Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

# *Condition monitoring of wooden railway sleepers using time-frequency techniques and pattern classification*

Siril Yella
Dept. of Computer Engineering
Dalarna University
78188, Borlänge, Sweden
sye@du.se

Samira Ghiamati
Dept. of Computer Engineering
Dalarna University
78188, Borlänge, Sweden
sgy@du.se

Mark Dougherty
Dept. of Computer Engineering
Dalarna University
78188, Borlänge, Sweden
mdo@du.se

***Abstract*—Railway sleepers are a key engineering element of all railways. Lack of much sophistication in monitoring railway sleepers makes it a key problem within the rail transportation domain. Current day condition monitoring applications involving wooden railway sleepers are mostly carried out through visual inspection and if necessary some impact acoustic examination is carried out. Decision making is largely based on intuition; moreover the process of manually inspecting sleepers is rather slow and expensive. Maintaining an even quality standard is another serious issue. In this article, a pattern recognition and classification approach is taken to automate such intuitive human skills for the development of more robust and reliable testing methods. Features were extracted from the impact acoustic emissions of wooden sleepers and were used for pattern classification. Time-frequency based feature extraction techniques such Short-time Fourier Transform and Discrete Wavelet Transform yielded good results. Multi-layer perceptron, Radial Basis Function Neural Networks and Support vector machine classifiers have been tested and compared. Further classifier fusion was investigated by considering the output of single best classifiers as input to a new classifier with an aim of improving performance. Results obtained experimentally demonstrate a classification accuracy of around 84%.**

***Keywords*—Rail transportation, Wooden railway sleepers, Condition monitoring, Short-time Fourier Transform, Discrete Wavelet Transform, Pattern classification, Classifier fusion**

## I. INTRODUCTION

Railway sleepers are an important engineering element of all railways. The sleepers both support and fasten the rails into position. Other relevant functionalities of a sleeper include, spreading wheel load to ballast, transmit lateral and longitudinal forces, insulating the rails electrically etc [1]. Sleeper failure could cause derailment of train with possibly catastrophic results. Depending on the circumstances, several types of sleeper construction can be used. Wood and concrete are the most common types; however, steel and plastic are also being used. Though it has been realized that railway sleepers significantly contribute to efficiency and safe operations not much sophistication has evolved in monitoring railway sleepers; with most of the monitoring routines being carried out manually. At this stage it is worth mentioning that discussion in the current article has been limited to condition monitoring of wooden railway sleepers.

Wooden railway sleeper (see Fig.1) inspections in Sweden (and to a large extent elsewhere in the world) are carried out manually; where a sleeper inspector walks along the track manually examining each sleeper. Where necessary some deeper inspection may be carried out on site, for example using an axe to strike and judge how hard a wooden sleeper is. Trained personnel have the ability to intuitively classify the condition of the sleeper. Such intuitive skills are quite simple in a way that sleepers in good condition do not bear wide cracks on them and emit a crisp sound when struck with an axe. In contrast, sleepers in bad condition bear wide cracks on them and emit a dull sound [2].

(a)

(b)

Figure 1. Wooden railway sleepers in different conditions, (a) Wooden sleeper in good condition. (b) A wooden railway sleeper in bad condition

This process of manually inspecting each sleeper in turn is rather slow, expensive and also requires skilled and trained staff. Human error together with maintaining an even quality standards are yet another serious issues. The Swedish Rail authority (Banverket[1]) is very much concerned that, as older employees retire, much of the necessary skills base to manage this type of infrastructure will disappear. The fact that many

[1] www.banverket.se



978-1-4244-2794-9/09/$25.00 ©2009 IEEE

sections (Inlandsbanan[2] is a typical example) of the railways in Sweden use wooden sleepers throughout motivates a serious need for an automatic inspection system that is capable of replacing manual inspection regime.

Automating condition monitoring of wooden railway sleepers has already been researched [2]. Studies based on emulation of the human inspection process have been considered a promising route of enquiry. The emulation process is achieved by selecting and evaluating two non-destructive testing methods. The first method (vision analysis) has been aimed at developing an appropriate machine vision algorithm to replicate the visual examination. The second method (impact acoustic analysis) has been aimed at building an automatic procedure to replace the usage of an axe for distinguishing sounds. Past work on impact acoustic analysis for the problem has researched the use of frequency based feature extraction techniques together with pattern classification to be able to classify the condition of wooden railway sleepers based on sound. Such work reports a classification accuracy of around 76%. A complete discussion detailing past work is out of the scope of this article but could be found elsewhere [2].

