Available online at www.sciencedirect.com
ScienceDirect
Journal of Sound and Vibration 304 (2007) 497–529
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions

J. Antoni

Laboratoire Roberval de Mécanique, Centre de Recherche de Royallieu, BP 20529 60205 Compiègne Cédex, France

Received 25 January 2006; received in revised form 22 February 2007; accepted 27 February 2007
Available online 7 May 2007

Abstract

Rolling-element bearing vibrations are random cyclostationary. This property is so symptomatic when an incipient fault develops that it can be exploited for diagnostics. This paper discusses which cyclic spectral tools should be considered for that purpose. Specifically, it demonstrates the optimality of the cyclic coherence, which can not only evidence the presence of a fault in high levels of background noise, but can also return a relative measure of its severity. The estimation issue of the cyclic coherence is addressed in detail, as well as its use in a statistical test, and sub-optimal simplifications. Eventually, it is shown that the familiar squared-envelope spectrum happens to be a special case of the cyclic coherence with very similar diagnostic capabilities.

1. Introduction

1.1. Context

Because they are fragile mechanical parts, rolling-element bearings are the focus of a special attention in condition monitoring programs. The experience gained over the years in monitoring rolling-element bearings has led to the elaboration of specific technologies, among which vibration-based practices enjoy a particular acknowledgement due to their non-invasive nature and their high reactivity to incipient faults. Vibration monitoring essentially consists in analysing some relevant indicators that may be either scalar (RMS value, kurtosis, or some complex combination of both), vectorial (signal spectrum, cepstrum or envelope spectrum), or even matricial (time–frequency or waterfall representations) [1–4]. In general the diagnosis capability of such techniques is proportional to their complexity. We are presently arriving at a point where modern computing capacities allow the use of very sophisticated monitoring techniques that reach the upper-right corner of the complexity/optimality curve. In this context, the exploitation of cyclostationarity has recently been proved extremely fruitful to conceive high accuracy diagnostics tools.

E-mail address: Jerome.Antoni@utc.fr.

0022-460X/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2007.02.029

1.2. Exploiting cyclostationarity in bearing diagnostics

Cyclostationarity encompasses a subclass of non-stationary signals which exhibit some cyclical behaviour. A typical example of a cyclostationary signal is a random noise that is amplitude modulated by a periodic function. A more versatile example is where different periodic modulations envelope each frequency component of the random noise. Cyclostationarity has been shown to ideally fit the property of many rotating and reciprocating machine vibrations, due to the inherent periodic modulations that these sustain during operation [5]. The recognition that a vibration signal is cyclostationary affords much more information than the usual and simplistic assumption of stationarity, since it provides the tools to simultaneously analyse the content of a signal (e.g. spectral content) along with the characterisation of how this content evolves periodically in time (e.g. within the machine cycle).

The cyclostationarity of rolling-element bearing vibrations was implicitly admitted in Refs. [6,7], and it was first formally demonstrated in Ref. [8]. More will be said on the subject in Section 2, but the fundamental observation to keep in mind from now onward is that cyclostationarity is symptomatic to the presence of faults, owing to the occurrence of repetitive shocks when a defect impacts a rolling surface (a series of repetitive shocks may be seen as a signal periodically amplitude modulated in time). The rate of repetition of these shocks—formally coined the *cyclic frequency*—then points out the origin of the fault (i.e. the rotating part of the bearing on which the fault is located), and the cyclostationarity "intensity" (as measured by some spectral quantities to be introduced later) may serve to indicate its severity.

1.3. Objectives of the paper

The diagnostics potentialities allowed by the cyclostationary approach have actually been foreseen in a number of precursory works concerned either explicitly with rolling-element bearings or with related mechanisms [5–10]. These works all proposed to exploit cyclostationarity by means of some "cyclic" tools such as the so-called cyclic spectrum or spectral correlation which are matrix indicators, or some simplified scalar version of them such as the "degree-of-cyclostationarity index". Although such indicators are very good at characterising the cyclostationary nature of a signal, their optimality for diagnostics purposes has not been proved so far, and the statistical questions associated with their estimation has been most often eluded. It is the ambition of this paper to fill in these gaps. In particular, we address such simple questions as

(1) which cyclic spectral quantity[1] turns out to be optimal for diagnostics purposes?
(2) how can such a quantity be efficiently computed from finite length vibration measurements?
(3) which statistical test should be used in accordance with it in an effort towards automated diagnostics?

The answers to these questions are progressively developed through Sections 2–5, and will eventually lead to the proposal of new methodologies for cyclostationarity-based diagnostics of rolling-element bearings. These methodologies will be extensively illustrated on actual bearing signals in Section 6.

Let us finally emphasise that, in order to provide an unified presentation of the subject, this paper necessarily involves a number of results that have been published in earlier works; however, we have deemed useful to detail them here once again, in particular to correct some misconceptions that seem to have hold out in the specialised literature.

[1]It is admitted in this paper, as is often the case in vibration analysis, that *spectral* quantities are those of interest as opposed to dual *temporal* quantities such as moments or cumulants. Indeed the formers enjoy an obvious physical interpretation that is usually missing in the latters.

2. Rolling-element bearing vibrations are cyclostationary

Before we proceed to answer these questions, we first need to give a detailed review of the cyclostationary modelling of rolling-element bearing vibrations.[2] Much of the material presented in this section is borrowed from Refs. [8,10,13].

2.1. Vibration models

Incipient faults in rolling-element bearings are usually the consequence of a local loss of material (pitting, spalling, corrosion, rubbing, contamination) on a matting surface (inner/outer race, rolling elements). When a rolling surface contacts the fault, this produces a short duration impulse which excites some structural resonance of the bearing or of the vibration transducer itself.[3] The repetition of these impacts when the bearing is operating results in a series of impulse responses whose temporal spacing depends on the type of fault and on the geometry of the bearing. Table A.1 in the Appendix gives some typical fault frequencies, from which a diagnosis can be carried out.

On top of that, the series of impulse responses produced by an incipient fault are possibly amplitude modulated due to the passing of the fault into and out of the load zone. Typically, for a stationary outer race and in the presence of a radial load, an outer race fault would experience an uniform amplitude modulation, an inner race a periodic amplitude modulation at the period of the inner race rotation, and a rolling-element fault a periodic amplitude modulation at the period of the cage rotation. Those modulation frequencies are also passing reported in Table A.1 of the Appendix—see e.g. Ref. [4] for a complete discussion on that topic.

The above observations are well-known and have led in the past to the proposal of a simple harmonic[4] model for the vibrations produced by single localised faults. Namely, let $h(t)$ be the impulse response to a single impact as measured by the sensor, $q(t) = q(t + P)$ the periodic modulation of period P due to the load distribution,[5] T the inter-arrival time between two consecutive impacts on the fault; then the vibration signal $x(t)$ was modelled as [14]:

$$x(t) = \sum_{i=-\infty}^{+\infty} h(t - iT)q(iT) + n(t), \tag{1}$$

where index i denotes the occurrence of the ith impact and $n(t)$ accounts for an additive background noise that embodies all other vibration sources. Clearly, the power spectral density (signal-units^2/Hz) of signal (1) gives rise to a *discrete* spectrum

$$S_x(f) = \frac{1}{T}|H(f)|^2 \sum_{k,l=-\infty}^{+\infty} q_l \delta\left(f - \frac{k}{T} - \frac{l}{P}\right) + S_n(f), \tag{2}$$

wherein q_l denotes the lth Fourier coefficient of the modulating function $q(t)$, $\delta(f)$ is the Dirac delta pulse and $S_n(f)$ is the power spectral density of the noise. This model predicts that classical spectral analysis could serve the purpose of diagnostics simply by searching for the presence of those symptomatic harmonics located at the fault frequency $f = 1/T$ and multiples $f = k/T$, $k = \pm 1, \pm 2, \ldots$, and their possible sidebands at $f = k/T + l/P$, $l = \pm 1, \pm 2, \ldots$ due to load modulations. Indeed, the spectral pattern $\sum_{k,l=-\infty}^{+\infty} q_l \delta(f - k/T - l/P)$ uniquely defines the *fault signature*.

[2]The objective of this section is not to give a review on the theory of cyclostationary processes. Readers interested by the theoretical aspects of cyclostationarity are invited to consult, e.g., Refs. [11,12].

[3]Bearings vibrations are typically measured with an accelerometer in a frequency range that includes the resonance frequency of the sensor itself in order to turn it into a shock-sensing device.

[4]A *harmonic* signal is understood here as resulting from a summation of sinusoidal components. Note that a harmonic signal may not be periodic in the general case, although it accepts a Fourier series expansion. This is actually the case with the signal generated by model (1) wherein periods T and P are possibly incommensurable—i.e. their ratio is not a rational. Harmonic signals are also coined *pseudo* (or *quasi*) periodic in the mathematical literature.

[5]The modulating function $q(t)$ may also account for the periodic changes in the impulse response as the distance and orientation of the impacts moves towards and backwards the sensor, as well as for possible bearing unbalance or misalignment.

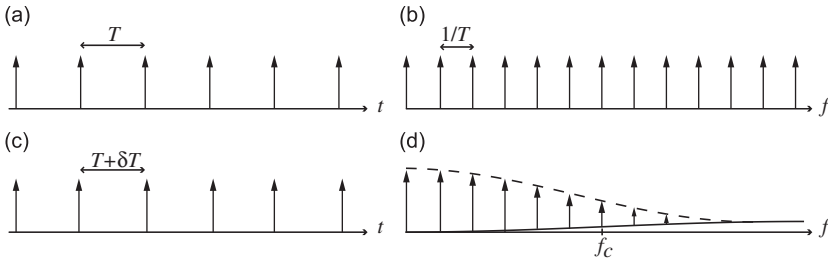

Fig. 1. Illustration as how a slight random fluctuation on the (inter-)arrival times of the impacts destroys the discrete structure in the frequency domain: an ideal train of Dirac pulses (impact forces) in the time-domain (a) produces another train of Dirac pulses in the frequency domain (b). The same train of Dirac pulses with random fluctuations (c) has a spectrum (d) where Dirac pulses rapidly vanish with frequency and turn into a continuous baseline. The same phenomenon happens as well with respect to random fluctuations in the magnitude of the impacts. The cut-off frequency f_c is as defined in Eq. (7).

Unfortunately, it appears that this naive spectral analysis is often prone to failure when applied to actual rolling-element bearing vibration signals. This is because the harmonic model (1), despite being pedagogically appealing, is generally wrong. First, the impacts on the fault do not occur exactly periodically because of necessary random slips,[6] possible speed fluctuations, and variations of the axial to radial load ratio. Second, the magnitude of the impulses are likely to experience random fluctuations as well due to the not exactly reproducible microscopic conditions when the fault impacts a rolling surface. Overall, very slight random fluctuations can completely destroy the harmonic structure of model (1), and result in a vibration signal that is essentially *random* in the frequency range of interest—see Fig. 1. To see this, let us switch to the following more realistic model for the vibration signal produced by a faulty bearing [8,10,15,16]:

$$x(t) = \sum_{i=-\infty}^{+\infty} h(t - iT - \tau_i)q(iT)A_i + n(t), \tag{3}$$

where τ_i and A_i accounts for the uncertainties on the arrival time (jitters) and on the magnitude of the ith impact, respectively. From the preceding discussion, $\{\tau_i\}_{-\infty}^{\infty}$ and $\{A_i\}_{-\infty}^{\infty}$ are modelled as two random sequences with expected values $E\{\tau_i\} = 0$ and $E\{A_i\} = 1$ (without loss of generality since the average magnitude is assigned to $q(iT)$). For simplicity, we will further assume that $\{\tau_i\}_{-\infty}^{\infty}$ and $\{A_i\}_{-\infty}^{\infty}$ are mutually independent, white,[7] stationary sequences, such that

$$\begin{aligned} E\{\tau_i\tau_j\} &= \delta_{ij}\sigma_\tau^2, \\ E\{A_i^2\} &= 1 + \delta_{ij}\sigma_A^2 \end{aligned} \tag{4}$$

with σ_τ and σ_A the standard deviations, and δ_{ij} the Kronecker symbol. Although these assumptions are somewhat idealised, they will allow us to gain sufficient insight into the analysed phenomenon from simple calculations. Indeed, model (3) and its implications have been thoroughly studied in Refs. [8,10,13]. Our main finding in those references was that even in the presence of very slight random fluctuations—i.e. typically on the order of a few percents—model (3) produces a harmonic structure that rapidly turns into a random signal. Precisely, the so-generated vibration signal $x(t)$ happens to expand as a mixture

$$x(t) = x_H(t) + x_R(t) \tag{5}$$

[6]The contact angle is not the same inside and outside the load zone so that rolling elements that are diametrically opposed w.r.t the radial load tend to roll at different speeds in spite of the cage trying to force them to maintain a similar speed.

[7]It was shown in Ref. [13] that having uncorrelated (white) inter-arrival time differences $\tau_{i+1} - \tau_i$ is physically more realistic than imposing uncorrelated (white) jitters τ_i on the arrival times, so that $E\{(\tau_{i+1} - \tau_i)(\tau_{j+1} - \tau_j)\} = \delta_{ij}\sigma_\tau^2$ should really be the correct assumption to use in numerical simulations. However when it comes to theory this leads to extra complications while not fundamentally changing the resulting properties of the vibration signal. This is why we will stick to assumption (4) in our theoretical analysis.

of:

(1) *a weak harmonic component $x_H(t)$ located in the lower-frequency range, and*
(2) *a dominating random cyclostationary component $x_R(t)$ located in the higher-frequency range.*

We now address the spectral characteristics of these two components one by one, in more details.

2.2. Assessment of the harmonic contribution

The harmonic contribution $x_H(t)$ in signal $x(t)$ has a power spectral density (signal-units^2/Hz) given by [8]

$$S_{x_H}(f) = \frac{1}{T}|H(f)|^2|\Phi(f)|^2 \sum_{k,l=-\infty}^{+\infty} q_l \delta\left(f - \frac{k}{T} - \frac{l}{P}\right), \tag{6}$$

where $\Phi(f)$ stands for the Fourier transform of the probability density function $\phi(\tau)$ of the random variable τ_i. Note that as expected, this is a *discrete spectrum* very similar to Eq. (2). Please keep in mind that it is the pattern of spectral lines $\sum_{k,l=-\infty}^{+\infty} q_l \delta(f - k/T - l/P)$ that carries the diagnostic information and that it assigns uniquely a distinctive spectral signature to each bearing fault. At least this is the idea behind the ideal model (1) *without* random fluctuations. Now the difference between the power spectral densities (2) and (6) is the presence of the weighting function $|\Phi(f)|^2$. Unfortunately this weighting is not innocuous since $\Phi(f)$ is the transfer function of a low-pass filter—as necessary for the Fourier transform of a probability density function. Therefore, the larger the random fluctuations—i.e. the standard deviation probability density function $\phi(\tau)$—the shorter the bandwidth of $\Phi(f)$. More precisely, assuming a Gaussian[8] probability density function $\phi(\tau)$ with standard deviation σ_τ, the corresponding filter has a 3 dB cut-off frequency given by

$$f_c \simeq \frac{0.19}{\sigma_\tau} \tag{7}$$

and a frequency gain that decreases sharply as $\Phi(kf_c)/\Phi(0) = 2^{-k^2}$ so that for any frequency $f > 2f_c$ (12 dB drop) the harmonic structure of $x_H(t)$ has almost completely vanished—see Fig. 1(d). For instance, with 1% fluctuation which is quite typical ($\sigma_\tau = 0.01T$), model (3) yields no more than about 40 significant harmonics. *In short, even slight random fluctuations are prone to severely low-pass filter the harmonic content of signal $x_H(t)$.* Now, considering the fact that the structural transfer function $H(f)$ will be typically of a high-pass or high-frequency band-pass nature (as expected from acceleration measurements possibly including the sensor resonance), it makes no doubt that the frequency gain $|H(f)\Phi(f)|^2$ will strongly attenuate the harmonic structure of signal $x_H(t)$. The presence of a strong additive background noise $n(t)$ will then finish the job by completely masking the signal, together with the diagnostic information, as illustrated in Fig. 2.

