Markov Modeling of Signal Condition Transitions for Bearing Diagnostics Under External Interference Conditions

Peng Chen, *Member, IEEE*, Yuhao Wu, Chaojun Xu, Yaqiang Jin, and Chengning Zhou

***Abstract*— In real-world scenarios, the performance of rolling bearings can be significantly affected by frequent environmental disturbances, including high-energy fluctuations, transient noises, unintended collisions, or sudden variations in loading. These disturbances have the potential to obscure signs of damage, posing a challenge to the identification of failures. Traditional techniques for selecting demodulation band encounter challenges due to inherent complexity and parameter sensitivity. These methods, which heavily rely on postprocessing signal analysis involving interference components, prove ill-suited for complex scenarios, especially faced with multivariate random pulse noise. To address these challenges, this article proposes a method for Markov modeling of signal condition transitions to mitigate transient noise interference at its source. This method aims to improve insensitivity for external transient noise, enabling a more accurate and reliable selection of demodulation bands. The proposed method employs a signal transition model grounded in a Markov transition matrix, coupled with wavelet transforms and reconstruction. It improves the identification of anomalous signal components and leverages the transition matrix to track temporal state changes. In addition, an amplitude interference-limiting mechanism is designed to identify and mitigate transient noise that may adversely affect the demodulation band selection process. The study's results demonstrate that this mechanism allows typical Kurtogram techniques to accurately identify bearing fault-characteristic bands, and the experimental results validate the effectiveness of the proposed methodology for fault diagnosis of rolling bearings.**

***Index Terms*— Anomaly detection, fault diagnosis, rolling bearings, transient noise, vibration signal.**

Manuscript received 16 December 2023; revised 29 January 2024; accepted 6 February 2024. Date of publication 1 April 2024; date of current version 12 April 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 52105111, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2022A1515010859, in part by the Guangdong Provincial Science and Technology Special Fund Project under Grant STKJ2021171, in part by the Shantou University (STU) Scientific Research Initiation under Grant NTF21029, and in part by the Natural Science Foundation of Sichuan Province under Grant 2023NSFSC0861. The Associate Editor coordinating the review process was Dr. Ke Feng. *(Corresponding author: Peng Chen.)*

Peng Chen is with the College of Engineering and the Key Laboratory of Intelligent Manufacturing Technology, Ministry of Education of China, Shantou University, Shantou, Guangdong 515063, China (e-mail: dr.pengchen@foxmail.com).

Yuhao Wu and Chaojun Xu are with the College of Engineering, Shantou University, Shantou, Guangdong 515063, China.

Yaqiang Jin is with the Center for Structural Acoustics and Machine Fault Diagnosis, Qingdao University of Technology, Qingdao 266520, China, and also with Qingdao Mingserve Tech, Qingdao 266041, China.

Chengning Zhou is with the Nuclear Power Institute of China, Chengdu 610213, China.

Digital Object Identifier 10.1109/TIM.2024.3383497

I. Introduction

ROLLING bearings plays a crucial role as components in various rotating systems, including wind turbines, electric motors, and machine tools. The variety of speeds, loads, and environmental disturbances they endure can result in damage, potentially leading to system downtime, reduced reliability, and economic losses [1].

Currently, there are two main approaches to diagnose rolling bearing faults: vibration signal processing [2] and deep learning [3]. Deep learning methods utilize general network architectures, including convolutional neural networks (CNNs) [4], [5], residual networks (ResNets) [6], Transformers [7], and others [8], [9], to extract intricate patterns and nonlinear correlations not easily observed in vibrations. However, these methods still face challenges, including but not limited to dependence on extensive datasets, model opacity, and high-computational demands. In contrast, vibration signal processing methods focus on the inherent characteristics of bearing vibration, analyzing potential features hidden in the vibrations that indicate bearing degradation performance. By revealing typical bearing faults in the time domain, frequency domain, and even time–frequency domain, these methods benefit from the intuitiveness and interpretable nature of vibration signal processing. In addition, they are capable of delivering robust fault diagnosis, particularly in scenarios where massive datasets are not available.

In the field of bearing fault diagnosis, the detection of damage signals, such as those from fatigue, cracks, and wear, is crucial, as they appear as repetitive peaks in the time domain and distinct frequencies in the frequency domain, referencing the specific bearing components [10]. Identifying these fault-characteristic frequencies amid noise is essential. Spectral kurtosis (SK) [11] is a key advancement for analyzing vibration signals, with methods, such as the Kurtogram by Antoni and Randall, for visualizing SK, although initially limited by computational demands. The development of the Fast Kurtogram [12] improved on this with a faster processing method without sacrificing accuracy, enhancing its industrial utility. Alternative approaches include Lei et al.'s [13] wavelet packet transform, improving time–frequency analysis, and Li et al.'s [14] multiscale clustering for optimized signal feature analysis.

However, the current methods that utilize bandwidth division based on the SK are overly sensitive to external disturbances. This sensitivity results in difficulty distinguishing between faults and external transient noise, potentially leading

to misdiagnosis. Furthermore, these techniques may not effectively adapt to rapid changes in complex conditions, which heightens the risk of false alarms or overlooked diagnoses. To tackle these challenges and combat issues with non-Gaussian noise and its high-energy pulses, Barszcz and JabŁo'nski [15] proposed the Kurtogram algorithm, which focuses on cyclostationarity rather than impulsiveness, resulting in more stable outcomes. In a similar vein, Borghesani et al. [16] forwarded a technique analyzing the amplitude peaks of the squared envelope spectrum (SES) for frequency band selection, based on the ratio of cyclic contents (RCCs), which offers refined precision. Antoni [17] introduced the SES Infogram, using negentropy to improve kurtosis plots, enabling better analysis of noisy signals. Moshrefzadeh and Fasana [18] developed the Autogram, which assesses the kurtosis of autocorrelation functions and combined it with the combined SES (CSES) metric to enhance detection accuracy in the presence of noise and non-Gaussian features. Finally, Mauricio and Gryllias [19] presented the IESFOgram, a method that focuses on cyclospectral correlation to prioritize and integrate the most informative frequency bands, culminating in the CIES for improved detection of bearing faults. For the optimization of demodulation frequency bands, Wang et al. [20] presents Psscgram, a fault diagnosis method that improves frequency band division models. Utilizing AR power spectrum and NE, it robustly extracts fault-related frequency bands and demonstrates its effectiveness with both simulated and experimental signals.