In the recent years research with the area of sound recognition (both speech and non speech) has demonstrated that time-frequency based feature extraction techniques produce reliable and robust results [3], [4] and [5]. Hence in the current article it is desired to extend the work done so far by investigating mainly two time-frequency based feature extraction techniques namely Short term Fourier transform and Discrete wavelet transform. We hypothesis that time-frequency based methods are needed because the acoustic signal is non-stationary in nature. The non-stationary nature of the signal is due to reflections and interference within the sleeper. After extracting the necessary features, the feature vectors were then presented to the pattern classifiers for further classification task. Pattern classifiers such as Multi-layer perceptron (MLP), Radial basis function neural networks (RBFNN), Support vector machines (SVM) have been tested and results in the current case indicate that SVM has achieved good classification rate (see section IV and V).

The rest of the paper is organized as follows. Section II details experimental setup and briefly describes design issues. Section III describes feature extraction employed in the current work. Section IV describes pattern classification and presents the results of the classifiers. Section V further discusses classifier fusion and presents results of fusion. The paper finally presents concluding remarks.

## II. Experimental setup and Design

The task of sound recognition and classification necessarily starts in the presence of acoustic data. In the current case impact acoustic signals were recorded by making experiments on 200 wooden sleepers in different conditions. After significant experimentation concerning the impact source, a metal hammer weighing 1.5 kg dropped from a height of 50 cm was used as the impact source. Best acoustic emissions have resulted on using such an impact source. The hammer was used to strike the wooden beams and the impact signals were recorded using a high directional microphone and a 16 bit A/D card with 44.1 KHz sampling rate. All the acoustic signals were saved on a computer in a WAV format (see Fig. 2). The WAV format for digital audio is simply the left and right stereo signal samples. Such an impact system generates a simple input with suitable characteristics for further signal interpretation.

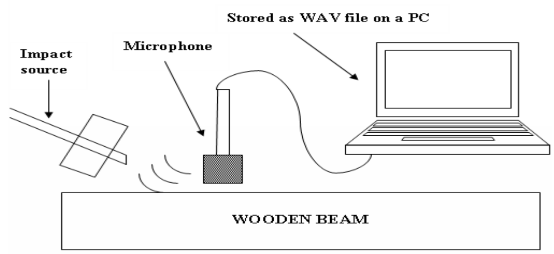


Figure 2. Experimental setup

As mentioned above impact acoustics signals have been acquired from 200 sleepers of which 144 were in good condition and 56 in bad condition. Data was collected by making impact acoustic tests on both (left and right) ends of the railway sleeper. As a result, 400 (200×2) acoustic signals were acquired from 200 sleepers. At this point it should be noted that sleeper conditions (good or bad) are judged by an experienced manual inspector. Also note that though 200 sleepers is not a huge number; the limited number of sleepers tested in the current work is due to the operational constraints in the rail transportation domain. Since collecting impact acoustic signals on railway sleepers demands re-routing or even cancellation of traffic operations and is an expensive procedure. Moreover, the difference in the number of sleepers in each class (good and bad) was due to fact that only a limited stretch of railway track could be allocated for closure, which has limited the scope for handpicking the number of sleepers in each class [2].

### A. *Design*

A pattern classification and fusion approach was considered appropriate for the purpose of automating the intuitive sound recognition skills of a human inspector. The key idea of pattern classification is to optimally extract patterns based on certain conditions and separate the data into different classes without having any idea about the distribution of the measurements in different groups [6]. As mentioned above (see Section II), acoustic signals have been acquired by making impact acoustic tests on both left and right ends of the sleeper. It is thereby evident that acoustic data from either ends of the sleeper needs to be combined to be able to arrive at a single conclusion concerning the condition of the sleeper. After necessary feature extraction

[2] www.inlandsbanan.se/

and reduction (see Section III), individual features from both left and right sides of a sleeper are combined to form a single feature vector. The feature vector is then presented to the classifiers for further pattern classification task. Further the output of single best classifiers is used as input to a new classifier (classifier fusion) with an aim of achieving more reliable and robust results (see section V).

In this article, pre-processing and feature extraction were performed using signal processing toolbox in MATLAB [3]. Pattern classification was performed using LNKNET [4]. LNKNET provides a good chance of testing several classifiers and also promises good graphical representation of classification.

