Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Journal of Sound and Vibration 284 (2005) 567–582

JOURNAL OF SOUND AND VIBRATION

www.elsevier.com/locate/jsvi

Selection of wavelet packet basis for rotating machinery fault diagnosis

B Liu*

Department of Mathematics, West Virginia University, Morgantown, WV 26506, USA

Received 5 December 2003; accepted 23 June 2004
Available online 28 December 2004

Abstract

Wavelet packets provide a library of orthonormal bases and can therefore represent a signal in many different ways. This paper presents a method to select the orthonormal basis to represent vibration signals for rotating machinery fault diagnosis. The selected basis is formed by two sets of basis functions for representation of the transients excited by localized faults and other signal components of interest, respectively. This method overcomes the limitation of the widely used Coifman and Wickerhauser's best basis algorithm in detection of transients hidden in large background vibration and is more effective in rotating machinery fault diagnosis especially in the early stage of fault development. In addition, differing from some existing studies, this method does not need training samples and is relatively easy to implement in practice. Applications in the analysis of gearbox and bearing vibration signals show that the proposed method performs better than the best basis algorithm.

1. Introduction

Rotating machinery covers a broad range of critical facilities. Fault diagnosis of such equipment is of particular importance in industries. Vibration analysis has been the major approach for this task. Its success depends largely on the techniques used in processing the vibration signals.

*Corresponding author.
E-mail address: bliu@math.wvu.edu (B. Liu).

0022-460X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2004.06.047

Owing to its advantage of direct association with the rotating characteristics, spectral analysis has been the most used signal processing tool in the vibration analysis of rotating machinery. However, despite the success in many situations, robust early detection of localized defects such as cracking teeth in gearboxes and surface spalling of rolling element bearings still remains a difficult problem. Localized defects often excite additional localized components into the background vibration. As a result, the picked-up signals may consist of not only harmonics which are mainly related to rotating activities but also transients generated due to the presence of the defects. Spectral analysis is excellent in analyzing harmonics but ineffective for handling transients because it represents a signal with complex exponentials that are not localized in time. To better deal with the problem, some other techniques have been developed such as kurtosis analysis [1], resonance demodulation [2,3], time domain averaging [1], etc. These techniques have improved the detection considerably but still have certain limitations. For example, some of them may require to bandpass-filter the signal for enhancement of the transients. This is unfortunately sometimes difficult since there may not exist a reliable way to know the frequency range of the transients.

In the last decade, time–frequency analysis has received great attention in the area of machinery fault diagnosis. Particularly, the recent advances of wavelet analysis offer a new set of techniques of time–frequency analysis that have been shown to be powerful in transient detection. They thus provide highly promising tools for dealing with localized defects of rotating machinery. Owing to its advantage of being adaptive and computationally efficient, wavelet packet transform (WPT) [4,5] has been one of the several most studied wavelet techniques in machinery diagnosis. Its effectiveness has been investigated by a number of researchers. To name a few, in Ref. [6], Geng and Qu showed the advantages of WPT over spectrum analysis and Wigner–Ville distribution in handling the vibration transients generated in rotating machine elements. Liu et al. [7] represented vibration signals of different conditions with different sets of wavelet packet vectors and obtained good results in bearing failure detection. Yen and Lin [8] employed statistical criteria to select the features represented with wavelet packets, which was found to be able to improve the performance of the associated neural network classifier significantly in bearing fault diagnosis. More recently, Altmann and Mathew [9] proposed a method for detection and diagnosis of low-speed rolling-element bearing faults. They used an adaptive network-based fuzzy inference system to extract the featured wavelet packets which cover multiple frequency bands and yield a much higher signal-to-noise ratio than the conventional bandpass or highpass filtering approaches. In another latest study [10], Nikolaou and Antoniadis suggested to take advantage of the modulation property of bearing vibration in applying WPT to detect localized defects. Their method needs little interventions of the user.

Wavelet packets can form a great number of orthonormal bases. Each of them is capable of representing a signal completely. This means that by using wavelet packets one can represent a signal in many different ways. So far, most of the existing studies use the well-known ‘‘best basis’’ algorithm developed by Coifman and Wickerhauser [4,5] to select the basis. Such a basis can well characterize the time–frequency composition of a signal but is not necessarily best suitable for feature detection in machinery diagnosis. This is because the selection process of the best basis is dominated by the signal components that are relatively large in a frequency band. In fault diagnosis, especially in the early stage of fault development, the featured components such as the transients produced by localized defects however are usually much smaller than the background components. If these two types of components overlap in frequency, the best basis algorithm may not be able to detect the transients.

In this paper, we propose a basis selection method to deal with the above problem. This method first selects the wavelet packet basis functions for the subspace that contains the transients excited possibly by localized defects. These basis functions are capable of representing the transients relatively clearly and make the detection of localized defects relatively easy. The method then selects the wavelet packet basis functions for the complementary subspace so that other signal components, such as harmonics, can be represented efficiently. This method improves the results of transient detection using Coifman and Wickerhauser's best basis algorithm while retaining its advantage in representing other types of signal components. In addition, the proposed method does not need training samples and is thus convenient for practical use. The effectiveness of this method is tested with applications to the detection of gear tooth crack and bearing surface spalling.

The paper is organized as follows: In the next section, the principle of wavelet packets and the best basis algorithm is reviewed briefly. Then, the algorithms of the proposed method are presented in Section 3. In Section 4, this method is applied to the detection of localized gear and ball bearing defects. Comparisons with the best basis wavelet packet transform are also presented. Concluding remarks are finally given in Section 5.