The prior approaches for demodulation band selection encounter challenges due to inherent complexity and sensitivity to parameters. These approaches predominantly depend on intricately crafted postprocessing signal strategies that encompass interference components. Such strategies may prove incapable of yielding effective results for complex scenarios, particularly when confronted with multivariate random pulse noise. To tackle these issues, this article proposes a novel approach aimed at suppressing transient noise interference at its source. The proposed method thoroughly explores the intrinsic characteristics of nonstationary signals in bearings when confronted with transient noise interference. Its objective is to enhance insensitivity to external transient noise, facilitating a more precise and dependable selection of demodulation bands. Initially, the proposed method integrates a signal transition model based on a Markov transition matrix, in conjunction with wavelet transforms and reconstruction. This combination enhances the identification of anomalous signal components and utilizes the transition matrix to track temporal state changes. In addition, a mechanism for limiting amplitude interference is devised to recognize and alleviate transient noise that could adversely impact the process of demodulation band selection. Subsequently, this mechanism enables typical Kurtogram techniques to accurately pinpoint bearing fault-characteristic bands. Finally, the effectiveness of the method in detecting and diagnosing bearing faults under transient noise interference conditions is validated through experimental signals.

The research framework is as follows. First, Section II reviews the fundamental theory of identifying abnormal patterns or features in signals. This is followed by Section III, which introduces the proposed Markov modeling for signal condition transitions. In Section V, a case study for bearing diagnosis under transient noise interference conditions is presented. The research concludes with Section VI, which summarizes the key results and provides insights into the implications of the proposed method for bearing diagnosis.

II. Related Theory

This section revisits the identification of abnormal patterns or features within signals, emphasizing factors, such as local peak ratios, energy levels, and zero-crossing rates. Noteworthy techniques, including short-time condition local peak ratio (STCLPR) [21], short-time energy (STE) [22], short-time kurtosis (STK) [23], and short-time zero-crossing rate (STZCR) [22], [24], have been developed for capturing sharp peaks in signals.

Examining a real-world discrete signal, such as a time sequence denoted by $[x_n]$, where n ranges from 1 to $N \in \mathbb{N}$, representing the sample number. The signal can be expressed as follows:

$$x_f[m] = x[m + f \cdot h] \cdot w[m] \tag{1}$$

where m denotes the time index, f is the frame index, and h represents the sliding frame length. The window function, typically a rectangular window denoted as w, ensures uniform processing weight for frame samples during analysis.

Subsequently, the signals within each sliding window undergo local peak detection. The identification of local peaks relies on a specific criterion, whereby a sample is designated as a peak only if its value surpasses that of its neighboring sample. This criterion can be calculated as follows:

$$\begin{aligned} x[m] - x[m-1] &> 0 \\ x[m+1] - x[m] &\leq 0. \end{aligned} \tag{2}$$

Local peaks, which signal critical variations, are identified by evaluating both the amplitude of the signal and its rate of change. Relying solely on amplitude, however, may lead to overlooked variations in the rate. In particular, in cases of transient or non-Gaussian noise, employing a new threshold, denoted as m_d, can enhance the detection of local peaks, assist in monitoring signal condition transitions, and accurately pinpoint anomalous elements

$$\begin{aligned} x[m] - x[m-1] &> m_d \\ x[m+1] - x[m] &\leq m_d. \end{aligned} \tag{3}$$

A filtering procedure enables the calculation of the total number of conditional local peaks (denoted by N_p) within a sliding window. Subsequently, the STCLPR is established as the ratio of this total number of local peaks to the total number of sampled points in the window

$$\text{STCLPR} = \frac{N_p}{M} \tag{4}$$

where M represents the length of the window function.

The determination of local peak density within each sliding window is achieved through the application of (4). This metric

objectively quantifies abnormal signal variation within the specified window range, thereby effectively capturing the local characteristics of the signal.

III. Markov Modeling for Signal Condition Transitions

The accuracy of STLPR is significantly reliant on the threshold m_d. Although m_d plays a crucial role in ensuring accurate detection, the thresholds based on the aforementioned methods (refer to Section II), which rely solely on manual selection, may lead to inconsistent performance. Given that m_d directly influences the trade-off between false alarms and missed detections, adopting a dynamic adjustment approach for this threshold becomes imperative. This is especially crucial in real-world scenarios where signals are susceptible to interference from random transient or non-Gaussian noise. To resolve these issues, a Markov modeling for signal condition transitions is proposed. This approach involves three essential modules, signal preprocessing analysis, Markov modeling for monitoring temporal state changes, and suppressing interference to distinguish signal types and sources.

To begin with, the Daubechies wavelet transform (DWT) [25] is employed to mitigate the impact of low-SNR environmental noise. This step is crucial in preventing the misclassification of high-amplitude random peaks as noteworthy signal signatures in the proposed method. For a given vibration signal, denoted as $x(t)$, it can be decomposed into a set of detail coefficients ($cD_{j,k}$) and approximation coefficients ($cA_{j,k}$)

$$x(t) = \sum_{k=-\infty}^{+\infty} cA_{j,k}\phi_{j,k}(t) + \sum_{k=-\infty}^{+\infty} cD_{j,k}\psi_{j,k}(t) \tag{5}$$

where j represents the scale factor of the wavelet and k indicates the wavelet's shift factor determining its central position. $\phi_{j,k}$ is the wavelet function, and $\psi_{j,k}$ represents the scale function.