2.3. Assessment of the random contribution

The random contribution $x_R(t)$ in signal $x(t)$ has a power spectral density (signal-units^2/Hz) given by [8]

$$S_{x_R}(f) = \frac{1}{T}|H(f)|^2\left(1 + \sigma_A^2 - |\Phi(f)|^2\right)\mathcal{Q}_0, \tag{8}$$

where $\mathcal{Q}_0$ stands for the power—i.e. the mean-square value—of the modulating function $q(t)$. Note that contrary to the harmonic contribution spectrum (6), this is now a *continuous* spectrum and it is weighted by a high-pass[9] function $(1 + \sigma_A^2 - |\Phi(f)|^2)$. Hence, the random contribution is likely to dominate the vibration signal in the frequency range spanned by the transfer function $H(f)$. However, due to the continuous nature of its power spectral density, this contribution is unable to display the diagnostic information by means of classical spectral analysis.

[8]The Gaussian assumption is used for simplicity, and provides anyhow very similar results than the more realistic gamma assumption that is constrained to deliver positive inter-arrival times [13].

[9]From probability theory, $|\Phi(f)| \leqslant \Phi(0) = 1$, because $\phi(\tau) \geqslant 0$.

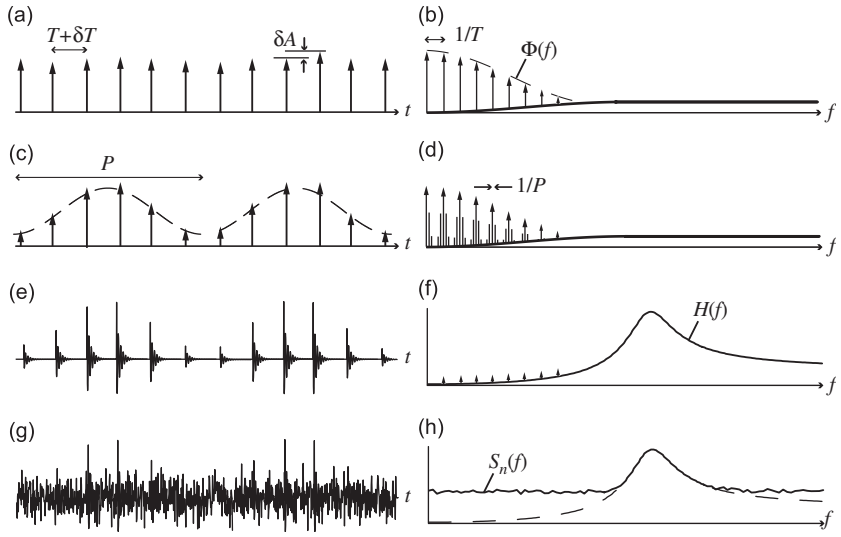

Fig. 2. Synthesis of the vibration signal produced by a localised fault as seen in the time domain (left) and in the frequency domain (right): (a,b) a series of impact forces with slight random fluctuations in their inter-arrival times (mean value = T) and in their magnitudes; (c,d) possible modulation by shaft or cage rotation (period = P); (e,f) filtering by a structural or transducer resonance; (g–h) additive noise from other vibration sources.

2.4. Insufficiency of classical spectral analysis

Putting all results together, the power spectral density of the vibration signal (3) reads

$$S_x(f) = \frac{1}{T}\underbrace{|H(f)|^2}_{\text{(i)}}\left[\underbrace{|\Phi(f)|^2}_{\text{(ii)}}\underbrace{\sum_{k,l=-\infty}^{+\infty} q_l\delta\left(f-\frac{k}{T}-\frac{l}{P}\right)}_{\text{(iii)}} + \underbrace{(1+\sigma_A^2-|\Phi(f)|^2)\mathcal{Q}_0}_{\text{(iv)}}\right] + \underbrace{S_n(f)}_{\text{(v)}}, \tag{9}$$

wherein:

- the transfer function (i) accounts for the high-pass or high-frequency band-pass filtering of the metallic structure (possibly including the sensor resonance),
- the weighting function (ii) accounts the low-pass filtering as a result of the random jitter on the impacts (inter-)arrival times,
- the discrete spectrum (iii) accounts for the fault signature,
- the continuous spectrum (iv) accounts for a high-pass distribution of energy due to the inclusion of randomness in the model,
- and the spectral power (v) accounts for the presence of background noise.

Note that $S_x(f)$ decomposes as the spectrum of a forcing term—(ii) + (iii) + (iv)—multiplied by the system transfer function (i). This is illustrated in Fig. 2. As clearly evidenced by this figure, the spectral signature of the fault (iii) has a weak energy that is likely to be completely masked by that of the background noise; indeed the noise is often found to have very large power in the lower frequency range, especially when it contains the harmonics parts of other rotating components such as gears. This explains why classical spectral analysis often fails to detect rolling-element bearing faults in practice.

Table 1
Bearing characteristics of Example 1

Speed of shaft (Ω) (rev/min)	650
Bearing roller diameter (d) (mm)	21.4
Pitch circle diameter (D) (mm)	203
Number of rolling elements (n)	23
Contact angle (θ) (deg.)	9.0

2.5. Illustrative Example 1

Our argumentation is illustrated by means of numerical simulations. A synthetic signal that simulates the vibrations (m/s^2) produced by an outer-race fault was generated according to model (3) with a sampling frequency $F_s = 20\,\text{kHz}$ and bearing characteristics as indicated in Table 1. From Table A.1, the expected fault frequency is $1/T = 111.6\,\text{Hz}$. The standard deviation σ_τ of the jitter[10] was set to $0.04T$ and that of the magnitudes of the impulses (i.e. σ_A) to 0. In order to simulate the excitation of a structural resonance, the signal was further filtered with the impulse-response a second-order system with natural frequency $f_n = 6\,\text{kHz}$ and damping ratio $\zeta = 5\%$. Finally, a stationary white Gaussian noise was added so as to produce a signal-to-noise ratio of 0 dB. Fig. 3(a,b) displays the signal in the time domain before and after adjunction of the background noise. At first glance it may seem surprising that such a signal is essentially random but not harmonic, even before the adjunction of noise. Nevertheless, according to formula (7), the cut-off frequency

$$f_c \simeq \frac{0.19}{0.04T} = 530\,\text{Hz} \tag{10}$$

states that the harmonic structure has almost completely vanished after $f > 2f_c \simeq 1.1\,\text{kHz}$. This is checked in Fig. 3(c) which displays the power spectral density. It is seen that, just as predicted by the theory, harmonics die off after 1.1 kHz whilst, at the time, the remaining lower-order ones below $2f_c$ are strongly masked by the background noise.

2.6. The cyclic spectral analysis of bearing vibration signals

We now proceed to show that the diagnostic information contained in the random signal $x_R(t)$—that classical spectral analysis was found hardly capable to reveal in the presence of masking noise—can be distinctly recovered by means of *cyclic spectral analysis*. This actually amounts to exploiting the cyclostationarity of $x_R(t)$. Indeed, one way to investigate whether a signal is cyclostationary or not is to check for the presence of statistical correlation between its spectral components spaced apart by some frequency shift $\alpha \neq 0$. Whereas such correlations do not exist for stationary signals (because their spectral components are independently and randomly phased), they exist for cyclostationary signals in which wave packets occur in a structured repetitive pattern (i.e. spectral components are necessarily synchronously phased). More precisely, by denoting $X_L(f)$ the Fourier transform of a cyclostationary signal $x(t)$ evaluated over an interval of length L, it can be shown that

$$S_x(f;\alpha) = \lim_{L\to\infty} \frac{1}{L} E\left\{ X_L\left(f + \frac{\alpha}{2}\right) X_L\left(f - \frac{\alpha}{2}\right)^* \right\} \tag{11}$$

will produce a non-zero correlation for any cyclic frequency α that characterises the cyclostationarity of signal $x(t)$. By analogy with the power spectral density which is defined by Eq. (11) with $\alpha = 0$, the quantity $S_x(f;\alpha)$ is coined the *cyclic power spectrum* (signal-units^2/Hz).[11] Intuitively, the cyclic power spectrum may be interpreted as providing the distribution of the frequency content of signal $x(t)$ that statistically repeats itself

[10] As indicated in the footnote of page 7 the correct assumption for realistic numerical simulations is $E\{(\tau_{i+1} - \tau_i)(\tau_{j+1} - \tau_j)\} = \delta_{ij}\sigma_\tau^2$ which is slightly different than the simplistic assumption (4). Again this does not fundamentally change the theory.

[11] This quantity is also sometimes referred to as the *spectral correlation*, although the latter is more logically defined by Eq. (11) without the $1/L$ normalisation so as to produce a surface-wise (signal-units^2/Hz^2) spectrum [17].

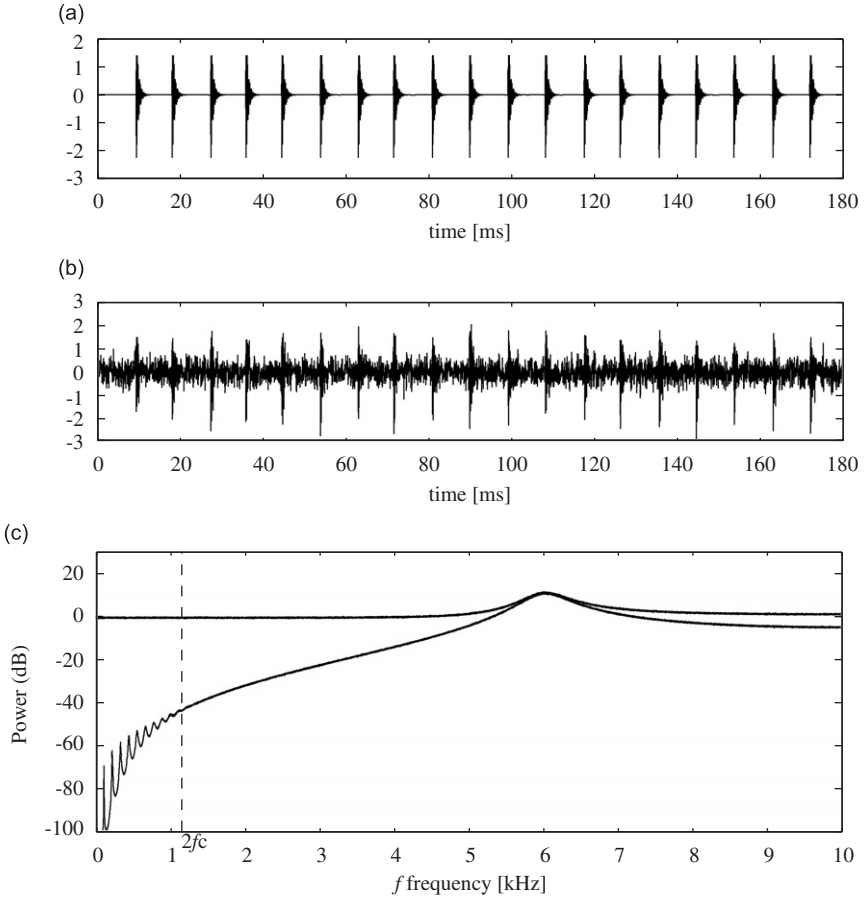

Fig. 3. Synthesised vibration signal of Example 1 (a) before and (b) after adjunction of background noise. (c) The corresponding power spectral densities (resp. lower and upper curves) measured from 1000 independent averages with frequency resolution $\Delta f = 0.3$ Hz.

with rate α. For instance a white random noise amplitude modulated by a periodic function of period T would produce a non-zero, constant-valued, cyclic power spectrum at each cyclic frequency $\alpha = k/T$, $k = 0, \pm 1, \pm 2, \ldots$.

Coming back to our rolling-element bearing signal (3), it can be shown that its cyclic power spectrum reads [8]:

$$S_x(f;\alpha) \simeq \frac{1}{T} H\left(f+\frac{\alpha}{2}\right) H\left(f-\frac{\alpha}{2}\right)^* \left(\Phi(\alpha)(1+\sigma_A^2) - \Phi\left(f+\frac{\alpha}{2}\right)\Phi\left(f-\frac{\alpha}{2}\right)^*\right) \times \sum_{k,l=-\infty}^{+\infty} \mathscr{Q}_l \delta\left[\alpha - \frac{k}{T} - \frac{l}{P}\right] + \delta[\alpha] S_n(f), \tag{12}$$

where the $\simeq$ sign comes from the fact that we have now neglected the contribution from the weak harmonic contribution (which is truly justified, at least after twice the cut-off frequency f_c (7), as explained above), where $\delta[\alpha]$ (with brackets) denotes the Kronecker delta pulse such that $\delta[\alpha] = 1$ if $\alpha = 0$ and 0 elsewhere, and where $\mathscr{Q}_l$ is the lth Fourier coefficient of $q(t)^2$. Inspection of the cyclic power spectrum (12) calls for some important remarks:

(1) setting $\alpha = 0$ in Eq. (12) produces as a particular case the continuous power spectral density (9),
(2) setting $\alpha = k/T + L/P \neq 0$ in Eq. (12) produces yet another *continuous spectrum* in the f frequency variable,

(3) more importantly, looking at $S_x(f;\alpha)$ as a function of the α frequency variable now fully discloses the hidden diagnostic information through the *discrete spectrum* $\sum_{k,l=-\infty}^{+\infty} \mathcal{Q}_l \delta[\alpha - k/T - l/P]$. Indeed, this pattern of spectral lines is found very similar to that appearing in spectra (2) or (9), with the Fourier coefficients $\mathcal{Q}_l$ of $q(t)^2$ simply substituted for the Fourier coefficients q_l of $q(t)$. As a consequence it uniquely defines each bearing-fault type and it can just as well serve as a distinctive *fault signature* to be used for diagnostic purposes.

The squared-magnitude cyclic power spectrum of the bearing vibration signal bears a very interesting interpretation when it is normalised by the signal power, that is when the following spectral quantity is considered:

$$|\gamma_x(f;\alpha)|^2 = \frac{|S_x(f;\alpha)|^2}{S_x(f+(\alpha/2))S_x(f-(\alpha/2))}, \tag{13}$$

which is known as the *(squared-magnitude) cyclic coherence*. By combining definitions (11) and (13), it is seen that the (squared-magnitude) cyclic coherence is actually like a correlation coefficient between the spectral components at frequencies $f+\alpha/2$ and $f-\alpha/2$. As a consequence it is normalised between 0 (no spectral correlation) and 1 (perfect linear relationship between two spectral components).

Now, let us assume that (i) the cyclic coherence is considered within a (f,α) frequency domain[12] such that

$$0 < \alpha < 2f_c < f, \quad \alpha \ll f \tag{14}$$

with f_c as defined in Eq. (7), and that (ii) the transfer function and the noise power density are smooth enough so that, for a small α as just specified in Eq. (14), $H(f \pm \alpha/2) \simeq H(f)$ and $S_n(f \pm \alpha/2) \simeq S_n(f)$. Then, noting in passing that $\Phi(f \pm \alpha/2) \simeq 0$ under (14), it is easy to show that:

$$|\gamma_x(f;\alpha)|^2 \simeq \underbrace{\left|\frac{\mathrm{SNR}(f)}{1+\mathrm{SNR}(f)}\right|^2}_{\text{(i)}} \underbrace{|\Phi(\alpha)|^2}_{\text{(ii)}} \underbrace{\sum_{k,l=-\infty}^{+\infty} \left|\frac{\mathcal{Q}_l}{\mathcal{Q}_0}\right|^2 \delta\left[\alpha - \frac{k}{T} - \frac{l}{P}\right]}_{\text{(iii)}}, \tag{15}$$

where

$$\mathrm{SNR}(f) = \frac{S_{x_R}(f)}{S_n(f)} = \frac{(1+\sigma_A^2)\mathcal{Q}_0}{T}\frac{|H(f)|^2}{S_n(f)} \tag{16}$$

stands for the signal-to-noise ratio of the fault. Hence, the interpretation of the cyclic coherence of the bearing vibration signal is the following:

(1) when read as a function of the f frequency, the cyclic coherence is essentially *a measure of the signal-to-noise ratio of the fault* normalised between 0 and 1 (quantity (i) in Eq. (15)); the latter increases with the modulation intensity $\mathcal{Q}_0$, with the modulation randomness σ_A^2, and with the impacts frequency $1/T$,
(2) when read as a function of the cyclic frequency α, the cyclic coherence *displays the fault signature* (iii); the latter is all the more low-pass weighted by function (ii) as the arrival times of the impacts undergo randomness.