## III. Feature Extraction and Dimension Reduction

Although the raw acoustic signals between good and bad sleepers exhibited significant differences (see Fig.3); the signal was not used directly for classification keeping in view the computational constraints that will be laid on the classifiers. Hence feature extraction was researched.

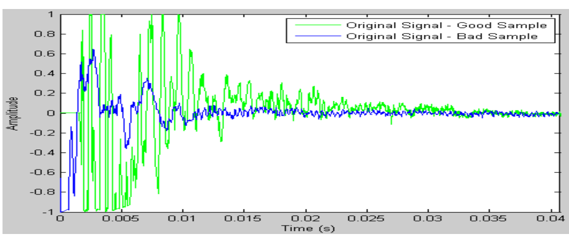


Figure 3. Acoustic signals from good and bad sleepers

Feature extraction is aimed at obtaining a compact representation of the input acoustic signal by manipulating the signal to produce a set of features that are characteristic of the signal. Feature extraction techniques on sounds can be neatly split into two categories as frequency-based feature extraction techniques and time-frequency based feature extraction techniques. An underlying difference between the two categories lies in the fact that frequency-based techniques produces an overall result detailing the frequencies contained in the entire signal, without any focus on where these frequencies occurred in the signal. In contrast, time-frequency based techniques splits the signals into discrete time units making it possible to identify where the frequencies have occurred in the signals thereby aiding in proper understanding of the signal [5]. A few popular time-frequency techniques mentioned in the literature [7] and [8] are as follows:

- Short -time Fourier Transform (STFT)
- Discrete wavelets Transform (DWT)
- Continuous wavelet Transform (CWT)
- Wigner-Ville distribution (WVD)

As a starting point of investigation on the problem only two feature extraction techniques namely STFT and DWT have been used in the current work. STFT has been most general time-frequency technique for audio signal processing. However the main limitation of STFT is the tradeoff between frequency and time resolution i.e. good time resolution requires short windows and good frequency resolution requires long ones [4]. This limitation of STFT has been complemented by introducing a second technique, DWT.

Before proceeding with any feature extraction, the data was initially pre-processed by first normalizing the signal data to a peak value of 1. Next, the mean was subtracted to remove any DC component. Finally, the data was tapered with a Hamming window using the standard hamming window equation (1):

$$h(k) = 0.54-0.46 \cos ( 2\pi k / N-1) \quad k = 1,..,N \qquad (1)$$

A Hamming window was selected, since it gave good side-lobe suppression. What follows next is a brief discussion concerning STFT and DWT.

### A. Short-time Fourier Transform (STFT)

Short-time Fourier Transform (STFT) is a powerful general-purpose tool for audio signal processing. It defines a particularly useful class of time-frequency distributions [9] which specify complex amplitude versus time and frequency for any signal. This means that it obtains the energy density of a function, simultaneously in the time and frequency. The Short Time Fourier Transform of a signal x(t) using a window function g(t) is defined as follows [9]:

$$STFT(f,s) = \int_{-\infty}^{\infty} x(t) g(t-s) e^{-j2\pi ft} dt \qquad (2)$$

The window g(t) is sliding along the signal x(t) and for each shift g(t-s) the usual Fourier Transform of the product function x(t)g(t-s) is computed. In this particular work a window size of 300 samples with 50% overlapping has been chosen. It should be noted that size of the window has been determined experimentally. The number of STFT features extracted in the current work can be found in Table I. (see table I). A good discussion concerning STFT can be found elsewhere [9], [10] and [11].

### B. Discrete Wavelet Transform (DWT)

The Discrete Wavelet Transform (DWT) is a special case of the Wavelet Transform that provides a compact representation of a signal in time and frequency that can be computed efficiently. The DWT is defined by the equation (3):

$$W(j,k) = \sum_j \sum_k x(k)\, 2^{-j/2}\, \psi(2^{-j} n-k) \qquad (3)$$

Where $\psi(t)$ is a time function with finite energy and fast decay called the mother wavelet. The discrete wavelet transform (DWT) analyzes the signal by decomposing it into its approximate and detailed information [11]. Such

