The spectral amplitude modulation: A nonlinear filtering process for diagnosis of rolling element bearings

The spectral amplitude modulation: A nonlinear filtering process for diagnosis of rolling element bearings

Ali Moshrefzadeh[a,*], Alessandro Fasana[a], Jérôme Antoni[b]

[a] *DIMEAS - Politecnico di Torino*
Corso Duca degli Abruzzi, 24 - 10129 Torino, Italy
Ali.Moshrefzadeh@polito.it
Alessandro.Fasana@polito.it
[b] *Laboratoire Vibrations Acoustique, Univ Lyon, INSA-Lyon*
LVA EA677, F-69621 Villeurbanne, France
Jerome.Antoni@insa-lyon.fr

Abstract

Rolling element bearings are the critical parts of every rotating machinery and their failure is one of the main reason of the machine downtime and even breakdown. For this reason, various methods have been proposed in the past for their early diagnosis. Among them, envelope analysis and cyclic spectral analysis (CSA) are the most effective and widely used approaches, working according to the principle of linear filtering process of signals to remove undesirable components. However, in many cases machine structures are excited in different frequency ranges by impacts generated when defects are engaged, hence, the signal should be filtered for multiple frequency bands to completely extract the defect signals. Also, finding the proper frequency bands for demodulation is not always a simple task. To overcome these challenges, in this paper an empirical and automated nonlinear filtering process will be proposed in which different components of a signal are decomposed based on their powers. Then, their squared envelope spectra are computed to seek the presence of bearings characteristic frequencies. Therefore, this method can be seen as complementary to the narrowband amplitude demodulation techniques. The phase of each filtered component is similar to the phase of the original signal but the magnitude is transformed. The idea is that reconstruction of a signal only by its phase preserves many useful features. A similar approach is employed by cepstrum

pre-whitening for bearings diagnosis to reduce the effect of powerful exogenous sources which mask the bearings signals and, despite its simplicity, has achieved noteworthy outcomes. Finally, the performance of the proposed method is validated on real case data recorded from two test rigs. Also, its effectiveness is investigated under constant and variable speed regimes and in presence of various level of Gaussian and non-Gaussian (highly impulsive) background noise.

Keywords: Rolling Element Bearing, Spectral Amplitude Modulation (SAM), Cepstrum Pre-Whitening (CPW), Diagnosis, Defect Frequency, Nonlinear Filtering, Magnitude Order

---

1. Introduction

Diagnosis of Rolling Elements Bearings (REBs) as a crucial component of every rotation machinery is an important task. Since, detection of faults prior to any catastrophic failure can save time and money.

Diagnostic information related to REBs faults are often masked by interferences from more powerful sources. Envelope analysis [1] is the best-known method to extract the bearings signals from raw signals and it has become the benchmark technique for their diagnostics. It was first proposed in [2] and named High-frequency resonance technique (HFRT) but now it is mainly known with the former name. But the main challenge has been finding the effective frequency band for demodulation. Various methods such as fast Kurtogram [3], Protrugram [4] and Autogram [5, 6] have been developed since then to overcome this challenge and valuable results have been achieved. These approaches select frequency bands in which the filtered time signal, Squared Envelope Spectrum (SES) of the filtered signal and autocorrelation of the filtered time signal's envelope have the highest kurtosis respectively. The main disadvantage of these techniques is related to selecting a single frequency band for demodulation. This is problematic especially when the signatures of multiple defects are present in different frequency bands because in this case only one defect is detected and

Nomenclature

$\mathcal{A}$	Analytic Signal
$\mathcal{H}$	Hilbert Transform
f_{rot}	Shaft Frequency
O_{rot}	Shaft Order
x_m	Modified Signals
ACEP	Automated Cepstrum Editing Procedure
BPFI	Ball Pass Frequency Inner race
BPFO	Ball Pass Frequency Outer race
BPOI	Ball Pass Order Inner race
BSF	Ball Spin Frequency
CPW	Cepstrum Pre-Whitening
CSA	Cyclic Spectral Analysis
CSES	Combined Squared Envelope Spectrum
CWRU	Case Western Reserve University
EES	Enhanced Envelope Spectrum
EMI	Electro-Magnetic Interference
FT	Fourier Transform
FTF	Fundamental Train Frequency
IFT	Inverse Fourier Transform
LES	Log-Envelope Spectrum
LSES	Log Squared Envelope Spectrum
MO	Magnitude Order
MSES	Maximum Squared Envelope Spectrum
PWM	Pulse-Width Modulated
REB	Rolling Element Bearing
SAM	Spectral Amplitude Modulation
SC	Spectral Correlation
SES	Squared Envelope Spectrum
SNR	Signal to Noise Ratio
TSA	Time Synchronous Averaging
VFD	Variable Frequency Drives

the others are overlooked. The combined squared envelope spectrum (CSES) has been proposed by [5] as an attempt to solve this problem.

Another family of approaches for condition monitoring of rotary machinery is Cyclic Spectral Analysis (CSA) [7]. This is a powerful method which separates and describes different 2nd order cyclostationary components in terms of the spectral (carrier) frequency and the cyclic (modulation) frequency variables. Spectral Correlation (SC) is one of the main tool for the CSA of machine signals which is defined as the two-dimensional Fourier Transform (FT) of the instantaneous autocovariance function. Ref. [8] discusses the various proposed methods to estimate the SC. High computational cost is the main drawback of these methods which hindered their application for condition monitoring of rotary machinery. Ref. [9] gives a new approach to include Cyclic Modulation Spectrum (CMS), a fast alternative to SC which is a cascade of envelope spectra in all possible frequency bands [10], in the unified form of cyclostationary estimators as an estimator of the SC. However, CMS suffers from a major drawback because cyclic frequencies larger than the frequency resolution of the STFT cannot be detected. Recently, a fast algorithm was introduced to compute the SC which substantially increases the computational efficiency [11]. Also, it does not suffer from the CMS limitation.

All the above-mentioned techniques could be considered as linear transformations. Cepstral analysis, as an alternative nonlinear transformation, has been proposed by Ref. [12] and it has been used for several purposes since then such as speech analysis, modal analysis, gears and REBs diagnosis [13]. This method is advantageous in detecting periodicities in the spectrum of a signal by reducing a whole family of harmonics into a single cepstral line. Randall and Swahili [14] and Borghesani et al. [15] took advantage of cepstral analysis features and proposed cepstrum editing and cepstrum pre-whitening approaches to separate periodic and random components and pre-whitening of a signal, respectively.

In the following section 2 the theoretical background, advantages and disadvantages of cepstral analysis, cepstrum editing and cepstrum pre-whitening will be explained. Afterwards, based on the idea behind these techniques a new

empirical method will be proposed in section 3. It includes a nonlinear filtering process, and in this respect, this is completely different from existing REBs diagnosis approaches. In section 4 the method is implemented on six cases from two different experiments. Each data is selected based on some unique characteristics such as presence of high Gaussian\non-Gaussian background noise or variable speed regimes. The detail explanation of these aspects in addition to the performance of the proposed method under these circumstances will also be discussed. Finally, the conclusion is drawn in section 5.

2. Theoretical background

2.1. Cepstral analysis

Cepstrum can be considered as the "spectrum of a spectrum". The word cepstrum is generated by manipulating the word spectrum and, in similar manner, quefrency, rahmonics and lifter are also warped versions of the corresponding spectrum terms of frequency, harmonics and filter.

Time signals are transformed to the frequency domain by Fourier Transform (FT) in terms of magnitude $A(f)$ and phase $\phi(f)$ as follows:

$$X(f) = \mathrm{FT}\{x(t)\} = A(f)\ e^{j\phi(f)} \tag{1}$$

where j is the imaginary unit.

The definition of the complex cepstrum is given by

$$C_c \approx \mathrm{IFT}\Big\{\log\big(\mathrm{FT}\{x(t)\}\big)\Big\} = \mathrm{IFT}\Big\{\log\big(A(f)+\epsilon\big) + j\phi(f)\Big\} \tag{2}$$

where IFT refers to Inverse Fourier Transform and an arbitrary small positive quantity of ϵ is added to $A(f)$ because $\log(A(f))$ has non-numeric value when $A(f) = 0$. It is worth noting that, in contrast to spectrum, the x-axis of cepstrum has the unity of time.

The real cepstrum is calculated by setting the phase to zero and using only the magnitude as follows:

$$C_r = \mathrm{IFT}\Big\{\log\big(A(f)+\epsilon\big)\Big\} \tag{3}$$

The major distinction between cepstrum and spectrum roots in the logarithmic conversion of the former, which leads to a couple of important characteristics. First, the level of periodic components, such as families of harmonics and sidebands, is boosted by the logarithm in comparison to the linear spectrum. Therefore, the IFT of log-spectrum will effectively manifest the presence of the periodic components in the spectrum. Second, multiplicative effect, e.g. effect of transmission path on bearing defect signals, is changed into an additive one; therefore, the two parts can be separated effectively [16].

It should be mentioned that cepstral analysis, despite some interesting features, is not a popular approach in diagnostic of REBs. One reason might be that its computation and interpretation are not as straightforward as the envelope analysis which is a well-known and widely used tool in condition monitoring of rotating machinery. Another reason is related to the non-periodic and cyclostationary nature of bearings defect signals, because their signatures in the spectra are likely to be masked by the background noise due to their weak energy [17]. Also, as the slippage increases the signals deviate from cyclostationarity and the peaks will be smeared in the magnitude spectra [18]. As a result, cepstral analysis will not be able to detect the presence of faults when the slippage rate and/or level of noise is high.

2.2. Cepstrum editing and pre-whitening

Cepstrum editing approach is based on editing the real cepstrum, presented in Eq. 3. Manually selected discrete quefrencies in the real cepstrum, corresponding to the families of harmonics and sidebands in the magnitude spectrum, are removed and then the time signal, related to the residual part, is generated by using the edited log magnitude and the actual phase. For limited speed fluctuation, this process does not necessarily need order tracking, and is able to remove selected discrete frequency components just in one operation while other desired parts are preserved.

Motivated by this methodology, a simpler and slightly different approach is proposed in [15]. The Fourier transform of a signal is divided by its norm

and subsequently is transformed into time domain by means of inverse Fourier transform:

$$x_{cpw}(t) = \mathrm{IFT}\left\{\frac{\mathrm{FT}\{x(t)\}}{|\mathrm{FT}\{x(t)\}| + \epsilon}\right\} \tag{4}$$

where x_{cpw} is the pre-whitened signal, $|\blacksquare|$ denotes the mathematical norm. This operation, called cepstrum pre-whitening (CPW) [15], is equivalent to setting amplitudes of all quefrencies in the real cepstrum to zero. Also, it is shown in [15] that using CPW after Time Synchronous Averaging (TSA) is beneficial in cases with variable speeds as the application of TSA is not able to completely remove the effect of deterministic components, such as gear coupling and misalignments. Since this approach is fast, easy to implement and can be performed without any additional input parameters, it is suited for practical and automatic applications.

It is worth noting that this technique has been well-known for long time as the "phase-only reconstruction" technique [19, 20] but Borghesani et al. [15], for the first time, used this approach for REBs diagnosis.

Oppenheim and Lim [20] thoroughly investigated, both empirically and analytically, and illustrated the importance of phase and its superiority over spectral magnitude in numerous cases, e.g. images, audios and crystallography signals. They also explained conditions under which a signal could be recreated completely or partially just by its phase. One possible explanation could be the fact that the information about the location of events are retained in the phase.

As an example, reconstruction of images by either their magnitude or phase is helpful to illustrate this statement. Fig. 1(a) shows the original images (combination of magnitude and phase). Moreover, Fig. 1(b) and (c) display the magnitude-only (phase = 0) and the phase-only (magnitude = 1) reconstructed images. By comparing Fig. 1(b) and (c), it can be inferred that significant features of the original images are preserved in the phase-only images. In contrast, there is no resemblance between the original and the magnitude-only images.

In general, the high frequency components of images have lower magnitude than their low frequency components; therefore, reconstruction of signals by

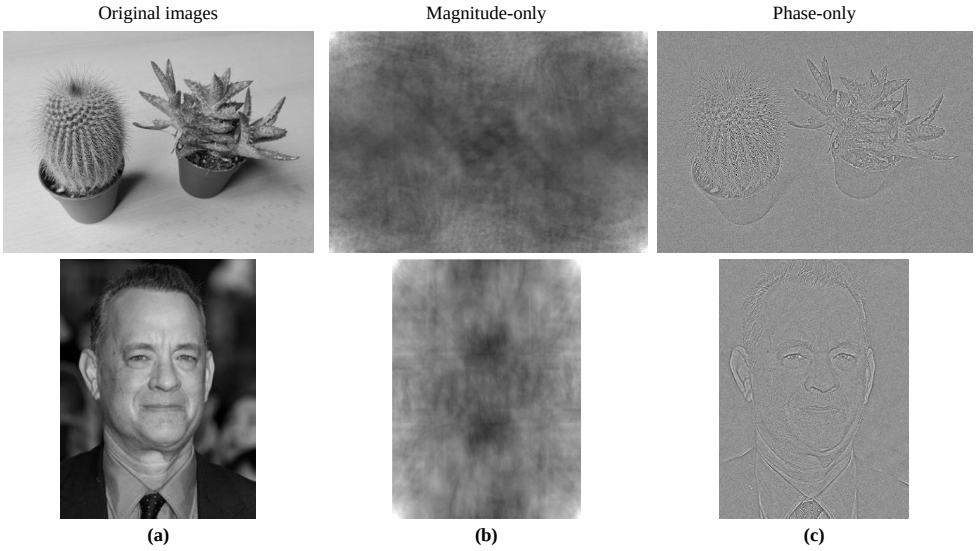

Figure 1: **(a)** Original images **(b)** Magnitude-only reconstructed images **(c)** Phase-only reconstructed images

using only their phases accentuates the effect of their high frequencies. As a result, pre-whitening highlights lines, edges and other narrow events. Additionally, the location of events which have crucial influence on visibility of images are preserved [20] (see also Fig. 1).

Another technique based on the manipulation of cepstral lines was proposed in [21]. Automated Cepstrum Editing Procedure (ACEP) sets to zero certain automatically selected peaks in the real cepstrum, contrary to CPW where the whole real cepstrum is eliminated. The consequent comparison of the performance of ACEP and CPW assesses the potential advantages/disadvantages of each method.

3. Proposed method

This section shows how the concept of cepstrum editing and CPW can be extended to introduce a more comprehensive approach for REBs diagnosis. The method is first explained intuitively and then it is investigated from a theoretical

point of view.

3.1. Spectral Amplitude Modulation (SAM)

A generalized version of CPW is developed in this section. CPW is an effective approach and has achieved valuable results for diagnosis of REBs [22] but suffers from two major drawbacks.

