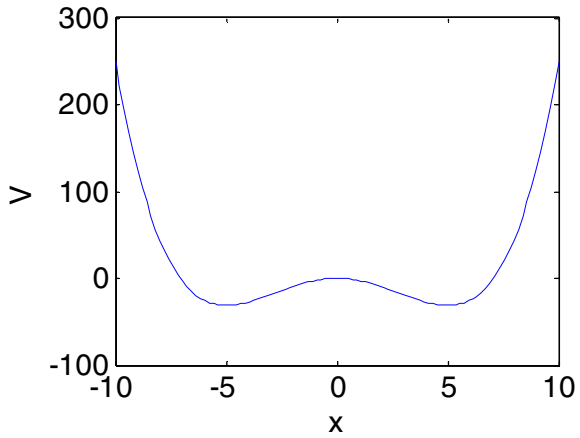

Figure 2. Bi-stable potential.

function.

$$\frac{dx}{dt} = -\frac{dV}{dx} + s(t) + n(t). \quad (1)$$

The quantity *x(t)* is the system output and denotes the trajectory of the Brownian particle in the potential function $V(x)$. In case there is no external excitation the position is only determined by the initial conditions and never changes. If a periodic input function is given as input the potential function is modulated and changes periodically, and in case there is also noise in the input, the particle will jump between the potential wells with a period corresponding to that of the input function. So, by properly tuning the potential to the noise present in the signal it is possible to detect weak signals by simply solving the above first-order differential equation using the discrete Runge–Kutta method.

SR may be implemented in several ways by using different types of potential functions and enhancement methods. The cases that will be taken into account in the following sections make use of a bi-stable (BSSR) and Woods-Saxon (WSSR) potentials: each has its own characteristic features and limitations in the use.

2.1. BSSR

Classical SR [3,4] uses the following polynomial expression as a potential function:

$$V(x) = -\frac{a}{2}x^2 + \frac{b}{4}x^4. \quad (2)$$

It represents a bi-stable symmetric system where a and b are positive real parameters, whose two stable points are located at $x_m = \pm\sqrt{a/b}$ (Fig. 2).

Substituting this in the Brownian particle motion equation and considering a periodic signal $s(t)$ of

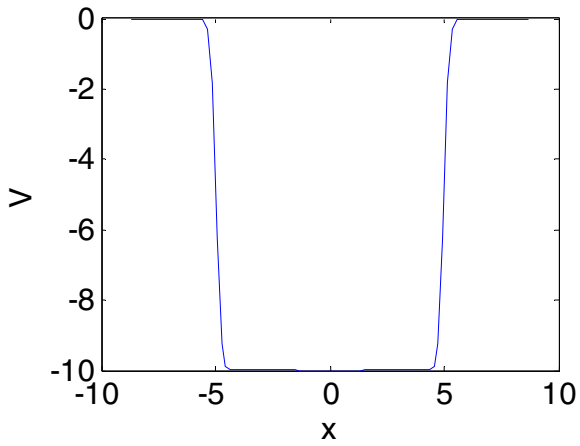

Figure 3. Woods-Saxon potential.

amplitude A and driving frequency f_0 and a Gaussian white noise $n(t)$ with zero mean and given variance, the main equation of the process is obtained:

$$\frac{dx}{dt} = ax - bx^3 + A_0\cos(2\pi f_0 t + \varphi) + n(t). \quad (3)$$

At this stage it is necessary to find the correct values of a and b potential parameters so that $x(t)$ takes the form of a square wave with the same oscillation frequency as the driving frequency of the periodic signal. In fact, the system output $x(t)$, which represents the motion of a Brownian particle inside the potential $V(x)$, should oscillate between the two potential wells at a transition rate that matches the period of the input signal. Consequently the periodic input signal is enhanced only by adjusting the dynamic system parameters.