2. Wavelet packets and best basis algorithm

Wavelet packets are a collection of functions $\{2^{-j/2}W_n(2^{-j}t-k), n \in N, j, k \in Z\}$ generated from the following sequence of functions [4,5]:

$$W_{2n}(t) = \sqrt{2}\sum_l h_l W_n(2t-l),$$

$$W_{2n+1}(t) = \sqrt{2}\sum_l g_l W_n(2t-l), \tag{1}$$

where h and g are the quadrature mirror filters, $W_0(t)$ and $W_1(t)$ are the scaling function and basic wavelet, respectively. The wavelet packet $2^{-j/2}W_n(2^{-j}t-k)$ is a localized function of unit energy with scale 2^j, translation 2^jk, and an oscillation parameter of n.

For a discrete signal, the decomposition coefficients of wavelet packets can be computed iteratively by

$$x^k_{2n,j+1} = \sum_l h_{l-2k} x^l_{n,j},$$

$$x^k_{2n+1,j+1} = \sum_l g_{l-2k} x^l_{n,j}. \tag{2}$$

The original signal can be reconstructed iteratively by

$$x^l_{n,j} = \sum_k h_{l-2k} x^k_{2n,j+1} + \sum_k g_{l-2k} x^k_{2n+1,j+1}. \tag{3}$$

The wavelet packet functions as well as the corresponding decomposition coefficients can be organized as a binary tree as shown in Fig. 1. Each node corresponds to a frequency band. The leaf nodes of any connected subtree that has the same root node as the full tree form an orthonormal basis and can represent a signal of finite energy completely. The best basis algorithm developed by Coifman and Wickerhauser [4,5] provides an approach to choose the so-called "best basis" to represent a given signal adaptively.

In order to select the best basis, Coifman and Wickerhauser introduced the concept of information cost function that is used as a measure to search for its minimum over all the wavelet packet bases. The most used information cost function is Shannon entropy. For a discrete signal X of unit energy, the Shannon entropy of the decomposition coefficient sequence $X_{n,j} = \{x_{n,j}^k\}$ of the nth node at level j of the wavelet packet tree is computed by [4,5]

$$M(X_{n,j}) = -\sum_k (x_{n,j}^k)^2 \log(x_{n,j}^k)^2, \tag{4}$$

where $(x_{n,j}^k)^2 \log(x_{n,j}^k)^2$ is taken as 0 if $x_{n,j}^k = 0$. $M(X_{n,j})$ reflects the degree of energy concentration of the coefficient sequence. It is large if the elements of the sequence are roughly the same and small if all but a few elements are negligible.

Let $B_{n,j}$ be the basis of vectors of the nth node at the jth level of the wavelet packet tree, and $A_{n,j}$, the best basis for the signal restricted to the span of $B_{n,j}$. The best basis algorithm starts by setting $A_{n,J} = B_{n,J}$ at the bottom level of the tree, and continues upwards from $j = J-1$ to 0 by setting recursively

$$A_{n,j} = \begin{cases} B_{n,j} & \text{if } M(B_{n,j}X) \leqslant M(A_{2n,j+1}X \cup A_{2n+1,j+1}X), \\ A_{2n,j+1} \oplus A_{2n+1,j+1} & \text{otherwise.} \end{cases} \tag{5}$$

The basis $A_{0,0}$ thus selected minimizes the information cost function and is referred to as the best basis of the signal.

The best basis can well characterize the time–frequency composition of a nonstationary signal. It however may not be the best for other tasks. The method given in the

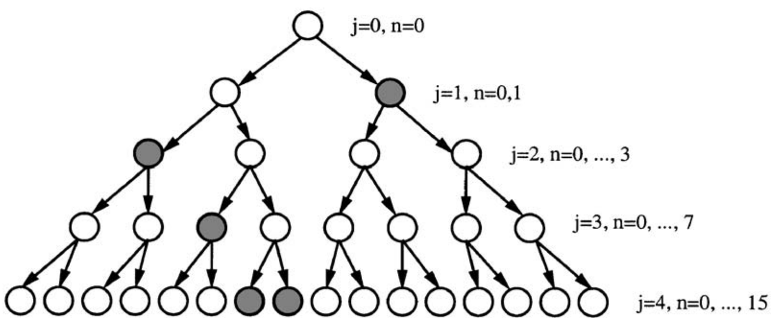

Fig. 1. Wavelet packet tree. The nth node at level j includes the wavelet packet functions $\{2^{-j/2}W_n(2^{-j}t-k), 0 \leqslant k < 2^{-j}N\}$. As an example, the wavelet packet functions of the filled nodes form an orthonormal basis. The decomposition coefficients are organized as the same tree, in which the nth node at level j includes the decomposition coefficients with respect to $\{2^{-j/2}W_n(2^{-j}t-k), 0 \leqslant k < 2^{-j}N\}$.

next section provides a way to select a basis that is more suitable for rotating machinery fault diagnosis.

3. Basis selection for fault diagnosis

3.1. Basis selection for the subspace containing fault excited transients

For convenience of description, we denote the subspace that contains transients excited possibly by localized faults by U and the complementary subspace by V with $U \oplus V = R^N$, where N is the length of the signal. This subsection addresses how to select the basis for U. Our basic approach is first to select a basis for the space R^N with the aim of enhancing the transient components of the analyzed signal. Then we identify from this basis the nodes for which the impulses in the corresponding decomposition coefficient sequences are possibly produced by localized faults. The subspace spanned by the wavelet packets of these nodes is specified as the subspace U and these wavelet packets form the desired orthogonal basis for U.