First, as the entire real cepstrum is set to zero or in other words, the magnitude of a signal in whole frequency domain is set to one, the Signal-to-Noise Ratio (SNR) for the reconstructed signal is decreased and the frequency components of noise have the same magnitude as the frequencies linked to the defect signals (carrier frequencies).

Second, assuming bearings defect signals only as 2nd order cyclostationary leads to the conclusion that the peaks related to damages have not considerable amplitude in the real cepstrum [15]. This assumption is not precise and always acceptable, since in many cases the random slippage is not high and the discrete components related to defects can be detected in both frequency and quefrency domains.

Moreover, even by assuming second order cyclostationarity of bearings defect signals, or pseudo-cyclostationarity as a more realistic process [18], for a typical percentage of slippage in REBs, peaks (discrete frequencies) are generated in the magnitude spectra. When the level of background noise is increased, it will likely mask the peaks completely. It is nonetheless worth mentioning that certain peaks are amplified in resonance frequencies ranges of structures and, even under this circumstance, they are detectable in the spectrum.

The two mentioned disadvantages are actually responsible for the false negative diagnosis of some cases in the benchmark study of Case Western Reserve University (CWRU) bearing data set [22] in which the defect frequencies are present in the SES of the raw signals but they are eliminated by CPW.

As it was mentioned, in the CPW method the whole real cepstrum is set to zero to reduce the effect of deterministic parts. In addition, by using Eq. 4 this approach is directly applicable in the frequency domain; therefore, there is no

need for transformation to cepstral domain. This equation can be rewritten in the following form:

$$x_{cpw}(t) = \mathrm{IFT}\Big\{A(f)^0 e^{j\phi(f)}\Big\} \tag{5}$$

This method introduces some disturbance as the edited amplitude is inconsistent with the original amplitude but this drawback is often insignificant [14]. In fact the phase relatively contains more information of a signal than the magnitude and to completely reconstruct a signal, within a scale factor, the information preserved in phase is adequate [20]. Consequently, many of the significant features of a signal, in this case valuable diagnostic information, are preserved after this process.

As the assumption of unity magnitude in CPW (phase-only reconstruction) is rather arbitrary, a better approach could be manipulation of the original magnitude so that it can represent the overall behavior of the spectrum. Giving different weights to different frequency components may lead to a nonlinear filtering process where the signals from different sources are separated. For this purpose, a novel method is proposed to generalize Eq. 5:

$$x_m(t,n) = \mathrm{IFT}\Big\{A(f)^n \; e^{j\phi(f)}\Big\} \tag{6}$$

where x_m will be called *modified signals* and are constructed by the "representative" magnitudes and the original phase of the signal. In fact, power of zero of the magnitude in Eq. 5 is replaced by n, while $a \leqslant n \leqslant b$, a and b are two arbitrary numbers. Hereafter, the variable n will be referred to as *Magnitude Order* (MO). It is worth mentioning that raw signals and signals after CPW happen to be two particular cases of the modified signals x_m, with MO equal to 1 and 0 respectively. Also, $\log\big(A(f)\big)$ is approached by a small positive value of MO.

The flowchart of the proposed method is shown in Fig. 2. Similarly to cepstrum editing and CPW, disturbance of signal is inevitable as the magnitude spectrum is altered. Nonetheless, the main interest for diagnosis of REBs is the repetition rate of the impulses generated by defects; therefore, if MO is

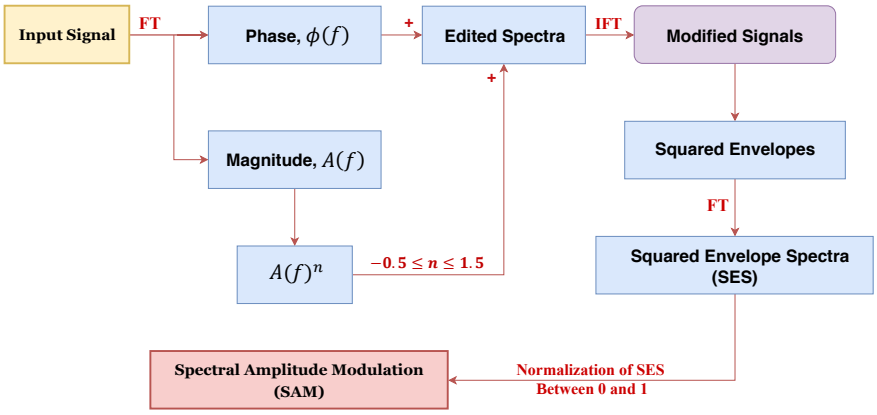

Figure 2: Flowchart of the proposed method

selected appropriately, the effects of these disturbances are not problematic as the repetition rate of the impulses will remain unaffected.

After the modified signals are obtained for a range of MO (-0.5 to 1.5 is recommended in this paper), the SES of each modified signal is calculated. Analytic signal ($\mathcal{A}$) and SES for the modified signal $x_m(t,n)$ by using Hilbert transform ($\mathcal{H}$) can be calculated as follows:

$$\mathcal{A}\{x_m(t,n)\} = x_m(t,n) + j * \mathcal{H}\{x_m(t,n)\} \tag{7a}$$

$$\mathrm{SES}\{x_m(t,n)\} = \left|\mathrm{FT}\left\{\left|\mathcal{A}\{x_m(t,n)\}\right|^2\right\}\right| \tag{7b}$$

In addition, a new envelope spectrum, namely Log-Envelope Spectrum (LES), has been proposed by [23] to investigate 2nd order cyclostationary signals contaminated by highly non-Gaussian and impulsive background noise:

$$\mathrm{LES}\{x(t)\} = \left|\mathrm{FT}\left\{\log\left(x(t)^2\right)\right\}\right| \tag{8}$$

In presence of sizable and non-periodic impulses, LES can be employed as a more robust alternative to SES for computation of modified signals' envelopes spectra. The only drawback of LES is related to the lower SNR of $\log\left(x(t)^2\right)$ in comparison to $\left|\mathcal{A}\{x(t)\}\right|^2$. Therefore, it is recommended to use this indicator whenever impulsive noise exists in the signals.

In this paper, rather than LES, the Fourier transform of logarithm of squared envelope signal is computed and it is called "LSES" (Eq. 9).

$$\mathrm{LSES}\{x(t)\} = \left|\mathrm{FT}\left\{\log\left(\left|\mathcal{A}\{x(t)\}\right|^2\right)\right\}\right| \tag{9}$$

The two approaches in Eq. 8 and Eq. 9 provide almost comparable results. Due to aliasing, there is a possibility that desired information are masked in LES as the sampling frequency is not doubled prior to squaring. In addition, squaring operation in Eq. 8, which doubles the frequency content of signals, is not followed by a low-pass filtering process.

Moreover, to counterbalance the effect of increased noise for MO values lower than 1, the spectra of the modified signals envelopes/log-envelopes are computed by Fourier transforming the autocorrelation functions, which are effectively equal to SES^2 or LSES^2 (hereafter for the sake of simplicity they will be referred as SES and LSES).

Next, these spectra (SES or LSES) are normalized between 0 and 1, which is essential to equalize the results for different MO values and thus provide explicit and comparable results. In addition, to avoid large peaks at the beginning of the SES/LSES the first few Hertz are not considered for the purpose of normalization.

Finally, the outcome is demonstrated by a 3-dimensional plot, where x, y, z axes represent the modulation (cyclic) frequency, MO and normalized amplitude respectively. Similarly, results can be presented in a more straightforward form with a 2-dimensional colormap where the color is proportional to the normalized amplitude. Since this approach investigates the signature of amplitude modulation of signals in frequency domain, it will be called "*Spectral Amplitude Modulation*" or "SAM" in an abbreviated form. This term is substituted by "log-SAM" when LSES is exploited instead of SES.

Implementation of the proposed method is simple, fast and no prior knowledge of signals is needed. Moreover, it is not required to set any parameter in advance or find a frequency band for demodulation. In fact, the procedure works as a nonlinear filtering process and automatically decomposes a signal for differ-

ent resonance and discrete frequencies based on their amplitudes. Consequently, this approach could be seen as a complementary to CSA, i.e. CSA/SAM maps signals in terms of carrier frequency/power and modulation frequency.

3.2. Physical interpretation of SAM and the effect of MO

Based on the values of MO, the effect of the proposed method on the modified time signal $x_m(t, n)$ could be grouped in three categories:

First, for $n > 1$ the effects of frequencies with higher magnitude are amplified. Also, as the value of n increases, part of the time signal represented by frequencies with lower magnitude will be gradually masked by other parts containing frequencies with higher level of energy. These values of MO may also have denoising effects as the distinction between resonance frequencies of the structure excited by defect impacts and other frequencies associated with the noise will be better highlighted.

Second, for $0 \leqslant n \leqslant 1$ the dominancy of frequencies with larger amplitude will gradually reduce in comparison to frequencies with lower amplitude as the value of n varies from 1 to 0. In contrast to CPW, all frequency components are not set to the same magnitude. As a result, the effect of both high and low energy parts, as well as cyclostationary and periodic components will be preserved to some extent. This is mostly important because in real cases the acquired signals are combination of several heterogeneous vibration components and the bearing defect signal is not often the dominant part. For instance, when the bearing defect signal is weak and masked by signals from other sources such as gears meshing, the impact of this "discrete frequency noise" will be lessened, meanwhile the role of bearing defect signal will be improved and the defect signal will be more effectively detected.

Third, for $n < 0$ this process has a different and unique outcome as the effect of magnitude will be reversed. Amplitudes of more dominant frequency components now become negligible which is analogous to filtering out these frequencies. On the contrary, the influence of previously insignificant frequencies will be boosted. Negative values of n are advantageous especially for diagnosis

of bearings with multiple defects and/or when a rotating machinery has more than one faulty bearing and there is a considerable difference among their energy levels. When the level of noise is not so significant to dominate the modified signal, the information associated to the defect with very low level of energy, which was previously masked by the high level of energy of other defect signals, can be extracted more effectively.

3.3. Mathematical interpretation of SAM

The Fourier series of a periodic signal $x(t)$ with zero mean value can be written as follows:

$$x(t) = \sum_{i=1}^{q} D_i \sin(\omega_i t + \phi_i) \tag{10}$$

where D is known as Fourier coefficient and q is length of x.

The Hilbert transform of $x(t)$ is defined by an integral transform:

$$\mathcal{H}\{x(t)\} = \pi^{-1} \int_{-\infty}^{\infty} \frac{x(\tau)}{\tau - t}\, d\tau \tag{11}$$

which shifts the phase of each element of the Fourier series in Eq. 10 by -90°, therefore

$$\mathcal{H}\{x(t)\} = \sum_{i=1}^{q} D_i \cos(\omega_i t + \phi_i) \tag{12}$$

As a result, the squared envelope or the norm of the analytic signal is calculated as follows:

$$x(t)^2 + \mathcal{H}\{x(t)\}^2 = \sum_{i=1}^{q}\sum_{j=1}^{q} D_i D_j \cos\left((\omega_i - \omega_j)t + \phi_i - \phi_j\right) \tag{13}$$

which can be rewritten in the equivalent form:

$$\begin{aligned} x(t)^2 + \mathcal{H}\{x(t)\}^2 = &\sum_{i=1}^{q}\sum_{j=1}^{q} D_i D_j \cos\left(\phi_i - \phi_j\right) \cos\left((\omega_i - \omega_j)t\right) \\ &- \sum_{i=1}^{q}\sum_{j=1}^{q} D_i D_j \sin\left(\phi_i - \phi_j\right) \sin\left((\omega_i - \omega_j)t\right) \end{aligned} \tag{14}$$

As the modified signal x_m, given by Eq. 6, is constructed by the original phase and the modified magnitude of the signal $x(t)$, similarly to Eq. 10 it can be represented by

$$x_m(t, n) = \sum_{i=1}^{q} D_i^n \sin(\omega_i t + \phi_i) \tag{15}$$

So its analytic signal can be also computed by

$$x_m(t,n)^2 + \mathcal{H}\{x_m(t,n)\}^2 = \sum_{i=1}^{q}\sum_{j=1}^{q} D_i^n D_j^n \cos\big(\phi_i - \phi_j\big)\cos\big((\omega_i - \omega_j)t\big) - \sum_{i=1}^{q}\sum_{j=1}^{q} D_i^n D_j^n \sin\big(\phi_i - \phi_j\big)\sin\big((\omega_i - \omega_j)t\big) \tag{16}$$

Consequently, SESs of modified signals are calculated as

$$\mathrm{SES}(f_k,n) = (2-\delta(k))\left(\Big(\sum_{r=1}^{q-k} D_r^n D_{r+k}^n \cos(\phi_k - \phi_{k+r})\Big)^2 + \Big(\sum_{r=1}^{q-k} D_r^n D_{r+k}^n \sin(\phi_k - \phi_{k+r})\Big)^2\right)^{\frac{1}{2}} \tag{17}$$

where $f_k = \omega_k/2\pi$, $k = 0, ..., q-1$ and δ is defined as

$$\delta(k) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$$

Assuming a (carrier) wave $c(t) = C\cos(\omega_c t + \phi_c)$ is modulated by a (modulation) wave $m(t) = M\cos(\omega_m t + \phi_m)$. The spectrum of the resultant signal contains a peak at ω_c with two sidebands at $\omega_c + \omega_m$ and $\omega_c - \omega_m$. In addition, the phases at ω_c, $\omega_c + \omega_m$ and $\omega_c - \omega_m$ are ϕ_c, $\phi_c + \phi_m$ and $\phi_c - \phi_m$ respectively.

Whenever in Eq. 14 (or Eq. 16) ω_i or ω_j is equal to ω_c and $\omega_i - \omega_j = \omega_m$, the value of $\phi_i - \phi_j$ is equal to ϕ_m and $D_i D_j$ (or $D_i^n D_j^n$) is non-zero and positive. On the other hand, when $\omega_i - \omega_j \neq \omega_m$ the product $D_i D_j$ (or $D_i^n D_j^n$), when there is no noise, is equal to zero. Even in presence of noise in which the value $D_i D_j$ (or $D_i^n D_j^n$) is not equal to zero, the phase difference $\phi_i - \phi_j$ is not equal to ϕ_m any more and will change randomly from sample to sample. As a result, for each specific component of the SES of the raw or modified signals in Eq. 17 ($f_k = \frac{\omega_i - \omega_j}{2\pi}$), the summation of corresponding Fourier series coefficient which are also equivalent to the amplitude of each frequency element in the SES, will generally approach zero by utilizing a sufficient number of samples and a suitable range for n.

To conclude, Eq. 17 shows that the mathematical background is the same for the classical SES and the proposed SESs of modified signals in this paper.