As a consequence of point (1), the magnitude of the cyclic coherence in the f frequency domain may serve as a relative measure of the fault severity. As a consequence of point (2), the diagnostics information that was previously masked by other vibration components in the f frequency domain is now fully revealed in the α frequency domain (Fig. 4).

[12]Condition (14) will always be fulfilled in practice; it will be seen in the examples of Section 6 that the range of interest of the cyclic frequency α is typically one order of magnitude below that of the f frequency.

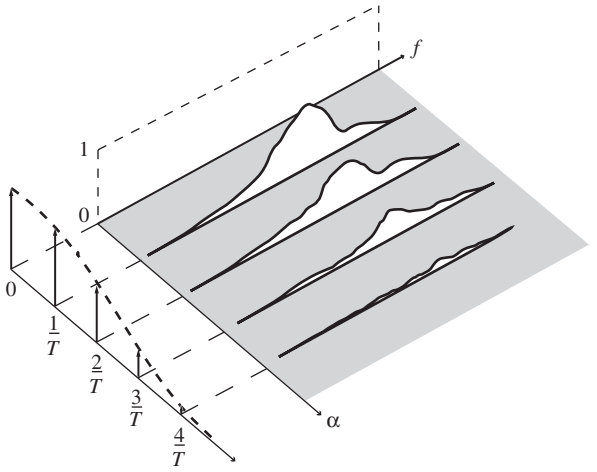

Fig. 4. Typical (squared-magnitude) cyclic coherence of a bearing vibration signal (without sidebands, e.g. as in the case of an outer-race fault). It is *continuous* in the f frequency and *discrete* in the α frequency. The spectral signature of the fault is fully characterised by the pattern of lines in the α frequency domain.

3. Experimental cyclic spectral analysis

The preceding section has demonstrated that the diagnostic information is fully revealed in the cyclic frequency domain whatever the degree of randomness of the fault signal and the level of background noise. The next section will show how this information can be efficiently extracted from finite-length vibration measurements. In the meantime, it remains to discuss how the cyclic power spectral density (11)—and therefore the cyclic coherence (13)—can be computed in practice. We believe that this fundamental question has been disregarded in several earlier works, and yet it is the key for any successful exploitation of cyclostationarity in bearing diagnostics.

The issue we are addressing here is how to get an accurate estimate of $S(f;\alpha)$, say $\hat{S}^{(L)}(f;\alpha)$, from a finite-length record $\{x[n]\}_{n=0}^{L-1}$ of L samples acquired at the sampling rate F_s (by convention $x[n]=x(n/F_s)$ with n an integer value). This issue has been thoroughly studied in Ref. [17], whose we resume here the main results.

3.1. The averaged cyclic periodogram

Probably the most practical estimator $\hat{S}^{(L)}(f;\alpha)$ is obtained from the *averaged cyclic peridogram*. Let $\{w[n]\}_{n=0}^{N_w-1}$ be a positive and smooth N_w-long data-window and let $w_k[n]=w[n-kR]$ be its shifted version by R samples so that $w_k[n]x[n]$ selects a segment of $x[n]$ at times $kR,\ldots,kR+N_w-1$. The increment R is set between 1 and N_w so as to allow possible overlap between adjacent segments. Then the averaged cyclic periodogram is defined as

$$\hat{S}_x^{(L)}(f;\alpha)=\frac{1}{KF_s\|w\|^2}\sum_{k=0}^{K-1}X_{N_w}^{(k)}\left(f+\frac{\alpha}{2}\right)X_{N_w}^{(k)}\left(f-\frac{\alpha}{2}\right)^*, \tag{17}$$

where

$$X_{N_w}^{(k)}\left(f\pm\frac{\alpha}{2}\right)=\sum_{n=kR}^{kR+N_w-1}w_k[n]x[n]\mathrm{e}^{\pm\mathrm{j}\pi\alpha n/F_s}\mathrm{e}^{-\mathrm{j}2\pi fn/F_s} \tag{18}$$

is the discrete Fourier transform (DFT) of the kth weighted sequence $w_k[n]x[n]\mathrm{e}^{\pm\mathrm{j}\pi\alpha n/F_s}$, $K=\lfloor(L-N_w)/R\rfloor+1$ (where $\lfloor x\rfloor$ stands for the greatest integer smaller than or equal to x) is the total number of averaged segments, and $\|w\|$ stands for the window RMS value. The principle is illustrated in Fig. 5. Furthermore, the

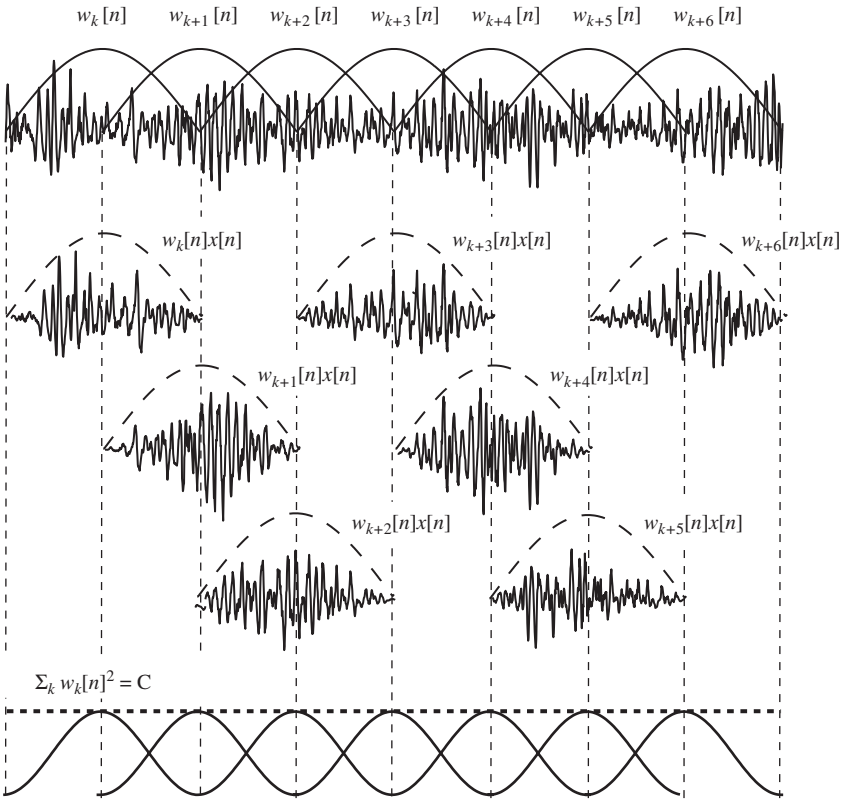

Fig. 5. Principle of the averaged cyclic periodogram: segmenting, windowing, Fourier transforming, averaging. The window shape and overlap must be chosen according to the condition illustrated in the bottom of the figure.

estimation variance associated with this estimator is given by, for large L:

$$\mathrm{Var}\{\hat{S}_x^{(L)}(f;\alpha)\} \simeq \mathscr{E} \cdot S_x\left(f+\frac{\alpha}{2}\right) S_x\left(f-\frac{\alpha}{2}\right) \tag{19}$$

almost everywhere (i.e. except on a finite number of frequency points), where the variance reduction factor $\mathscr{E}$ reads

$$\mathscr{E} = \sum_{k=-K+1}^{K-1} R_w[kR]^2 \cdot \frac{K-|k|}{K^2} \tag{20}$$

with $R_w[n]$ the autocorrelation function of the data-window $w[n]$.

Note that this simple estimator[13] is a direct transcription of the formal definition (11) where the expected value (ensemble average operator) has been replaced by a (finite) average over segments, and where the limit to infinity of the integration interval has been omitted. Note also that it does not require any key phasor signal as sometimes advocated, and that the choice of N_w is completely independent of the cyclic frequency α. Moreover, as compared to other potential spectral estimators, we claim that the averaged cyclic periodogram has the advantage of (i) ease of implementation and (ii) enjoying a fast and efficient computation by means of several short FFT's of fixed size.

[13]When $\alpha = 0$, it also boils down to the classical averaged periodogram—or Welch's estimator—of the power spectral density.

3.2. Practical guidelines

3.2.1. Setting the window shape and the amount of overlap

We have shown in Ref. [17] that the window shape and the amount of overlap must be jointly chosen so that the condition

$$\sum_{k=-\infty}^{+\infty} w_k[n]^2 = C \tag{21}$$

is met for some arbitrary constant C—see bottom of Fig. 5. If that condition was not met, the estimator $\hat{S}_x^{(L)}(f;\alpha)$ would possibly yield non-zero values at incorrect cyclic frequencies, an effect that has been coined *cyclic leakage*. Typical configurations that will ensure condition (21) are:

- a hanning or a hamming window with 2/3 overlap between adjacent segments,
- a halfsine window with 1/2 overlap between adjacent segments—as illustrated in Fig. 5.

It is noteworthy that these recommendations also bring the variance reduction factor (20) to a near minimal value without increasing too much the computational demand.

3.2.2. Setting the cyclic frequency resolution $\Delta\alpha$

The cyclic frequency resolution $\Delta\alpha$ is the minimum spacing between two cyclic frequencies at which the cyclic power spectrum is evaluated. It must be fine enough so that no cyclic information is lost, but not excessively fine so as to avoid unnecessary computations. It is shown in Ref. [17] that the critical (i.e. coarser allowable) cyclic frequency resolution is given by

$$\Delta\alpha = \frac{F_s}{L}, \tag{22}$$

where L is the record length. Hence, the cyclic frequency axis should be sampled like $\alpha_k = kF_s/L$, $k = 0, \pm1, \pm2, \ldots$ and the cyclic power spectrum evaluated at each of these cyclic nodes. As a corollary, the longer the recording time, the finer the available cyclic frequency resolution. Consequently, in vibration-based monitoring of rolling-element bearings, the user should be careful to record long enough signals to as to be able to resolve between the different types of faults, and between the various types of sidebands—see Table A.1.

3.2.3. Setting the frequency resolution Δf

Given a data-window shape $w[n]$, the frequency resolution may be evaluated as

$$\Delta f = \frac{\sum_{n=0}^{N_w-1} w[n]^2}{\left|\sum_{n=0}^{N_w-1} w[n]\right|^2} F_s. \tag{23}$$

For instance, $\Delta f \simeq 1.5F_s/N_w$ with a hanning data-window, and $1.2F_s/N_w$ with a halfsine data-window. Setting the frequency resolution Δf then amounts to choosing the data-window length N_w. In turn, given N_w the frequency axis is sampled[14] like $f_k = kF_s/N_{\mathrm{FFT}}$, $k = 0, 1, \ldots, N_{\mathrm{FFT}} - 1$ where $N_{\mathrm{FFT}} \geqslant N_w$ is typically the smallest power of 2 greater than or equal to N_w.

Clearly, the shorter the data-window, the coarser the resolution Δf. We claim that imposing *a coarse resolution is necessary* for at least two reasons. First, it is the main control lever to pull down the estimation variance (19) to reasonably small values since the variance reduction factor (20) is a decreasing function of N_w. In order to achieve sufficient statistical stability of the estimates, the user should be careful to ensure that $N_w/L \ll 1$, or equivalently that

$$\Delta\alpha \ll \Delta f, \tag{24}$$

[14]This frequency sampling is accomplished automatically when computing the FFT of the windowed data (possibly zero-padded if $N_{\mathrm{FFT}} > N_w$) over N_{FFT} lines.

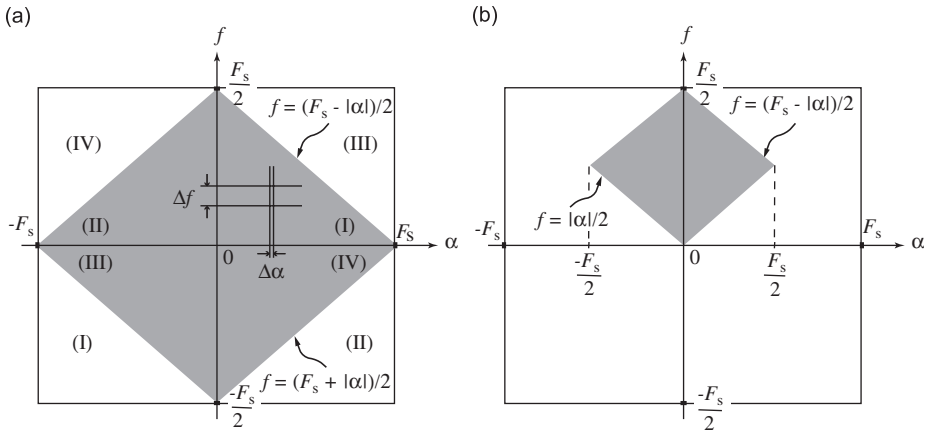

Fig. 6. (a) Principal domain (grey area) of the cyclic power spectrum for a signal with sampling frequency F_s, along with the symbolisation of the frequency resolutions Δf and $\Delta\alpha$. Values outside the principal domain (white area) are obtained by periodisation of domains (I–IV) and should be disregarded from the analysis. (b) Non-zero values (grey area) of the cyclic power spectrum of the analytic signal: they are totally included in the principal domain.

where the ratio $\Delta\alpha/\Delta f$ can be shown to be on the order of the variance reduction factor (20). It is our feeling that this condition has been missed in several earlier works where the images obtained for the cyclic spectra displayed as functions of f and α hardly exhibited the expected discrete structure along the cyclic frequency axis.

The second reason why the resolution Δf should be kept coarse enough is to reduce the memory allocation required to store the collection of N_{FFT}-long cyclic spectra computed at several cyclic frequencies α_k.

3.2.4. The analytical signal

Because the discrete Fourier transform is a periodic function of frequency, it results that $X_{N_w}^{(k)}(f+\frac{\alpha}{2})X_{N_w}^{(k)}(f-\frac{\alpha}{2})^*$ in Eq. (17) truly measures the correlation between spectral components spaced apart by α (and not $\alpha \pm kF_s$, $k \neq 0$) *only* within the principal frequency domain—see Fig. 6(a):

$$|f| \leqslant \frac{F_s - |\alpha|}{2}. \tag{25}$$

All values of the cyclic power spectrum outside this domain should be disregarded as they produce artifacts. One straightforward solution to ensure that condition (25) is met is to compute the cyclic power spectrum on the *analytical signal*[15] [8]—see Fig. 6(b).

3.3. Illustrative Example 2

The estimation of the cyclic power spectrum is briefly illustrated here on the synthetic signal of Example 1. All simulation parameters are as specified in Section 2.5—see Table 1—except that an inner-race fault was now generated. According to Table A.1, the signal is thus expected to exhibit cyclostationarity at the inner-race element passing frequency and its multiples, that is at $\alpha = k/T = k \times 137.5\,\mathrm{Hz}$, $k = 0, \pm 1, \pm 2 \ldots$ as well as at $k/T + l/P = k \times 137.5 + l \times 10.8\,\mathrm{Hz}$, $l = 0, \pm 1, \pm 2 \ldots$ due to modulations. The estimation of the cyclic power spectrum was carried out using the averaged cyclic periodogram (17) with the parameters as indicated in Table 2.Different window lengths were tried, so as to produce different frequency resolution Δf and variance reduction factors $\mathscr{E}$.