In case of impact signals given as input *s(t)* to the system, the output *x(t)* will be made of a series of impulses located in the exact position as the original signal. For example, by considering only a single impact event, the equation of motion becomes:

$$\frac{dx}{dt} = ax - bx^3 + Ae^{-Dt}\cos(2\pi f_0 t) + n(t). \quad (4)$$

In this case the Brownian particle can jump between the potential wells just in a few oscillation periods or just in one by properly tuning the potential parameters, while in the remaining parts of the signal it will get stuck inside the potential well because no sufficient energy is provided by noise.

2.2. WSSR

Recently a new method [5] has been used for SR which makes use of Woods-Saxon potential. It is used in physics to approximately describe the forces applied on each nucleon in the nuclear shell model.

$$V(x) = -\frac{V_0}{1 + e^{\frac{|x|-r}{c}}}. \quad (5)$$

It is a mono-stable potential (Fig. 3) dependent on three parameters. Each of them influences directly one feature

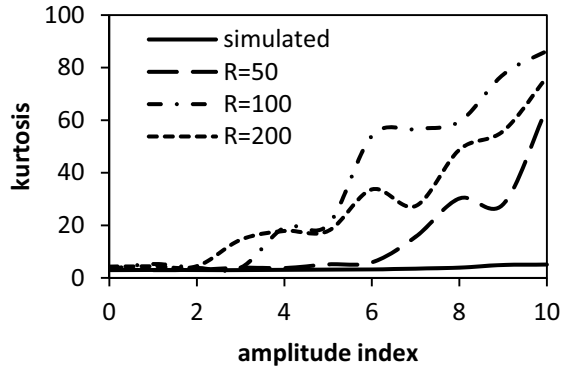

Figure 4. BSSR: Kurtosis versus amplitude index for different R values.

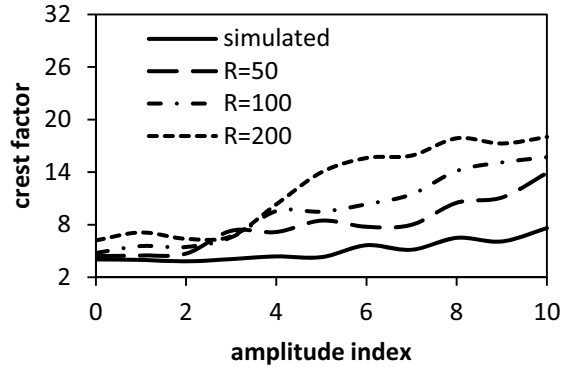

Figure 5. BSSR: Crest Factor versus amplitude index for different R values.

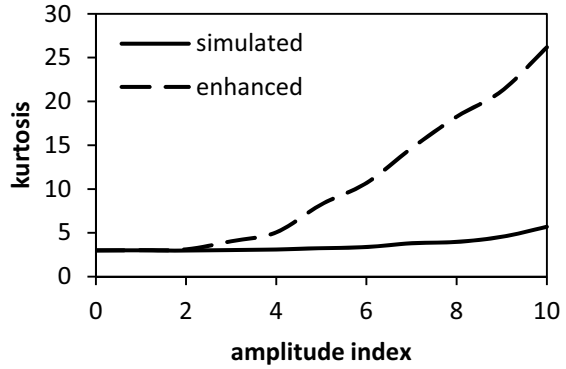

Figure 6. WSSR: Kurtosis versus amplitude index.

of the curve: V_0 affects the depth, r the width and c the steepness of the potential. By deriving this expression, it is possible to get the fundamental equation of SR in the two cases, respectively for periodic sinusoidal function and for impact signal:

$$\frac{dx}{dt} = \frac{V_0}{c} sign(x) \frac{e^{\frac{|x|-r}{c}}}{\left(1 + e^{\frac{|x|-r}{c}}\right)^2} + A_0 \cos(2\pi f_0 t + \varphi) + n(t) \quad (6)$$