[3] www.mathworks.com
[4] www.ll.mit.edu/IST/lnknet/

decomposition is accomplished by the use of successive high-pass and low-pass filtering and sub-sampling operations, on the basis of the equations (4) and (5):

$$yhigh[k] = \sum_{n} x[n].g[2k-n] \quad (4)$$

$$ylow[k] = \sum_{n} x[n].h[2k-n] \quad (5)$$

Where *yhigh(k)* and *ylow(k)* are the outputs of high-pass and low-pass filters with impulse response *g* and *h*, respectively, after sub-sampling by 2 (decimation). This procedure is repeated for further decomposition of the low-pass filtered signals. Starting from the approximation and detailed coefficients, the inverse discrete wavelet reconstructs the signal, inverting the decomposition step by inserting zeros and convolving the results with the reconstruction filters. In this work, decomposition of the signal was done into 3 levels using popular Daubechies (db10) wavelet family to obtain the approximation and detailed coefficients [12]. The number of DWT features extracted in the current work can be found in Table I. (see table I). A good discussion concerning DWT could be found elsewhere [11].

### *C. Dimension reduction*

After the feature extraction phase, 1500 STFT features and 1800 DWT features have been extracted respectively for either ends of the sleeper (see table I.). However the number of features is still large and any classification scheme would require a lot of data and computational resources. Hence, dimension reduction using principal component analysis (PCA) has been applied for a more compact representation of the feature vectors. PCA is based on the concentration of the relevant data variance into a small number of new variables called principal components (PCs) by means of a suitable mathematical transformation. As a result, the first PC describes the maximum amount of information from the data; the second PC describes the amount of the residual variance and is orthogonal to PC1, and so on. PCA feature provided by LNKNET has been used to reduce the dimensionality [3].

TABLE I. DIMENSION REDUCTION FIGURES

| | **Acoustic signal (left)** | | **Acoustic signal (right)** | |
|---|---|---|---|---|
| **Feature Extraction Technique** | ***Total number of features*** | ***Number of features after performing PCA*** | ***Total number of features*** | ***Number of features after performing PCA*** |
| **STFT** | 1500 | 50 | 1500 | 50 |
| **DWT** | 1800 | 70 | 1800 | 70 |

## IV. PATTERN CLASSIFICATION

After feature extraction the choice of the right classifier might lead to the achievement of successful results. In the current work Multi-layer Perceptron (MLP), Radial basis Function Neural Network (RBFNN) and Support Vector Machine (SVM) have been selected as classifiers due to their contrasting nature. As an example, MLP is based on empirical risk minimization principle where as SVM is based on structural risk minimization principle from statistical learning theory [6] and [13]. It is desired to compare the efficiency of such contrasting techniques and report the most suitable technique for similar problems. Before proceeding any further it is worth mentioning that data were partitioned into training and test sets respectively. As indicated earlier in Section II, impact acoustic signals were collected by making experiments on 200 sleepers. Separation of data into training and test sets has been tabulated in Table II for the sake of clarity.

TABLE II. PARTITION OF DATA INTO TRAINING AND TEST SETS

| **Class** | **Training Set (75% of data)** | **Testing Set (25% of data)** | **Total** |
|---|---|---|---|
| **Good** | 108 | 36 | 144 |
| **Bad** | 42 | 14 | 56 |
| **Total** | 150 | 50 | 200 |

The implementation details of the classifiers used in the current work are:

- A three layer back-propagation neural network, consisting of 100 nodes in each hidden layer and 2 nodes in the output layer with step size of 0.1.
- A SVM based on a Gaussian kernel with standard deviation of 1.5 and a Lagrange Multiplier upper bound of 20 has been used.
- A RBFNN initialized by a clustering procedure known as adaptive k-means algorithm with k = 2.

Initial classification task has investigated usage of one feature at a time against each classifier (see tables III and table IV).

TABLE III. RESULTS OF USING STFT AS A FEATURE WITH THE CLASSIFIERS

| **Classifier** | **Class** | **Total number of patterns** | **Number of errors** | **% Correct classification** |
|---|---|---|---|---|
| **MLP** | Bad (0) | 14 | 8 | 82 |
| | Good (1) | 36 | 1 | |
| | Overall | 50 | 9 | |
| | | | | |
| **RBFNN** | Bad (0) | 14 | 11 | 74 |
| | Good (1) | 36 | 2 | |
| | Overall | 50 | 13 | |
| | | | | |
| **SVM** | Bad (0) | 14 | 8 | 82 |
| | Good (1) | 36 | 1 | |
| | Overall | 50 | 9 | |

TABLE IV. RESULT OF USING DWT AS A FEATURE WITH THE CLASSIFIERS

| Classifier | Class | Total number of patterns | Number of errors | % Correct classification |
|---|---|---|---|---|
| **MLP** | Bad (0) | 14 | 9 | 78 |
| | Good (1) | 36 | 2 | |
| | Overall | 50 | 11 | |
| | | | | |
| **RBFNN** | Bad (0) | 14 | 11 | 72 |
| | Good (1) | 36 | 2 | |
| | Overall | 50 | 14 | |
| | | | | |
| **SVM** | Bad (0) | 14 | 11 | 74 |
| | Good (1) | 36 | 3 | |
| | Overall | 50 | 13 | |

A good discussion concerning the classifiers has not been included in this article due to space limitations. A detailed description concerning pattern classification can be found elsewhere [6].