To select the transient enhancement basis for R^N, it is needed to design an information cost function that should be capable of reflecting the impulse intensity of a sequence. Note that increase of the magnitude of the impulse components in a sequence means increase of the energy concentration of the sequence. Shannon entropy, as a measure of energy concentration, thus can be used to design the cost function. In more detail, for the decomposition coefficient sequence $X_{n,j} = \{x_{n,j}^k\}$ of the nth node at level j, we define the information cost function as

$$T(X_{n,j}) = -2^j \sum_k \frac{(x_{n,j}^k)^2}{2^j \|X_{n,j}\|^2} \log \frac{(x_{n,j}^k)^2}{2^j \|X_{n,j}\|^2}. \tag{6}$$

It is essentially the Shannon entropy of a new sequence generated from $X_{n,j}$ by appending to each coefficient $x_{n,j}^k$ itself $2^j - 1$ times. The denominator $2^j \|X_{n,j}\|^2$ is introduced to normalize the sequence. From Eq. (4), it can be easily shown that

$$T(X_{n,j}) = -\frac{1}{\|X_{n,j}\|^2} M(X_{n,j}) + j + \log \|X_{n,j}\|^2. \tag{7}$$

Evidently, the computation cost of $T(X_{n,j})$ is only slightly more than that of $M(X_{n,j})$ in Eq. (4).

There exists the situation where the coefficient sequence of a node contains impulses and the information cost T is low but the average energy of the sequence, i.e., the energy of the sequence divided by its corresponding frequency bandwidth or equivalently by the sequence length, is extremely small. Detection of such impulses is of no practical value since they may not be produced by localized defects. To avoid the influence of such nodes, a simple way is to assign them a very large information cost value, e.g., $T_{\inf} = e^{1000}$. In the present study, a node is processed this way if the ratio $\eta = E_{\text{node}}^a / E_{\text{sig}}^a$, where E_{node}^a denotes the average energy of the coefficient sequence of the node, and E_{sig}^a, the signal energy divided by the signal length, is less than $\eta_{\min} = 0.05$.

Boundary effect is another issue that needs to be considered since it results in impulses at the ends of the coefficient sequence of a node at a small scale and consequently a low information cost. In the present research, we overcome this problem by windowing the original signal in the preprocessing stage with a Tukey window defined as [11]

$$w[k] = \begin{cases} 0.5\left[1+\cos\left(\dfrac{2\pi(k-1)}{r(N-1)}-\pi\right)\right], & k<\dfrac{r}{2}(N-1)+1, \\ 1, & \dfrac{r}{2}(N-1)+1\leqslant k\leqslant N-\dfrac{r}{2}(N-1), \\ 0.5\left[1+\cos\left(\dfrac{2\pi}{r}-\dfrac{2\pi(k-1)}{r(N-1)}-\pi\right)\right], & N-\dfrac{r}{2}(N-1)<k, \end{cases} \tag{8}$$

where N is equal to the length of the signal and the parameter $r\,(0<r<1)$ is the ratio of the taper to the constant section. The taper section reduces the boundary effect but causes only a little change in the time–frequency composition of the signal if r is small ($r = 0.1$ for the application examples in Section 4).

The selection of the transient enhancement basis for the space R^N uses a recursive procedure similar to Coifman and Wickerhauser's best basis algorithm. However, besides using different information cost functions, the comparison in the recursive procedure is also different. While the best basis algorithm compares the information cost of a mother node with the sum of the information costs of its children as shown in Eq. (5), the comparison here is performed between the information cost of a mother node and the minimum information cost of its children. This is because our purpose is to find the node which best enhances the transient components and the remaining nodes in each comparison are of little importance. A more rigorous description of this process will be given in Algorithm 1 at the end of this subsection.

The basis selected in the above way is suitable for transient enhancement but not necessarily as good as the basis selected using Coifman and Wickerhauser's algorithm for characterization of some other types of signal components such as harmonics. It therefore appears better for rotating machinery fault diagnosis if we could select a basis that combines the advantages of both the methods, that is, to represent the transients with the basis functions selected using the method described above while other components with the basis functions selected using Coifman and Wickhauser's algorithm. This is the reason that we divide the space R^N into two subspaces and represent them in different ways.

Now we address how to identify from the transient enhancement basis the nodes for which the impulses in the corresponding decomposition coefficient sequences are produced possibly by localized faults. As mentioned earlier, we specify the subspace spanned by the wavelet packets of these nodes as U and these wavelet packets form the desired orthogonal basis for U. To do the identification, it is needed to choose a measure and a threshold. Although the information cost function defined in Eq. (6) could be a choice of the measure, determination of its threshold value is however difficult. In the proposed method, kurtosis [1] is employed since it is well known in the area of vibration monitoring and there are empirical quantitative criteria one can consult to evaluate the condition of typical rotating machines. For example, in condition monitoring of gears, if the kurtosis of raw vibration signals is considered, the threshold can be specified as 3.5–4.5. The kurtosis usually increases if the signals are properly filtered. For a sequence $S = \{s_k\}$, the kurtosis is computed by

$$K(S) = \frac{E\{(S-\mu)^4\}}{\sigma^4}, \tag{9}$$

where μ and σ are the mean and standard deviation of S and $E\{\cdot\}$ is the expectation.

In the process of the identification, we first reconstruct the signal restricted to the span of each node in the selected transient enhancement basis and compute its kurtosis. Then the kurtosis is compared with a threshold $K_{\min}$. A node is picked up if the kurtosis is larger than $K_{\min}$. There may be more than one such nodes. The wavelet packets of all of them generate the subspace U and form its desired basis. In the case where all the kurtosis values are less than $K_{\min}$, the subspace U is empty.

It should be pointed out that the purpose of the above identification is only to determine which basis selection method is used rather than decision making of fault diagnosis. The latter is carried out based on the representation with the union of the bases for U and its complement V, which can provide much more detailed and reliable information for fault diagnosis. The value of the threshold $K_{\min}$ thus is not of decisive importance and can be determined easily. We should also indicate that instead of Eq. (6), one may define the information cost function as a function of kurtosis, e.g., $T(X_{n,j}) = 1/K(X_{n,j})$, to select the transient enhancement basis for the space R^N using the same algorithm. This however will increase the total computation load if compared with the proposed method.

We now summarize the whole procedure described above as the following algorithm.