The results are displayed in Fig. 7. It is seen that the theoretical "discrete in α and continuous in f" property of the cyclic power spectrum shows off correctly only after sufficient averaging is performed, i.e. when $\mathscr{E} \leqslant .0076$ (at least 132 equivalent number of averages) or, equivalently when $\Delta f \geqslant 100\Delta\alpha$. The spectral

[15]The analytic signal may be simply obtained from the DFT by applying the transformation $X(f) \to (1 + \mathrm{sign}(f))X(f)$.

Table 2
Estimation parameters of Example 2

Fixed parameters				
Sampling frequency (F_s) (kHz)	20			
Record length (L)	16 384 samples			
Window type	Hanning			
Overlap	2/3			
α Frequency resolution ($\Delta\alpha$) (Hz)	1.2			
Scrutinised cyclic frequency range (Hz)	[0;366]			
Variable parameters				
Window length (N_w samples)	16 384	4096	1024	256
f Frequency resolution (Δf in Hz)	1.8	7.3	29.3	117.2
Variance reduction factor ($\mathscr{E}$)	1	.1401	.0312	.0076
Equivalent number of independent averages ($1/\mathscr{E}$)	1	7	32	132

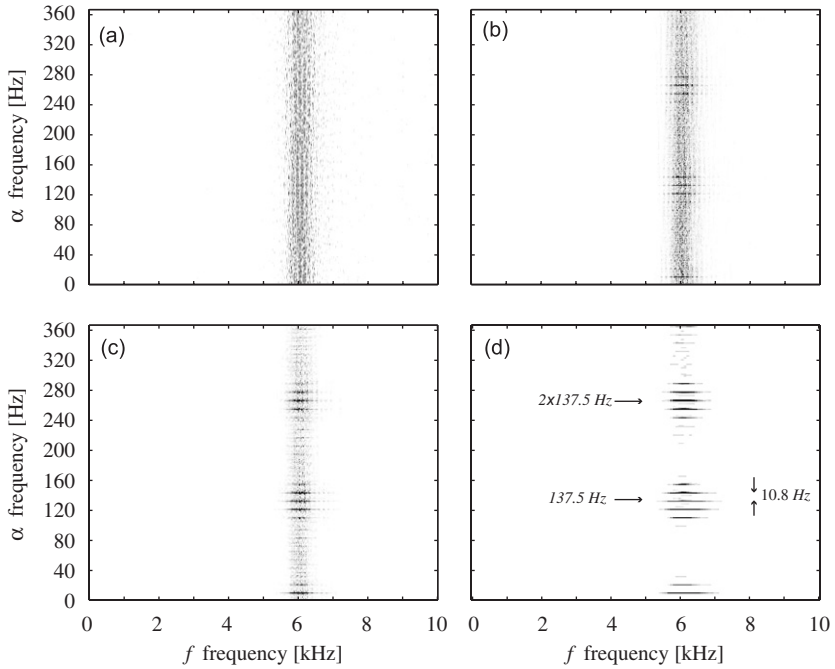

Fig. 7. Estimated cyclic power spectrum of Example 1 displayed as a grey-level image. (a) Variance reduction factor $\mathscr{E} = 1$, (b) $\mathscr{E} = 0.1401$, (c) $\mathscr{E} = 0.0312$, (d) $\mathscr{E} = 0.0076$. The corresponding numbers of independent averages are 1, 7, 32 and 132.

signature of the fault is then clearly recognisable, with its harmonics and sidebands (horizontal lines) at the expected positions. Very similar results have been obtained with various signal-to-noise ratios.

3.4. Some common pitfalls

We insist on the fact that the inherent assumption when computing the cyclic power spectrum or the cyclic coherence of a cyclostationary signal is that the signal is *random*. This is why the estimation process requires a large amount of averaging in the f direction, as is always the case in the experimental spectral analysis of random vibrations. As a consequence it is erroneous to try to compute the cyclic power spectrum of a harmonic signal [18,19]. Indeed, purely harmonic signals should be computed with other dedicated tools (e.g. a

very long DFT which allows preserving a very fine frequency resolution) which are outside the scope of this paper.[16] Nevertheless, in practice the signal of interest is likely to be of a mixed nature, that is a mixture of random and harmonic components. Ideally, all harmonic components should therefore be removed before cyclic spectral analysis in order to avoid interferences that would mask the cyclostationary structure of the random contribution [5,10]. Ref. [20] provide some algorithms to blindly filter out all the random components from a signal, which have been purposely devised for bearing diagnostics. However, provided that the harmonic and the random components happen to fall in different frequency regions, the removal of the former may be needless. This is actually the situation that frequently occurs with rolling-element bearing vibrations, where the fault excites some high-frequency resonance that is well beyond the lower frequency region where masking harmonics (from gears, unbalance, blade frequency, etc) take place.

4. Cyclic spectral analysis for bearing diagnostics

It has been explained in Section 2 that classical spectral analysis may fail to detect bearing faults because of masking noise and of the marked randomness of the vibrations. Yet, the diagnostic information has been shown to rise up intact in the cyclic frequency domain in the form of symptomatic discrete spectral signature. The natural idea is therefore to devise a diagnostic strategy based on the detection of such signatures. Moreover, in order for this strategy to be credible and possibly implemented in an automated monitoring system, its sensitivity must be assessed by means of a (user-specified) rate of false alarm.

Several strategies have actually been proposed in the specialised literature in order to detect the cyclostationary signature of rolling-element bearings or related mechanical components [6–9]. Most of them are based on the direct visual inspection of the magnitude of the cyclic power spectral density displayed as an image over the (f,α) frequency domain. Obviously the sensitivity of such an approach is rather subjective as it strongly depends on the diagnostician expertise. Moreover nothing justifies that the cyclic power spectrum is the best spectral quantity to be considered. We prove below that the *cyclic coherence* should be considered indeed, for several reasons.

4.1. An optimal test

Let us first recast the diagnostics issue into a statistical decision test. Namely, given a characteristic cyclic frequency α at which a fault is likely to occur, a probability p of false alarm (i.e. of wrongly concluding to the presence of a fault), and a finite-length record $\{x[n]\}_{n=0}^{L-1}$ of L samples, one must decide between to the two alternative hypotheses:

$$\begin{aligned} &H_0 : \text{“The vibration signal does not contain a fault at the cyclic frequency } \alpha.\text{”} \\ &H_1 : \text{“The vibration signal contains a fault at the cyclic frequency } \alpha.\text{”} \end{aligned} \tag{26}$$

It is proved in the Appendix that the following decision rule is optimal:

$$\text{“Reject the null hypothesis } H_0 \text{ if } \quad |\hat{\gamma}_x^{(L)}(f;\alpha)|^2 = \frac{|\hat{S}_x^{(L)}(f;\alpha)|^2}{\hat{S}_x^{(L)}(f+\frac{\alpha}{2})\hat{S}_x^{(L)}(f-\frac{\alpha}{2})} \geqslant \frac{\mathscr{E}}{2}\cdot\chi^2_{1-p,2}\ \text{”}, \tag{27}$$

where $\chi^2_{1-p,2}$ is the percentile of the chi-square law with two degrees-of-freedom and $\mathscr{E}$ is the variance reduction factor defined in Eq. (20).

The optimality of the decision rule (27) is understood in the sense that:

(1) it is *invariant* under any linear filtration of the signal,
(2) it *maximises the probability of detection* given a fixed record length L and a fixed probability of false alarm p or, put differently, it minimises the necessary record length L to not exceed a given probability of error (erroneous detection+missed detection).

[16] This discussion is reminiscent to recognising the fundamental difference between *pure* and *impure* second-order cyclostationarity [5].

Table 3
Estimation parameters of Example 3

Record length (L)	22 050 samples
Window type	Hanning
Overlap	2/3
Window length (N_w)	128 samples
f Frequency resolution (Δf) (Hz)	234.4
α Frequency resolution ($\Delta\alpha$) (Hz)	0.9
Scrutinised cyclic frequency range[a]	[0.9;600]
Variance reduction factor ($\mathscr{E}$)	0.0028

[a]The zero cyclic frequency is excluded as it produces a constant unit-valued cyclic coherence.

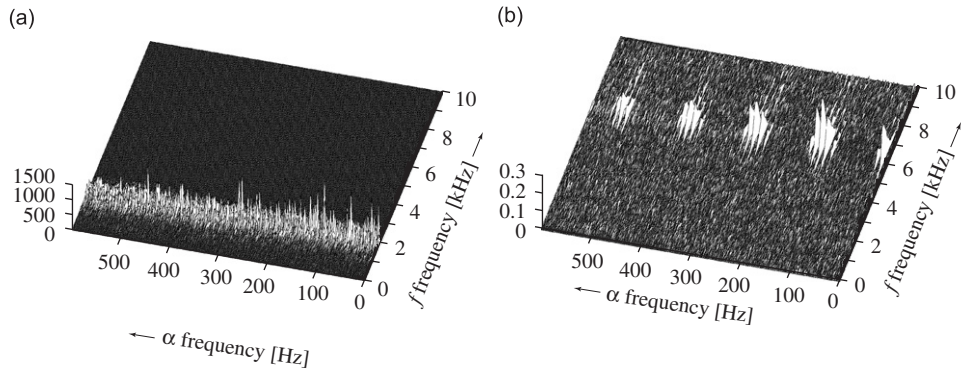

Fig. 8. (a) Estimated cyclic power spectrum (magnitude in signal-units^2/Hz) of Example 2. (b) Estimated cyclic coherence (squared-magnitude)—the 0.1%-statistical threshold is 0.024.

The optimality holds true for the (squared-magnitude) cyclic coherence $|\hat{\gamma}_x^{(L)}(f;\alpha)|^2$ but not for the cyclic power spectrum. Yet another considerable advantage linked with the cyclic coherence—notwithstanding its sound interpretation as a signal-to-noise ratio as explained in Section 2—is that it involves an easy-to-compute threshold (27) that is independent of frequency. This is illustrated in the following example.

4.2. Illustrative Example 3

The outer-race fault signal of Example 2 is considered again. However, it was further corrupted by a narrow-band noise so as to simulate the contribution of another dominant vibration source. This extra noise source was generated by filtering a white Gaussian noise with a second-order system filter with natural frequency $f_n = 2\,\text{kHz}$ and damping ratio $\zeta = 5\%$, such as to produce an overall signal-to-noise ratio of $-24\,\text{dB}$. Such a small signal-to-noise ratio is not uncommon in practice—the corresponding signal really looks stationary in the time-domain—and it illustrates well a typical situation where cyclic spectral analysis, if properly applied, will succeed.

The cyclic power spectrum and the cyclic coherence were computed according to the algorithm outlined in Section 3, with the parameters given in Table 3.

The results are displayed in Figs. 8(a,b) as functions of f and α. It is seen that in the cyclic power spectrum the fault signature is now completely masked by the presence of the narrow-band noise around $f \sim 2\,\text{kHz}$ and whatever the cyclic frequency α. This is because from Eq. (19) the estimation variance of the cyclic power spectrum is roughly proportional to the square of the power spectrum, so that regions where the noise power predominates can completely absorb the picture dynamics at the expense of other regions where the (weak) diagnostic information lies. Obviously, performing a statistical test on that cyclic power spectrum would be

hardly feasible because the threshold would have to follow the spectral variations of the estimation variance. On the other hand, the cyclic coherence perfectly exhibits the expected fault signature in the frequency range [5;8] kHz where the signal-to-noise ratio is the greatest, in spite of the extreme background noise. This is because the dynamics of its estimation variance is equalised throughout the whole f frequency domain.[17] For the same reason it easily allows performing a statistical test: according to the decision rule (27), any value of the (squared-magnitude) cyclic coherence that is greater than or equal to the constant threshold $\frac{1}{2}\mathscr{E}\cdot\chi^2_{1-p,2}$ will point out the emergence of the fault signature in $100(1-p)\%$ of the cases. For instance, with a probability of false alarm $p=0.001$ (0.1%), this is a threshold of $\frac{1}{2}\mathscr{E}\cdot\chi^2_{.999,2}=0.024$ whatever f and α.

5. Substantial simplifications

The statistical test introduced in the previous section is optimal in several respects and would certainly benefit to a diagnostics protocol as such. However, there are at least two reasons why further simplifications may be deemed desirable.

First, the computation time and the memory allocation required to evaluate the cyclic coherence over a large (f,α) domain with a fine $\Delta\alpha$ resolution may still be too onerous in certain industrial applications where near *real-time* processing is required (for instance, the cyclic coherence of example 2 took about 2 min to be computed with a ×86, 996 MHz processor and was stored as a 660×128 pixel image). Second, the degree of expertise required by the user to interpret the cyclic coherence displayed as an image—and any matrix diagnostic indicator in general—may be too demanding. In such cases, substantial simplifications are still possible. They essentially rely on designing sub-optimal indicators that do not consist in a complete exploration of the (f,α) plane, and that can be displayed in a vectorial form. Two different situations should be distinguished whether the cyclic frequencies at which the diagnostic information lies is (partially) known or totally unknown. The first situation leads to considering the cyclic coherence at only one cyclic frequency. The second one leads to the *integrated cyclic coherence* and the *squared-envelope spectrum*.

5.1. The cyclic coherence at only one cyclic frequency

This first simplification trivially consists in computing the cyclic coherence at the cyclic frequency that coincides with the expected characteristic fault frequency. In practice, the fault frequency is only known approximately (it depends on rough estimates of the rotation speed and of the angle of contact) so that the cyclic coherence should actually be computed in a narrow cyclic frequency band around the expected fault frequency, and its greatest value retained.

Although the proposed strategy involves the cyclic coherence, it is sub-optimal in the sense that only one cyclic frequency is taken into consideration and that the full spectral fault signature (fundamental + harmonics + sidebands) is not exploited. This might rise some difficulties to discriminate fault types if it happens that the cyclic frequency resolution $\Delta\alpha$ is set too coarse to properly resolve between different characteristic frequencies.

5.2. The integrated cyclic coherence (ICC)

The preceding simplification has led to a vectorial diagnostic indicator by considering only one single cyclic frequency of interest. When the latter is unknown, or when the full spectral signature (fundamental + harmonics + sidebands) is to be preserved, another strategy consists in scrutinising the whole cyclic frequency domain while shrinking the f frequency domain. Surely the most natural way to do so is to integrate the (squared-magnitude) cyclic coherence along the f variable. In order to retain only the relevant diagnostic information and not to pollute it with other interferences (e.g. harmonic interferences from the lower-frequency range), the integration should be carried out in the frequency band where the signal-to-noise ratio is the highest, say $[F_1;F_2]$ $(0\leqslant F_1<F_2\leqslant F_s)$. Practically, by denoting $f_k=k/N_{\mathrm{FFT}}$ the sampled frequency

[17] The cyclic coherence (13) may be seen as the cyclic power spectrum (11) normalised by its variance (19).