## V. CLASSIFIER FUSION

According to this, data is initially pre-processed and features are extracted. The resultant features from each method are then passed on to a pattern classifier (or several pattern classifiers) and classification results are obtained and finally fused. Classifier fusion methodology employed in the current work has been illustrated in Fig.4.

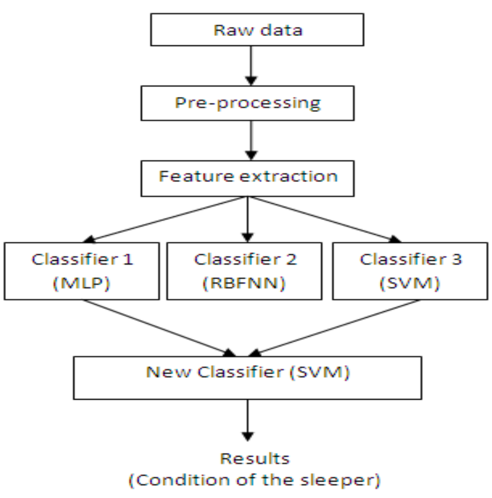


Figure 4. Classifier fusion

Fusion at this level can be achieved in several ways depending on the circumstances. Maximum method, minimum method, T-norm operators, arithmetic mean, majority vote etc serve as typical examples of such fusion strategies. The fact that every classifier produces a different kind of error on a different region of the input space motivates classifier fusion. It is hoped that combining the information of more than one classifier might result in better classification rates for a given problem. In this particular work classifier fusion has been achieved by training another classifier (SVM) using the output of the single best classifiers (MLP and SVM in the current case) as new features. Past work by the authors has employed a similar approach and reported successful results; also motivate the choice of the approach in the current work [14]. Results on the current problem indicate that classifier fusion has achieved better classification accuracy (see tables V and VI) in comparison to using single classifier at a time. A good discussion concerning classifier fusion could be found elsewhere [15].

TABLE V. RESULT OF USING SVM TO CLASSIFY THE OUTPUTS OF SINGLE BEST CLASSIFIERS USING STFT FEATURES

| Classifiers | Class | Patterns | Number of errors | %Correct classification |
|---|---|---|---|---|
| Using SVM as new classifier to classify the outputs of single best classifier (MLP and SVM) | Bad (0) | 14 | 8 | 84 |
| | Good (1) | 36 | 0 | |
| | Overall | 50 | 8 | |

TABLE VI. RESULT OF USING SVM TO CLASSIFY THE OUTPUTS OF SINGLE BEST CLASSIFIERS USING DWT FEATURES

| Classifiers | Class | Patterns | Number of errors | %Correct classification |
|---|---|---|---|---|
| Using SVM as new classifier to classify the outputs of single best classifier (MLP and SVM) | Bad (0) | 14 | 8 | 78 |
| | Good (1) | 36 | 3 | |
| | Overall | 50 | 11 | |

## VI. DISCUSSION

In the current case, the outcome of the each classifier was reported as labels (i.e. condition of the sleeper, 1-Good and 0-Bad). Given that this is a two class problem, resultant errors can be attributed to one of the two cases as:

- False-negative (Bad sleeper being classified as a good sleeper)
- False-positive (Good sleeper being classified as bad sleeper)

Technically, false-negatives could have disastrous consequences; where as a false-positive is fairly benign. On the other hand, false-positives are not very economical due to the fact that sleepers in good condition have to be changed unnecessarily. From the results we can see that the choice of classification (without classifier fusion and with classifier fusion) affects the ratio of false negatives to false-positives as well as having an impact on the overall classification rate. For example, SVM as a new classifier (which takes outputs of single best classifiers as input) has achieved the best overall classification rate of around 84% (see tables V and VI). Results achieved in the current work indicates that STFT features have achieved better classification accuracy in spite of the frequency and time resolution properties (see section III.). In contrast DWT performed slightly worse than STFT. The fact that DWT is mostly used for encoding and decoding signals might be responsible for the poor performance in the current case [5].