The wavelet transformation operates through a hierarchical decomposition, methodically exposing features across various frequency bands and levels. This involves calculating successive approximation coefficients and then decomposing these coefficients into higher order and finer detail coefficients. The discrete wavelet transform (DWT) can be expressed mathematically as follows:

$$\begin{aligned} cA_{j,k} &= \sum_n h[n-2k]x_{j-1,n} \\ cD_{j,k} &= \sum_n g[n-2k]x_{j-1,n} \end{aligned} \tag{6}$$

where $h[n]$ and $g[n]$ represent the low-pass and high-pass filters, respectively. The variable $x_{j-1,n}$ corresponds to the low-frequency component of the current approximation coefficient.

Following coefficient acquisition, thresholding and filtering processes are applied. Approximate coefficients cA' are obtained via a Butterworth low-pass filter, and it can be calculated as follows:

$$cA'_n = \sum_{i=0}^{M} b_i \cdot cA_{n-i} - \sum_{j=1}^{N} a_j \cdot cA'_{n-j} \tag{7}$$

where b_i and a_i are the feedforward and feedback coefficients of the filter, respectively. j starts from 1, because a_0 is typically normalized to 1.

The detail coefficients cD' undergo essential soft thresholding for signal characteristic extraction, as outlined below

$$cD_i{}' = \begin{cases} 0, & \text{if } |cD_i| \le \varepsilon \\ cD_i - \varepsilon, & \text{if } |cD_i| > \varepsilon \\ cD_i + \varepsilon, & \text{if } |cD_i| > \varepsilon. \end{cases} \tag{8}$$

The threshold ε is defined as a fraction of the maximum value of cD_i and can be represented as follows:

$$\varepsilon = k \cdot \max(cD_i). \tag{9}$$

The signal is constructed using the inverse DWT (IDWT). In addition, this signal reconstructed following wavelet decomposition successfully mitigates noise while maintaining fidelity in comparison with the original signal. The calculation is expressed as follows:

$$\hat{x}(t) = \sum_{j=-\infty}^{\infty} \left(cA'_{j,k}\phi_{j,k}(t) + \sum_{k=-\infty}^{\infty} cD'_{j,k}\psi_{j,k}(t) \right) \tag{10}$$

where $\phi_{j,k}$ and $\psi_{j,k}$ denote the wavelet function and scale function, respectively.

Following that, a Markov model [26] is implemented to track temporal state changes for the preprocessing of signals after applying DWT. For the purpose of this investigation, a first-order Markov model is employed, which can be represented by the following formula:

$$P(x_i \mid x_{i-1}, x_{i-2}, \ldots, x_1) = P(x_i \mid x_{i-1}) \tag{11}$$

where x_{i-1} denotes the current state, while x_i represents the subsequent state.

In a Markov chain, as expressed in (11), state transitions are determined at each step through a probability distribution, where the current state is influenced by the preceding one. To assess this influence objectively, an indicator, such as the transition probability matrix (TPM), is employed. This matrix signifies the conditional probabilities of state transitions. Notably, given the presence of n possible states at any given moment, there exist n^2 potential transition scenarios. The matrix is calculated as follows:

$$M_{TPM} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1j} & \cdots \\ p_{21} & p_{22} & \cdots & p_{1j} & \cdots \\ \vdots & \vdots & & \vdots & \\ p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots \\ \vdots & \vdots & \vdots & \vdots & . \end{bmatrix}. \tag{12}$$

The current methods commendably enhance signal comparability by sorting amplitudes and evenly distributing signals across fixed time intervals, thereby facilitating more coherent statistical analysis. However, these strategies often overlook

the innate temporal sequencing of signals. This oversight can result in the exclusion of pivotal, time-sensitive information and a heightened susceptibility to noise disruption. To address these limitations, this study adopts a new method of Markov signal discrete modeling. The signals obtained through wavelet reconstruction, as referred in (10), effectively differentiate atypical from typical states by grouping magnitudes within a certain range into the same interval denoted as j and designating them as signal states S_i. This approach enables the entire complex continuous signal into a simplified discrete framework (10) without losing the state transitions contained within the original time series. A designated formula, presented in (13), is used to convert signal magnitudes to interval index markers

$$S_i = \begin{cases} j, & x_n[i] \in S_j \\ j, & x_n[i] = \max(x_n),\ j = n_{\text{bin}} - 1. \end{cases} \tag{13}$$

The original signal is partitioned into distinct states, and the S_j of each state is determined by the following formula:

$$\begin{aligned} S_j &= \left[\min(x_n) + (j-1)\cdot\Delta, \min(x_n) + j\cdot\Delta\right] \\ \Delta &= \frac{\max(x_n) - \min(x_n)}{n_{\text{bin}}} \end{aligned} \tag{14}$$

where Δ represents the uniform interval length for partitioning, whereas n_{bin} indicates the predetermined number of states, exerting a direct influence on the density of states after discretization.

After obtaining state sequences, a statistical analysis is conducted to determine the frequency of transitions between states. This process involves using the Markov transition matrix (MTM) to create a transition frequency matrix referred to as N_{ij}^{MTM}

$$N_{ij}^{\text{MTM}} = \sum_{n=1}^{N-1} \delta(S_n = i, S_{n+1} = j) \tag{15}$$

where $\delta(x, y)$ is the Kronecker delta function, yielding 1 when $x = y$ and 0 otherwise. N represents the length of the state sequence S.