Another point about SES which is worth mentioning here is that, based on Eq. 17 and the above explanation, two parts are responsible for the amplitude

of each frequency component of the SES. Magnitude of spectral lines of the signal D, has the first contribution by $D_i D_j$ as the sidebands are spaced at ω_m around each ω_c. The second contribution is by phase in terms of modulation frequency ($\cos(\phi_m)$ and $\sin(\phi_m)$). From this point of view, CPW only takes advantage of the signature of modulation in the phase of a signal. On the other hand, the influence of magnitude and phase are considered in the SES of a raw signal simultaneously. Similarly, the proposed method uses the effect of both parts but manipulates the magnitude spectrum while keeping the signature of modulation in the phase unchanged. This is advantageous since the positive role of magnitude spectra is not discarded like in CPW. On the other hand, contrary to the SES of raw signals, its negative role, i.e. large peaks in the magnitude spectra from exogenous sources, is moderated.

3.4. Practical consideration for selecting the limits of MO

To implement SAM, the last step is to select the limits of MO. For MO $\geqslant 0$ no modulation source which already does not exist in $x(t)$ will be added into $x_m(t, n)$. In other words, no undesirable and false peak will be generated in the SESs because the signals have similar phases and also no new peak is added to magnitude spectra of modified signals, only existing peaks in the magnitude spectrum of raw signal are given different weights (see Eq. 17). Thus, there is no theoretical restriction for choosing upper limit of MO. However, it is reasonable to avoid very high values of b, since only a few large peaks, mostly not linked to REBs signals, will then dominate the magnitude spectrum and masks other components of the signal.

On the other hand, new false components will be added to SESs of modified signals for very low and negative values of MOs. Because very low amplitude components in the magnitude spectrum of raw signal now dominate the magnitude spectra of modified signals and at one point the adverse effect of these newly generated peaks ($D_i^n D_j^n$) cannot be removed even by the randomness of $\phi_i - \phi_j$ (see Eq. 17); therefore, new false components begin to emerge in the SESs of modified signals.

For practical reason and in order to keep the approach as general as possible for REBs diagnosis, the values of -0.5 and 1.5 for a and b are suggested. These quantities are recommended intuitively, and their validity will be examined in the following section.

Additionally, a simple yet effective approach is proposed to find the lowest acceptable value of MO. This point will be called critical point since false peaks in the SESs of modified signals begin to appear for lower values than this point. As all the SESs are normalized between 0 and 1 it can be deduced that the mean value of normalized SES of a signal related to noise will be very high. The reason is that many components in the normalized SES of noise are equal to one or have very high values. In contrast, presence of peaks in a normalized SES will reduce its mean value. As the value of MO decreases the effect of noise will gradually growth and at the critical point the modified signal is completely noisy and therefore has the highest mean. For lower MO than critical point, false peaks start to emerge in the SESs and the mean value of normalized SESs decreases subsequently. This characteristic of normalized SESs could be used to find the minimum acceptable value of negative MOs. In fact, the critical point is revealed as a maximum in a mean-MO plot.

4. Results and discussion

This section exemplifies SAM for diagnostics of rolling element bearings on six real case data. The first five experimens are selected from CWRU bearings data set [24]. The last application referes to the data recorded under variable operating conditions, available as supplementary material of Ref. [25].

The effect of various ranges of MO was discussed in subsection 3.2 and practical consideration for their selection has been given in subsection 3.4. The recommended values of -0.5 and 1.5 for lower (a) and upper (b) limits of MOs will be employed for all cases and the effectiveness of these suggestions will also be discussed.

Each test case is presented with a particular purpose: in case 1, the capa-

bility of the method in presence of multiple defects is shown. Also, a detailed explanation is given regarding its performance. The effect of Electro Magnetic Interference (EMI) is studied in case 2, where the SES of the raw signal and CPW provide negative and positive diagnostic outcome respectively. In contrast to case 2, in case 3 the SES of the raw signal contains the defect signature but CPW erases the defect signal. In case 4, neither SES of the raw signal nor CPW include valuable diagnostic information. Case 5 shows that SAM can be also used to deal with signal contaminated with highly impulsive noise. Case 6 presents the performance of SAM while signals are acquired under variable regime.

As it was mentioned the proposed method includes the SES of raw and CPW signals; therefore, its results are automatically compared with these two approaches. Furthermore, the first five cases are chosen from CWRU data set which has been widely examined in various papers. For instance, a benchmark study by Smith and Randall [22] investigated all the signals in this data set and CPW, as a subcase of the proposed technique, was shown to be the most successful method for diagnosis. Case 1 is also investigated in Ref. [5] by using the Autogram. The result of the spectral correlation for case 6 and an identical data to case 1 is given in Ref. [11]. As a result, a comparison with other techniques is only provided for case 4 which is also the most difficult case.

4.1. Case 1

For this example, record 275 DE of CWRU is utilized. The defected bearing with seeded inner race fault is located on the fan-end of the test rig. The transducer acquires the accelerations from the drive-end so that the transmitted signal attenuates through the apparatus, which makes it difficult to investigate the defect signature. Moreover, the unexpected ball pass frequency of inner race (BPFI) for drive-end in addition to its sidebands are also present in the SES, which might be related to the misalignment of the drive-end bearing and not to a defected bearing [11].

All things considered, this record is an interesting and challenging case to

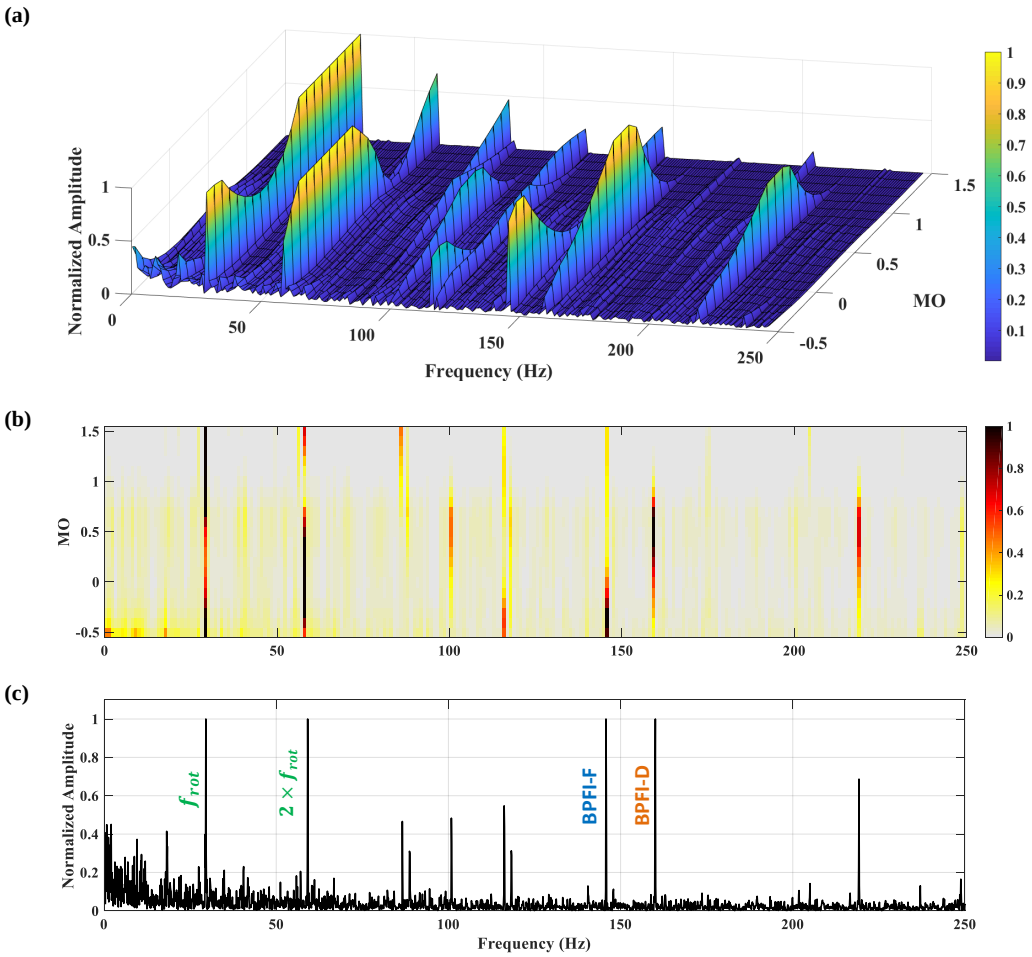

Figure 3: Case 1: 275 DE, Spectral Amplitude Modulation (SAM) **(a)** 3D plot **(b)** Above view **(c)** View along the MO-axis (MSES)

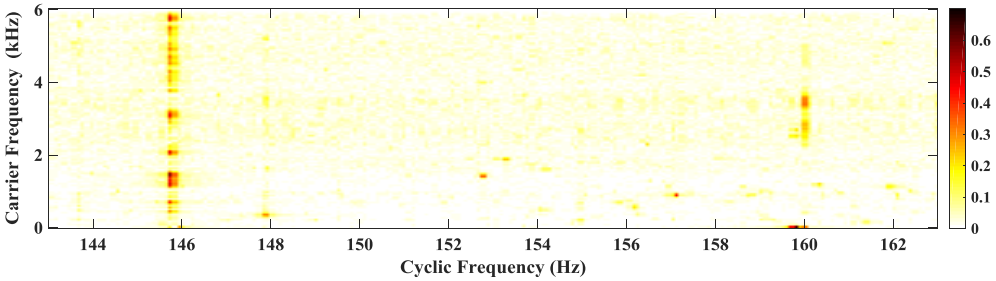

Figure 4: Case 1: Spectral coherence

evaluate the performance of each novel bearing diagnostic technique, such as the proposed SAM.

The outcome of the proposed method, is illustrated with a three-dimensional graph in Fig. 3(a) where the x, y and z axis represent cyclic frequency, MO and normalized amplitude respectively, i.e. cascades of normalized SES for modified signals generated by varying n. As already stated, the values of MO ranges from -0.5 to 1.5 and the increment of 0.1 is utilized in this paper. Furthermore, before normalization of SES/LSES the first 3 Hertz which mainly includes large peaks are dismissed to avoid their adverse effect.

Fig. 3(b) displays the 2D view of Fig. 3(a) , where the colors are proportional to the normalized amplitude values. The variation of each cyclic frequency intensity based on the MO values can be followed more clearly in this figure. Moreover, the MO value in which a specific cyclic frequency reaches its maximum could be considered as an indicator of energy of carrier frequencies associated to that cyclic frequency.

Fig. 3(c) depicts the view along the y-axis (MO) of Fig. 3(a). In fact, this plot shows the maximum normalized amplitude value of each cyclic frequency, which will be called "*Maximum Squared Envelope Spectrum* (MSES)". MSES could be considered as a useful diagnostic tool since one cyclic frequency might be the largest peak in the SES for a certain MO, conversely does not exist for some other MO but this plot displays the entire cyclic frequencies of a signal just in a single figure.

For this case, the shaft frequency (f_{rot}), BPFI of fan-end (BPFI-F) and BPFI of drive-end (BPFI-D) bearings are equal to 29.5 Hz, 146.1 Hz and 159.9 Hz respectively. Both damage frequencies in addition to the shaft frequency and its harmonics can be noticed in Fig. 3(a-c) and in fact, all the cyclic frequencies related to the signal are revealed. This is a great advantage of this method over other approaches which try to find a single effective frequency band for demodulation.

Based on the result presented in Fig. 3, the BPFI-F has high values for MO from -0.5 to 0, with a maximum at -0.4. On the other hand, the BPFI-D is

the dominant frequency in the SES for MO of 0.5 and 0.6; also, it has high values throughout -0.5 to 1.1. As the maximum value of BPFI-F is achieved for lower MO compared to the BPFI-D, it indicates that the signal from the fan-end bearing has lower energy level than the drive-end bearing signal. This deduction is in good agreement with the fact that the transducer is positioned near the drive-end bearing and far from the fan-end bearing.

In Fig. 3(a) and (b) the normalized amplitude will not change in some range against different MO. This phenomenon happens because for different ranges of MO different frequency bands will become the main contributor to the corresponding modified signal. Therefore, the cyclic frequency related to that specific frequency band will become the most dominant peak in SAM.

In order to have a better understanding of carrier frequencies related to each cyclic frequency, particularly the BPFI-F and BPFI-D, the spectral coherence (normalized version of Spectral Correlation (SC)) computed by using Fast-SC algorithm [11] is visible in Fig. 4. Several carrier frequencies are responsible for the BPFI-F (with nominal value of 146.1 Hz) which makes it difficult to extract all parts of this defect signal. Even knowing the exact carrier frequencies, they correspond to a series of filtering processes for multiple frequency bands and central frequencies. The carrier frequencies related to the BPFI-D are mainly concentrated in the mid-range. Although, this band is not completely continuous and in this range some carrier frequencies are still linked to the BPFI-F.

To investigate in more details the performance of the method and its diagnostic capabilities, the SES of the modified signals for MO equal to 1, 0, -0.4, 0.5 and 1.5 are depicted in Fig. 5(a-e) respectively. In addition, the magnitude spectra of the modified signals are illustrated in Fig. 6(a-e).