Algorithm 1. Suppose that we have a discrete signal of length N and have chosen the wavelet and the maximum level J for wavelet packet decomposition. The basis for the subspace U that possibly contains fault excited transients can be selected as follows:

Step 1: Windowing the signal with the Tukey window given in Eq. (8). Denote by X the windowed signal.

Step 2: Decompose X into wavelet packets. Let $B_{n,j}$ be the basis of vectors of the nth node at the jth level, and $A_{n,j}$, the best transient enhancement basis for the signal restricted to the span of $B_{n,j}$.

Step 3: Compute the average energy of the coefficient sequence of each node E^a_{node}. Set the information cost of a node as $T_{\inf}$ if the ratio $\eta = E^a_{\text{node}}/E^a_{\text{sig}}$ is smaller than $\eta_{\min}$. Compute the information cost of the remaining nodes using Eq. (7). Denote by $T^B_{n,j}$ and $T^A_{n,j}$ the information cost corresponding to $B_{n,j}$ and $A_{n,j}$, respectively.

Step 4: Set $A_{n,J} = B_{n,J}$ and $T^A_{n,J} = T^B_{n,J}$ for $n = 0, 1, \ldots, 2^J - 1$.

Step 5: Determine $A_{n,j}$ for $j = J-1, J-2, \ldots, 0,\ n = 0, 1, \ldots, 2^j - 1$ as follows:

if $T^B_{n,j} \leqslant \min\{T^A_{2n,j+1}, T^A_{2n+1,j+1}\}$, *then* $A_{n,j} = B_{n,j}$ and set $T^A_{n,j} = T^B_{n,j}$, *else* $A_{n,j} = A_{2n,j+1} \oplus A_{2n+1,j+1}$ and set $T^A_{n,j} = \min\{T^A_{2n,j+1}, T^A_{2n+1,j+1}\}$. The finally obtained basis $A_{0,0}$ is the desired transient enhancement basis of the signal.

Step 6: Compute the kurtosis of the reconstructed signal restricted to each selected node with Eq. (9) and compare it with the threshold $K_{\min}$. The wavelet packets of the nodes with a value of kurtosis greater than $K_{\min}$ then generate the subspace U and form the desired orthogonal basis for U.

3.2. Basis selection for the complementary subspace

Now we address the selection of the basis for representation of the complementary subspace V. As an example, Fig. 2(a) shows a subtree obtained by pruning away the descendants of the nodes that form the selected basis for U, where these nodes are marked by "A". To select the basis for V, we first pick up along each branch the highest subtree disjoint with the nodes marked by "A". In Fig. 2(a), the root nodes of such subtrees are marked by "Q". These nodes generate the subspace V (but not necessarily form its best basis). The method for picking up such subtrees is given later in Algorithm 2.

Fig. 2(b) gives the time–frequency phase plane of the subtree in Fig. 2(a), where the correspondence between the marked nodes in the subtree and their frequency bands is displayed schematically. Selection of the basis for V is equivalent to selection of a partition of the frequency bands marked by "Q".

Once the above-mentioned subtrees whose root nodes generate the subspace V are picked up, we apply Coifman and Wickerhauser's algorithm to select the best basis for each of them. The recursive process now is carried out up to the root node of a subtree, i.e., those marked by "Q" in Fig. 2(a). The linear combination of the selected basis functions of all the subtrees then forms the best basis for the complementary subspace V.

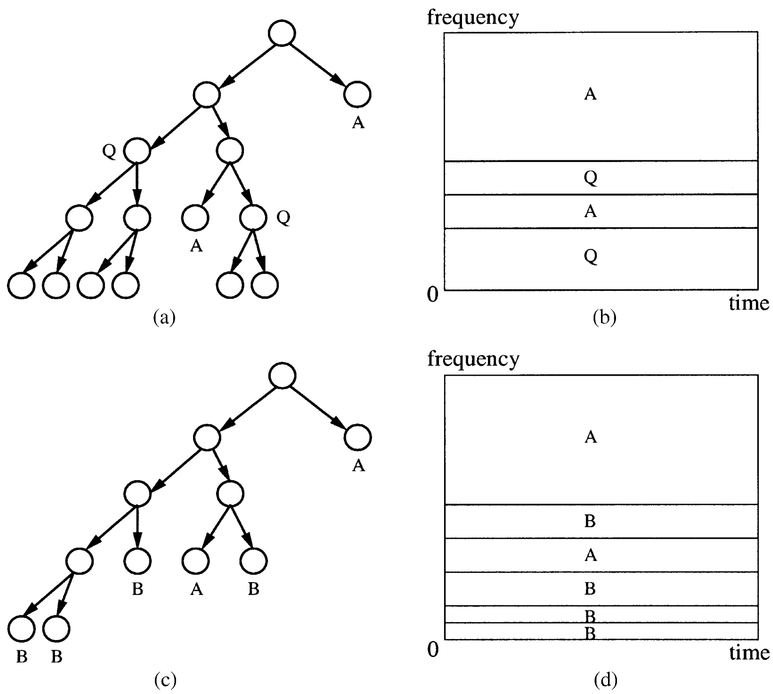

Fig. 2. (a) A subtree obtained by pruning away the descendants of the nodes (marked by "A") that form the basis for U. The nodes marked by "Q" are the root nodes of the highest subtrees disjoint with the nodes marked by "A". (b) The time–frequency phase plane corresponding to the nodes marked by "A" and "Q". (c) The subtree whose leaf nodes form the selected basis for fault diagnosis. The nodes marked by "A" and "B" form the basis for U and V, respectively. (d) The time–frequency phase plane corresponding to the selected basis.

The following algorithm describes the details for selection of the basis.