Thereafter, the state matrix undergoes normalization, yielding the Markov TPM $P_{i,j}$. This matrix indicates the probability of transitioning from state i to state j

$$P_{ij} = \frac{N_{ij}^{\text{MTM}}}{\sum_{j=1}^{N} N_{ij}^{\text{MTM}}}. \tag{16}$$

Furthermore, matrix normalization is critical after calculating Markov transition probabilities to ensure the elements sum to one. In this study, signals were discretized using a fixed number of intervals ($n_{\text{bin}} = 10$). The interval boundaries with the greatest probability deviation are identified as critical regions for pattern transitions, and their associated probabilities are represented by P_m. The probability of overall extreme value transfer, denoted as P_t, can be computed as follows:

$$P_{\text{t}} = \sum_{m} P_m. \tag{17}$$

Thus, the new threshold θ_{Markov} can be determined as follows:

$$\theta_{\text{Markov}} = \frac{k}{P_{\text{t}}} \times x(t)_{\max} \tag{18}$$

where $x(t)_{\max}$ represents the maximum amplitude of the absolute value of the signal and k is a coefficient.

The original SRLPR in (4) can be enhanced by integrating the proposed short-time Markov peak rate (STMPR) as follows:

$$\begin{aligned} \text{STMPR} = \frac{1}{M}\sum_{n=1}^{M-1} & H(x[n] - x[n-1] - \theta_{\text{Markov}}) \\ & \cdot H(x[n] - x[n+1] - \theta_{\text{Markov}}) \end{aligned} \tag{19}$$

where H is the Heaviside step function, which has a value of 1 when the expression inside it is positive; otherwise, it is 0.

Algorithm 1 Proposed Markov Modeling Algorithm Tracks Signal Condition Transitions

Require: Signal sequence $x[m]$, threshold θ_{Markov}
Ensure: STMPR
  Determine frame length M
  Initialize peak counter $N_p \leftarrow 0$, start and end indices $start \leftarrow 1$, $end \leftarrow M$
  while $end \leq$ length(x) **do**
    for $m = 0$ **to** $M - 1$ **do**
      Initialize current frame peak counter $P_c \leftarrow 0$
      for $n =$ start **to** end **do**
        if $n > 1$ **and** $n <$ length(x) **then**
          if $x[n] > x[n-1] + \theta_{\text{Markov}}$ **and** $x[n] > x[n+1] + +\theta_{\text{Markov}}$ **then**
            $P_c \leftarrow P_c + 1$
          end if
        end if
      end for
    end for
    $N_p \leftarrow P_c + N_p$, $end \leftarrow end + 1$, $start \leftarrow start + 1$
  end while
  STMPR $\leftarrow N_p/M$
  return STMPR

After a comprehensive analysis of the entire signal's STMPR, an average threshold is established to identify abnormal frames accurately. The sections in the original signal corresponding to these frames are subsequently modified to restrict amplitude spikes. The objective of this procedure is to minimize the impact of anomalies and improve signal consistency, thereby facilitating more effective analysis, all while preserving the fundamental traits of the core signal. These processes can be calculated as follows:

$$\mu = \frac{1}{K}\sum_{k=0}^{K-1} \text{STMPR}[k] \tag{20}$$

$$B = \{k \mid \text{STMPR}[k] > 2\mu\} \tag{21}$$

$$y_{\text{ais}}^{i} = L \cdot \bar{x}, i \in [k\cdot M, k\cdot M + h) \quad \forall k \in B \tag{22}$$

where y_{ais}^{i} represents the amplitude interference-imitating signal. B denotes the initial boundaries for anomalies, and L is

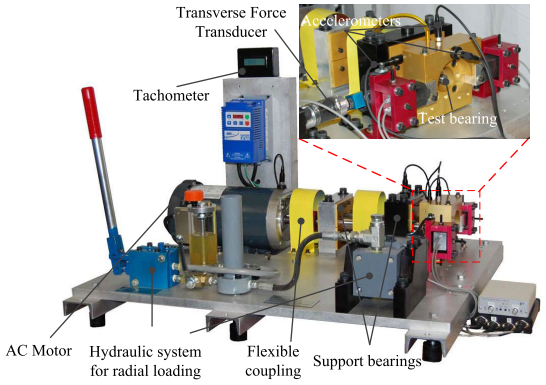

Fig. 1. SpectraQuest bearing diagnostics and prognostics simulator.

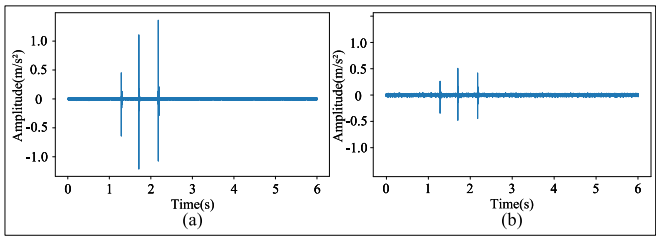

Fig. 2. (a) Raw signal. (b) Level-1 approximation coefficient in DWT.

the coefficient used for amplitude limitation. $\bar{x}$ represents the mean amplitude of the signal, while M and h denote the frame length and frame shift, with the typical values of 256 and 128, respectively.

IV. Test Rig and Data Overview

In this investigation, a bearing diagnostics and prognostics simulator provided by SpectraQuest Inc. served as a crucial tool. The test rig, illustrated in Fig. 1, consists of an ac motor, hydraulic system, test bearing, accelerometers, and other components. Vibration measurement is conducted using one horizontally positioned PCB 352C04 accelerometer on the side and one vertically positioned accelerometer at the top. The investigation employed a single-row deep groove ball bearing with nine rolling elements, each having a diameter of 7.94 mm and a pitch diameter of 38.51 mm. The vibration signals are analyzed under constant rotating speed conditions, with a sampling frequency set at 51 200 Hz.