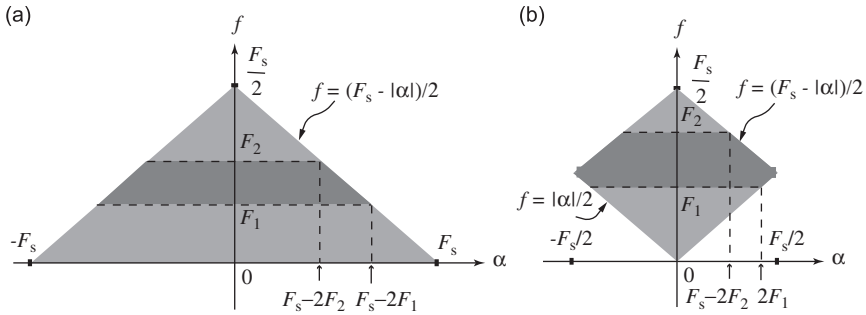

Fig. 9. Domain of integration (dark area) of the integrated cyclic coherence (28) in the case of (a) a real signal and (b) an analytic signal.

variable—see Section 3.2.3—this amounts to computing the indicator

$$\mathrm{ICC}_x^{(L)}(\alpha) = \frac{1}{K_2(\alpha) - K_1(\alpha) + 1} \sum_{k=K_1(\alpha)}^{K_2(\alpha)} |\hat{\gamma}_x^{(L)}(f_k;\alpha)|^2, \tag{28}$$

where $\{K_1(\alpha) \leqslant k \leqslant K_2(\alpha)\}$ stands for the set of frequency nodes that fall within the intersection of the frequency band $[F_1; F_2]$ and of the principal domain (25), that is such that, for each α—see Fig.9:

$$\begin{aligned} K_2(\alpha) &= \lfloor N_{\mathrm{FFT}} \cdot \min(F_2, F_s/2 - |\alpha|/2) \rfloor, \\ K_1(\alpha) &= \lfloor N_{\mathrm{FFT}} \cdot \max(F_1, |\alpha|/2) \rfloor \quad \text{(analytic signal)}, \\ K_1(\alpha) &= \lfloor N_{\mathrm{FFT}} \cdot F_1 \rfloor \quad \text{(real signal)}. \end{aligned} \tag{29}$$

It is proved in the Appendix that the corresponding decision rule reads:

$$\text{“Reject the null hypothesis } \mathrm{H}_0 \text{ if } \quad \mathrm{ICC}_x^{(L)}(\alpha) \geqslant \frac{\mathscr{E}}{2} \cdot \chi^2_{1-p,v} \text{ ”}, \tag{30}$$

where $v = \lfloor 2(K_2(\alpha) - K_1(\alpha) + 1) N_{\mathrm{FFT}}/N_w \rfloor$.

Note that, in spite of being vectorial, the integrated cyclic coherence (28) does not solve the computation impediment for real-time applications since it first requires evaluating the cyclic coherence over the whole (f, α) domain before condensing the information by integration. A further simplification arises with the squared-envelope spectrum.

5.3. The squared-envelope spectrum (SES)

If $[F_1; F_2]$ is a narrow-band selection (as is often the case in practice), then the integrated cyclic coherence (28) is very well approximated[18] by the discrete Fourier transform of the squared-magnitude signal pre-filtered in the $[F_1; F_2]$ band (up to an unimportant scaling factor). Namely,

$$\begin{aligned} \mathrm{ICC}_x^{(L)}(\alpha) &\propto \left| \frac{1}{L} \sum_{n=0}^{L-1} |x[n] * g[n]|^2 \mathrm{e}^{-\mathrm{j}2\pi n\alpha/F_s} \right|^2 \\ &= |\mathrm{DFT}\{|x[n] * g[n]|^2\}|^2 \\ &= \mathrm{SES}_x^{(L)}(\alpha), \end{aligned} \tag{31}$$

[18]Equality would hold exactly if the complex values of the cyclic coherence were averaged rather than their squared magnitudes [8].

where $x[n] * g[n]$ stands for the signal convolution with an equivalent impulse response $g[n]$ that encompasses the following cascade of linear operations:

(1) *whitening* of the signal (to account for the power normalisation that distinguishes the cyclic *coherence* as opposed to the cyclic *spectrum*),
(2) *analytic signal* transformation,
(3) *band-pass filtering* in the $[F_1; F_2]$ band.

The so-obtained vectorial indicator $\mathrm{SES}_x^{(L)}(\alpha)$ is known as the *squared-envelope spectrum*[19] and has been thoroughly studied in Refs. [16,22]. It is noteworthy that it involves only elementary linear transformations (all of which that can be implemented by means of fast algorithms[20]) which makes it a good candidate for real-time applications.

Based on Eq. (27), the decision rule relative to the squared-envelope spectrum can be shown to take the following form (irrespectively of whether the pre-whitening operation and/or the analytic signal transformation are actually performed or not[21]):

$$\text{“Reject the null hypothesis } H_0 \text{ if } \quad \mathrm{SES}_x^{(L)}(\alpha) \geqslant \frac{\sigma_{x*g}^4}{2L}\frac{F_s}{F_2-F_1} f(\alpha)\cdot \chi^2_{1-p,2}\text{”}, \tag{32}$$

where $\sigma_{x\star g}$ stands for the standard deviation of the *filtered* signal $x[n]*g[n]$, and

$$f(\alpha) = \begin{cases} 1-|\alpha|/(F_2-F_1), & |\alpha| < F_2-F_1, \\ 0 & \text{elsewhere.} \end{cases} \tag{33}$$

Note that if the analytical signal is not considered (which we warn against), it is essential that at least the demodulation condition

$$F_2 - F_1 < \min(2F_1, F_s - 2F_2) \tag{34}$$

is met in order to avoid spectral leakage when squaring the signal.

5.4. A few remarks

The idea of condensing the information displayed over the $(f;\alpha)$ by integrating along the f variable has been proposed in several earlier works [6,7,9]. However, we strongly advise the reader to perform this integration on the cyclic *coherence* rather than the cyclic *spectrum* (the reason of which is very obvious after comparing Figs. 8(a) and (b)). In particular the frequently advocated “degree of cyclostationarity index”[22] is not optimal in this respect. Moreover, if the integration is to be carried out explicitly as in Eq. (28), it does not alleviate the need of properly estimating the cyclic power spectrum, i.e. by being careful of imposing that $\Delta\alpha \ll \Delta f$.

On the other hand, the squared-envelope spectrum (31) eludes completely this estimation issue. Very interestingly the squared-envelope spectrum[23] (also named the “high-resonance frequency technique” or “demodulate resonance analysis”) was recognised as being one of the most effective diagnostic tool for rolling-element bearings well before the introduction of cyclostationarity [23–25]. Yet cyclostationarity probably provides for the first time the theoretical framework to formally justify the effectiveness of this empirical technique [8]. Among other things, such a framework has allowed the derivations of the decision rule (32) associated with the squared-envelope.

[19]Note this is a *squared-amplitude* spectrum (signal-units^2) as opposed to the *power* spectra (signal-units^2/Hz) introduced in Section 2.
[20]A fast solution to whiten a signal is by means of the Levinson–Durbin algorithm.
[21]Pre-whitening may be needless when the $[F_1; F_2]$ selection is very narrow.
[22]The degree of cyclostationarity index is defined as [21]:

$$\mathrm{DCS}(\alpha) = \int_0^{F_s/2} |S_x(f;\alpha)|^2\,\mathrm{d}f \Big/ \int_0^{F_s/2} |S_x(f;0)|^2\,\mathrm{d}f.$$

[23]In the past the envelope was often computed by rectifying the signal. The advantage of taking the squared-magnitude signal instead is discussed in Ref. [22].

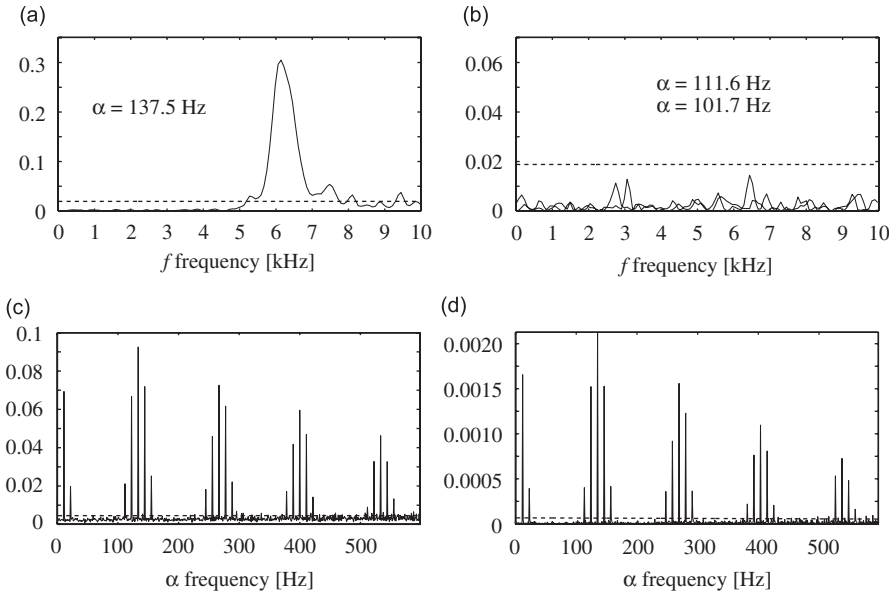

Fig. 10. Estimated cyclic coherences (squared-magnitude) of Example 2 at (a) α equal to the inner-race fault frequency and (b) α equal to the outer-race and rolling-element fault frequencies, together with the 0.1%-statistical thresholds (dotted lines). (b) Integrated cyclic coherence together with its 0.1%-statistical threshold (dotted line). (c) Squared-envelope spectrum of the same signal taken in the band $[5;8]$ kHz together with its 0.1%-statistical threshold (dotted line).

A few words must be said, finally, on the inherent difficulty associated with the selection of the frequency band $[F_1;F_2]$ where to demodulate the signal before computing the squared-envelope spectrum. Ideally this frequency band would be obtained as that which maximises the cyclic coherence values, i.e. the signal-to-noise ratio; but the computation of the cyclic coherence is specifically what we have tried to avoid by the introduction of the sub-optimal squared-envelope spectrum (31)! A sounder solution consists in comparing the actual vibration spectrum with a template spectrum constructed from historical data, and then to select the frequency band where the difference is the greatest. Another common solution is to directly demodulate around a marked structural resonance—or the accelerometer resonance—that is likely to carry the fault impacts. Other solutions have recently been proposed to automatically find the best demodulation band, without any a priori knowledge, based on the spectral kurtosis [26,27]. Their coupling with the proposed cyclic indicators is straightforward.

5.5. *Illustrative Example 4*

The signal of Example 3 is considered again. Fig. 10(a) shows the cyclic coherence evaluated at the cyclic frequency equal to the inner-race fault frequency—this actually amounts to taking a slice of Fig. 8(b) at $\alpha = 137.5$ Hz. The coherence largely overrun the 0.1% statistical threshold in the $[5;8]$ kHz band, thus indicating the presence of a fault in that band. For sake of completeness, Fig. 10(b) also show the cyclic coherence evaluated at the theoretical outer-race and rolling-element fault frequencies ($\alpha = 111.6$ and $\alpha = 101.7$ Hz, respectively). As expected it is well confined below the statistical threshold in both cases. Fig. 10(c) shows the integrated cyclic coherence obtained by summing the cyclic coherence of Fig. 8(b) along the f-axis. It is noteworthy that this way of doing perfectly preserve the diagnostic information: the symptomatic fundamental, harmonics, and sidebands of the inner-race fault nicely appear, with a significant overrun of the 0.1% statistical threshold. Finally, Fig. 10(d) shows the squared-envelope spectrum computed after band-pass filtering the signal in the $[5;8]$ kHz band where the signal-to-noise ratio was found the highest. As argued above, the squared-envelope spectrum is very comparable to the integrated cyclic coherence.

6. Experimental results

This section illustrates the use of the cyclic coherence, and of the simplified vectorial indicators that have been derived from it, on experimental data. All data have been collected at the Laboratory of Vibration and Acoustics of the University of New South Wales (Sydney), under the supervision of Professor Bob Randall.

6.1. First experiment

This first experiment is concerned with bearing signals recorded on rail vehicle bearings, which have been overloaded on a test rig so that faults have developed. Details of the geometry of the bearings with the

Table 4
Bearing characteristics of Experiment 1

Dimensions	
Speed of shaft (Ω) (rev/min)	650
Bearing roller diameter (d) (mm)	21.4
Pitch circle diameter (D) (mm)	203
Number of rolling elements (n)	23
Contact angle (θ) (deg)	9.0
Expected fault frequencies	
Inner-race fault frequency (Hz)	120.74
Outer-race fault frequency (Hz)	97.76
Rolling-element fault frequency (Hz)	89.34
Cage frequency (Hz)	4.25

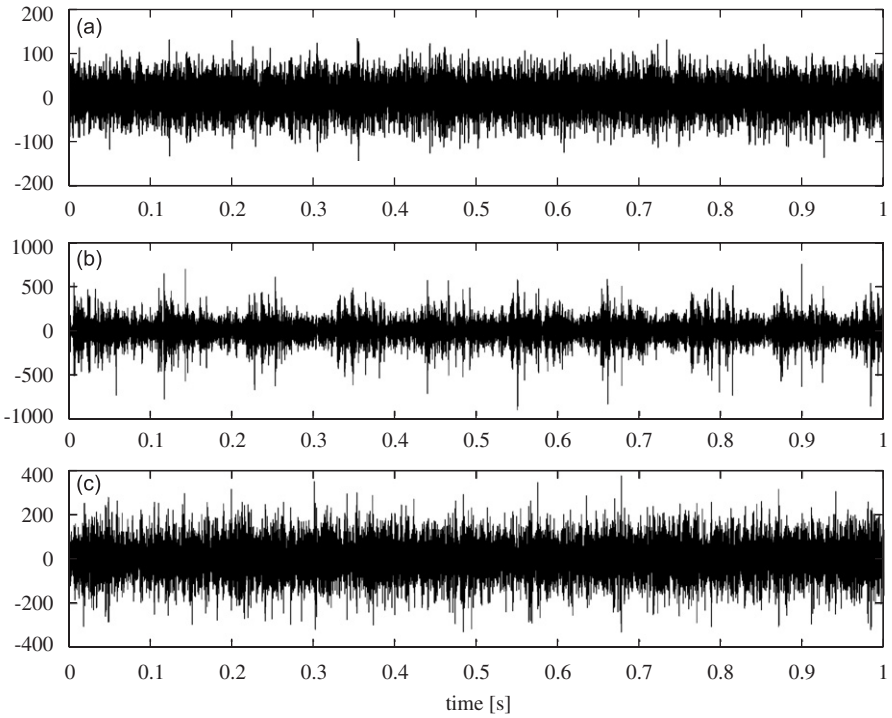

Fig. 11. Vibration signals of Experiment 1: (a) reference, (b) inner-race fault, (c) outer-race fault.

Table 5
Estimation parameters of Experiment 1

Sampling frequency (F_s) (Hz)	32 768
Record length (L)	32 768 samples
Window type	Hanning
Overlap	2/3
Window length (N_w)	128 samples
f frequency resolution (Δf) (Hz)	384
α frequency resolution ($\Delta\alpha$) (Hz)	1
Scrutinised cyclic frequency range[a]	[1;350]
Variance reduction factor ($\mathscr{E}$)	0.0019

[a]The zero cyclic frequency is excluded as it produces a constant unit-valued cyclic coherence.

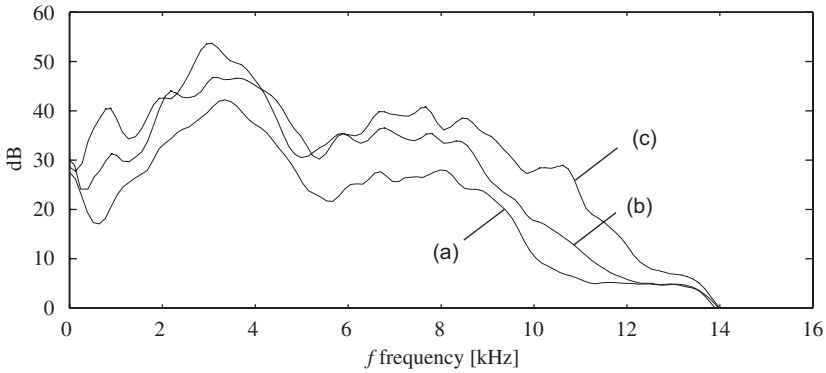

Fig. 12. Power spectral densities of (a) the reference, (b) the inner-race fault, and (c) the outer-race fault signals of Experiment 1.