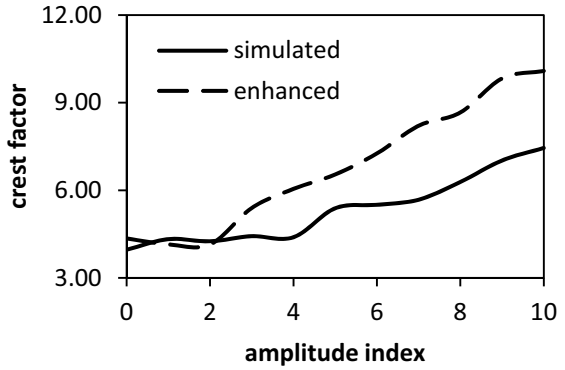

Figure 7. WSSR: Crest Factor versus amplitude index.

$$\frac{dx}{dt} = \frac{V_0}{c} sign(x) \frac{e^{\frac{|x|-r}{c}}}{\left(1 + e^{\frac{|x|-r}{c}}\right)^2} + Ae^{-Dt} \cos(2\pi f_0 t) + n(t). \quad (7)$$

Adjusting properly the potential model, the particle will oscillate between the two potential wells and it will be possible to amplify the oscillation of the input function.

2.3. The algorithm for impact signal detection

The main difficulty for an efficient implementation of SR is the selection of the parameters of the potential. In fact, it is necessary to define a criterion to determine if the selected set of values gives good results in the output, which means its capability of enhancing signal.

Several measurement indexes exist to assess the quality of the procedure, for example kurtosis, crest factor or others. The first is defined as the ratio between the fourth central moment and the square of the variance and it is a measure of the peakness of a probability distribution of a real-valued random variable. The more the peaks are narrow and sharp the more kurtosis is high, in contrast to the case in which there is a normal distribution when the kurtosis tends to 3:

$$kurt\,(x) = \frac{E[(x - \bar{x})^4]}{(E[(x - \bar{x})^2])^2}. \quad (8)$$

Crest factor is defined as the absolute peak value over the root mean square of the distribution. The higher are the peaks emerging from the background noise after the application of the method, the more this factor increases.

$$CF = \frac{|x|_{peak}}{x_{RMS}} = \frac{|x|_{peak}}{\sqrt{E[x^2]}}. \quad (9)$$

Whatever index is selected, the parameters of the potential have to be selected in order to maximize it.

The first step after signal pre-processing is the initialization of the range of the coefficients in terms of minimum and maximum values but also step size. By substituting them in the SR equation, the output signal is computed with a fourth-order Runge–Kutta algorithm and then evaluated through the criterion previously selected. Generally after each equation is solved it is necessary

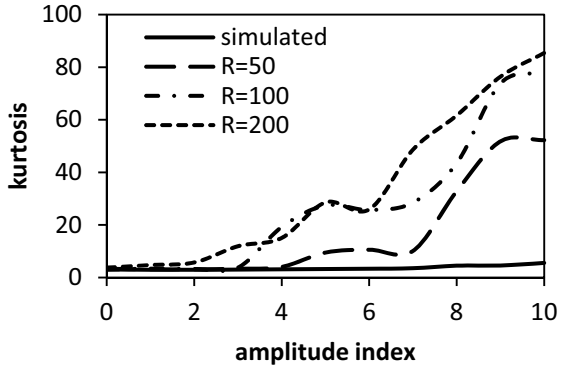

Figure 8. BSSR with impulse period fluctuations: Kurtosis versus amplitude index for different R values.

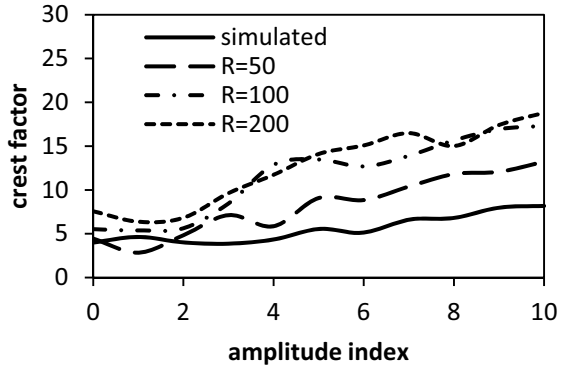

Figure 9. BSSR with impulse period fluctuations: Crest Factor versus amplitude index for different R values.

to remove the initial transient of its solution in order to compute correctly the corresponding criterion function, otherwise numerical peaks could be included in its evaluation. Once all possible values of the coefficients have been used, the maximum value of the criterion is chosen together with the corresponding values of the coefficients. The improved waveform is then computed and, from this, it is possible to get all the information on the characterization of the impacts.