Algorithm 2. Suppose that we have selected the basis for U and obtained the subtree with the descendants of the nodes that form this basis pruned away. The basis for the complementary subspace V can be selected as follows:

Step 1: Let Γ be the index set of the nodes that form the selected basis for U. Assign each node in the pruned tree an initial value:

$$P^0_{n,j} = \begin{cases} 1 & \text{if } (n,j) \in \Gamma, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Step 2: Let Λ be the index set of the leaf nodes of the pruned tree. Start from the bottom of each branch and reassign each node in the tree a value

$$P^1_{n,j} = \begin{cases} P^0_{n,j} & \text{if } (n,j) \in \Lambda, \\ P^0_{2n,j+1} \cup P^0_{2n+1,j+1} & \text{otherwise.} \end{cases} \tag{11}$$

Step 3: Start from the top of the pruned tree and search downwards. The first encountered node along each branch with a reassigned value $P^1_{n,j} = 0$ is the root node of the highest subtree along its branch included in the subspace V.

Step 4: For each of these subtrees, select the best basis using Coifman and Wickerhauser's algorithm. The union of all the selected bases then forms the best basis for V.

By combining the selected bases for the two subspaces U and V, we finally obtain a basis for adaptive representation of a signal. Corresponding to Fig. 2(a), Fig. 2(c) shows a subtree obtained by further pruning away the descendants of the nodes that form the selected basis for V, where these nodes are marked by "B". All the leaf nodes in this subtree form the selected basis of the signal. Fig. 2(d) shows the time–frequency phase plane corresponding to this subtree. The frequency bands marked by "A" are assumed to be suitable for representation of transients. The frequency bands formed by the nodes marked by "Q" in Fig. 2(a) are now partitioned by the nodes marked by "B" corresponding to the best basis for V. Since the basis is signal dependent and is selected with the aim of effective representation of both transients and other features such as harmonics of a vibration signal, it is thus suitable for fault diagnosis of rotating machinery.

4. Applications

To demonstrate the performance of the proposed method, this section presents several application examples for detection of localized gear and bearing failures using vibration signals. In all the examples, Daubechies wavelet db10 is used.

The vibration signals analyzed in the first two examples were collected from a life testing experiment of an automobile gearbox. The transmission train in the test was: $Z28/Z48 \rightarrow Z20/Z44 \rightarrow Z30/Z36 \rightarrow Z15/Z42$. The rotation speed of the input shaft was 1600 rpm. A load of 880 kgm was applied to the output shaft. At the end of the test, one tooth of the driving gear ($Z15$), which ran at 5.89 Hz, in the last meshing pair was broken. The vibration signals were picked up at the bearing seat of the output shaft. They were lowpass-filtered at 1.8 kHz and digitized at a sampling frequency of 4 kHz.

We first analyze a vibration signal picked up near the time when the tooth was broken. Figs. 3(a) and (b) show its waveform and Fourier transform. Fig. 4 shows the time–frequency phase plane of the signal obtained using the proposed method, where the threshold K_{min} is set as 6

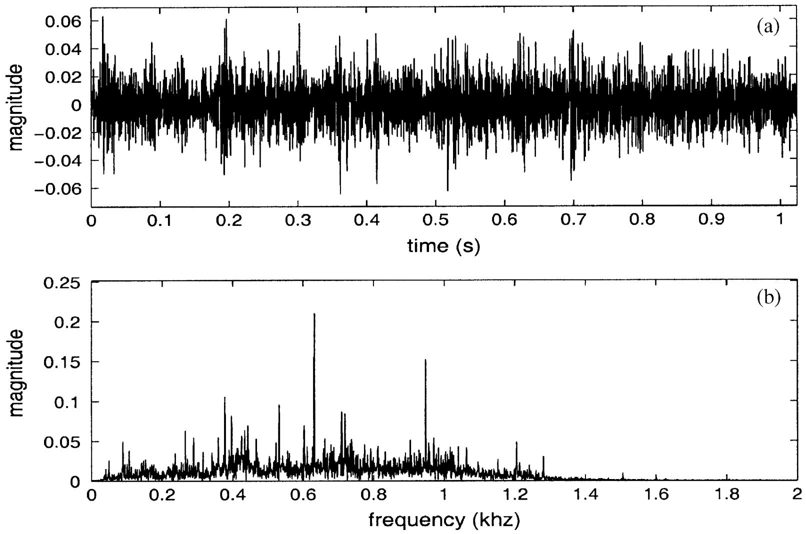

Fig. 3. (a) A vibration signal collected from the gearbox before the tooth breakage, and (b) its Fourier transform.

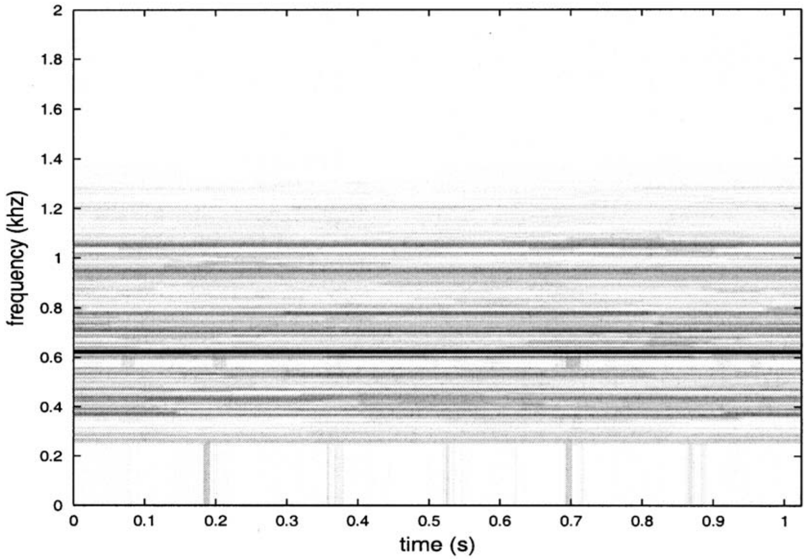

Fig. 4. Time–frequency plane of the signal in Fig. 3(a) obtained using the proposed method.

