Noise Diagnostics of Scooter Faults by Using MPEG-7 Audio Features and Intelligent Classification Techniques

Mingsian R. Bai*, Meng-Chun Chen and Jian-Da Wu

Department of Mechanical Engineering, National Chiao-Tung University, 1001 Ta-Hsueh Road, Hsin-Chu 300, Taiwan

Abstract: A scooter fault diagnostic system that makes use of feature extraction and intelligent classification algorithms is presented in this paper. Sound features based on MPEG (Moving Picture Experts Group)-7 coding standard and several other features in the time and frequency domains are extracted from noise data and preprocessed prior to classification. Classification algorithms including the Nearest Neighbor Rule (NNR), the Artificial Neural Networks (ANN), the Fuzzy Neural Networks (FNN), and the Hidden Markov Models (HMM) are employed to identify and classify scooter noise. A training phase is required to establish a feature space template, followed by a test phase in which the audio features of the test data are calculated and matched to the feature space template. The proposed techniques were applied to classify noise data due to various kinds of scooter fault, such as belt damage, pulley damage, etc. The results reveal that the performance of methods is satisfactory, while varying slightly in performance with the algorithm and the type of noise used in the tests.

1. INTRODUCTION

Taiwan is one of the major scooter manufacturers in the world. Fault diagnosis is an important element in the manufacturing and maintenance process of scooters. In the past, the diagnosis of scooter fault generally relied on well-trained technicians. However, this experience-oriented approach is not only inefficient but also prone to human errors and inconsistencies. It is highly desirable to identify and classify scooter faults in a systematic, automatic and reliable fashion, using machines instead of humans. To this end, this paper proposes a noise diagnostic system on the basis of sound features and intelligent classification techniques for scooter faults.

Fault diagnostics for machines can be largely divided into two categories: the model-based methods and the signal-based methods. For example, Bai *et al.* [1] developed an online fault detection and isolation technique for the diagnosis of rotating machinery. The system consists of a signal-based feature generation module and a model-based fault inference module. Bai *et al.* [2] also proposed an order tracking system for the diagnosis of rotating machinery. The recursive least squares (RLS) algorithm and the Kalman filter are exploited to extract the order amplitudes of vibration signals as features, followed by fault classification using the fuzzy state inference module. On the other hand, Jia *et al.* [3] reported a noise diagnosis technique based on wavelet and fuzzy C-clusters for connecting rod bearing fault. The system is capable of classifying four kinds of bearing fault. Wavelet packet proved to be robust to background noise while extracting fault characteristics. Wang *et al.* [4] suggested a method to diagnose misfire of internal combustion engines based on singular value decomposition (SVD) and artificial neural networks (ANN).

*Address correspondence to this author at the Department of Mechanical Engineering, National Chiao-Tung University, 1001 Ta-Hsueh Road, Hsin-Chu 300, Taiwan; E-mail: msbai@mail.nctu.edu.tw

Extraction of features is an important step prior to a classification system. Since we are mainly dealing with noises due to various types of scooter faults, sound features are derived from the measured noises. In this research, the sound features are based on the MPEG (Moving Picture Experts Group)-7 standard [5]. Although MPEG-7 was originally intended for audio signals, we found it useful to use 11 of the MPEG-7 descriptors for describing noise content. Slightly different in purpose, Crysandt [6] used MPEG-7 descriptors to classify music of multimedia contents. Peeters [7] conducted a comprehensive survey on audio features for sound description in the Content-based Unified Interfaces and Descriptors for Audio/music Databases available Online (CUIDADO) project. Some of the noise descriptors in this research are also derived from his work. Nineteen auditory features including Audio Spectrum Centroid, Linear Predictive Coding (LPC) adopted in this work are summarized in APPENDIX.

Given the abundant set of sound features, one needs to choose only features that are most effective to the problem at hand. There are ways for one to select features according to the nature of problem. Xavier *et al.* [8-10] suggested several pre-selection techniques including Discriminant Analysis (DA) [8], Mutual Information (MI), and Gradual Descriptor Elimination (GDE) [9-10]. Zongker *et al.* [11] evaluated the quality of feature subsets and compared their computational requirements. The Sequential Forward Selection (SFS) algorithm was found to be quite effective for selecting features.