2.4. Re-scaling ratio

The main problem of classical bi-stable stochastic resonance is the small parameters restriction, and this is in contrast with bearing faults diagnosis. Essentially BSSR focus on low frequency and weak periodic signal submerged in small noise, small in this case means that the values of the frequency and amplitude of periodic signal and noise intensity are less than 1. But defect frequencies are usually much higher than 1 Hz and as a consequence it is necessary to adapt SR algorithm to large parameters signals.

The approach that will be followed here makes use of a re-scaling factor applied to the sampling frequency f_S in order to make it much lower by linearly compressing the frequency scale. R is a rescaling ratio that satisfy the requirement of small parameters [6].

$$f_{Sr} = f_S/R. \tag{10}$$

Table 1. Amplitude index for each value A of the impulses amplitude and the corresponding ratio A/σ.

index	A	A/σ
0	0	0
1	0.05	0.71
2	0.10	1.43
3	0.15	2.14
4	0.20	2.86
5	0.25	3.57
6	0.30	4.29
7	0.35	5.00
8	0.40	5.71
9	0.45	6.43
10	0.50	7.14

3. Numerical simulation

The methods have been tested on several simulated data. The first set has been generated by summing a series of impulses, in the form of impulse responses, to a vector of Gaussian white noise:

$$s(t) + n(t) = Ae^{-D(t-T_i)}\cos(2\pi f_0(t - T_i)) + \mathcal{N}(0, \sigma). \tag{11}$$

Five equally spaced impulse responses have been added [4], whose frequency is $f_0 = 16\,\text{Hz}$ and attenuation rate is $D = 12\,\text{s}^{-1}$. $T = [4, 9, 14, 19, 24]\,\text{s}$ is the vector of instants at which the impulses appear. Sampling frequency is set at $f_S = 500\,\text{Hz}$ and $N = 14500$ is the number of samples considered. A is the signal amplitude and the $n(t)$ is Gaussian white noise with zero mean value and variance σ^2.

The purpose of this first analysis is to show what is the sensitivity of the algorithm to different levels of impact amplitude, fixed the variance of Gaussian background noise. In particular, it is necessary to evaluate what is the minimum ratio A/σ it which the impact signal is well detected and enhanced and also to verify that when SR is applied to pure Gaussian signals it does not detect any (false) impulse.

Noise variance is set at $\sigma^2 = 0.07^2$ and the amplitude was changed between 0 and 0.5 with a spacing of 0.05 (Table 1).

At first, for all levels of the ratio A/σ, data were elaborated by BSSR. Parameter are selected in the ranges: $a \in [0.6\ 1.4]$, $b \in [5\ 15]$ with steps 0.1 and 1 respectively and the iteration is made for every couple (a, b). The choice is made with both criteria: kurtosis and crest factor; and a comparison is made between the optimization criteria respectively of the signal of measure and of the output signal for each intensity level of the input impulses. Three values of rescaling factor R are used, as shown in Table 2.

Fig. 4 refers to the case in which kurtosis is used as cost function. It is evident that for larger values of re-scaling factor, which corresponds to a larger sampling period, the identification of the peaks happens at lower levels of A/σ. This means that impulses may be efficiently recognized even when they are submerged inside background noise. It should be also noticed that when zero-amplitude impulses

Table 2. Re-Scaling factors used for BSSR t_s original sampling period and t'_s caled sampling period.