## VII. Conclusion and Future work

In this work, a promising approach for automating the process of condition monitoring for the problem of wooden sleeper inspection has been presented. Past work has researched usage of frequency based feature extraction techniques together with pattern recognition to solve the problem. The fact that frequency-based techniques produces an overall result detailing the frequencies contained in the entire signal, without any focus on where these frequencies occurred in the signal has really motivated the use of time-frequency techniques in the current problem. STFT and DWT have been chosen as the main feature extraction techniques in the current case. Pattern classification using classifiers such as MLP, RBFNN and SVM has been tested and compared. Further classifier fusion has been investigated by presenting the output of single best classifiers as inputs to another classifier as features. Results of using one feature at a time together with the classifiers showed that STFT has performed slightly better than DWT. Classification results indicate that both STFT and DWT features performed well; thereby supporting the choice of time-frequency techniques over frequency based techniques where a maximum classification efficiency of around 76% . Usage of SVM as a new classifier that takes outputs of the single best classifiers as new inputs has achieved best rates with an overall classification accuracy of around 84%. This can be regarded as good performance, given that even humans disagree on certain judgments.

We aim to extend the work further by considering other popular time-frequency feature extraction techniques such as Continuous Wavelet Transform and Wigner-Ville Distribution. Another interesting aspect that could be worth investigating is to fuse the features of different feature extracting techniques together to form a single feature vector before presenting it for further classification task. Another future goal is to automate the whole setup of data collection, by placing relevant sensors onboard an automatic vehicle that is capable of running on the tracks. The automatic vehicle should be aimed at acquiring impact acoustic signals and other relevant data (if necessary) automatically. For such a setup to become reality several key issues such as sensor positions for data acquisition, vehicle dynamics, vehicle speed and sensor response etc should be well focused. Issues concerning feature extraction and classification should also be onboard the vehicle to be able to completely replace the manual inspection procedures.

## References


[1] C F Bonnett, 'Practical railway engineering', Imperial College Press, Cop. London, pp. 207, 1996.

[2] S Yella, "Pattern Recognition for Automating Condition Monitoring of Wooden Railway Sleepers", PhD Thesis, Napier University, Edinburgh, U.K. June 2008.

[3] S Yella, N K Gupta, M Dougherty, 'Comparison of pattern recognition techniques for the classification of impact acoustic emissions', Transportation Research: Part C-Emerging technologies, ISSN: 0968-090X, Vol. 15C, Issue 6, pp.345-360, Elsevier, 2007.

[4] M Kepesi , L Weruaga , "Adaptive chirp-based time-frequency analysis of speech signals " Journal of Speech Communication, Vol. 48 No.5 pp. 474-492, 2006.

[5] M Cowling, R Sitte, 'Comparison of techniques for environmental sound recognition', Pattern Recognition Letters, Vol. 24, No. 15, pp. 2895-2907, 2003.

[6] R Duda, P E Hart, D G Stork, 'Pattern classification', Second edition, New York, Wiley, 2001.

[7] L Cohen, Time–Frequency Analysis. Prentice Hall, Englewood Cliffs, NJ, 1995

[8] B Hubbard, 'The World According to Wavelets: the Story of a Mathematical Technique in the Making', Wellesley, Mass, USA, 1995.

[9] J B Allen, "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," IEEE Transactions on Acoustics, Speech, Signal Processing, vol. ASSP-25, pp. 235-238, June 1977.

[10] D Gabor, "Theory of communication", Journal IEE (London), Vol. 93 ,Part III, No. 26, pp. 429-45, 1946.

[11] M Vetterli, J Kovacevic, "Wavelets and Subband Coding", Prentice Hall, Englewood Cliffs, NJ, 1995.

[12] I Daubechies "Orthonormal Bases of Compactly Supported Wavelets" Communications on Pure and Applied Math. Vol.41 pp.909-996, 1988.

[13] V Vapnik, 'The Nature of Statistical Learning', Theory, Springer, 1995.

[14] S Yella, S M Pasha, M Dougherty, N K Gupta, "Classifier fusion for condition monitoring of wooden railway sleepers", Engineer IT, pp-31-35 May 2008.

[15] J Kittler, M Hatef, R P W Duin, J Matas, 'On combining classifiers', IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 3, pp. 226-239, 1998.