The SES of the raw signal and its magnitude spectrum (MO = 1) are shown in Fig. 5(a) and Fig. 6(a). The BPFI-F and BPFI-D and their harmonics are present in the SES but they are not the most dominant frequencies and the maximum amplitude in the SES corresponds to the shaft frequency. It was discussed that the resonance frequencies of BPFI-D have a high level of energy,

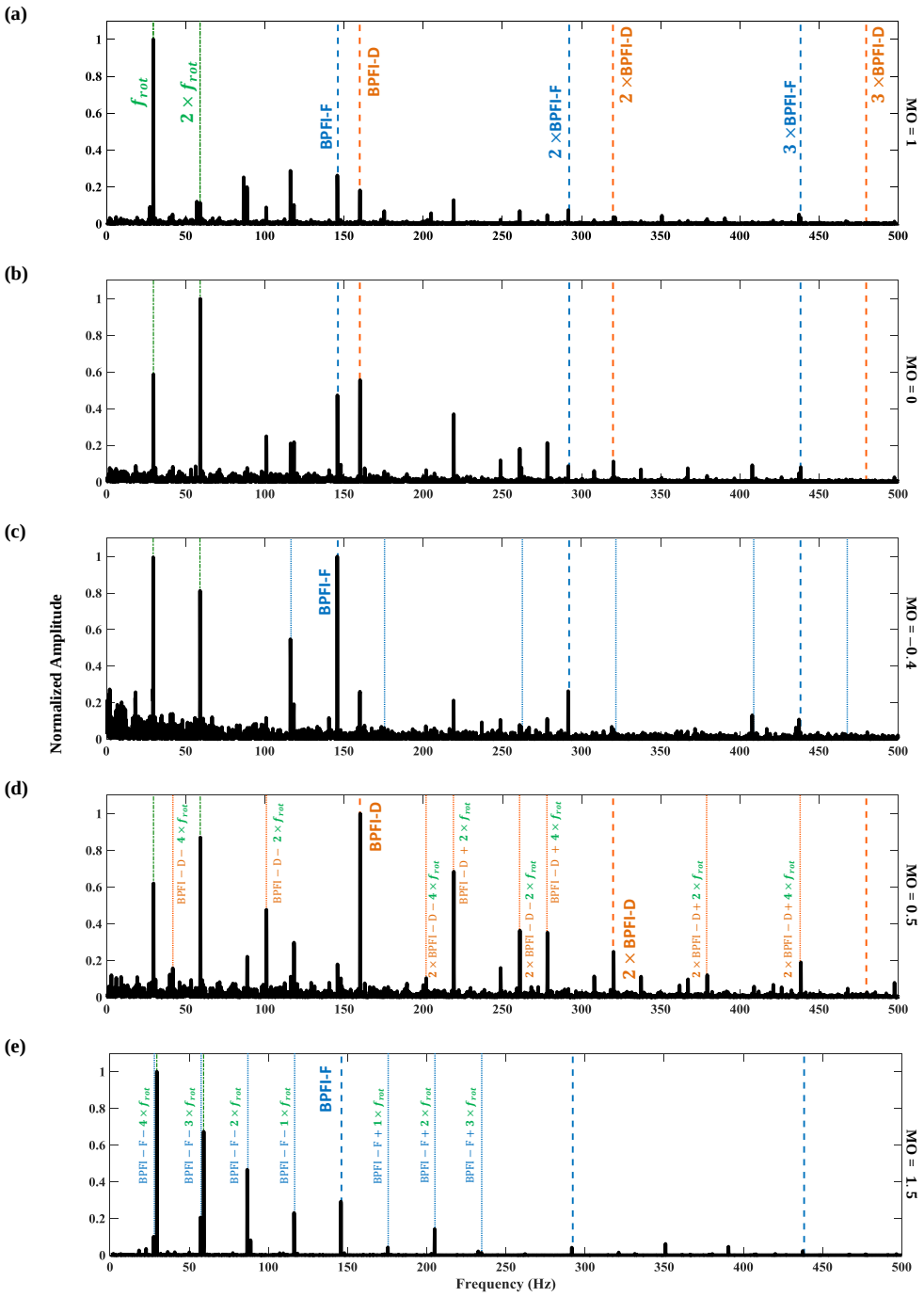

Figure 5: Case 1: SES of the modified signals for **(a)** MO = 1 (raw signal) **(b)** MO = 0 (CPW) **(c)** MO = −0.4 **(d)** MO = 0.5 **(e)** MO = 1.5

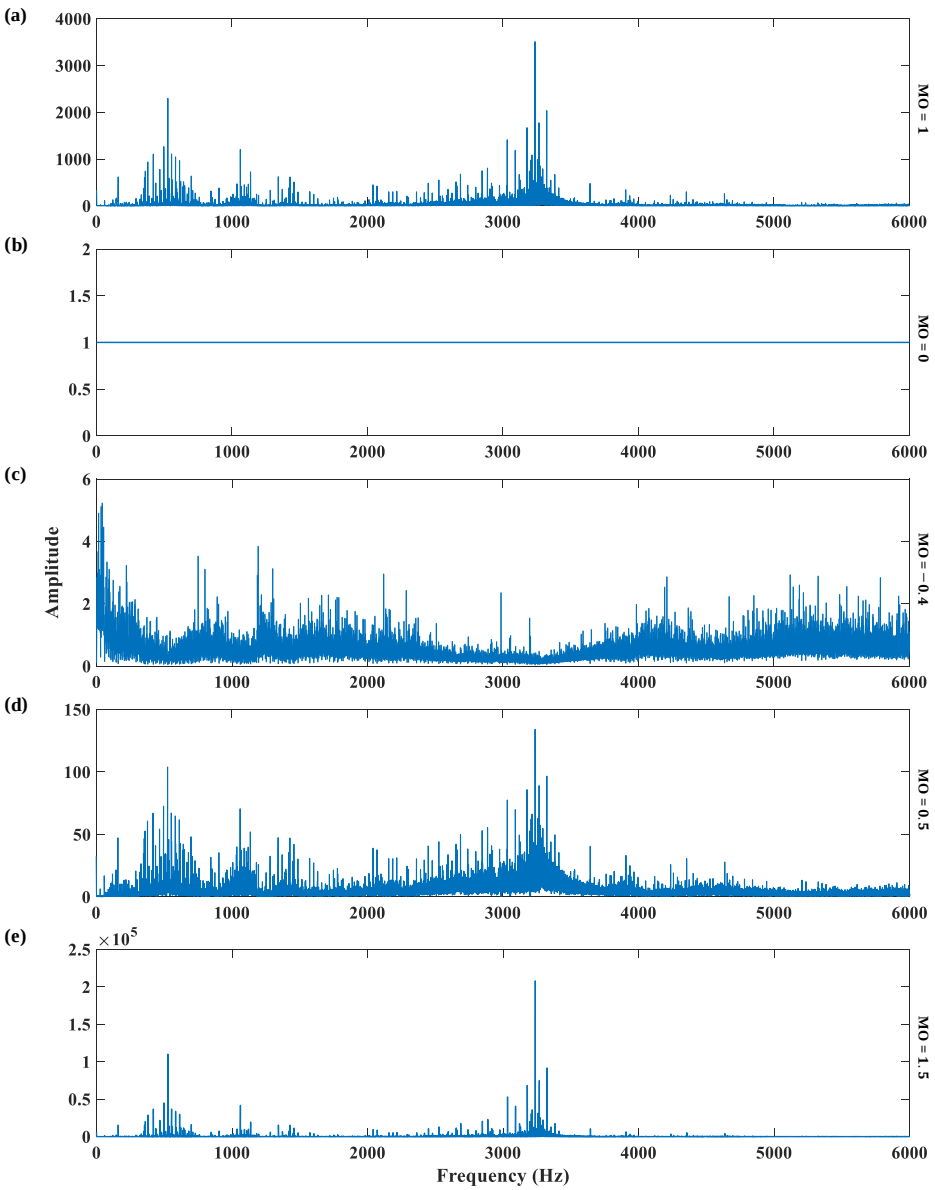

Figure 6: Case 1: Magnitude spectra of the modified signals for **(a)** MO = 1 (raw signal) **(b)** MO = 0 (CPW) **(c)** MO = −0.4 **(d)** MO = 0.5 **(e)** MO = 1.5

which can be spotted around 3200 Hz in Fig. 6(a). Two resonance frequency bands around 500 Hz and 1100 Hz have also sizable magnitudes but amplitudes of other frequencies are very small and negligible. Another observation is that the discrete frequencies which are amplified by the resonances are present in the magnitude spectrum.

The SES and spectrum of the acceleration signal after CPW are depicted in Fig. 5(b) and Fig. 6(b) respectively. As expected, the magnitude of whole frequency range is one. Consequently, the dominance of the shaft frequency is reduced and the two ball pass frequencies as well as their sidebands are now more noticeable, compared to the SES of the raw signal.

It should be noted that sidebands related to the BPFI-F are spaced at the shaft frequency. But the signal related to the drive-end defect is modulated at twice the rotation speed, thus sidebands of BPFI-D are spaced at this frequency. This might be the reason behind the large magnitude of the second harmonic of the shaft frequency in the SES.

The SES and magnitude spectra for MO equal to -0.4 and 0.5 are shown in Fig. 5(c-d) and Fig. 6(c-d). As it was discussed, for these two values the BPFI-F and BPFI-D are the most dominant frequencies in the SES of the corresponding modified signals.

Fig. 5(c) clearly reveals three harmonics of the BPFI-F as well as the shaft frequency and the resultant sidebands. Since MO is negative, the dominant resonance frequencies which are related to the BPFI-D and shaft frequency are canceled (see Fig. 6(a) and (c)). As a result, amplitude of the BPFI-D is decreased in the SES (see Fig. 5(a) and (c)). Moreover, the influence of other carrier frequencies with negligible amplitude, mainly including the signature of the fan-end bearing defect signal (see Fig. 4 and Fig. 6(c)), are enhanced.

For MO equal to 0.5, the SES is displayed in Fig. 5(d). Three harmonics of BPFI-D, accompanied by their sidebands which are clearly spaced at twice and forth of the shaft frequency, are highlighted. As the value of MO is positive, the overall behavior of the magnitude spectrum (Fig. 6(d)) is comparable to the magnitude spectrum of the raw signal (Fig. 6(a)). Furthermore, the value of

MO is lower than one; hence, the effect of the frequencies with lower amplitude, which in this case also include the resonance frequencies around 3200 Hz and linked to the BPFI-D (see Fig. 4), is improved and dominates the time signal. Furthermore, the effect of large discrete frequencies in the same range is weakened, which reduces the amplitudes of the shaft frequency and BPFI-F in the SES.

It can be noticed that in Fig. 3(a) and (b), in addition to the absolute maximum of the BPFI-F for MO = −0.4, the value of BPFI-F has another relative maximum at 1.5, i.e. it starts to increase again for MO greater than 1. The SES of the modified signal for this value of MO is shown in Fig. 5(e). Although the shaft frequency has the largest amplitude, presence of the defect frequency BPFI-F in addition to its harmonics provides valuable diagnostic information. The corresponding time signal is highly modulated by the rotation of the shaft; therefore, the shaft frequency and the subsequent sidebands around BPFI-F and its harmonics can be identified evidently.

The magnitude spectrum of the modified signal for MO = 1.5 is displayed in Fig. 6(e). Since MO > 1, frequencies with large amplitude dominate the spectrum. These frequencies mainly consist of discrete frequencies enhanced by the two resonance frequency bands adjacent to 500 Hz and 3200 Hz. As it was explained, these discrete frequencies are mostly related to the shaft frequency but also contain symptoms of BPFI-D (see Fig. 4).

Comparison of SES depicted in Fig. 5(c) and (e) discloses two dissimilar patterns for the shaft frequency and the defect frequency BPFI-F. This is attributed to the contribution of two phenomena with different characteristics in the defect signal. The signal associated with the fan-end bearing defect includes two main parts, one cyclostationary and the other periodic. The cyclostationary part, presented by multiple flows of energy with very low magnitude in the spectrum (see Fig. 4 and Fig. 6(c)), has been extracted by setting the value of MO to -0.4. On the other hand, the periodic part, with multiple discrete frequencies in the spectrum, has been obtained by setting the value of MO to 1.5. This is also the reason why BPFI-F and the shaft frequency have two relative

maxima in the SAM.

As it was mentioned, there is an overlap between carrier frequencies related to BPFI-D and BPFI-F around 3200 Hz. However, this frequency band is still selected for MO = 1.5 but there is no sign of BPFI-D and only BPFI-F exists in the SES. It is due to the fact that the proposed method mostly separate components of the signal based on the energy levels of their carrier frequencies; therefore, the spectral lines linked to the BPFI-F which have greater amplitudes are effectively separated from carrier frequencies related to the BPFI-D in this frequency range.

4.2. Case 2

The second case will investigate data 318 BA with an outer race defect on the fan-end bearing. The acceleration signal, and a zoomed portion are depicted in Fig. 7(a). The signal is not impulsive and the defect signature neither can be spotted in the time signal nor in the SES of the raw signal (Fig. 7(b)), since the BPFO (88.2 Hz) is not present in the SES. The most dominant frequency is associated to the cyclic frequency 114.1 Hz and the shaft frequency (f_{rot} = 28.8 Hz) is also detectable in the SES. The magnitude spectrum of the raw signal is illustrated in Fig. 8(a). The spectrum is contaminated with some discrete frequency "noise" which has very high amplitudes. The source of this noise is not completely evident; however, it might be a result of strong Electro-Magnetic Interference (EMI) from Variable Frequency Drive (VFD). EMI could be explained as "any unwanted signal that is either radiated or conducted to electronic equipment and negatively affects the performance of the equipment" [26].

To investigate this assumption, a zoomed portion of Fig. 8(a) around the resonance frequency 3644 Hz and in logarithmic scale is depicted in Fig. 8(b). In detail, Ref. [26] explains the frequency representation of EMI interferences from a VFD with Pulse-Width Modulated (PWM) output. PWM waveforms contain switching frequency of the VFD and its harmonics, modulation frequency which are also called "pseudo line frequency" and sidebands around the

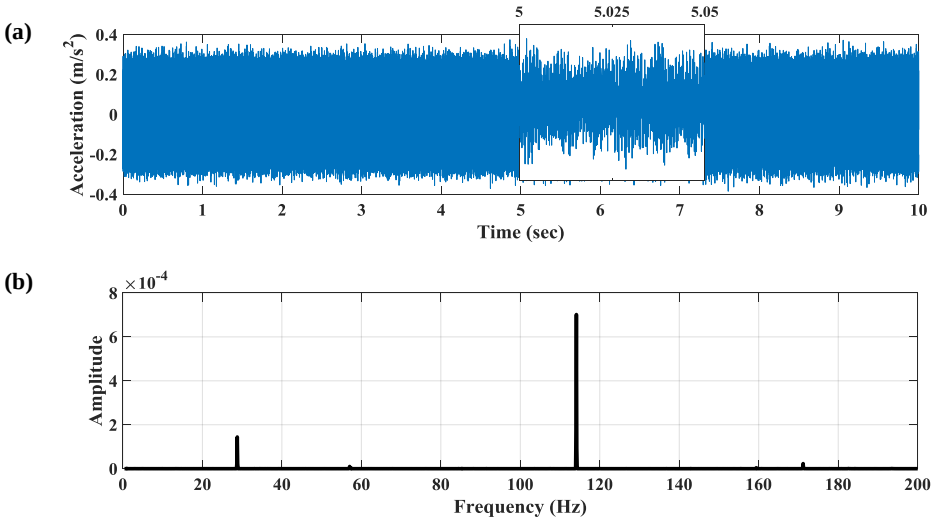

Figure 7: Case 2: 318 BA **(a)** Acceleration signal **(b)** SES of the raw signal

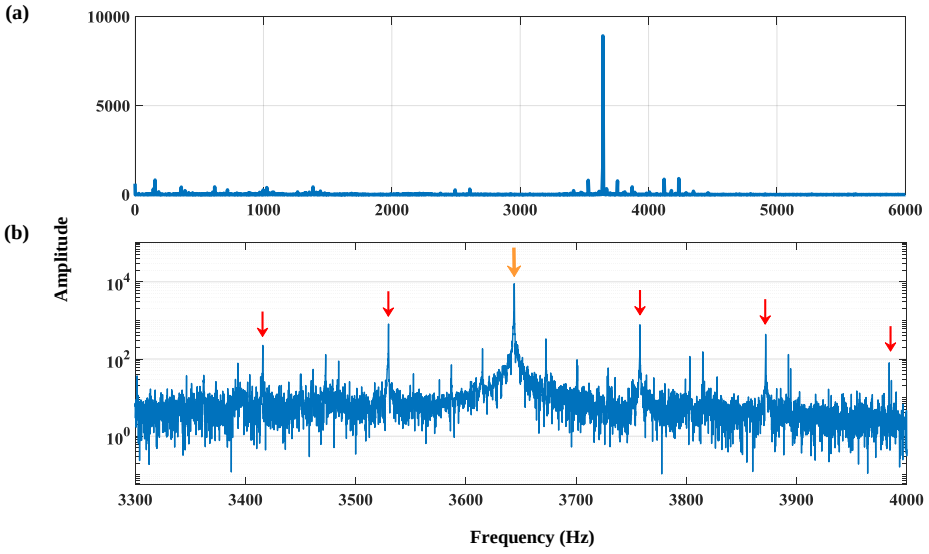

Figure 8: Case 2 **(a)** Magnitude spectrum of the raw signal **(b)** Zoomed magnitude spectrum in logarithmic scale. Carrier (switching) frequency of VFD (orange arrow) and sidebands spaced at pseudo line frequency of 114.1 Hz (red arrows)

carrier harmonics spaced at pseudo line frequency [26]. In Fig. 8(b) the carrier (switching) frequency of the VFD (3644 Hz, denoted by a orange arrow) and sidebands spaced at pseudo line frequency of 114.1 Hz (f_{EMI}, denoted by the red arrows) around the carrier frequency can be spotted distinctly.