R	t_s	t'_s
50	0.0020	0.1000
100	0.0020	0.2000
200	0.0020	0.4000

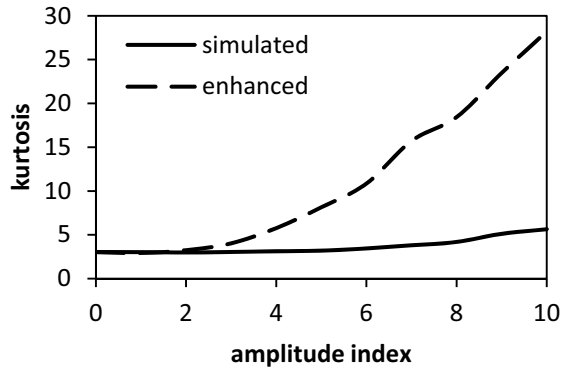

Figure 10. WSSR with impulse period fluctuations: Kurtosis versus amplitude index.

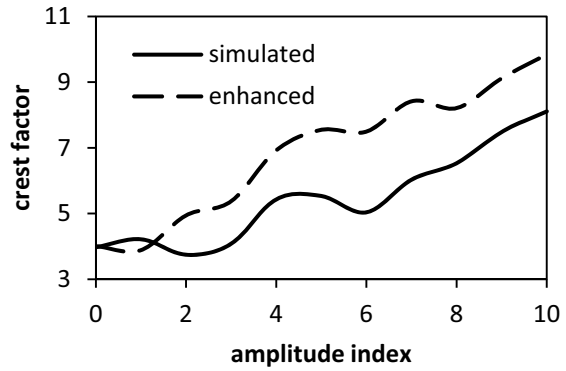

Figure 11. WSSR with impulse period fluctuations: Crest Factor versus amplitude index.

are provided to the system, the algorithm does not give false alarms as kurtosis remains the same as the initial signal.

When the crest factor is used, the more the re-scaling factor is high the more peaks are enhanced but there are some problems for zero-amplitude impulses signals.

As can be seen in Figs. 4 and 5, the more scaled sampling frequency tends to the unit or generally to lower values, the more the algorithm enhances better peaks from background noise because the small parameters assumption is true.

Then WSSR approach is used and parameters are selected in the ranges: $V_0 \in [10\,200]$, $r \in [0.12]$, $c \in [0.61]$ with steps respectively 5 for V_0 and 0.1 for the others. In Figs. 6 and 7 are the results for this case. Here it is evident that even for small amplitude peaks the enhancement of weak signal is well performed by using both criteria, but the variation in the criterion functions is lower than by employing the other dynamical system with bi-stable potential.

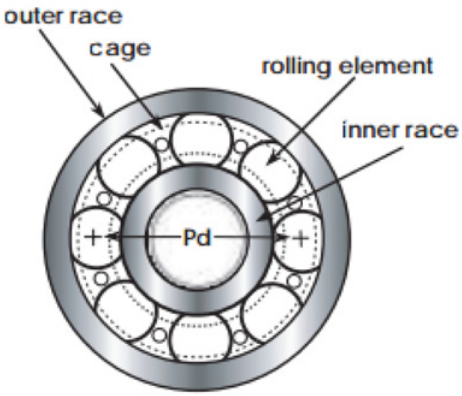

Figure 12. Bearing elements.

In order to increase the complexity of the data elaboration, some random fluctuations have been added to the period at which impulses appear, to see how the algorithm can deal with this. While in previous tests this period was fixed constant at the value $\Delta T = 5\,s$ here a fluctuation of $\pm 20\%$ around this value is allowed: in real applications this is an extreme occurrence, being the true fluctuation usually very low, but this is just for test purposes. The same procedure has been followed by applying both methodologies with the same parameters ranges and by varying the impulse amplitude: BSSR in Figs. 8 and 9 and WSSR in Figs. 10 and 11.

4. Bearing faults data simulation

Several kinds of damage could happen to rolling element bearings due to different causes, such as fatigue, wear, poor installation, improper lubrication and occasionally manufacturing faults. Defects could arise in all constituting elements and each has a distinct pattern in the time signal acquisition and could be identified by a good examination of that. Bearing is made of the following components: outer race, inner race, cage and rolling elements (Fig. 12).