There are also plenty of choices for classification techniques. In the present work, the Nearest Neighbor Rule (NNR) [12], the Artificial Neural Network (ANN) [13, 14], the Fuzzy Neural Network (FNN) [15-17], and the Hidden Markov Model (HMM) [18-20] were employed to classify the noises due to scooter faults. These methods have been extensively used in a variety of problems such as music classification [14], continuous Mandarin word recognition [19], and so forth. These methods achieve high detection rates by forming highly nonlinear decision boundaries in the feature

1874-8376/08

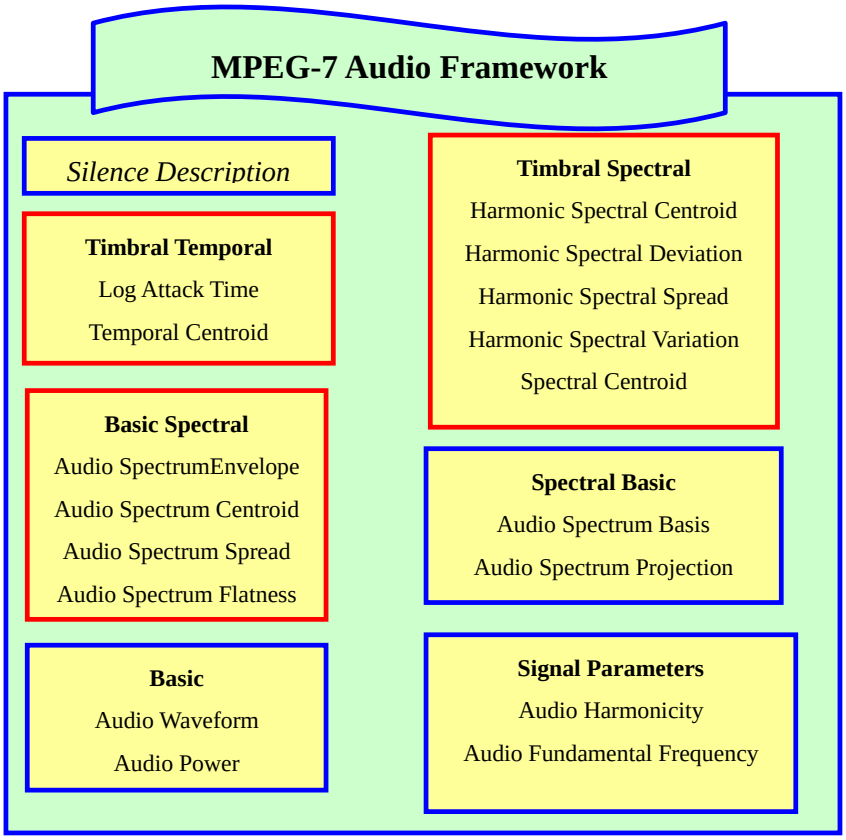


Fig. (1). The MPEG-7 audio framework. The audio framework consists of seventeen low level descriptors of temporal and spectral audio features that can be divided into six groups.

space. Each of these algorithms will be reviewed in the following sections in more detail. The overall architecture of the proposed diagnostic system is depicted in Fig. (**2**). Experiments were conducted for noise data measured on scooters to validate the proposed intelligent diagnostic system. Fault types such as belt damage and pulley damage were examined.

2. FEATURE EXTRACTION

Sound features generally fall into three categories: spectral features, temporal features and statistical features. Nineteen features are used in this study (see APPENDIX). The majority of the sound features are taken from MPEG-7 [5] which is an ISO/IEC standard for describing the multimedia content. Seventeen descriptors in MPEG-7 can be categorized into six groups: Timbral Temporal, Timbral Spectral, Basic Spectral, Basic, Signal Parameters, and Spectral Basis, as shown in Fig. (**1**). Only the first three groups are used in this research. Among the Timbral Temporal descriptors, the Log Attack Time (LAT) characterizes the attack of a sound, or the time it takes for the signal to rise from silence to the maximum amplitude. It signifies the difference between a sharply changed sound and a smoothly changed sound.

$$\mathrm{LAT} = \log_{10}(T_1 - T_0), \tag{1}$$

where T_0 is the time when the signal starts and T_1 is the time when it reaches its maximum.