V. Experimental Validation

In this section, the effectiveness of the proposed Markov modeling for signal condition transitions is validated using a vibration signal with random transient interference for analysis.

The vibration signal, as shown in Fig. 2 (a), includes several random external transient interferences. For a comprehensive analysis of the vibration signal amid environmental noise, applying a preprocessing method, such as the DWT, is recommended. According to (5), the signal can be decomposed to maintain the signal's global dynamics, and the level-1 approximation coefficient can be obtained, depicted in Fig. 2(b).

Furthermore, the results of the detail coefficients for levels 1–3 in the DWT are presented in Fig. 3. The high-frequency

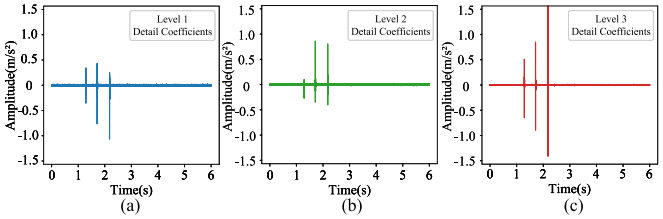

Fig. 3. Detail coefficient across levels 1–3 in DWT. (a) Level 1 cofficient in DWT. (b) Level 2 cofficient in DWT. (c) Level 3 cofficient in DWT.

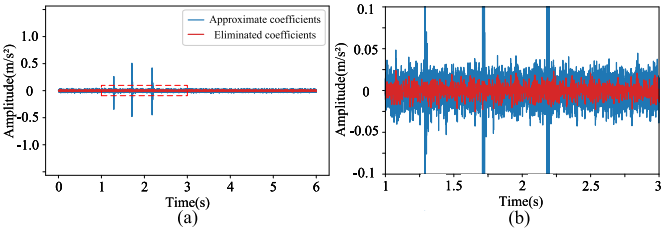

Fig. 4. (a) Low-pass filtered approximation coefficient. (b) Zoomed-in region from (a).

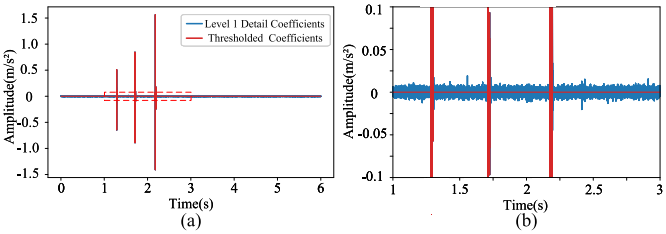

Fig. 5. (a) Thresholded detail coefficients. (b) Zoomed-in region from (a).

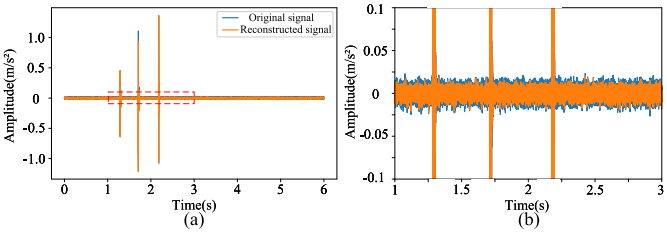

Fig. 6. (a) Raw signal. (b) Retrieve signal.

signal components, as depicted for each level, capture rapid changes, providing crucial early warnings for significant events. Concurrently, the low-frequency components illustrate the signal's fundamental structure and long-term trends. The application of low-pass filtering to approximation coefficients aids in clarifying signal representation by eliminating minor disturbances, thereby reducing complexity. Consequently, the long-term stable trend of the signal is enhanced, and minor disturbances are effectively suppressed, as demonstrated in Fig. 4.

In accordance with (9) and (10), the thresholded signal, as illustrated in Fig. 5, can be obtained. This process effectively minimizes noise in detail coefficients while concurrently preserving significant features in the signal.

From (10), the comparison between the original and retrieved signals is illustrated in Fig. 6. The retrieved signal, shown in Fig. 6(b), effectively eliminates noise while emphasizing significant external interferences for precise anomaly signal transition. This process enhances accuracy by making crucial information more prominent and facilitates the detection of potential anomalies.

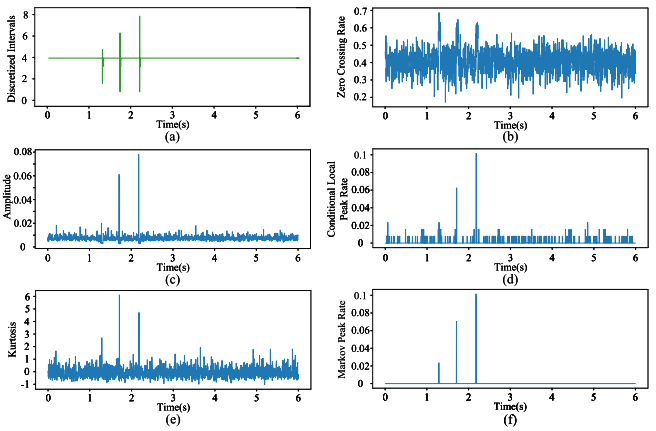

Fig. 7. (a) Equidistant signal discretization. (b) STZCR. (c) STE. (d) STCLPR. (e) STK. (f) STMPR.

Next, the retrieved signal should be partitioned into distinct states according to (14)–(16). The results are illustrated in Fig. 7(a), where the discretized signals clearly display distribution intervals for both normal and noise-affected signals.