This carrier frequency is the absolute maximum of the spectrum and has considerable amplitude comparing to other spectral lines. As a result, all the other frequencies of the signal, in particular the frequency band comprising the bearing defect signal, are masked by this interference.

The 3D and 2D presentations of the proposed method (SAM) result are shown in Fig. 9(a) and (b) and MSES is depicted in Fig. 9(c). It can be noticed from Fig. 9(a) and (b) that throughout the −0.5 to 0.4 values of MO the BPFI is the dominant cyclic frequency. The SES of the modified signal and its magnitude spectrum for MO equal to 0.4 are displayed in Fig. 10(a) and (b) respectively. Since the value of MO is smaller than 1, the effect of frequencies related to the EMI are lessened. In contrast, other carrier frequencies containing the defect signal are enhanced (see the difference between Fig. 8(a) and Fig. 10(b)) and now the presence of BPFO and its harmonic combined with the shaft frequency provide a successful diagnosis of the bearing, though the frequency band linked to the EMI is still part of the spectrum. It should be noted that the BPFO is the largest peak in the SES for MO from -0.5 to 0.4; therefore, CPW as a specific case of the proposed method for MO = 0 will also detect the presence of the defect. Although, the result is presented for MO = 0.4 to attain a higher SNR.

For MO > 0.5 the EMI signal will completely mask the bearing defect signal. As a consequence, the pseudo line frequency (modulation frequency) of 114.1 Hz which is produced by the VFD will dominate the SES and absolutely eradicate the signature of the defect frequency.

4.3. Case 3

The data 202 DE will be examined as the third case. The outer race of the drive-end bearing is reported to be defected with nominal BPFO = 106.9 Hz.

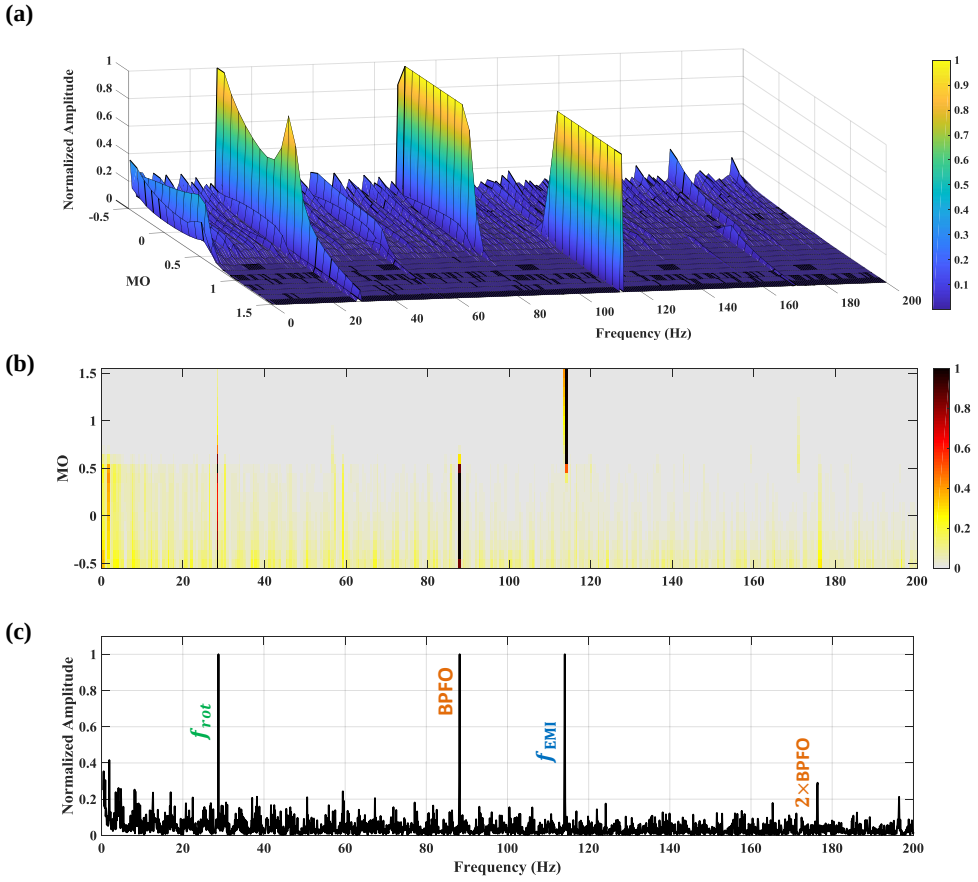

Figure 9: Case 2: SAM **(a)** 3D plot **(b)** Above view **(c)** View along the MO-axis (MSES)

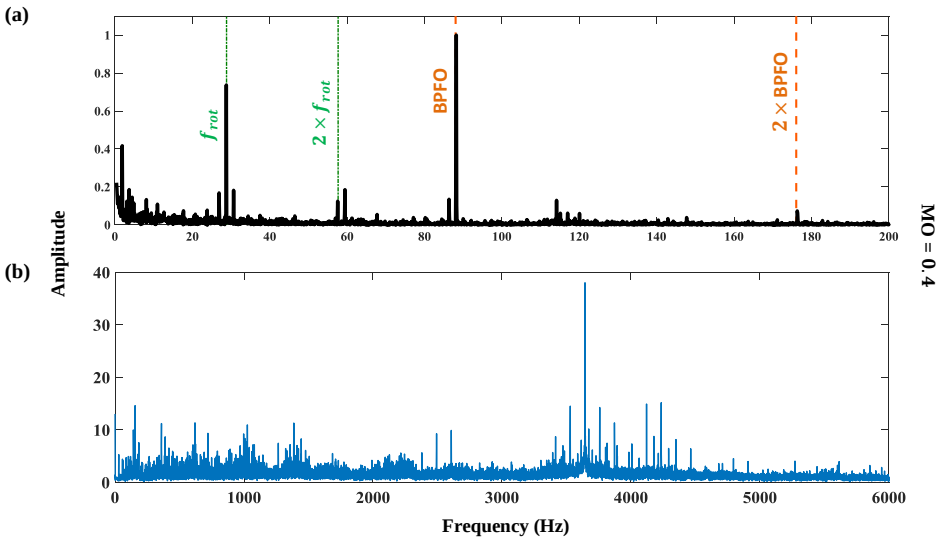

Figure 10: Case 2 **(a)** SES and **(b)** Magnitude spectrum of the modified signal for MO = 0.4

The bearing is partially diagnoseable by using the raw signal's SES and fast Kurtogram but CPW is not able to extract the defect signal [22].

The SAM is depicted in Fig. 11(a) where the BPFO is present for MO from 0.4 to 1.5. For MO < 0, the SES are almost flat and none of the cyclic frequencies has considerable magnitude. It can be concluded that, first, frequencies with very low magnitudes, which are magnified by the negative values of MO, are related to noise and not to the carrier frequencies of the bearing defect; therefore, noise has overwhelmed the modified signals. Second, the time signal contains some non-Gaussian and impulsive noise which dominates the modified time signals.

Fig. 11(b) shows the MSES which does not provide valuable diagnostic information like SAM. Since the SES for MO lower than 0 do not contain large peaks, when the SES are normalized between 0 and 1, most of the cyclic frequencies will have high values close to 1. As a result, SES for lower values of MO will mask other SES which may includes valuable diagnostic information. Thus, in this case, the SES for MO < 0.5 could be neglected. The MSES for

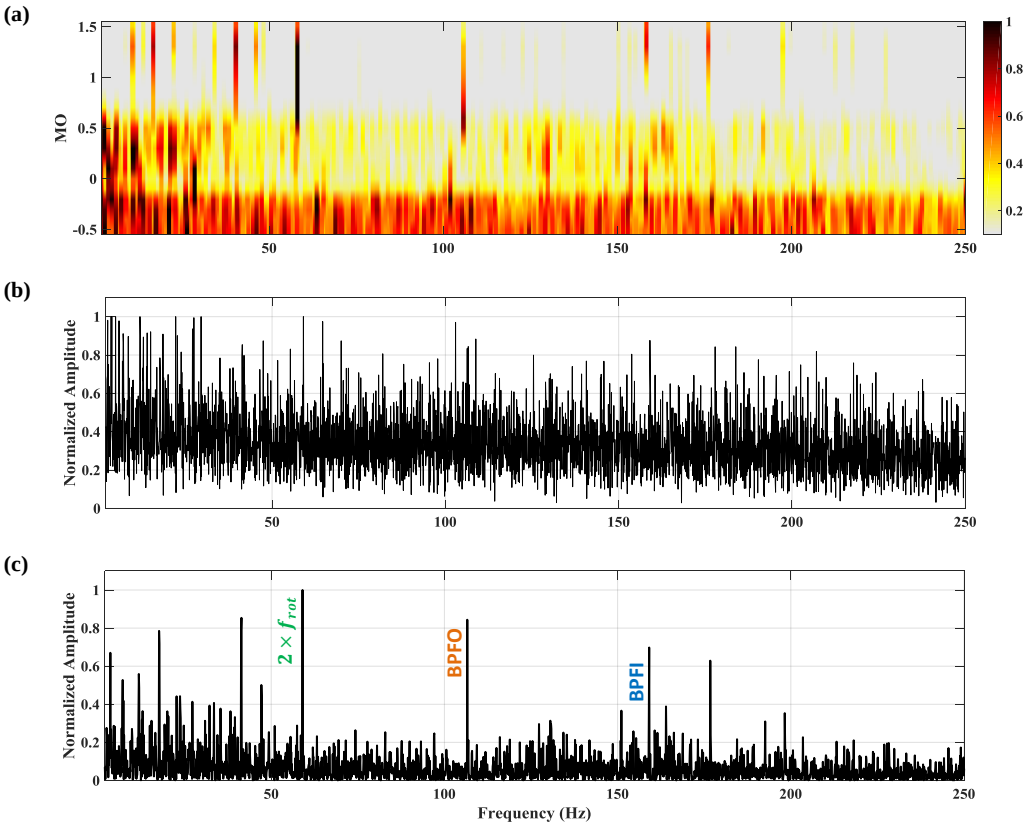

Figure 11: Case 3: 202 DE **(a)** SAM **(b)** MSES for $-0.5 \leqslant \text{MO} \leqslant 1.5$ **(c)** MSES for $0.5 \leqslant \text{MO} \leqslant 1.5$

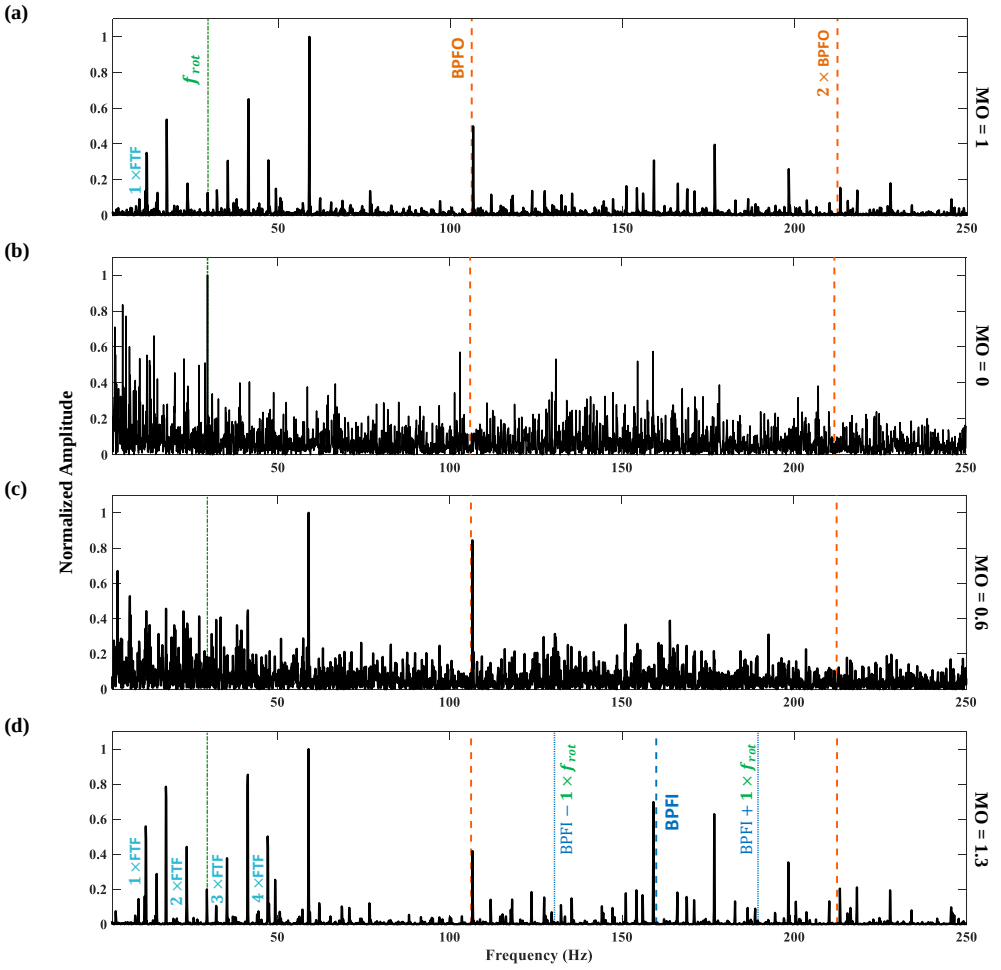

Figure 12: Case 3: SES of the modified signals for **(a)** MO = 1 (raw signal) **(b)** MO = 0 (CPW) **(c)** MO = 0.6 **(d)** MO = 1.3

$0.5 \leqslant \mathrm{MO} \leqslant 1.5$ is depicted in Fig. 11(c) and enhancement of the result, in comparison to Fig. 11(b), can be noticed evidently.