The Temporal Centroid (TC) characterizes the signal envelope, signifying where in time the energy of a signal concentrates. It is defined as follows:

$$\mathrm{TC} = \frac{\sum_{n=1}^{length(\mathrm{SE})} \frac{n}{\mathrm{SR}} \cdot \mathrm{SE}(n)}{\sum_{n=1}^{length(\mathrm{SE})} \mathrm{SE}(n)}, \tag{2}$$

where SE(n) is the signal envelope at time instant n, calculated using the Hilbert Transform [23] and SR is the sampling rate.

The Basic Spectral descriptors are obtained from the time-frequency analysis of the audio signal. The Audio Spectrum Envelope (ASE) describes the short-term power spectrum of an audio signal as a time-series of spectra on a logarithmic frequency scale in 1/4 octave resolution. It may be used to display a spectrogram, and is defined as follows:

$$\mathrm{ASE}(k) = \frac{|A(k)|^2}{lw \cdot NFFT} \quad k = \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8 \text{ kHz} \tag{3}$$

where $NFFT$ is the FFT size, lw is the window length, $A(k)$ is the magnitude of a component in the frequency range, 62.5 Hz and 8 kHz.

The Audio Spectrum Centroid (ASC) calculates the center of gravity of the log-frequency power spectrum. It indi-