To evaluate the effectiveness of the proposed method in tracking signal condition transitions, various existing methods, including STCLPR, STE, STK, and STZCR, are performed for comparative analysis. The results are presented in Fig. 7. The STZCR method, for instance, is unable to detect interference transient noise, as illustrated in Fig. 7(b). The STE and STCLPR methods, displayed in Fig. 7(b) and (c), respectively, both only detect two interference components, missing other components within the signal. Though STK, represented in Fig. 7(e), detects some interference components, it is vulnerable to strong interference from unidentified signal elements. In contrast, the STMPR methodology exhibits impressive performance. It consistently and accurately identifies all interference transient noises, with no discernible impact from any other unknown interferences.

Upon numbering each discretized sampling point, the Markov transition probabilities for the entire signal sequence can be computed using (1). Subsequently, a matrix illustrating the transition probabilities between signal states is presented in Fig. 8(a). This heat map, depicted in Fig. 8(a), facilitates a visual analysis of state transitions and provides insights into the intrinsic structure of the signal through the Markov transition probabilities. The areas on the heat map that appear brighter represent zones with higher transition probabilities, indicating that state changes occur more frequently there. When the diagonal looks particularly bright, it signifies stability, emphasizing that the signal is both predictable and continuous. On the other hand, if the rows have just a few very bright spots, this implies that certain state transitions are more likely and could indicate a higher chance of unusual activity. It is important to note, however, that these specific transitions occur with lower frequency within the comprehensive transition frequency matrix. This characteristic precisely facilitates the objective differentiation of various signal transition modes.

Finally, the proposed method effectively enables the identification of signal condition transitions across multiple states, as demonstrated by the data presented in Fig. 7(b). An analysis of this figure indicates that the peak rate measure of the signal

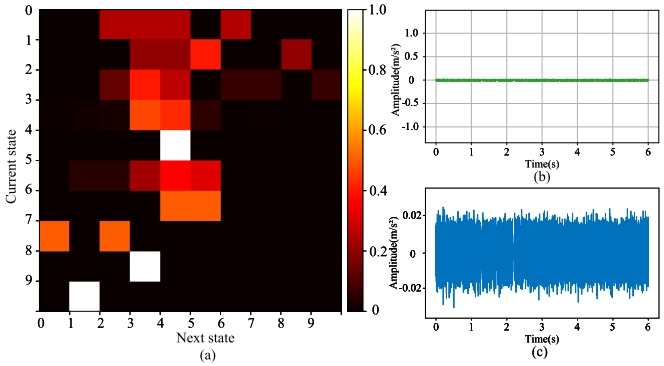

Fig. 8. (a) Markov signal probability heat map. (b) Vibration signal with interference suppression. (c) Zoomed-in view of amplitude interference limiting.

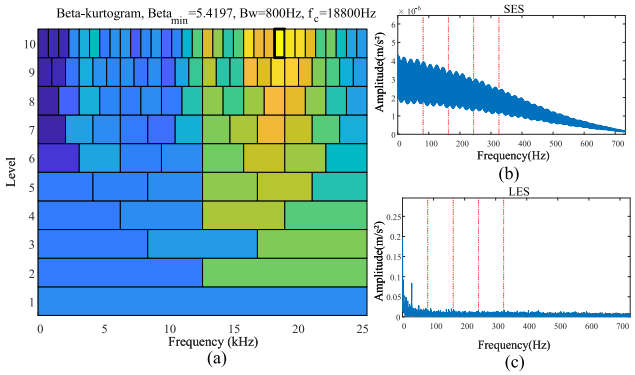

Fig. 9. Raw signal with interference from external transient noise. (a) Demodulation band selection through beta-Kurtogram. (b) SES. (c) LES.

reliably detects anomalies associated with external transient noises. The methodology's resilience to such disturbances is exhibited in Fig. 8(b), illustrating the vibration signal's immunity to external transients. To further mitigate the issue of amplitude interference, a mechanism has been established to limit the amplitude of such disturbances. This is accomplished by computing the mean absolute value across all signals and subsequently capping any excessive amplitude at 80%. In addition, the method involves widening the margin around the anomaly by several units in either direction, thereby facilitating more precise analysis. This technique is essential as it preserves the fidelity of the standard signal pattern while ensuring that anomalous signals are not disregarded. This is critical for improving the accuracy of fault diagnosis. The effectiveness of this amplitude interference limitation mechanism is illustrated in Fig. 8(c).

In this study, the benchmark methods, Kurtogram and Autogram, are employed for a comparative analysis. The raw bearing vibration signal, susceptible to interference from external random transient noise components, and the signal after postinterference suppression and amplitude interference limiting are used for validation through demodulation band selection and envelop analysis.

The demodulation band selection results for the raw signal are depicted in Figs. 9(a) and 10(a), while their corresponding envelop analyses, utilizing the SES and logarithmic envelope spectrum (LES), are provided in Figs. 9(b) and (c), and 10(b) and (c). The filtered central frequency and bandwidth are

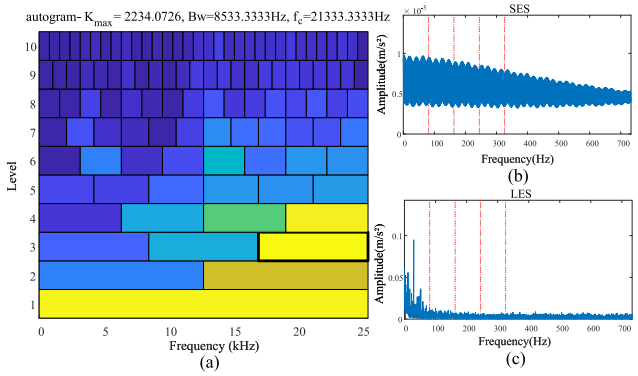

Fig. 10. Raw signal with interference from external transient noise. (a) Demodulation band selection through Autogram. (b) SES. (c) LES.