for identification of the subspace U. One can see that in the frequency band below 250 Hz, there exist a sequence of regularly spaced dark patches. They have a short time duration but a relatively large frequency band and thus represent impulsive components in the signal. The average time spacing between the neighboring impulses is about 0.17 s, corresponding to 5.9 Hz in frequency, which is very close to the rotation frequency of the damaged gear mentioned above. These features indicates clearly the existence of a tooth defect on this gear. Fig. 5 shows the time–frequency plane obtained using Coifman and Wickerhauser's best basis algorithm. It is the same as Fig. 4 in the frequency band above 250 Hz. In this band, most of the components have a large time scale and narrow frequency band and represent harmonic components in the signal. In the band below 250 Hz, Fig. 5 shows the existence of a harmonic component at the frequency of 88 Hz, which is close to 88.39 Hz, the meshing frequency of the last gear pair in the transmission path. There are several regularly spaced vertical patches below and above the frequency of 200 Hz, which indicates the existence of a localized defect. Compared with Fig. 4, this indication however is much less clearer and is liable to be neglected.

In the second example, we analyze a signal collected earlier than the one in the first example. The waveform and its Fourier transform are shown in Fig. 6. Fig. 7 shows the result obtained using the proposed method with $K_{\min} = 5$. It indicates that there exist impulsive components in the frequency band below 250 Hz. The time spacing between the neighboring impulsive components is close to either the rotation period of the damaged gear or its multiple. Although some of the impulses are not revealed, the obtained features are nevertheless sufficient for detection of the defect. Fig. 8 shows the time–frequency plane obtained using Coifman and Wickerhauser's best basis algorithm. Again, both the methods can well characterize the harmonic components above 250 Hz. However, unlike the first example, the best basis algorithm here does not show the existence of the impulsive components in the band below 250 Hz. This is because the

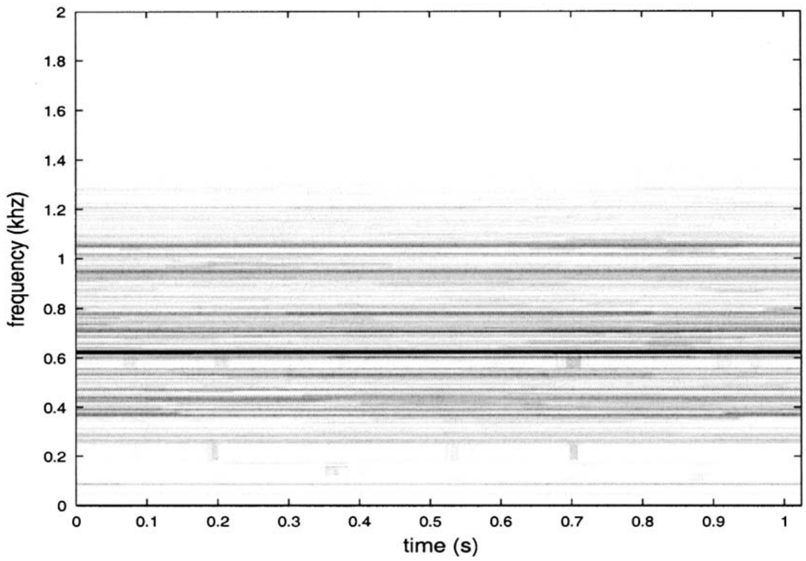

Fig. 5. Time–frequency plane of the signal in Fig. 3(a) obtained using Coifman and Wickerhauser's best basis algorithm.

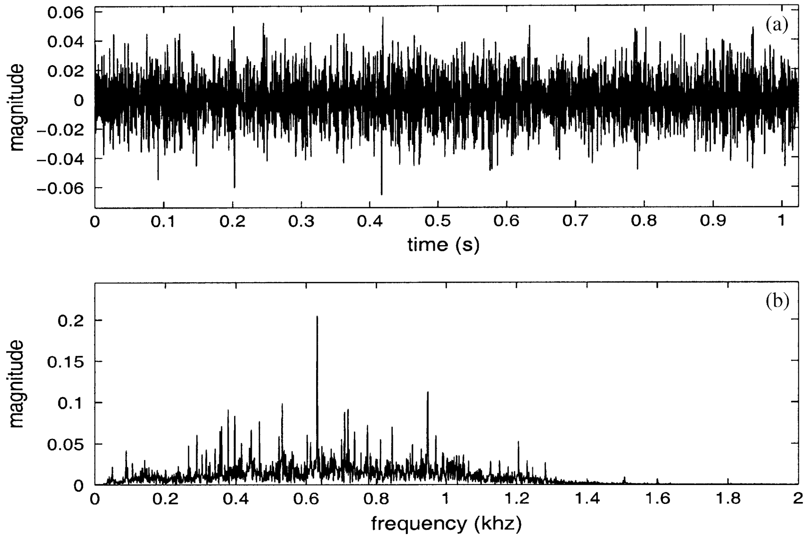

Fig. 6. (a) A vibration signal collected earlier than the one depicted in Fig. 3(a) from the same gearbox, and (b) its Fourier transform.