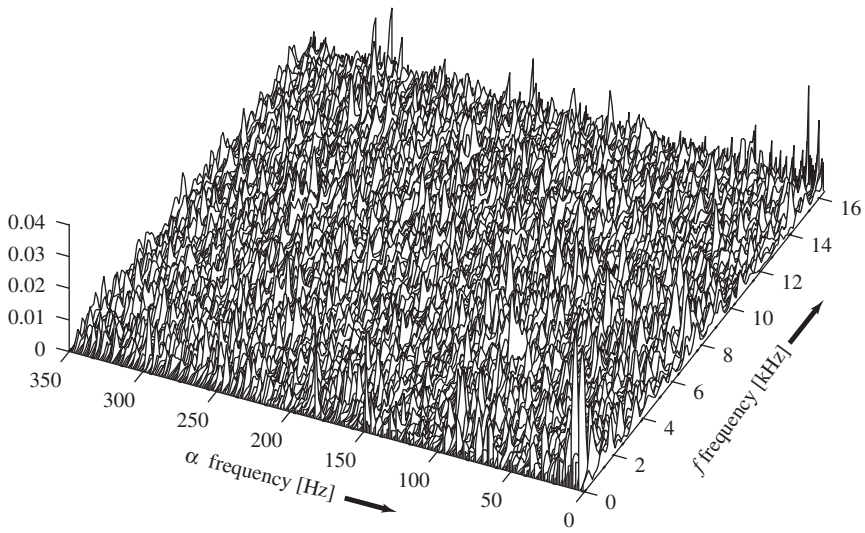

Fig. 13. Estimated cyclic coherence (squared-magnitude) of the reference vibration signal of Experiment 1 (0.1%-statistical threshold = 0.0131).

expected fault frequencies are given in Table 4. Fig. 11 displays the recorded time signals in the case of (a) a good bearing, (b) a bearing that has developed an inner-race fault, and (c) a bearing with an outer-race fault. Clearly, the temporal analysis of the vibrations does not help to provide a diagnosis: at first sight all signals

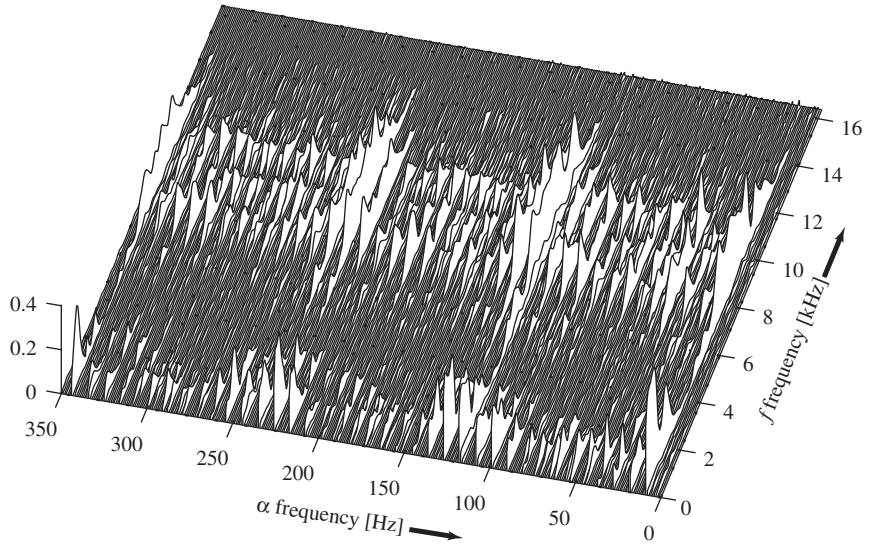

Fig. 14. Estimated cyclic coherence (squared-magnitude) of the vibration signal of Experiment 1 with an inner-race fault (0.1%-statistical threshold = 0.0131).

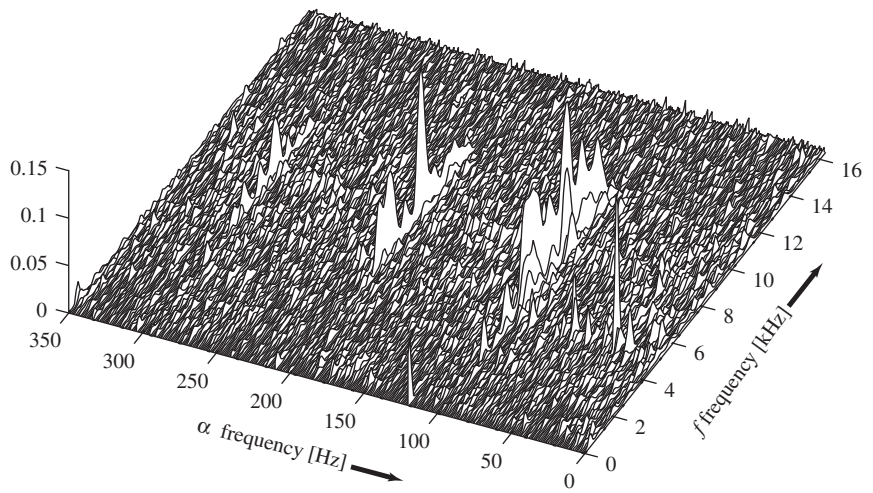

Fig. 15. Estimated cyclic coherence (squared-magnitude) of the vibration signal of Experiment 1 with an outer-race fault (0.1%-statistical threshold = 0.0131).

look gently stationary, and the slight modulation of signal (b) which coincides with the shaft rotation at 10.8 Hz is quite typical even in normal condition. The power spectral densities (computed with the parameters listed in Table 5) of the vibration signals are displayed in Fig. 12. It is seen that signals (b) and (c) evidence some significant changes in their spectral distribution, with a maximum relative increase of energy within the 6;12 kHz frequency band. These changes would tell the experienced diagnostician that something abnormal has happened, however, they are still not enough to diagnose the faults.

The cyclic coherences were then computed with the parameters displayed in Table 5. They are shown in Figs. 13–15. The signal issued from a good bearing—Fig.13—is seen to have a very stationary behaviour since its cyclic coherence is statistically zero: most of its values (in theory 99.9%) remain below the 0.1% statistical threshold of 0.0131. This is in agreement with our fundamental assumption that a good bearing should not produce significant cyclostationarity. On the other hand, the cyclic coherences of the damaged bearings—Figs. 14 and 15—evidence significant cyclostationary behaviours that are well above the 0.0131 threshold.

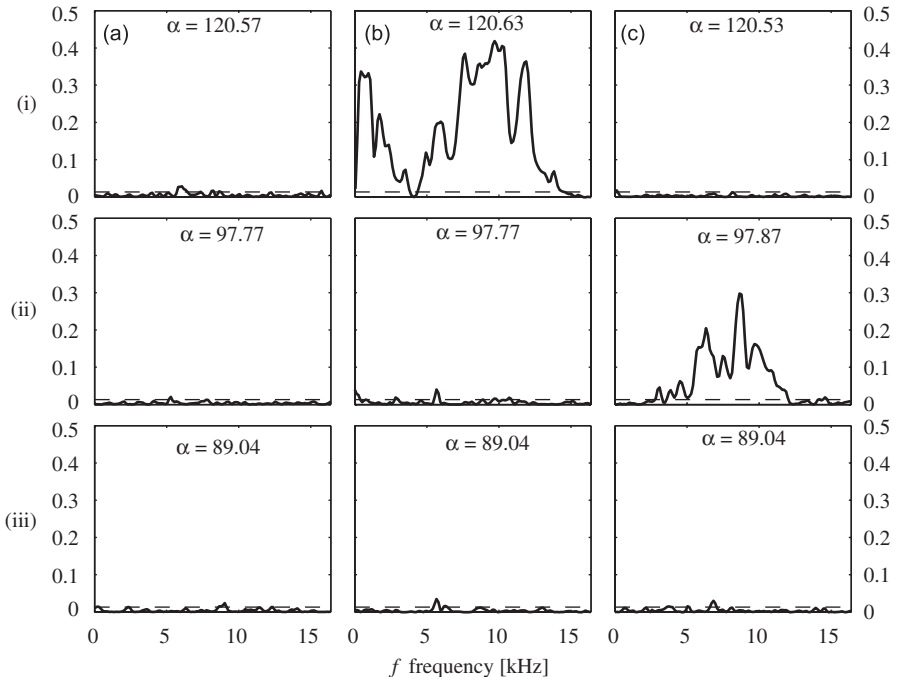

Fig. 16. Estimated cyclic coherence (squared-magnitude) of (a) the reference, (b) the inner-race fault and (c) the outer-race fault signal of Experiment 1 evaluated at the expected (i) inner-race, (ii) outer-race and (iii) rolling-element fault frequencies. In each case the cyclic coherence is returned at that cyclic frequency which maximises it in a narrow cyclic frequency band around the expected value. The 0.1%-statistical threshold (dotted lines) is 0.0131.

The cyclic coherence in Fig. 14 shows a fundamental spectral line at $\alpha = 120.63$ Hz plus its harmonics: according to Table 4 this tells the occurrence of an inner-race fault. Furthermore, as expected from the model of Section 2, the fault signal is strongly modulated by the shaft speed due to the periodic passing of the fault in the load zone; this explains the numerous symptomatic sidebands around the fault frequencies, that are spaced apart by 10.8 Hz. Besides, the cyclic coherence in Fig. 15 shows a fundamental spectral line at $\alpha = 97.87$ Hz plus its harmonics, without sidebands: this is the characteristic of an outer-race fault.

As explained in Section 2, when read as a function of f, the cyclic coherence points out the frequency bands where the signal-to-noise ratio of the fault is maximum. In the cyclic coherences of Figs. 14 and 15, such bands are located within [0;3] kHz and [6;12] kHz, and within [5;12] kHz, respectively. They surely correspond to structural resonances that are excited by the impacts of the rolling elements on the faults. These observations are coherent with the frequency regions of maximum variations that were observed in the power spectral densities of Fig. 12. However, it should be emphasised that these frequency bands are readily available in the cyclic coherence without requiring any comparison with historical data, as required for instance when comparing power spectral densities.

Fig. 16 displays in a matrix the cyclic coherences of (a) the reference, (b) the inner-race fault, and (c) the outer-race fault signals evaluated at the expected cyclic frequencies of (i) an inner-race, (ii) an outer-race, and (iii) a rolling-element fault. Because the characteristic fault frequencies are only predictable within certain limits, the cyclic coherence with maximum value in a narrow neighbourhood centred on the theoretical frequency was actually returned in each case. Inspection of these indicators leads, yet in a lighten way, to similar observations as obtained from the full (3D) cyclic coherences. In particular the inner-race fault in signal (b) and the outer-race fault in signal (c) are clearly detected from abnormally high values above the 0.1% statistical threshold at the corresponding

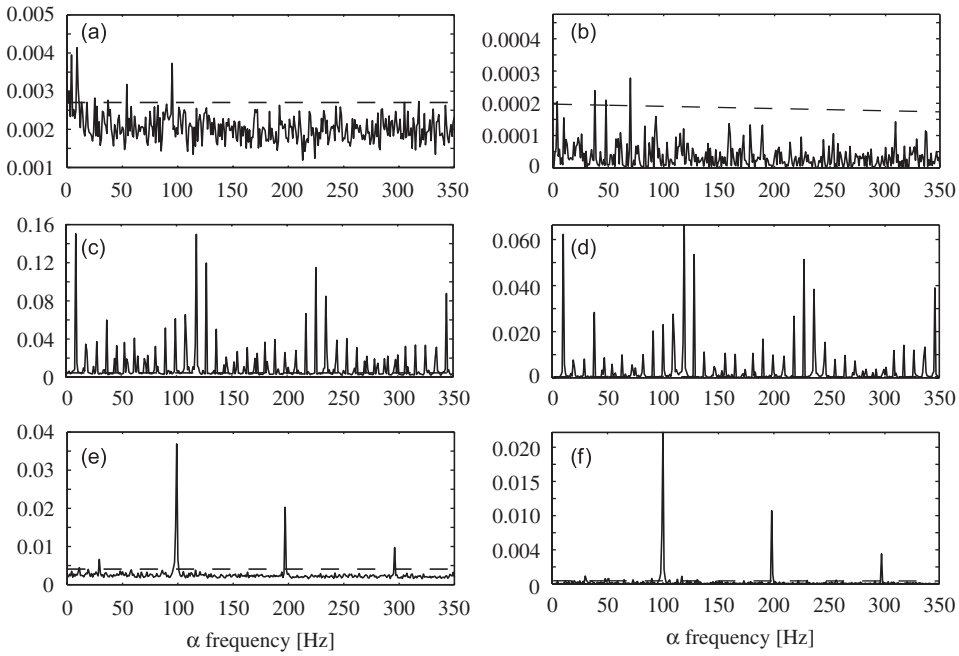

Fig. 17. Integrated cyclic coherence of (a) the reference signal, (b) the inner-race fault, (c) the outer-race fault (resp., squared-envelope spectrum in (d–f)) of Experiment 1. The frequency band of interest is [6; 12] kHz.

expected fault frequencies—see cells (b-i) and (c-ii). On the other hand, the reference signal (a) does not show any significant cyclostationary activity at any fault frequency.

Finally, the integrated cyclic coherence and the squared-envelope spectrum were computed. The former was considered here only to corroborate our claim that it gives nearly the same results as the latter. The frequency band of interest was set to [6;12] kHz because it is where the signal-to-noise ratio was previously found to be maximal (either from comparing the power spectra of Fig. 12, or from inspecting the cyclic coherences of Figs. 13–15 or 16). Fig. 17(a–f) shows that both indicators essentially lie below their statistical threshold when the bearing is in good condition,[24]whilst they display values well above it when a fault has developed. Here again the typical spectral signatures (fundamental + harmonics + possible sidebands) of the inner-race and outer-race faults are easily recognised. It is seen that the squared-envelope spectrum shows little difference with the integrated cyclic coherence so that it could very well replace it in actual applications, as it is much faster to compute. Indeed, if the relevant [6; 12] kHz frequency band was obtained by some other means than cyclic spectral analysis (e.g. by means of the spectral kurtosis [27]), the squared-envelope spectrum would certainly replace the cyclic coherence analysis as well (at least for diagnostics)!

6.2. Second experiment

This second experiment is concerned with bearing signals recorded on a two-stage gearbox where one of the bearings supporting the primary shaft was purposely damaged by introducing an inner-race fault, an outer-race fault, and a rolling-element fault. Details of the geometry of the bearings with the expected fault

[24]This is except a few isolated values that overrun the statistical threshold, however in the allowable 0.1% probability of false alarm—i.e. 0.1% of the spectral lines over the displayed frequency range will overrun the threshold on the average. Obviously much looser thresholds could be used in practice if diminishing the rate of false alarm was of concern: this would be readily achieved by increasing the statistical threshold to a higher value. Incidentally, from our experience the chi-square rules of Eqs. (27), (30) and (32) tend to produce conservative values, which means that they give slightly larger rates of false alarm than specified. This seems all the more true as the stationary assumption (no faults) under which they have been derived is not strictly respected in practice, such as in the presence of impulsive noise or other non-stationary sources. In such cases an *ad hoc* correction factor may be required.