The SES of modified signals for different MO are shown in Fig. 12. SES of the raw signal in Fig. 12(a) reveals the defect frequency. For MO = 0, the SES is exhibited in Fig. 12(b). Since the value of MO is decreased, the SNR is increased and now the defect frequency is completely eliminated. As it was mentioned in section 3, this is one of the main problems related to CPW. The highest density of BPFO in the SAM (Fig. 11(a)) is achieved for MO = 0.6 and the corresponding SES can be seen in Fig. 12(c). Moreover, for this case the BPFI and Fundamental Train Frequency (FTF) are also present for MO from 0.5 to 1.5 and their highest values are achieved for MO = 1.3 (Fig. 12(d)).

4.4. Case 4

In what follows, record 199 BA with defected outer race with nominal BPFO = 105.2 Hz, will be studied. Based on the study in Ref. [22], the bearing is not diagnoseable by calculating SES of the raw and signal after CPW but fast Kurtogram is partially successful to reveal the defect frequency.

SAM is depicted in Fig. 13(a) where the BPFO and its second harmonic is visible for $0.1 \leqslant \mathrm{MO} \leqslant 0.5$, with the highest value assigned to MO = 0.3. MSES is presented in Fig. 13(b) and the BPFO is also detectable in this figure. However, the shaft frequency and its second harmonics, which are present for $\mathrm{MO} > 0$ in Fig. 13(a), are masked by high spectral lines in the low frequency range. Similar to previous case, this problem can be solved by considering only SESs with $\mathrm{MO} > 0$. The SAM obtained by replacing SES by LSES is shown in Fig. 13(c). The aforementioned problem in the low frequency range is solved but the log-envelope has worsened the SNR of the modified signals and now the BPFI is barely detectable in the SAM. Therefore, as it was discussed in section 3, it is advantageous to employ LSES only when a signal contains non-periodic and highly impulsive contents.

The SES of signals for MO equal to 1, 0 and 0.3 are displayed in Fig. 14. The SES of raw signal in Fig. 14(a) does not provide a positive diagnosis as

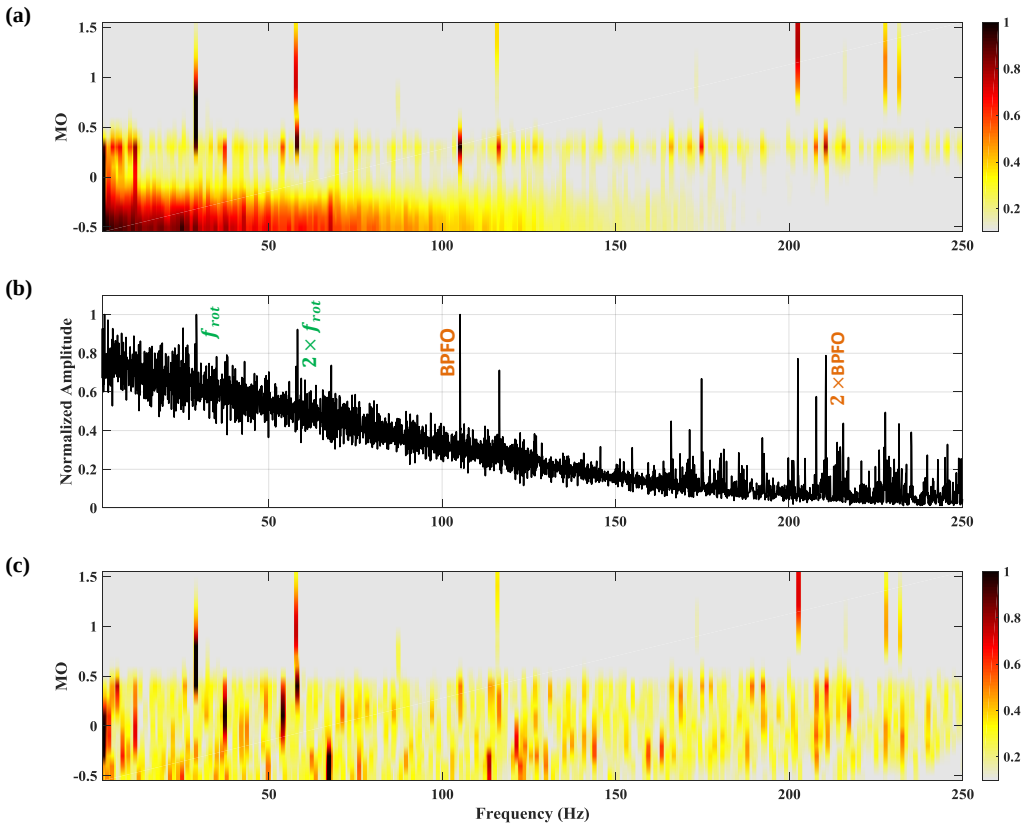

Figure 13: Case 4: 199 BA **(a)** SAM **(b)** MSES **(c)** log-SAM

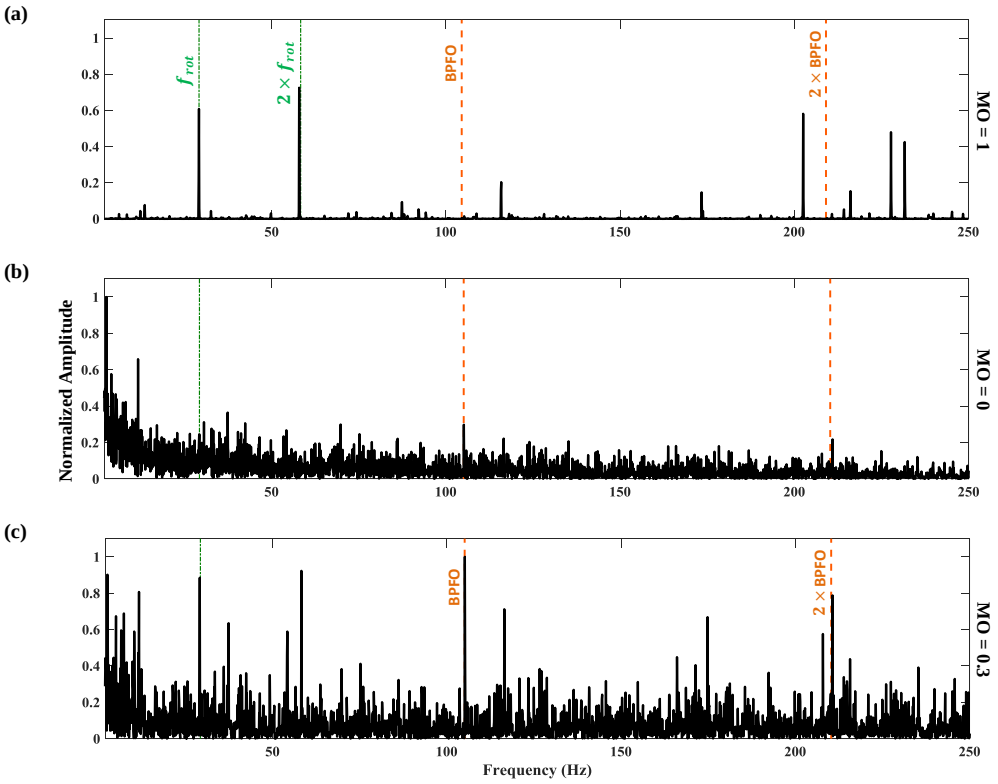

Figure 14: Case 4: SES of the modified signals for **(a)** MO = 1 (raw signal) **(b)** MO = 0 (CPW) **(c)** MO = 0.3

the amplitude of BPFO is insignificant and the shaft frequency and its second harmonics are the largest peaks in the SES. Fig. 14(b) demonstrates the SES of the signal after CPW. Small peaks can be detected at BPFO and $2 \times$ BPFO. This result is inconsistent with the diagnosis outcome of Ref. [22]. This is as a consequence of using SES^2 instead of SES in this paper.

For MO = 0.3, the SES of modified signal is presented in Fig. 14(c) where the BPFO is the dominant frequency. Consequently, the defect on the outer race is effectively detected by using the proposed method while two other approaches (SES of the raw and CPW signals) fail to highlight the bearing characteristic frequency.

For the sake of comparison, the fast Kurtogram and Autogram for this difficult case are depicted in Fig. 15(a) and (b). SESs of frequency bands selected by these methods are also presented in Fig. 15(c) and (d). The results show that both methods fail to diagnose the fault as the defect signal is weak (fast Kurtogrma failure is in contrast to the diagnosis report in Ref. [22]).

Upper Autogram is a more sensitive version of Autogram to repetitive transients in presence of noise. Combined Squared Envelope Spectrum (CSES) is also proposed in Ref. [5] to extract all the modulation sources in a signal. Combination of CSES and upper Autogram was found to be more effective for diagnosis of REBs [6]. The CSES by using upper Autogram is depicted in Fig. 16 and the BPFO can be hardly spotted for levels 3 to 5. Additionally, fast spectral correlation [11] is applied on this data. Fig. 17 displays the spectral coherence and Enhanced Envelope Spectrum (EES). The BPFO (105.2 Hz) is present in the EES but its amplitude is low. To enhance the result the signal should be filtered for the frequency band linked to the defect frequency and highlighted by spectral coherence. But this is challenging as the frequency band should be selected manually and, for this case, this range is hardly noticeable (see Fig. 17(a)).

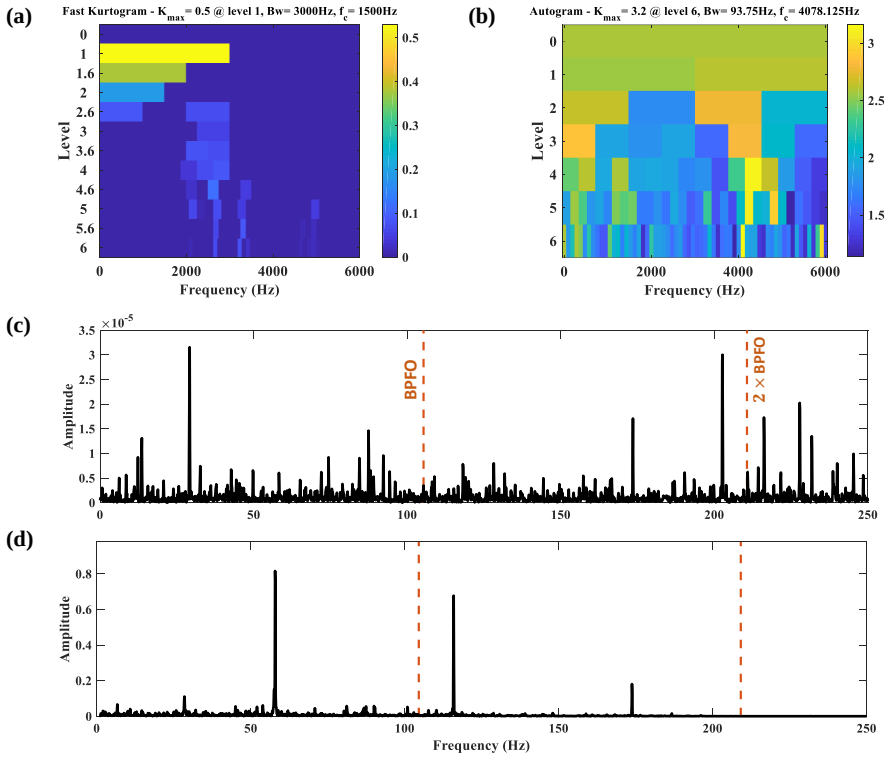

Figure 15: Case 4 **(a)** Fast Kurtogram **(b)** Autogram; squared envelope spectrum (SES) of the signal related to node with highest kurtosis in **(c)** Fast Kurtogram **(d)** Autogram

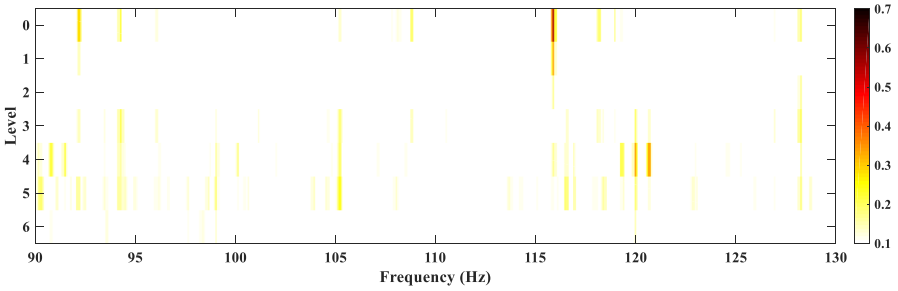

Figure 16: Case 4: Combined Squared Envelope Spectrum (CSES)

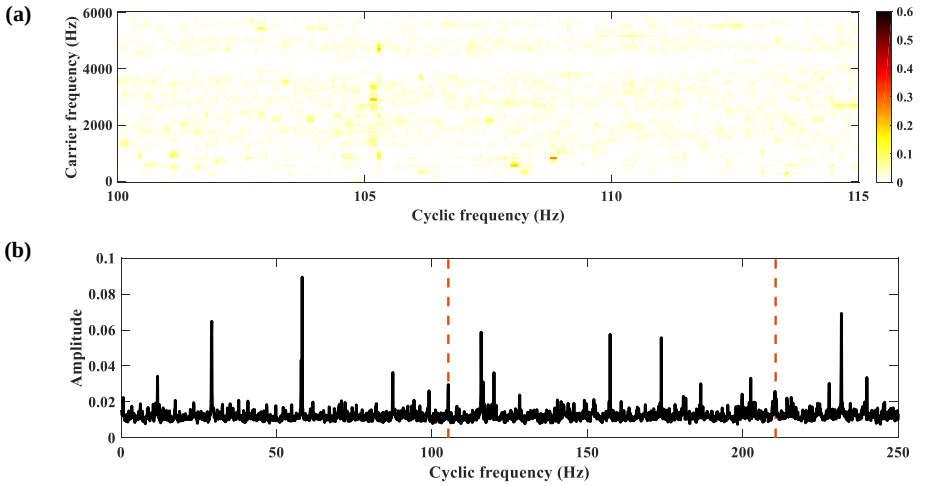

Figure 17: Case 4 **(a)** Fast Spectral Coherence **(b)** Enhanced Envelope Spectrum (EES)

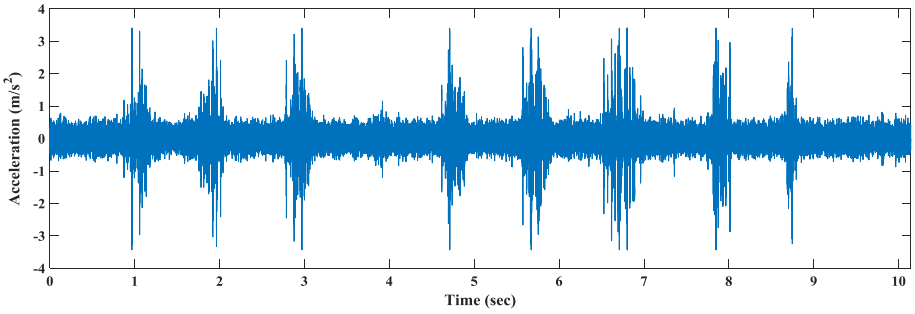

Figure 18: Case 5: The acceleration signal of record 229 DE

4.5. Case 5

In this part, data 229 DE related to a bearing with a reported defected ball is examined (BSF = 65 Hz). The depicted time signal in Fig. 18 is highly non-stationary and includes sizeable impulsive and random amplitude modulated contents. All the three approaches in Ref. [22] were not able to highlight the defect frequency and the bearing diagnosis was categorized as negative.