When a bearing spins, any irregularity in the surface of inner or outer race, or in the roundness of the rolling elements excites periodic frequencies called fundamental defect frequencies. These depend on the geometry of the bearing and clearly on the shaft speed. Below there is a list of these frequencies in which d is ball diameter, D is pitch diameter, ϕ is contact angle, f_r is shaft speed, n is the number of elements. It is assumed that outer race is fixed and inner race rotates.

1. Ball Pass Frequency of the Outer race (frequency created when all the rolling elements pass across a defect in the outer race)

$$BPFO = \frac{nf_r}{2}\left(1 - \frac{d}{D}cos\phi\right). \qquad (12)$$

2. Ball Pass Frequency of the Inner race (frequency created when all the rolling elements roll across a defect in the inner race)

$$BPFI = \frac{nf_r}{2}\left(1 + \frac{d}{D}cos\phi\right). \qquad (13)$$

Table 3. Bearing dimensions and defect frequencies as a function of shaft frequency.

d	9 mm
D	40.5 mm
ϕ	0°
n	10
$BPFO$	3.89 f_r
$BPFI$	6.11 f_s
FTF	2.14 f_s
BSF	0.39 f_r

Table 4. Properties of the structure.

resonace	5500 Hz
damping factor	0.05
mass	10 kg

3. Fundamental Train Frequency (cage speed)

$$FTF = \frac{f_r}{2}\left(1 - \frac{d}{D}cos\phi\right). \quad (14)$$

4. Ball Spin Frequency (circular frequency of each rolling element as it spins)

$$BSF = f_r \frac{D}{2d}\left(1 - \left(\frac{d}{D}cos\phi\right)^2\right). \quad (15)$$

All formulas previously listed are valid only in case there is pure rolling contact between balls, inner race and outer race, but actually there is always some random slip when a bearing is under load and after some wear and consequently frequencies are not located in the predicted position inside the spectrum.

Bearing data with faults are generated through a Matlab code that simulates and modulates all kinds of defects, like impulse response functions, inside background noise and then both algorithms for signal fault enhancement are applied. The simulated bearing has the characteristics listed in Table 3, together with its corresponding defect frequencies as function of shaft speed.

The variance of noise was set at $\sigma^2 = 0,07^2$ and the defect was simulated as a pulse modified by a passage through a transmission path with a long impulse response. Two cases are analyzed: the first with a lower shaft speed (30 Hz) and consequently lower defect frequency so that each impulse response do not overlap with the following, and the second with a higher speed (30 Hz) in which bearing fault frequencies are so high that the spacing between them is comparable with the length of each impulse response.

The structure is assumed to have the properties listed in Table 4, from which it is possible to evaluate the entity and the characteristics of the impulse responses.

4.1. Low speed shaft

In the first simulation shaft speed is 30 Hz, it is carried out at $f_S = 51200$ Hz for 0.1 s. The interval length is set this way in order to contain just a small number of peaks. In Tables 5 and 6 it is possible to see the results in terms of enhancement of the criterion functions for a simulated defect, on the outer and on the inner ring respectively. For BSSR a re-scaling factor R = 25600 is used in order to satisfy the small parameter requirement.

Table 5. Results for outer ring defect, low speed shaft.

	original	*enhanced*	
		BSSR	*WSSR*
kurtosis	3.65	12.65	5.00
crest factor	4.66	8.49	5.19

Table 6. Results for inner ring defect, low speed shaft.

	original	*enhanced*	
		BSSR	*WSSR*
kurtosis	3.52	16.15	6.45
crest factor	4.56	8.10	5.08