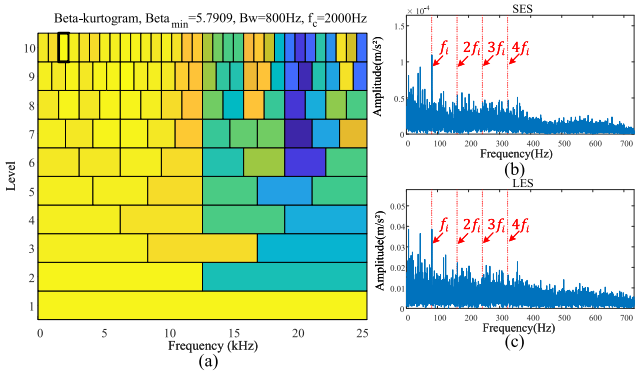

Fig. 11. Signal postinterference suppression and amplitude interference limiting. (a) Demodulation band selection through beta-Kurtogram. (b) SES. (c) LES.

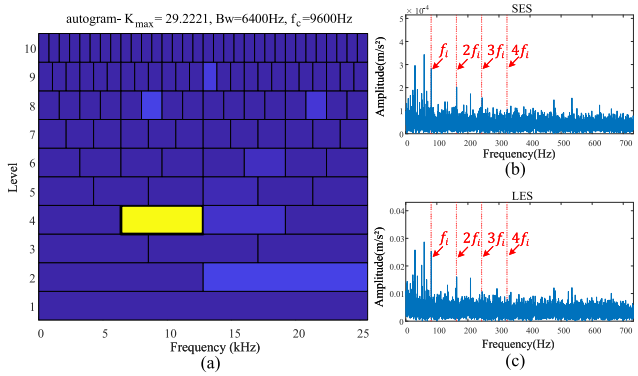

Fig. 12. Signal postinterference suppression and amplitude interference limiting. (a) Demodulation band selection through Autogram. (b) SES. (c) LES.

specified as [18800, 800 Hz] and [21333.3333, 8533.3333 Hz]. Analyzing the envelop analysis results in Figs. 9(b) and (c), and 10(b) and (c), no bearing characteristic frequency is identified, in contrast to the theoretical calculation of $f_i = 81.45$ Hz.

Conversely, Figs. 11(a) and 12(a) depict the results of demodulation band selection and envelope analysis for the signal subsequent to postinterference suppression and amplitude interference limiting. Meanwhile, the corresponding envelope analyses, utilizing the SES and LES, are presented in Figs. 11(b) and (c), and 12(b) and (c). The filtered central frequency and bandwidth are specified as [2000, 800 Hz] and [9600, 6400 Hz].

Upon analyzing the envelope analysis results in Figs. 11(b) and (c), and 12(b) and (c), characteristic frequencies, such as f_i, $2f_i$, $3f_i$, and $4f_i$, are easily identified. The validation of the proposed Markov modeling for signal condition transitions is crucial for bearing diagnosis, particularly in the presence of external interference scenarios.

VI. Conclusion

This article proposes a novel approach to enhance the diagnostic performance of rolling bearings in real-world scenarios by effectively addressing challenges posed by frequent environmental disturbances. The proposed method focuses on suppressing transient noise interference at its source, aiming to improve the insensitivity to external disturbances and consequently enhance the accuracy and reliability of demodulation band selection. By incorporating a signal transition model based on a Markov transition matrix, coupled with wavelet transforms and reconstruction, the method demonstrates an improved identification of anomalous signal components and the ability to track temporal state changes. Furthermore, this article introduces an amplitude interference-limiting mechanism designed to identify and mitigate transient noise that may, otherwise, adversely affect the demodulation band selection process. Through rigorous experimentation, the study's results indicate that this mechanism enables typical Kurtogram techniques to accurately identify bearing fault-characteristic bands. The validation of the proposed methodology for fault diagnosis of rolling bearings underscores its effectiveness in overcoming challenges associated with complex scenarios and multivariate random transient noise.

References

[1] T. Sibanda and S. Schmidt, "Cyclomap: A new phase-cycle analysis to study the kinematics of gears and bearings," *Mech. Syst. Signal Process.*, vol. 205, Dec. 2023, Art. no. 110832.

[2] P. Chen, K. Wang, M. J. Zuo, and D. Wei, "An ameliorated synchroextracting transform based on upgraded local instantaneous frequency approximation," *Measurement*, vol. 148, Dec. 2019, Art. no. 106953.

[3] Z. Wu, H. Jiang, H. Zhu, and X. Wang, "A knowledge dynamic matching unit-guided multi-source domain adaptation network with attention mechanism for rolling bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 189, Apr. 2023, Art. no. 110098.

[4] P. Chen, Y. Li, K. Wang, M. J. Zuo, P. S. Heyns, and S. Baggeröhr, "A threshold self-setting condition monitoring scheme for wind turbine generator bearings based on deep convolutional generative adversarial networks," *Measurement*, vol. 167, Jan. 2021, Art. no. 108234.

[5] P. Chen, Y. Li, K. Wang, and M. J. Zuo, "An automatic speed adaption neural network model for planetary gearbox fault diagnosis," *Measurement*, vol. 171, Feb. 2021, Art. no. 108784.

[6] X. Wang, C. Shen, M. Xia, D. Wang, J. Zhu, and Z. Zhu, "Multi-scale deep intra-class transfer learning for bearing fault diagnosis," *Rel. Eng. Syst. Saf.*, vol. 202, Oct. 2020, Art. no. 107050.