The SAM, by using SES, is presented in Fig. 19(a). A high density of the defect frequency and its continuity for a range of MO can be noticed, but the considerable values of normalized amplitudes for adjacent frequencies make the

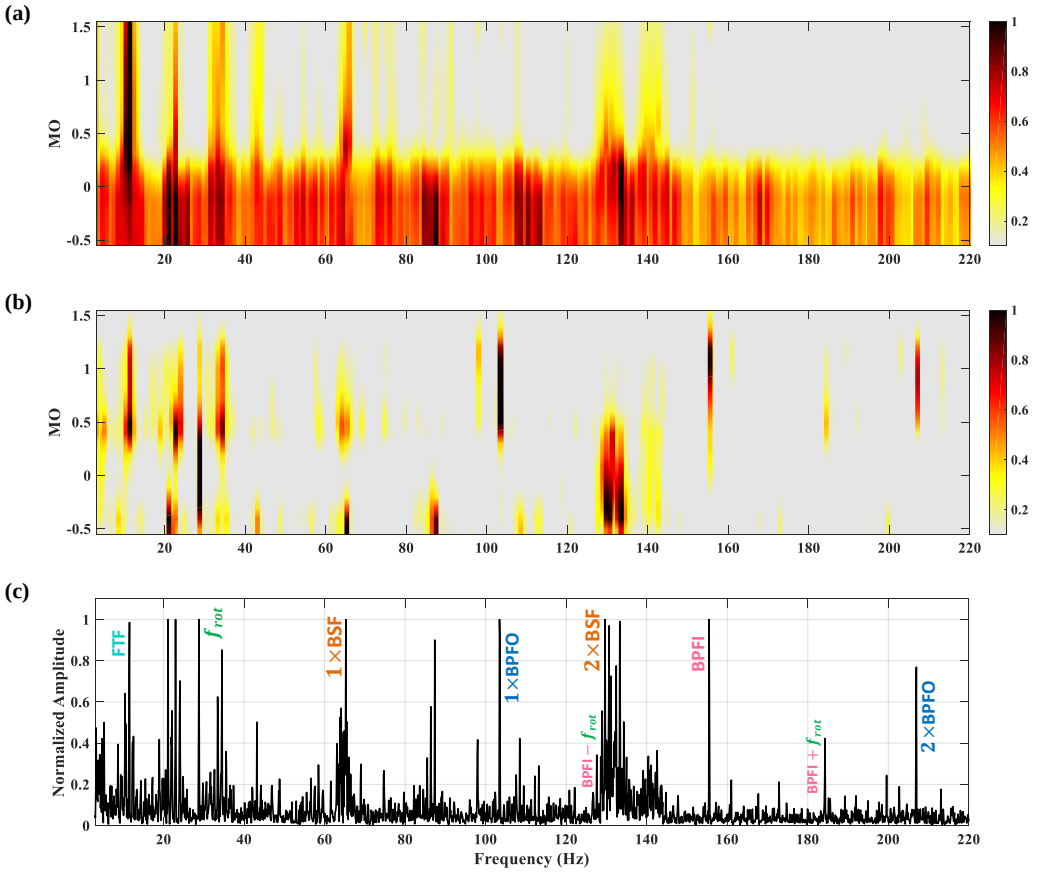

Figure 19: Case 5: 229 DE **(a)** SAM **(b)** log-SAM **(c)** log-MSES

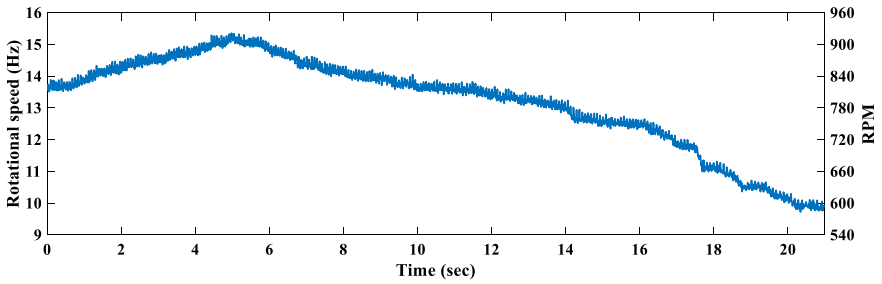

Figure 20: Case 6: Shaft rotational speed

diagnosis unreliable. The large pulses in the signal are responsible for the SES deterioration. To eliminate the destructive effect of these impulses, the SES is substituted by LSES. The result is demonstrated in Fig. 19(b) and now the FTF, BSF and 2×BSF can be clearly spotted. In addition, as a typical characteristic of rolling elements (balls) defects, the second (even) harmonic of the BSF (2×65 Hz) is larger than the first (odd) harmonic. This modulation frequency is present for $-0.5 \leqslant \mathrm{MO} \leqslant 0.5$ and it is the largest peak in the LSES for MO equal to -0.3 and -0.4. Moreover, the log-MSES is depicted in Fig. 19(c) which provides a comprehensive and valuable map of the signal's cyclic frequencies.

It is worth noting that the impulsive components of the signal are not completely random peaks and actually, they are impulse responses spaced at multiples of BSF. Also, it can be seen in Fig. 19(b) and (c) that the harmonics of BSF are smeared, which is a consequence of random amplitude modulation of the impulsive signal related to the ball defect.

Another interesting observation is that, in addition to the BSF, both the BPFI and BPFO, related to the same bearing (the latter with sidebands spaced at the shaft frequency), can also be detected in Fig. 19(c). Since these two frequencies are discrete, it indicates that their corresponding signals are related to the stationary part of the raw signal and, in contrast to the signal associated to the ball defect, their amplitudes are not randomly modulated.

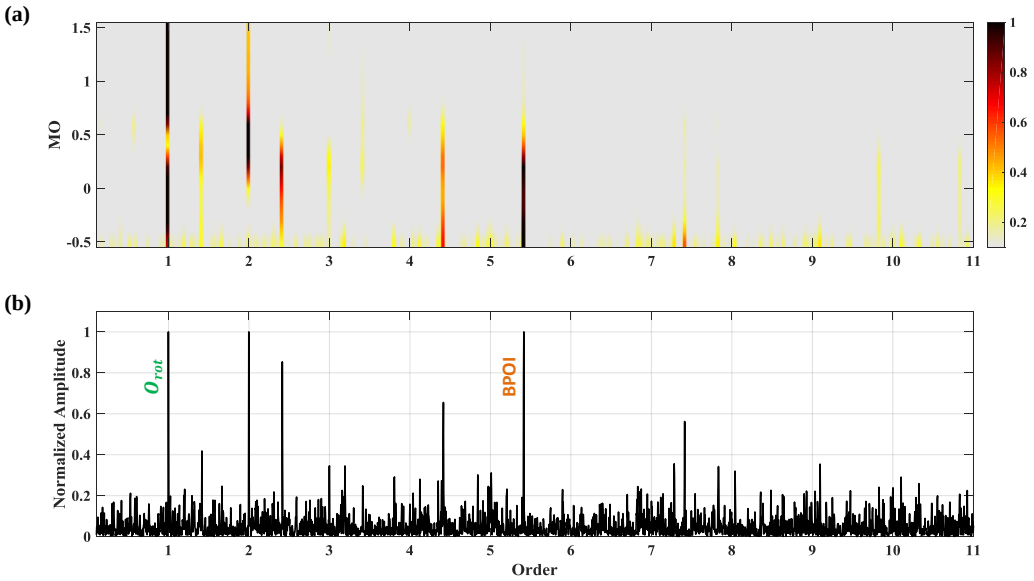

Figure 21: Case 6 **(a)** SAM **(b)** MSES

4.6. Case 6

For the last case the objective is to investigate the performance of the proposed method for diagnosis of a REB working under variable speed condition. In contrast to previously studied data in this paper, the time record is provided by Ref. [25]. Data are available for three different defective conditions: outer race, inner race and roller. Here, the signal related to the faulty inner race, which is the most difficult case, is examined. The acceleration signal is recorded for 21 seconds with sampling frequency 50 kHz and the shaft rotational speed varies in the range of 9.9 to 15.2 Hz (Fig. 20). The detailed explanation of the experiment and test rig are also given in Ref. [25].

Conventional bearings fault diagnosis techniques are developed by assuming constant shaft speed during the operations, i.e. the impulse generated by bearings faults are almost uniformly spaced in time. Accordingly, they are not directly applicable to variable speed conditions as the time interval between two successive impulses is a function of the shaft rotational speed. To be able to utilize these methods, a primary pre-processing step is required. For this

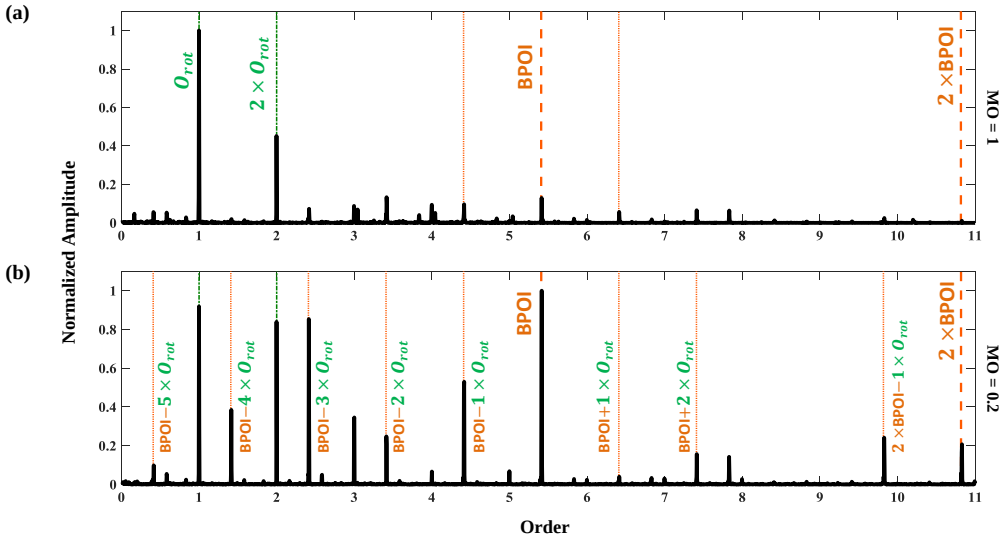

Figure 22: Case 6: SES of the modified signals for **(a)** MO = 1 (raw signal) **(b)** MO = 0.2

purpose, order tracking [1] is employed to map the recorded signal from time domain to angular domain, i.e. by using a tachometer signal the vibration signal is resampled at uniform increment of the shaft angular rotation.

The proposed method is applied on the resampled signal and the SAM and MSES (the x axis is harmonics or "orders" of shaft speed, O_{rot}) are exhibited in Fig. 21(a) and (b). The Ball Pass Order Inner race (BPOI) is clearly detectable for $-0.5 \leqslant \text{MO} \leqslant 0.7$ in Fig. 21(a) and for MO equal to -0.5, -0.4 and 0.2 the BPOI has its highest density and also it is the largest peak in the SES of the corresponding modified signals. In addition, it can be noticed that bearing defect signal is the only component of the signal since only BPOI, shaft order and sidebands around BPOI (spaced at the shaft order) are present.

The SES and magnitude spectrum of the raw signal (MO = 1) are depicted in Fig. 22(a) and Fig. 23(a) respectively. Although the BPOI is present, the amplitude is low and harmonics of the shaft order dominates the SES. Similarly, the discrete frequencies in the magnitude spectrum are related to the harmonics of the shaft order.

The SES and magnitude spectrum of the modified signal for MO = 0.2 (the

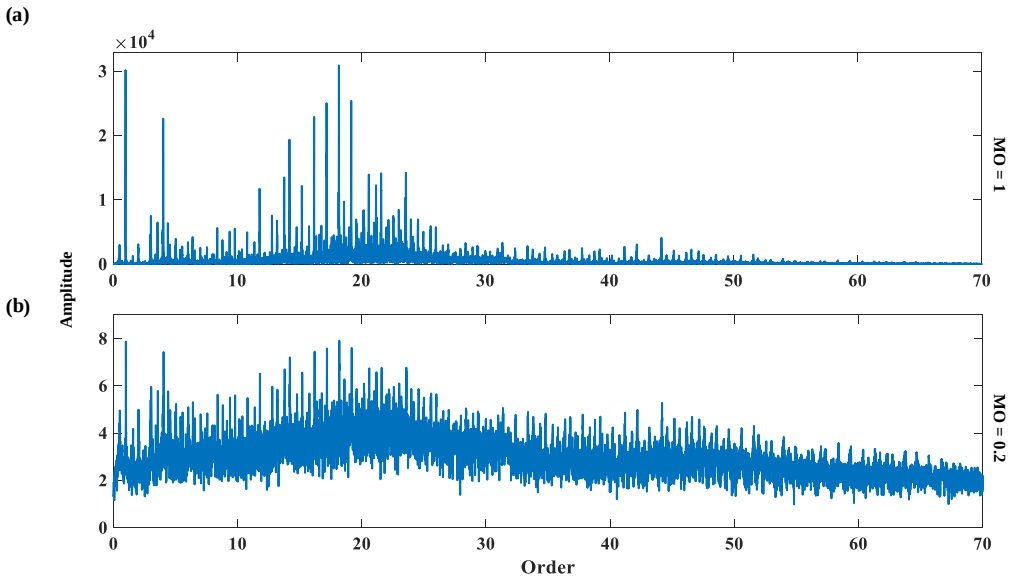

Figure 23: Case 6: Magnitude spectra of the modified signals for **(a)** MO = 1 (raw signal) **(b)** MO = 0.2

maximum value of MO in which the BPOI is equal to one) are presented in Fig. 22(b) and Fig. 23(b). In contrast to the SES of raw signal, first and second harmonics of the BPOI along with their sidebands are evidently detectable. The reason behind the improvement can be recognized by comparing Fig. 23(a) and (b). As the MO is smaller than one, the effect of frequencies with sizeable magnitude related to the shaft frequency is decreased. On the other hand, the influence of the resonance frequencies containing the information of the bearing defect signal is boosted. It is interesting that this worthwhile diagnostic outcome is achieved by reducing the effect of discrete frequencies and not by removing them.