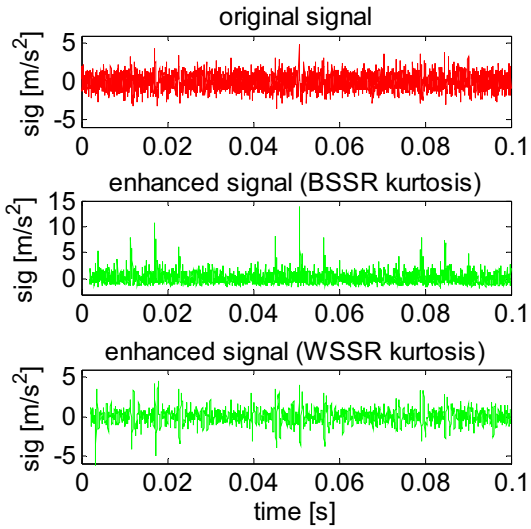

Figure 13. Comparison of enhanced signal with the two algorithms in reference to original signal for low speed shaft.

For both cases the two algorithms give good results, but the only difference, in terms of performance, is due to the nature of the defect in terms of peaks modulation. For an outer ring defect all peaks are almost all at the same level of amplitude and consequently SR enhance them all equally. Whereas for inner ring defects peaks are highly modulated by the rotation speed and consequently the algorithms can enhance better the larger ones but some others are not well detected because of their small amplitude. In any case the characteristic path of the defect is clearly seen in the enhanced signal and this is an advantage for evaluating its presence or not in the signal. In Fig. 13 there is a comparison between the enhanced signals by employing the two algorithms with kurtosis in reference to the original signal for the case of defect on the inner ring.

4.2. High speed shaft

In this case the shaft rotates at 300 Hz, the signal is still sampled at $f_S = 51200$ Hz but only 0.02 s are considered in order to have again a few peaks due to the simulated defect. Table 7 shows the results of the analysis for a defect on the outer ring while Table 8 is for the inner ring.

Table 7. Results for outer ring defect, high speed shaft.

	original	*enhanced*	
		BSSR	*WSSR*
kurtosis	3.98	10.48	2.65
crest factor	3.77	8.95	3.36

Table 8. Results for inner ring defect, high speed shaft.

	original	*enhanced*	
		BSSR	*WSSR*
kurtosis	3.13	10.48	3.00
crest factor	3.31	9.98	3.84

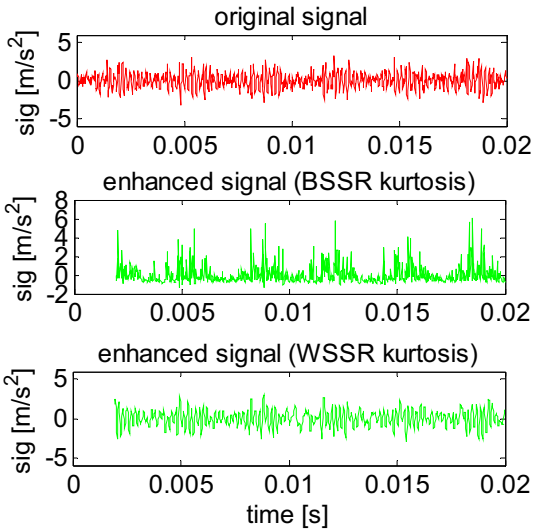

Figure 14. Comparison of enhanced signal with the two algorithms in reference to original signal for high speed shaft.

In this situation there is generally a good improvement by using the bi-stable system but employing Woods-Saxon potential does not give satisfactory results. This can be seen also in Fig. 14 where there is a comparison for the two cases for a defect in the inner ring using kurtosis as criterion function.

5. Experimental validation

At this point an experimental application of the proposed methodologies is presented in order to validate their applicability and effectiveness. The test rig, set up in the Laboratory of the Department of Mechanical and Aerospace Engineering of Politecnico di Torino in collaboration with Avio, is made of three bearings and a rotating shaft, see Fig. 15. One bearing holds the radial load while the other two serves as supports for the shaft. One of the latter is analysed for different levels of damage at different speeds of the shaft and levels of load by equipping the structure with triaxial accelerometers.