[7] Y. Ding, M. Jia, Q. Miao, and Y. Cao, "A novel time–frequency transformer based on self–attention mechanism and its application in fault diagnosis of rolling bearings," *Mech. Syst. Signal Process.*, vol. 168, Apr. 2022, Art. no. 108616.

[8] L. Tang, X. Wu, D. Wang, and X. Liu, "A comparative experimental study of vibration and acoustic emission on fault diagnosis of low-speed bearing," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 3529211.

[9] P. Chen, C. Xu, Z. Ma, and Y. Jin, "A mixed samples-driven methodology based on denoising diffusion probabilistic model for identifying damage in carbon fiber composite structures," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 3513411.

[10] R. F. Dwyer, "A technique for improving detection and estimation of signals contaminated by under ice noise," *J. Acoust. Soc. Amer.*, vol. 74, no. 1, pp. 124–130, 1983.

[11] J. Antoni and R. B. Randall, "The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines," *Mech. Syst. Signal Process.*, vol. 20, no. 2, pp. 308–331, Feb. 2006.
[12] J. Antoni, "Fast computation of the Kurtogram for the detection of transient faults," *Mech. Syst. Signal Process.*, vol. 21, no. 1, pp. 108–124, Jan. 2007.
[13] Y. Lei, J. Lin, Z. He, and Y. Zi, "Application of an improved kurtogram method for fault diagnosis of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 25, no. 5, pp. 1738–1749, Jul. 2011.
[14] C. Li, D. Cabrera, J. V. De Oliveira, R.-V. Sanchez, M. Cerrada, and G. Zurita, "Extracting repetitive transients for rotating machinery diagnosis using multiscale clustered grey infogram," *Mech. Syst. Signal Process.*, vols. 76–77, pp. 157–173, 2016.
[15] T. Barszcz and A. Jabłoński, "A novel method for the optimal band selection for vibration signal demodulation and comparison with the Kurtogram," *Mech. Syst. Signal Process.*, vol. 25, no. 1, pp. 431–451, Jan. 2011.
[16] P. Borghesani, P. Pennacchi, and S. Chatterton, "The relationship between kurtosis- and envelope-based indexes for the diagnostic of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 43, nos. 1–2, pp. 25–43, 2014.
[17] J. Antoni, "The infogram: Entropic evidence of the signature of repetitive transients," *Mech. Syst. Signal Process.*, vol. 74, pp. 73–94, Jun. 2016.
[18] A. Moshrefzadeh and A. Fasana, "The autogram: An effective approach for selecting the optimal demodulation band in rolling element bearings diagnosis," *Mech. Syst. Signal Process.*, vol. 105, pp. 294–318, May 2018.
[19] A. Mauricio and K. Gryllias, "Cyclostationary-based multiband envelope spectra extraction for bearing diagnostics: The combined improved envelope spectrum," *Mech. Syst. Signal Process.*, vol. 149, Feb. 2021, Art. no. 107150.
[20] X. Wang, J. Zheng, and J. Zhang, "A novel optimal demodulation frequency band extraction method of fault bearing based on power spectrum screening combination-gram," *Mech. Syst. Signal Process.*, vol. 174, Jul. 2022, Art. no. 109104.
[21] Z. Hristo and K. Nikolay, "A method for signal change detection via short-time conditional local peaks rate feature," *J. Elect. Electron. Eng.*, vol. 15, no. 2, pp. 106–109, 2022.
[22] P. A. Schirmer and I. Mporas, "Energy disaggregation from low sampling frequency measurements using multi-layer zero crossing rate," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 3777–3781.
[23] S. Alimi and O. Awodele, "Voice activity detection: Fusion of time and frequency domain features with a SVM classifier," *Comput. Eng. Intell. Syst.*, vol. 13, no. 3, pp. 20–29, 2022.
[24] B. Chen, Y. Hu, L. Wu, and H. Li, "Partial discharge pulse extraction and interference suppression under repetitive pulse excitation using time-reassigned multi-synchrosqueezing transform," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art. no. 3535609.
[25] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.*, vol. 4, no. 3, pp. 247–269, 1998.
[26] Y. Khalifa, D. Mandic, and E. Sejdić, "A review of hidden Markov models and recurrent neural networks for event detection and localization in biomedical signals," *Inf. Fusion*, vol. 69, pp. 52–72, May 2021.

Peng Chen (Member, IEEE) received the M.S. degree in mechatronic engineering and the Ph.D. degree in mechanical engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2016 and 2020, respectively, under the supervision of Prof. Ming J. Zuo.

From 2019 to 2020, he was an International Scholar with the Department of Mechanical Engineering, Katholieke Universiteit Leuven (KU Leuven), Leuven, Belgium. He was a Visiting Research Scholar with the Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa, in 2018. He is currently an Assistant Professor with the College of Engineering, Shantou University, Guangdong, China. His research interests include signal and acoustics processing, deep learning, intelligent interactions, and fault diagnosis and prognostics for advanced industrial equipment.

Yuhao Wu received the B.S. degree in computer science and technology from the East China Jiaotong University, Nanchang, China, in 2022. He is currently pursuing the M.S. degree in mechanical engineering with the College of Engineering, Shantou University, Guangdong, China.

His research interests include signal and acoustics processing.

Chaojun Xu was born in Chongzuo, Guangxi, China, in 1999. He received the B.S. degree in measurement and control technology and instrumentation from Beijing Jiaotong University, Beijing, China, in 2021. He is currently pursuing the M.S. degree in mechanical engineering with the College of Engineering, Shantou University, Guangdong, China.

His research interests include signal processing, and deep learning for fault diagnosis and prognostics for industrial equipment.

Yaqiang Jin, photograph and biography not available at the time of publication.

Chengning Zhou, photograph and biography not available at the time of publication.