4.7. Summarization of SAM performance

The performance of SAM and the comparison with other techniques are summarize as follows:

- The method is simple and fast. For instance, for cases 1, 3 and 6 with

120000, 480000 and 1 million samples the computational time by a laptop with a core i7 processor was 1.3, 4.3 and 7 seconds respectively

- No parameter should be set in advance based on the characteristic of system
- CPW and SES of raw signals are two subclasses of SAM. As a result, the performance of SAM is better than each of them individually. Case 4 shows that SAM is also successful while both CPW and SES of raw signal fail, so the overall performance of SAM is better than these two methods combined
- In contrast to the methods which try to find a proper frequency band for demodulation such as fast Kurtogram or Protrugram:
  - There is no need to multiple filtering process and further computation of any indicator such as kurtosis to find the most impulsive band
  - All modulation sources are revealed. Therefore, presence of multiple defects might not be neglected
- Log-SAM provides a robust technique for diagnosis of REBs in presence of highly impulsive (non-Gaussian) noise
- The completely original contribution of SAM is that different components of a signal are automatically separated based on their energy levels. These components may also be distributed in different frequency bands. Therefore, SAM could be considered as a complementary to methods that decompose a signal for different center frequencies and frequency bands
- Although CSA includes a powerful family of approaches to extract modulation frequencies, unlike SAM the proper frequency band for demodulation should be set manually which is sometimes challenging

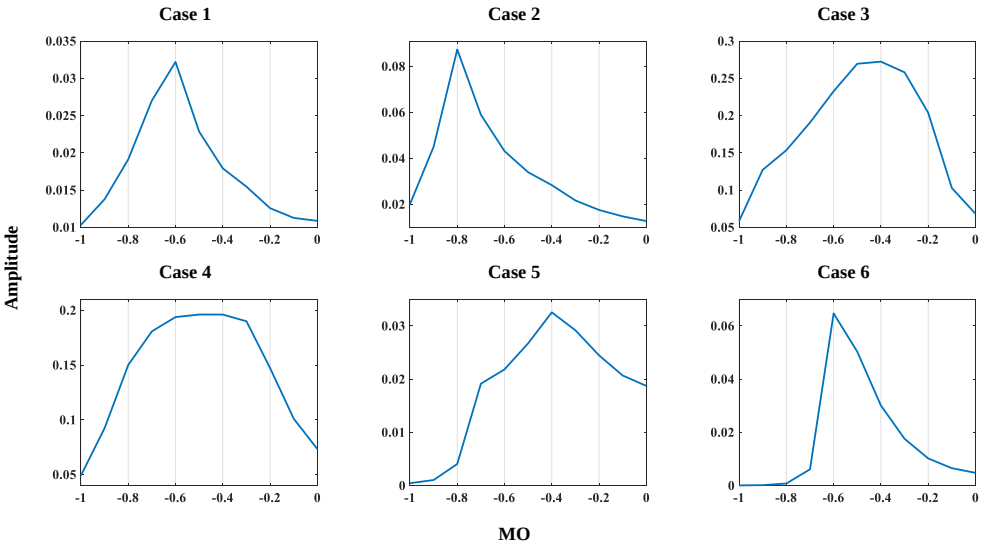

Figure 24: Mean-MO plots for six cases

4.8. Effect of the selected range for MO

By comparing the results for these six cases, it can be noticed that the most valuable diagnostic information are revealed by $0 \leqslant \mathrm{MO} \leqslant 1$.

In addition, as it was discussed $\mathrm{MO} > 1$ boost the more powerful parts of a signal which are represented by larger peaks in its magnitude spectrum. Those are not usually associated to bearings signals which are often weaker than signals generated by other parts of a machinery. Only for case 3 the $1 < \mathrm{MO} \leqslant 1.5$ was able to enhance the bearing signal. For $\mathrm{MO} > 1.5$ (results are not depicted here) the defect frequencies are not maximized in neither case. It was discussed that $\mathrm{MO} > 1$ does not generate any false peak in the SESs; therefore, there is no actual limitation to select even a very large value. But as it is shown by the results in this section, using $\mathrm{MO} > 1$ does not provide significant practical benefit due to lower energy level of bearings signals. As a result, choosing 1.5 as a not very large upper limit is suitable for practical purpose.

Negative MOs larger than -0.5 have enhanced the weaker defect signals for cases 1, 2, 5 and 6 as the effect of magnitude spectra are reversed. On the

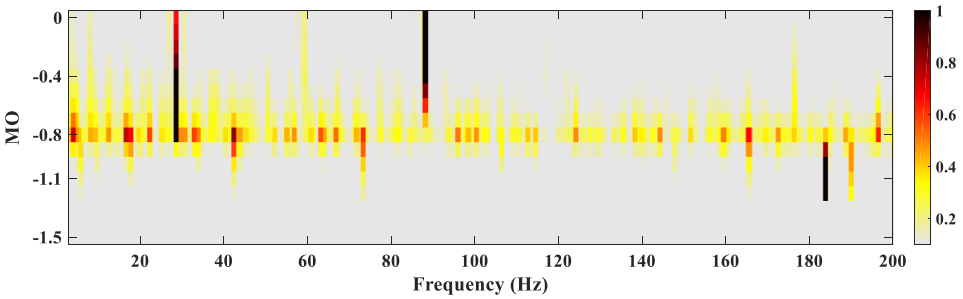

Figure 25: SAM for case 2 ($-1.5 \leqslant \mathrm{MO} \leqslant 0$)

other hand, for case 4 and 5 background noise has dominated the modified signals. Although for some low values of MO the SESs do not include any useful information, this does not hinder the effectiveness of SAM at all, as it presents the SESs for all values of MOs. Moreover, no false peak has been produced by negative MOs as low as -0.5 for either case. To further investigate this matter, mean-MO plots are depicted in Fig. 24 for all cases. Bases on the proposed technique to find the lower limit of MO in section 3, the lowest acceptable values of MO (value of MO related to the maximum in a mean-MO plot) for cases 1 to 6 are -0.6, -0.8, -0.4, -0.5, -0.4 and -0.6 respectively. This shows that selection of -0.5 was practically effectual for REBs diagnosis. Also, SAM for case 2, for a different range of MO ($-1.5 \leqslant \mathrm{MO} \leqslant 0$) is shown in Fig. 25 to exemplify the emergence of false peaks for $\mathrm{MO} < -0.8$.

To conclude, all things considered, the proposed $-0.5 \leqslant \mathrm{MO} \leqslant 1.5$ was shown to be an appropriate range for practical applications. Wider arbitrary ranges could be also employed but care must be taken for selection of the lower limit.

5. Conclusion

This paper introduces a new empirical method for rolling element bearings diagnosis which employs a nonlinear filtering process to separate components of a signal. Since it is achieved automatically, it offers a shortcut to traditional techniques which finds a single effective frequency band for demodulation.

The proposed method, the Spectral Amplitude Modulation (SAM), is inspired by and relies on the fact that the phase of a signal contains more information than its magnitude. First, magnitude and phase of a signal are computed by mean of Fourier transform. Afterwards, a series of signals, which are called modified signals, are reconstructed by giving different weights (the Magnitude Order (MO)) to the magnitude of the original signal, while keeping the same phase as the original signal. Then, the squared envelope spectrum of each modified signal is calculated, and the outcomes are normalized and presented in a three-dimensional graph. Moreover, LSES (Fourier transform of logarithm of squared envelope signals) has been employed as a more robust alternative to SES in presence of highly impulsive noise. Furthermore, physical and mathematical interpretation of the method has been provided. It has been detailed that, by setting different values of MO, components of a signal with different energy levels are extracted. From a mathematical point of view, it has been demonstrated how the effect of modulation on the phase and magnitude of a signal is responsible for the amplitude of each spectral line in the SES.

Finally, the performance of SAM has been validated and examined thoroughly on six real cases collected from two different experimental setups. In presence of multiple defects and also strong masking exogenous signals such as Electro Magnetic Interference (EMI), the superior performance of SAM to highlights all the modulation sources has been presented. In addition, it was clarified that SES of raw and cepstrum pre-whited signals may overlook valuable diagnostic information. The proposed method combines the advantages of these two techniques as it puts them in a unified and comprehensive concept, in other words, these two techniques can be considered as two particular subsets of SAM. Moreover, employing LSES provides an additional and effective tool to cope with highly contaminated signals with impulsive noise which is common in industrial application. The last case shows the high potential of SAM for diagnosis of bearing when rotational speed of shaft is variable.

External resource

The algorithm coded in Matlab is available on the following address: www.?????.com

References

[1] R. B. Randall, J. Antoni, Rolling element bearing diagnostics–A tutorial, Mechanical Systems and Signal Processing 25 (2) (2011) 485–520 (2011).

[2] M. S. Darlow, R. H. Badgley, G. Hogg, Application of high-frequency resonance techniques for bearing diagnostics in helicopter gearboxes, Tech. rep., Mechanical Technology Inc Latham NY (1974).

[3] J. Antoni, Fast computation of the kurtogram for the detection of transient faults, Mechanical Systems and Signal Processing 21 (1) (2007) 108–124 (2007).

[4] T. Barszcz, A. JabŁoński, A novel method for the optimal band selection for vibration signal demodulation and comparison with the kurtogram, Mechanical Systems and Signal Processing 25 (1) (2011) 431–451 (2011).

[5] A. Moshrefzadeh, A. Fasana, The Autogram: An effective approach for selecting the optimal demodulation band in rolling element bearings diagnosis, Mechanical Systems and Signal Processing 105 (2018) 294–318 (2018).

[6] A. Moshrefzadeh, A. Fasana, L. Garibaldi, Analysis of Autogram performance for rolling element bearing diagnosis by using different data sets, Advances in Condition Monitoring of Machinery in Non-Stationary Operations 15 (2019).

[7] J. Antoni, Cyclic spectral analysis in practice, Mechanical Systems and Signal Processing 21 (2) (2007) 597–630 (2007).

[8] W. Gardner, Measurement of spectral correlation, IEEE Transactions on Acoustics, Speech, and Signal Processing 34 (5) (1986) 1111–1123 (1986).

[9] P. Borghesani, The envelope-based cyclic periodogram, Mechanical Systems and Signal Processing 58 (2015) 245–270 (2015).

[10] J. Antoni, Cyclostationarity by examples, Mechanical Systems and Signal Processing 23 (4) (2009) 987–1036 (2009).

[11] J. Antoni, G. Xin, N. Hamzaoui, Fast computation of the spectral correlation, Mechanical Systems and Signal Processing 92 (2017) 248–277 (2017).

[12] B. Bogert, M. Healy, J. Tukey, The frequency analysis of time series for echoes: cepstrum, pseudo-auto covariance, cross-cepstrum, and shaft cracking, in: Proceedings of the Symposium on Time Series Analysis (M. Rosenblatt, Ed), 1963, pp. 209–243 (1963).

[13] R. B. Randall, A history of cepstrum analysis and its application to mechanical problems, Mechanical Systems and Signal Processing 97 (2017) 3–19 (2017).

[14] R. B. Randall, N. Sawalhi, A new method for separating discrete components from a signal, Journal of Sound and Vibration 45 (5) (2011) 6 (2011).

[15] P. Borghesani, P. Pennacchi, R. B. Randall, N. Sawalhi, R. Ricci, Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions, Mechanical Systems and Signal Processing 36 (2) (2013) 370–384 (2013).

[16] H. Konstantin-Hansen, H. Herlufsen, Envelope and cepstrum analyses for machinery fault identification, Journal of Sound and Vibration 44 (5) (2010) 10 (2010).

[17] J. Antoni, Cyclic spectral analysis of rolling-element bearing signals: facts and fictions, Journal of Sound and Vibration 304 (3) (2007) 497–529 (2007).

[18] J. Antoni, R. B. Randall, Differential diagnosis of gear and bearing faults, Journal of Vibration and Acoustics 124 (2) (2002) 165–171 (2002).

[19] A. V. Oppenheim, J. S. Lim, G. Kopec, S. Pohlig, Phase in speech and pictures, in: Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'79., Vol. 4, IEEE, 1979, pp. 632–637 (1979).

[20] A. V. Oppenheim, J. S. Lim, The importance of phase in signals, Proceedings of the IEEE 69 (5) (1981) 529–541 (1981).

[21] C. Peeters, P. Guillaume, J. Helsen, A comparison of cepstral editing methods as signal pre-processing techniques for vibration-based bearing fault detection, Mechanical Systems and Signal Processing 91 (2017) 354–381 (2017).

[22] W. A. Smith, R. B. Randall, Rolling element bearing diagnostics using the case western reserve university data: A benchmark study, Mechanical Systems and Signal Processing 64 (2015) 100–131 (2015).

[23] P. Borghesani, M. R. Shahriar, Cyclostationary analysis with logarithmic variance stabilisation, Mechanical Systems and Signal Processing 70 (2016) 51–72 (2016).

[24] Case western reserve university bearing data center website.
URL http://csegroups.case.edu/bearingdatacenter/home

[25] C. Mishra, A. Samantaray, G. Chakraborty, Rolling element bearing defect diagnosis under variable speed operation through angle synchronous averaging of wavelet de-noised estimate, Mechanical Systems and Signal Processing 72 (2016) 206–222 (2016).

[26] W. A. Smith, Z. Fan, Z. Peng, H. Li, R. B. Randall, Optimised spectral kurtosis for bearing diagnostics under electromagnetic interference, Mechanical Systems and Signal Processing 75 (2016) 371–394 (2016).