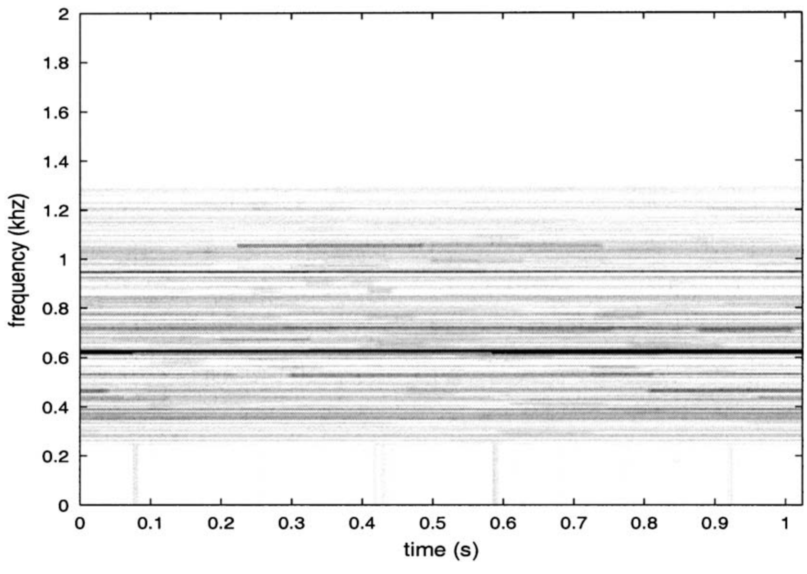

Fig. 7. Time–frequency plane of the signal in Fig. 6(a) obtained using the proposed method.

impulsive components here are very weak and the representation of this band using the best basis algorithm is dominated by the harmonic components.

The problem concerned in the next example is localized failure detection of a rolling element bearing. The specifications of the tested bearing in the experiment were as follows: number of

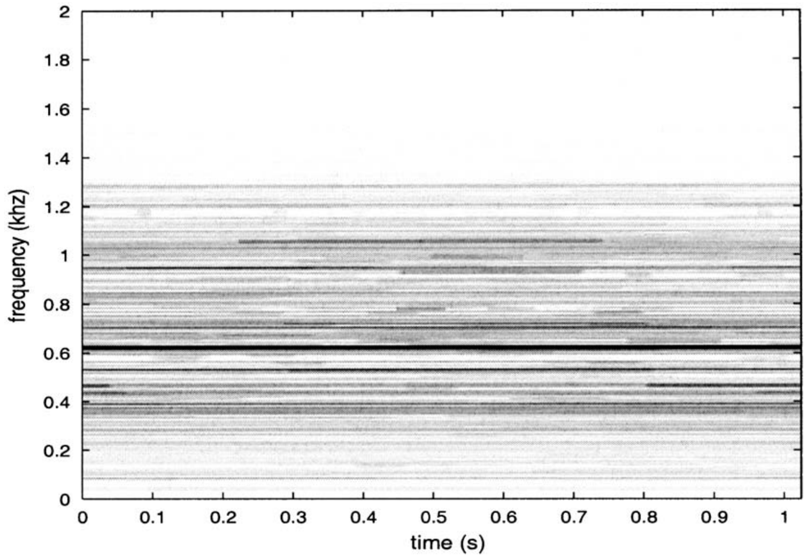

Fig. 8. Time–frequency plane of the signal in Fig. 6(a) obtained using Coifman and Wickerhauser's best basis algorithm.

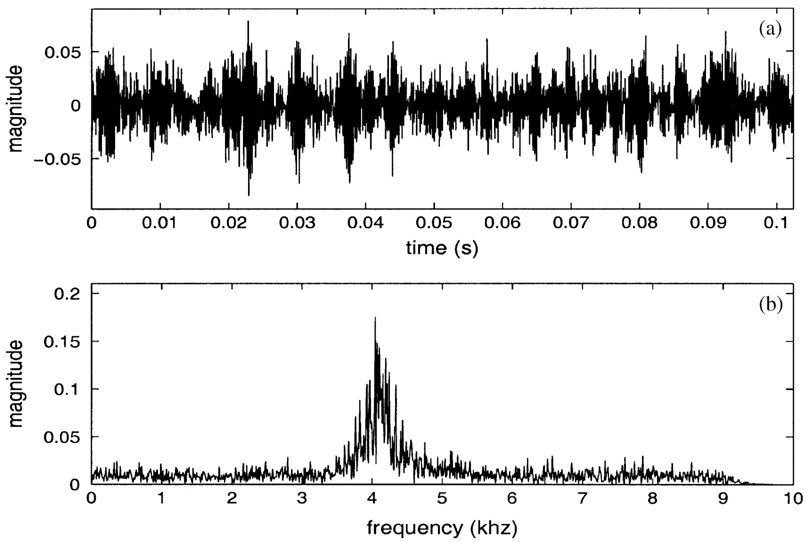

Fig. 9. (a) A vibration signal collected from the bearing with a localized defect on one rolling element, and (b) its Fourier transform.

rolling elements, 8; diameter of the rolling elements, 15 mm; medium diameter, 65 mm; and contact angle, 0°. The bearing carried a defect on one rolling element induced in the form of a dimple measuring about 0.1 mm in depth and 1 mm in diameter. The vibration signals were picked

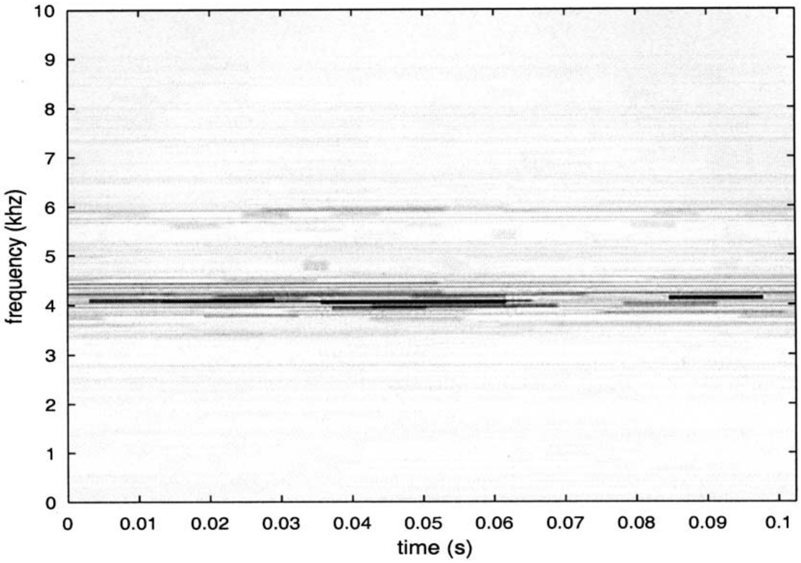

Fig. 10. Time–frequency plane of the signal in Fig. 9(a) obtained using Coifman and Wickerhauser's best basis algorithm.