Table 6
Bearing characteristics of Experiment 2

Dimensions	
Speed of shaft (Ω) (Hz)	10
Bearing roller diameter (d) (mm)	7.12
Pitch circle diameter (D) (mm)	38.5
Number of rolling elements (n)	12
Contact angle (θ) (deg)	0
Expected fault frequencies	
Inner-race fault frequency (Hz)	71.10
Outer-race fault frequency (Hz)	48.90
Rolling-element fault frequency (Hz)	52.22
Cage frequency (Hz)	4.08

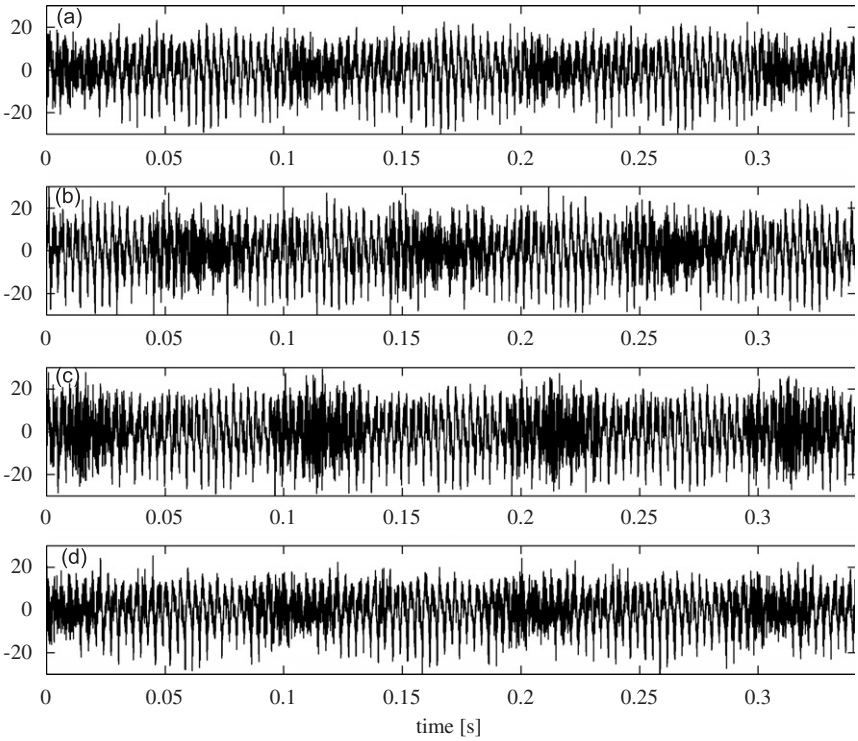

Fig. 18. Vibration signals of Experiment 2: (a) reference, (b) inner-race fault, (c) outer-race fault, and (d) rolling-element fault.

frequencies are given in Table 6. Fig. 18 displays the recorded time signals in the case of (a) the good bearing, (b) the inner-race fault, (c) the outer-race fault, and (d) the rolling-element fault. Just as in the first experiment, the time history of the vibration signals look rather stationary and are useless, as such, for diagnostics. Into addition, they seem to contain numerous harmonic interferences due to gears operation. Fig. 19 shows the power spectral densities computed with the parameters of Table 7. It is seen that the low-frequency part of the spectra is dominated by gear-mesh harmonics and sidebands (clearly evidenced by using a finer frequency resolution Δf), so that the diagnostic information is likely to be masked below 10 kHz. The dB-difference spectra between the faulty cases and the good case—lower panel of Fig. 19—confirm the former assertion.

The cyclic coherences were then computed with the parameters of Table 7. They are shown in Figs. 20–23. The signal issued from a good bearing—Fig. 20—is seen to have a fairly stationary behaviour in the frequency

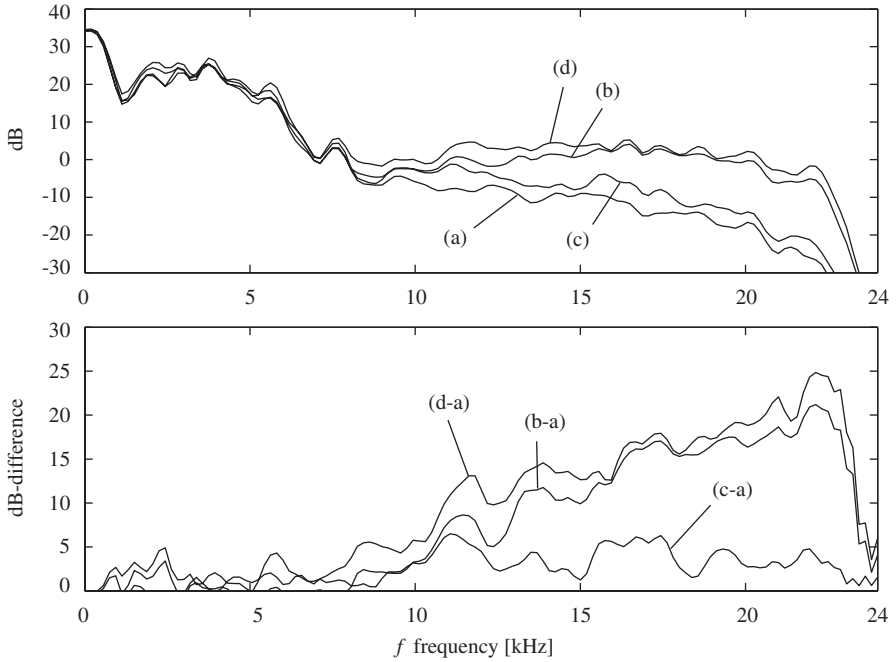

Fig. 19. Upper panel: power spectral densities of (a) the reference, (b) the inner-race fault, (c) the outer-race fault, and (d) the rolling-element fault signals of Experiment 2. Lower panel: dB-differences of the above power spectra w.r.t. the reference signal.

Table 7
Estimation parameters of Experiment 2

Sampling frequency (F_s) (Hz)	48 000
Record length (L)	65 536 samples
Window type	Hanning
Overlap	2/3
Window length (N_w)	128 samples
f frequency resolution (Δf) (Hz)	562.5
α frequency resolution ($\Delta\alpha$) (Hz)	0.73
Scrutinised cyclic frequency range[a] (Hz)	[0.73;256]
Variance reduction factor ($\mathscr{E}$)	0.0011

[a]The zero cyclic frequency is excluded as it produces a constant unit-valued cyclic coherence.

range above 10 kHz: the spectral lines in the lower-frequency range are recognised to be harmonics of the shaft speed and sidebands of the gear-mesh frequency. As explained in Section 3, they should ideally be removed from the signal before computing the cyclic coherence because they might produce interferences on the spectral signature of the fault; however since they fall in a different frequency range (<10 kHz) than where the diagnostic information is expected (>10 kHz), this precaution is needless here. The cyclic coherence of the inner-race fault—Fig. 21—shows a fundamental spectral line at $\alpha = 71.10$ Hz plus its harmonics, with small sidebands at 10 Hz. This is in excellent agreement with the expected fault frequency given in Table 6. The fact that the side-bands have low amplitudes is because the axial load was small. Into addition the cyclic coherence shows very high values, close to one, at $\alpha = 71.10$ Hz and $f > 10$ kHz; this indicates a very high signal-to-noise ratio in the high-frequency range, where the operation of the gears has few effect. The cyclic coherence of the outer-race fault—Fig. 22—also shows spectral lines positioned exactly at the expected fault frequency (48.90 Hz+harmonics), as given in Table 6. Finally, the cyclic coherence of the rolling-element

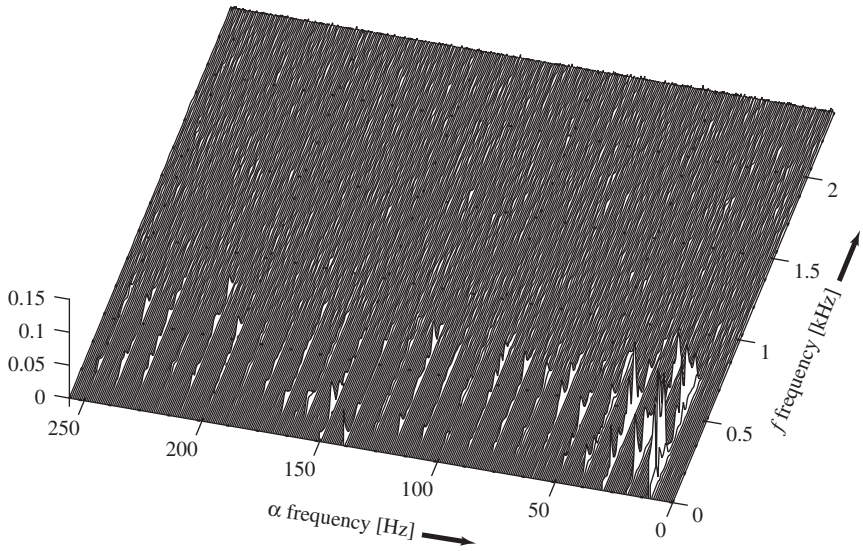

Fig. 20. Estimated cyclic coherence (squared-magnitude) of the reference vibration signal of Experiment 2 (0.1%-statistical threshold = 0.0073).

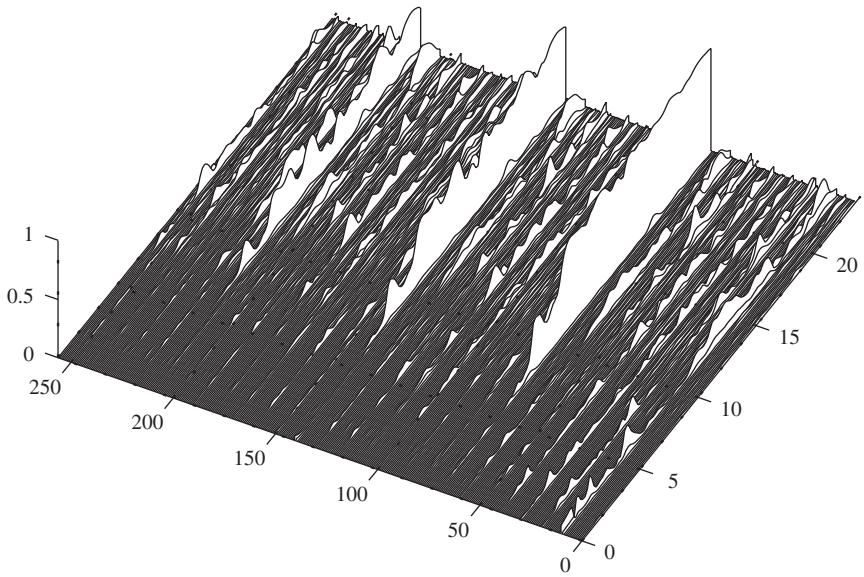

Fig. 21. Estimated cyclic coherence (squared-magnitude) of the vibration signal of Experiment 2 with an inner-race fault (0.1%-statistical threshold = 0.0073).

fault—Fig. 23—bears a very similar interpretation as that of the inner-race, where the fault frequency is 52.22 Hz and where the sidebands due to shaft modulation are replaced by numerous sidebands due to the cage modulation (4.08 Hz), again as expected from the theory.

The presence of these faults is similarly detected when the cyclic coherence is computed only at the characteristic cyclic frequencies where the faults are supposed to occur. This is illustrated in Fig. 24 which reads as Fig. 16. The inner-race fault in signal (b), the outer-race fault in signal (c) and the rolling-element fault in signal (d) are clearly detected from abnormally high values above the 0.1% statistical

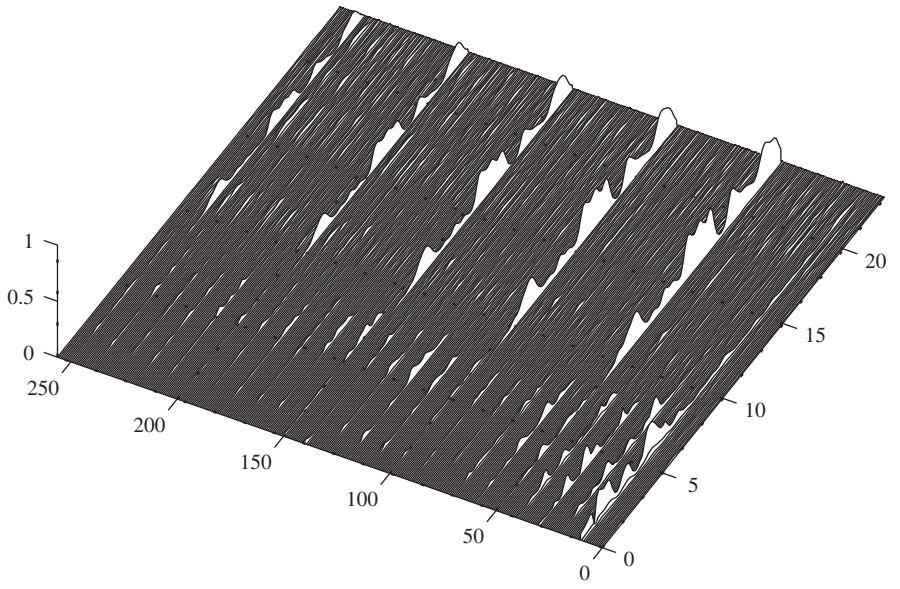

Fig. 22. Estimated cyclic coherence (squared-magnitude) of the vibration signal of Experiment 2 with an outer-race fault (0.1%-statistical threshold = 0.0073).

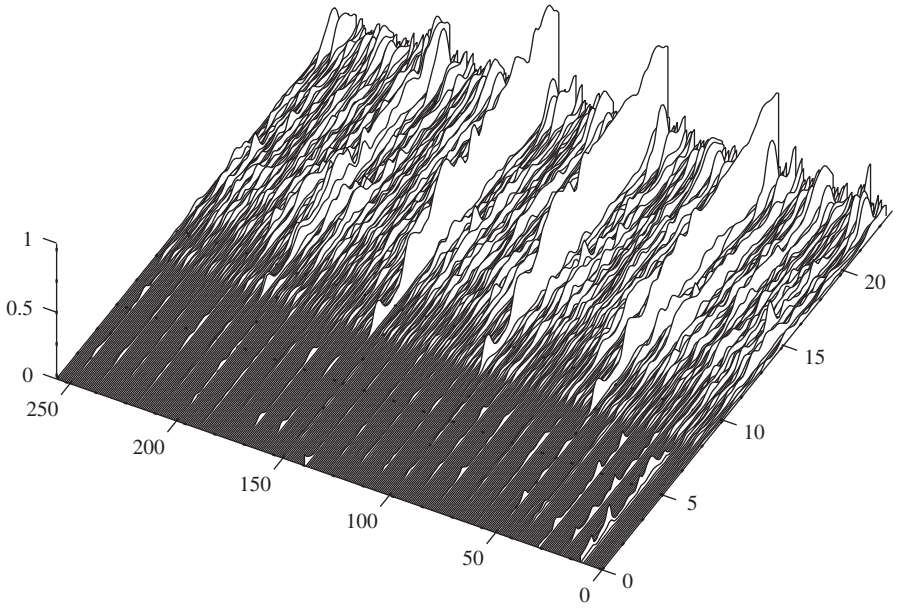

Fig. 23. Estimated cyclic coherence (squared-magnitude) of the vibration signal of Experiment 2 with a rolling-element fault (0.1%-statistical threshold = 0.0073).

threshold[25] at the corresponding expected fault frequencies—see cells (b-i), (c-ii) and (d-iii). It is noteworthy that the rolling-element fault also exhibit significant a cyclostationary activity near the inner-race and outer-race fault frequencies; this is because it produces so many tight sidebands due to the cage modulation that

[25]Note that in this experiment the 0.1% statistical threshold is more severe than in the previous experiment (given the same probability of false alarm). This is because longer records are analysed, which allows a lower estimation variance of the cyclic coherence.