The component in exams has the characteristics and defect frequencies as in Table 3. A bearing with a defect in the rolling elements is mounted on the structure, its characteristic pattern on the signal is given by impacts spaced with BPF and whose amplitude is modulated by FTF. The shaft rotates at a nominal speed of 278 Hz. The

Figure 15. The test rig.

Table 9. Results for experimental data.

	original	*enhanced (BSSR)*
kurtosis	4.98	7.73
crest factor	5.72	6.82

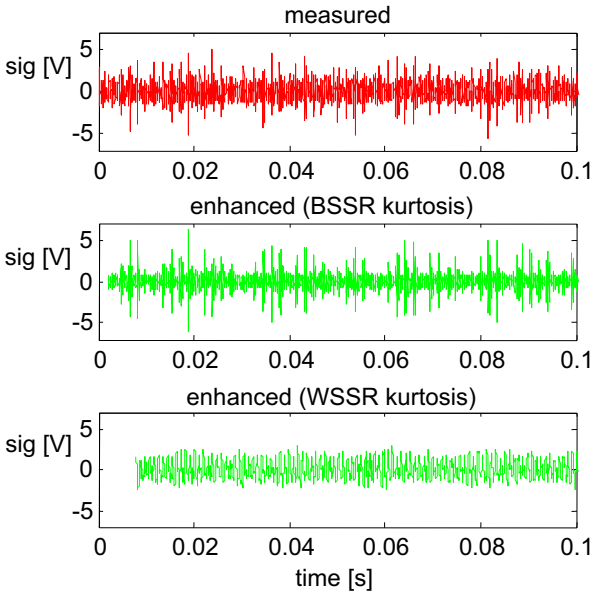

Figure 16. Comparison of original and enhanced signal for experimental data.

central support is loaded with 1800 N. Sampling frequency is $f_S = 51200$ Hz.

Only BSSR has given satisfactory results because the speed of rotation of the bearing is high, see Fig. 15. The theoretical frequencies for a defect on the rollers are $BPF = 105$ Hz and $FTF = 595$ Hz and from the spectrum of the enhanced signal the extracted frequencies are 106 Hz and 592 Hz.

6. Conclusions

The mathematics behind the mechanism of stochastic resonance is relatively simple and easy to be implemented. Theoretically SR with both potential systems works really well in the detection of pulses submerged inside background noise even with low levels of excitation.

The main limitation is the choice of the range of the parameters. When passing to simulated and experimental data of bearings there are some problems as the high rotational speed of the shaft causes the defect period to become smaller and consequently their impulse responses are overlapped one with each other. This makes the process harder as the algorithm may enhance not the pulses but also the resonance frequency of the structure that characterize the damped oscillation of impulse responses. The greatest enhancement is given in general by using the bi-stable system with both criteria.

This work has been carried out in the framework of the GREAT2020 project – phase II.

References

[1] R. Benzi, A. Sutera, A. Vulpiani A, The mechanism of stochastic resonance, Journal of Physics A: mathematical and general **14** (1981) L453–L457

[2] MD. McDonnell, D. Abbott, What Is Stochastic Resonance? Definitions, Misconceptions, Debates, and Its Relevance to Biology, PLoS Computational Biology **5**(5) (2009)

[3] J. Li, X. Chen, Z. He, Multi-stable stochastic resonance and its application research on mechanical fault diagnosis, Journal of Sound and Vibration **332** (2013) 5999–6015

[4] J. Li, X. Chen, Z. He, Adaptive stochastic resonance method for impact signal detection based on sliding window, Mechanical Systems and Signal Processing **36** (2013) 240–255

[5] S. Lu, Q. He, F. Kong, Stochastic resonance with Woods–Saxon potential for rolling element bearing fault diagnosis, Mechanical Systems and Signal Processing **45** (2014) 488–503

[6] J. Tan et al., Study of frequency-shifted and re-scaling stochastic resonance and its application to fault diagnosis, Mechanical Systems and Signal Processing **23** (2009) 811–822