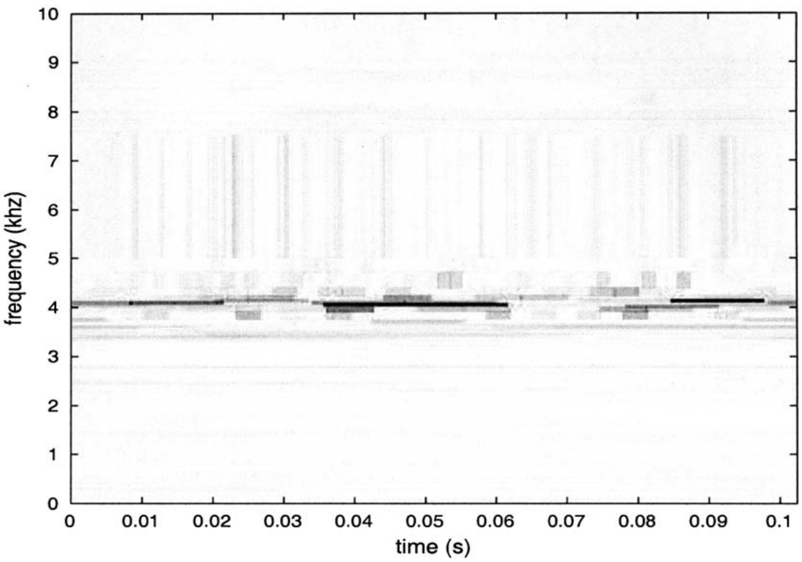

Fig. 11. Time–frequency plane of the signal in Fig. 9(a) obtained using the proposed method.

up at a constant inner race rotation speed of 2000 rpm and with a load of 20 kg applied to the outer race fixed on the test rig. The signals were lowpass-filtered at 9 kHz and digitized at a sampling rate of 20 kHz. Under these test conditions, the passing period, i.e., the average time spacing between the neighboring impact excitations generated by the defect [1], was 7.31 ms.

Fig. 9(a) shows a typical signal of the bearing. Its Fourier transform is given in Fig. 9(b). There are several apparent transients in the range from 0.02 to 0.05 s. The analysis result of Coifman and Wickhauser's best basis algorithm is presented in Fig. 10. It gives no meaningful information for detection of the failure. Fig. 11 shows the time–frequency plane obtained using the proposed method with $K_{\min} = 4$. Some transients that are not clear in Fig. 9(a) are enhanced here. The average time spacing between most of the neighboring dominant vertical patches in the band from 5 to 7.5 kHz is about 7.1 ms. Note that the passing period of the defect is 7.31 ms. It is thus determined that the bearing carried a localized rolling element defect. This example again shows that the proposed method is more effective than Coifman and Wickhauser's best basis algorithm in detection of incipient localized failures.

5. Conclusion

In this paper, we have proposed a method to select the wavelet packet basis for fault diagnosis of rotating machinery. This method divides the signal vector space into two subspaces and represents them in different ways. One of the subspaces contains the transient components excited possibly by localized defects and the other contains the remaining components. For the former subspace, a basis selection algorithm has been developed with the aim of enhancing the detection of localized defects while for the later, the well-known Coifman and Wickerhauser's best basis algorithm is employed to select the basis. The union of these two bases then gives the basis for representation of the analyzed signal. This basis is adapted to the composition of a signal and can well reveal the characteristics of the transients and other components such as harmonics. The proposed method was tested in the fault diagnosis of a gearbox and a rolling element bearing. The results showed that it can better detect incipient localized failures than Coifman and Wickerhuser's best basis wavelet packet transform that is one of the several wavelet methods used most successfully in machinery diagnosis. In addition, the proposed method does not require training samples and is thus relatively easy to implement in practice.

Acknowledgements

This work was supported in part by the National Science Foundation under Grant NSF0314742.

References

[1] S. Braun (Ed.), *Mechanical Signature Analysis*, Academic Press, London, 1986.
[2] D.R. Harting, Demodulated resonance analysis: a powerful incipient failure detection technique, *ISA Transactions* 17 (1978) 35–40.
[3] P.D. Mcfadden, J.D. Smith, Vibration monitoring of rolling element bearings by the high frequency resonance techniques—a review, *Tribology International* 17 (1984) 3–10.
[4] M.V. Wickerhauser, *Adapted Wavelet Analysis—From Theory to Software*, A.K. Peters, Welleslay, MA, 1994.

[5] R.R. Coifman, M.V. Wickerhauser, Entropy-based algorithms for best basis selection, *IEEE Transactions on Information Theory* 38 (1992) 713–718.
[6] Z. Geng, L. Qu, Vibrational diagnosis of machine parts using the wavelet packet technique, *British Journal of Non-Destructive Testing* 36 (1994) 11–15.
[7] B. Liu, S.F. Ling, Q. Memg, Machinery diagnosis based on wavelet packets, *Journal of Vibration and Control* 3 (1997) 5–17.
[8] G.G. Yen, K.C. Lin, Wavelet packet feature extraction for vibration monitoring, *IEEE Transactions on Industrial Electronics* 47 (2000) 650–667.
[9] J. Altmann, J. Mathew, Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnosis, *Mechanical Systems and Signal Processing* 15 (2001) 963–977.
[10] N.G. Nikolaou, I.A. Antoniadis, Rolling element bearing fault diagnosis using wavelet packets, *NDT & E International* 35 (2002) 197–205.
[11] F.J. Harris, On the use of windows for harmonic analysis with the discrete Fourier transform, *Proceedings of the IEEE* 66 (1978) 66–67.