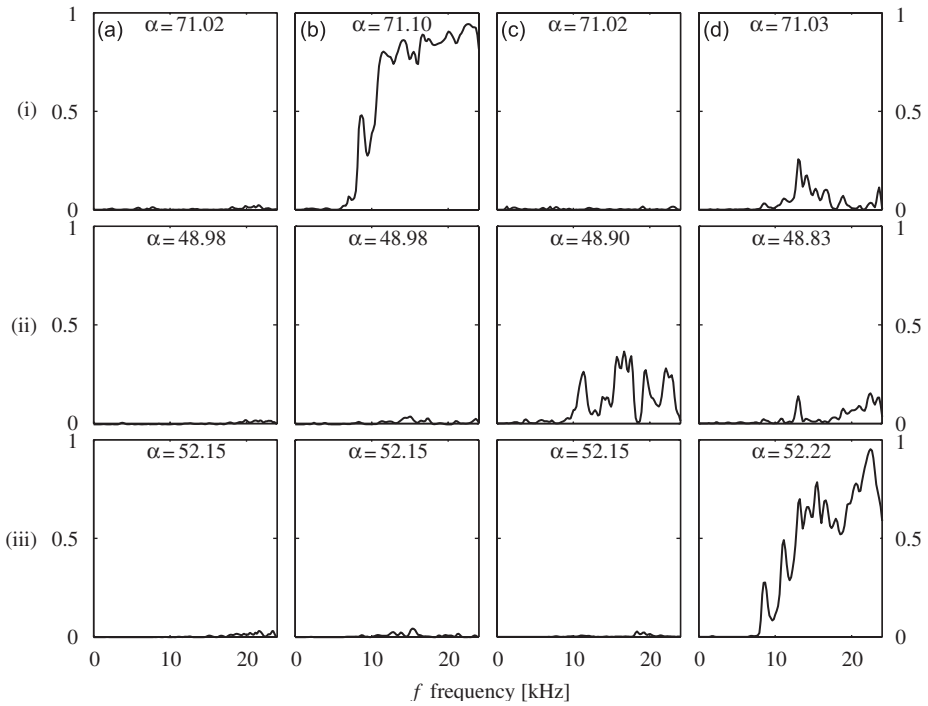

Fig. 24. Estimated cyclic coherence (squared-magnitude) of (a) the reference, (b) the inner-race fault, (c) the outer-race fault and (d) the rolling-element fault signal of Experiment 1 evaluated at the expected (i) inner-race, (ii) outer-race and (iii) rolling-element fault frequencies. In each case the cyclic coherence is returned at that cyclic frequency which maximises it in a narrow cyclic frequency band around the expected value. The 0.1%-statistical threshold is 0.0073.

some of them happen to fall very close to those latter frequencies. This phenomenon was clearly visible in Fig. 23. On the other hand, the reference signal (a) does not show any significant cyclostationary activity at any fault frequency. Incidentally, the similitude in shape and in relative amplitude between the cyclic coherences of Fig. 24(a–c) and the dB-differences in spectra of Fig. 19(b) is striking. We insist again on the remarkable fact that the formers can be obtained without the need of historical data, contrary to the latters.

Finally, Fig. 25(a–h) compare the integrated cyclic coherence and the squared-envelope spectrum. Just as in the previous experiment, very similar results are obtained for both techniques in the frequency band $[10;24]$ kHz where the signal-to-noise is maximised (as observed either from comparing the power spectra of Fig. 19, or from inspecting the cyclic coherences of Figs. 20–23 or Fig. 24(a–c)). Due to the excellent signal-to-noise ratio in that frequency band, these indicators show values that are well above the statistical threshold. The characteristic patterns of the fault signatures are perfectly recognised in the three cases.

7. Conclusion

The intent of this paper was to gather a number of results concerning the cyclic spectral analysis of rolling-element bearings signals, for use in diagnostics. Some of these results are new and presented here for the first time, while others have been published in earlier works but are purposely repeated in order to correct a few misconceptions that seem to have survived in the specialised literature. First of all we have shown that vibrations produced by faulty rolling-element bearing are essentially random cyclostationary (instead of harmonic) in the high-frequency range where the signal-to-noise ratio is maximal, thus explaining why classical spectral analysis often turns out a failure. Second, we have proved the diagnostic information is perfectly preserved in the cyclic frequency domain in the form of a symptomatic pattern of spectral lines. We have then demonstrated that, given a finite-length record of data, the cyclic coherence (power-normalised

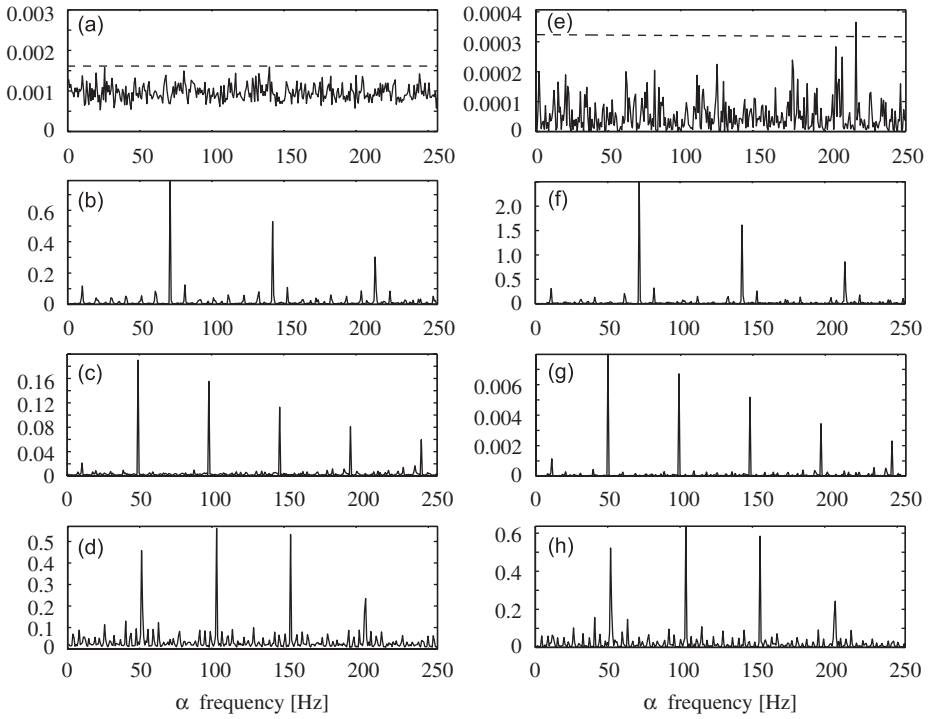

Fig. 25. Integrated cyclic coherence of (a) the reference signal, (b) the inner-race fault ($\alpha = 71.10$ Hz), (c) the outer-race fault ($\alpha = 48.90$ Hz), (d) the rolling-element fault ($\alpha = 52.22$ Hz) (resp., squared-envelope spectrum in (e–h)) of Experiment 2. The frequency band of interest is [10; 24] kHz.

Table A.1
Typical fault frequencies

Inner-race fault	$\frac{n}{2}\Omega(1+\frac{d}{D}\cos\theta)$
Outer-race fault	$\frac{n}{2}\Omega(1-\frac{d}{D}\cos\theta)$
Rolling-element fault	$\frac{\Omega d}{D}(1-(\frac{d}{D}\cos\theta)^2)$
Cage fault	$\frac{\Omega}{2}(1-\frac{d}{D}\cos\theta)$
Inner-race modulation	Ω
Cage modulation	$\frac{\Omega}{2}(1-\frac{d}{D}\cos\theta)$

where Ω =speed of shaft; d = bearing roller diameter; D =pitch circle diameter; n =number of rolling elements; θ =contact angle.

cyclic power spectrum) is the cyclic spectral quantity that offers the best detection capability of the fault. In short, when read as a function of the cyclic frequency α, the cyclic coherence reveals the discrete spectral signature of the fault, whilst when read as function of the f frequency, its magnitude serves as a relative measure of the fault severity. This has been formalised through the proposal of an optimal statistical test which consists in comparing the value of the estimated cyclic coherence against a statistical threshold that depends on the record length and on a user-defined probability of false alarm. This test can be used as such in an automated diagnostics system, yet we have also proposed simplified (but sub-optimal) versions of it that are computationally more advantageous by avoiding the exhaustive exploration of the (α,f) frequency domain. Indeed, one important conclusion of the paper is that the familiar squared-envelope spectrum turns out to be a special case of the cyclic coherence that returns just as good results, provided that it benefits from the adequate information to be properly computed (i.e. knowledge of the demodulation band where the signal-to-noise ratio is maximal). This observation has been supported by theoretical arguments as well as intensive experimental results. Therefore the reason of not using

the squared-envelope spectrum—on behalf of more complex indicators—is probably questionable in many instances.

Appendix A. Characteristic frequencies of rolling-element bearing faults

The following equations (Table A.1) assume the inner-race rotates at the shaft speed and the outer-race is stationary.

Appendix B. Proof of Eq. (26)

Let us denote by $\mathbf{X}_{N_w}(f)$ the $K \pm 1$ vector of spectral components $X^{(k)}_{N_w}(f)$, $k = 0, \ldots, K-1$ as defined in Eq. (18). Then, under H_0, $\mathbf{X}_{N_w}(f+\alpha/2)$ is uncorrelated with $\mathbf{X}_{N_w}(f-\alpha/2)$ for any f, whereas under H_1 the two vectors must share some correlation over a non-empty set of f frequency values. Testing for H_1 against H_0 then amounts to detecting the existence of a statistical correlation between $\mathbf{X}_{N_w}(f+\alpha/2)$ and $\mathbf{X}_{N_w}(f-\alpha/2)$. There are a variety of possible statistical tests to do so. The generalised likelihood ratio test is optimal in the sense that it maximises the probability of detection given a probability of false alarm, and the union intersection test in the sense that it is invariant under any scaling of the data $\mathbf{X}_{N_w}(f)$. Now considering that $\mathbf{X}_{N_w}(\cdot)$ tends to Gaussianity in virtue of the Central Limit theorem applied to the discrete Fourier transform—at least in the frequency range of interest where the random part of the signal was shown to predominate—both tests can be shown to consist in comparing[26]

$$\frac{|\mathbf{X}_{N_w}(f-\alpha/2)^{\mathrm{H}}\mathbf{X}_{N_w}(f+\alpha/2)|^2}{\|\mathbf{X}_{N_w}(f-\alpha/2)\| \cdot \|\mathbf{X}_{N_w}(f+\alpha/2)\|} \tag{B.1}$$

to some statistical threshold that depends on the estimation parameters. Note that Eq. (B.1) is nothing else than the averaged cyclic periodogram estimate of the cyclic coherence (13). The latter divided by $\mathscr{E}/2$ can be shown to be approximatively chi-square distributed with two degrees of freedom [17], so that the $100(1-p)\%$ threshold is as given in Eq. (27).

Appendix C. Proof of Eq. (29)

According to the preceding discussion, $2\mathrm{ICC}^{(L)}_x(\alpha)/\mathscr{E}$ is a sum of $K_2(\alpha) - K_1(\alpha) + 1$ chi-square random variables with two degrees of freedom each. However, if $N_{\mathrm{FFT}} > N_w$ in Eq. (18), then only N_{FFT}/N_w of the variables in the sum are independent due to frequency interpolation in the DFT. Hence the sum is chi-square distributed with $\lfloor 2(K_2(\alpha) - K_1(\alpha) + 1)N_{\mathrm{FFT}}/N_w \rfloor$ degrees of freedom. The values of $K_2(\alpha)$ and $K_1(\alpha)$ are obtained as illustrated in Fig. 9, where care is taken to perform the integration only over the principal domain (25).

Reproducible research

The algorithms used to obtain the figures presented in this paper are available upon request to the author (antoni@utc.fr), or downloadable at www.utc.fr\~antoni. So do the detailed proofs of our theoretical results.

References

[1] N. Tandon, A. Choudhury, A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings, *Tribology International* 32 (1999) 469–480.
[2] J.I. Taylor, Identification of bearing defects by spectral analysis, *Journal of Mechanical Design* 102 (1980) 199.
[3] J. Mathew, R.J. Alfredson, The condition monitoring of rolling element bearings using vibration analysis, *Journal of Vibration, Acoustics, Stress, and Reliability in Design* 106 (1984) 447.

[26] The symbol $^{\mathrm{H}}$ stands for the transpose conjugate operator, and $\|\cdot\|$ for the euclidian norm of a vector.

[4] Y.-T. Su, S.J. Lin, On the initial fault detection of a tapered roller bearing: frequency domain analysis, *Journal of Sound and Vibration* 155 (1) (1992) 75–84.
[5] J. Antoni, F. Bonnardot, A. Raad, M. El Badaoui, Cyclostationary modelling of rotating machine vibration signals, *Mechanical Systems and Signal Processing* 18 (6) (2004) 1285–1314.
[6] A.C. McCormick, A.K. Nandi, Cyclostationary in rotating machine vibrations, *Journal of Sound and Vibration* 12 (2) (1998) 225–242.
[7] I. Antoniadis, G. Glossiotis, Cyclostationary analysis of rolling-element bearing vibration signals, *Journal of Sound and Vibration* 248 (5) (2001) 829–845.
[8] R.B. Randall, J. Antoni, S. Chobsaard, The relationship between spectral correlation and envelope analysis for cyclostationary machine signals: application to ball bearing diagnostics, *Mechanical Systems and System Processing* 15 (5) (2001) 945–962.
[9] C. Capdessus, M. Sidahmed, J.-L. Lacoume, Cyclostationary processes: application in gear faults early diagnosis, *Mechanical Systems and Signal Processing* 14 (3) (2000) 371–385.
[10] J. Antoni, R.B. Randall, Differential diagnosis of gear and bearing faults, *ASME Journal of Vibration and Acoustics* 124 (2) (2002) 165–171.
[11] W.A. Gardner, Introduction to Random Processes with Applications to Signals and Systems, second ed., McGraw-Hill, 1990 (Chapter 12).
[12] G. Giannakis, Cyclostationary signal analysis, *The Digital Signal Processing Handbook*, CRC Press and IEEE Press, New York, 1998 (Chapter 5).
[13] J. Antoni, R.B. Randall, A stochastic model for simulation and diagnostics of rolling element bearings with localized faults, *ASME Journal of Vibration and Acoustics* 125 (3) (2003) 282–289.
[14] P.D. McFadden, J.D. Smith, Model for the vibration produced by a single point defect in a rolling element bearing, *Journal of Sound and Vibration* 96 (1) (1984) 69–82.
[15] D. Ho, R.B. Randall, Manifestation of bearing fault vibrations in gearboxes, *Sixth International Congress on Sound and Vibration*, 5–8 July 1999, Copenhagen, Denmark.
[16] D. Ho, R.B. Randall, Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals, *Mechanical Systems and Signal Processing* 14 (5) (2000) 763–788.
[17] J. Antoni, Cyclic spectral analysis in practice, *Mechanical Systems and Signal Processing* 21 (2) (2007) 597–630.
[18] J. Antoni, R.B. Randall, On the use of the cyclic power spectrum in rolling element bearing diagnostics, *Journal of Sound and Vibration* 281 (1–2) (2005) 463–468.
[19] L. Li, L. Qu, Cyclic statistics in rolling bearing diagnosis, *Journal of Sound and Vibration* 267 (2) (2003) 253–265.
[20] J. Antoni, R.B. Randall, Unsupervised noise cancellation for vibration signals, *Mechanical Systems and Signal Processing* 18 (1) (2004) 89–117.
[21] G. Ivanovi, W. Gardner, Degrees of cyclostationarity and their application to signal detection and estimation, *Signal Processing* 22 (1991) 287–297.
[22] R.B. Randall, Developments in digital analysis techniques for diagnostics of bearings and gears, *Fifth International Congress on Sound and Vibration*, December 15–18, 1997, Adelaide, Australia
[23] M.S., Darlow, R.H. Badgley, Applications for early detection of rolling element bearing failures using the high-frequency resonance technique, ASME Paper 75-DET-46, 1975.
[24] D. Dyer, R.M. Stewart, Detection of rolling element bearing damage by statistical vibration analysis, *Journal of Mechanical Design* 100 (1978) 229–235.
[25] P.D. McFadden, J.D. Smith, Vibration monitoring of rolling element bearings by the high frequency resonance technique—a review, *Tribology International* 117 (1) (1984) 3–10.
[26] J. Antoni, R.B. Randall, The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines, *Mechanical Systems and Signal Processing* 20 (2) (2006) 308–331.
[27] J. Antoni, Fast computation of the kurtogram for the detection of transient faults, *Mechanical Systems and Signal Processing* 21 (2) (2007) 108–124.