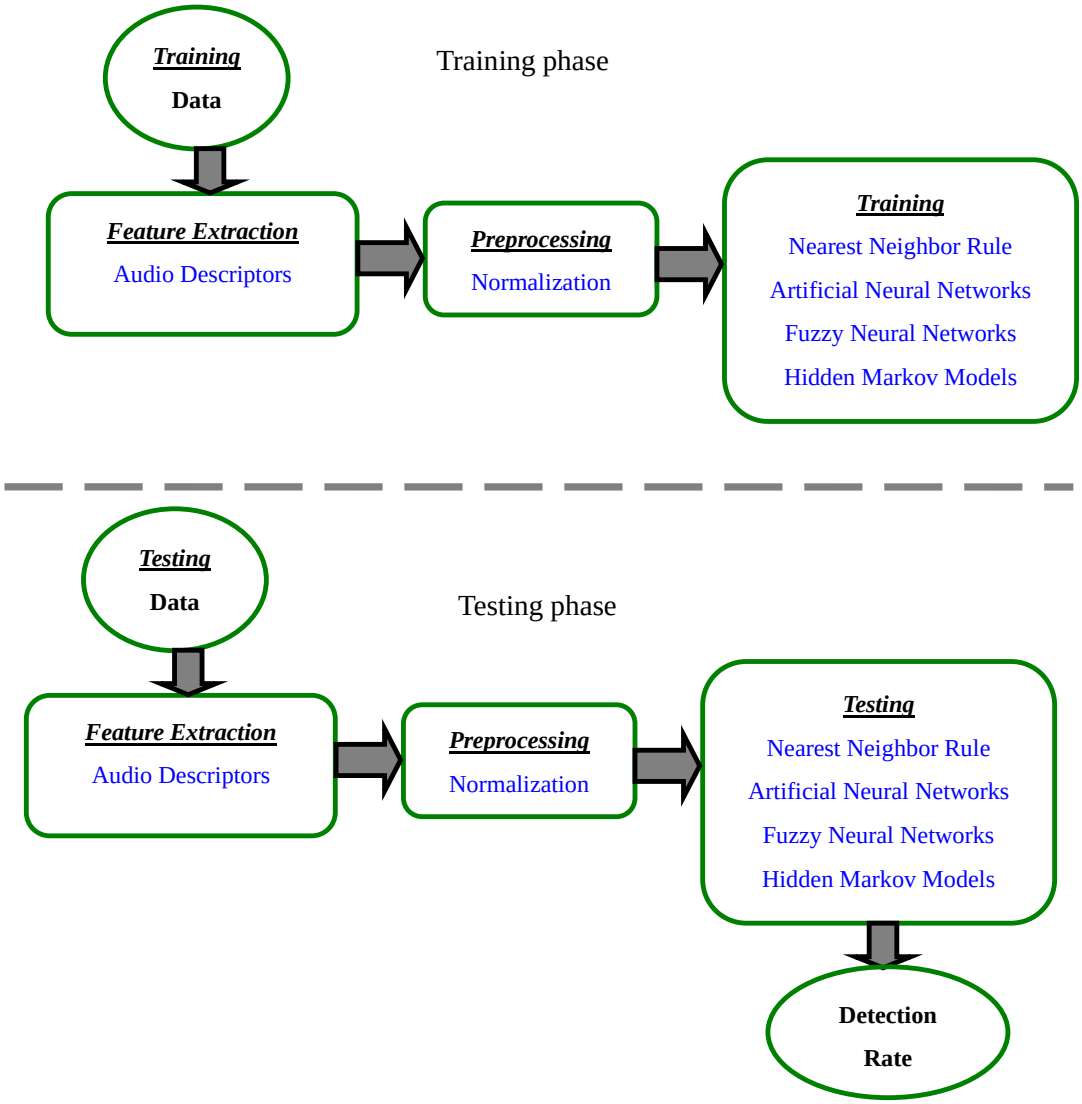


Fig. (2). The general architecture of the noise diagnostic system with feature extraction and classification.

cates whether the spectral content of the signal dominates in high or low frequencies.

$$\mathrm{ASC} = \frac{\sum_k \log_2\left(\frac{f(k)}{1000}\right) P(k)}{\sum_k P(k)}, \tag{4}$$

where $P(k)$ is the power spectrum magnitude associated with the frequency $f(k)$ in Hz.

The Audio Spectrum Spread (ASS) is the second moment of the log-frequency power spectrum, indicating the spread of the power spectrum with respect to the spectral centroid.

$$\mathrm{ASS} = \sqrt{\frac{\sum_k \left(\left(\log_2\left(\frac{f(k)}{1000}\right) - \mathrm{ASC}\right)^2 P(k)\right)}{\sum_k P(k)}} \tag{5}$$

The Audio Spectrum Flatness (ASF) signifies the flatness of the signal spectrum.

$$\mathrm{ASF} = \frac{\sqrt[kh(b)-kl(b)+1]{\prod_{k=kl(b)}^{kh(b)} P(k)}}{\frac{1}{kh(b)-kl(b)+1}\sum_{k=kl(b)}^{kh(b)} P(k)}, \tag{6}$$

where $kl(b)$ and $kh(b)$ are the lower and higher edges of the band b, respectively. An ASF value much greater than unity indicates the presence of tonal components.

The Timbral Spectral descriptors describe the spectral characteristics of sounds in a linear-frequency scale. The Spectral Centroid (SC) is the power-weighted average of the frequency bins in the linear power spectrum.

$$\mathrm{SC} = \frac{\sum_{i=1}^{N_f} \mathrm{ISC}(i)}{N_f}, \tag{7}$$

where

$$\mathrm{ISC}(i)=\frac{\sum_{k=1}^{length(S)} f_i(k)P_i(k)}{\sum_{h=1}^{length(S)} P_i(k)}, \tag{8}$$

where N_f is the number of frames and $\mathrm{ISC}(i)$ is the instantaneous spectral centroid of the *i*th frame

The Harmonic Spectral Centroid (HSC) is the amplitude-weighted mean of the harmonic peaks of a spectrum.

$$\mathrm{HSC}=\frac{\sum_{i=1}^{N_f}\mathrm{IHSC}(i)}{N_f}, \tag{9}$$

where

$$\mathrm{IHSC}(i)=\frac{\sum_{h=1}^{N_h} f_i(h)A_i(h)}{\sum_{h=1}^{N_h} A_i(h)}, \tag{10}$$

where N_h is the number of harmonic peaks, *i* is the frame index, $f_i(h)$ is the frequency in Hz of the *h*th harmonic, $A_i(h)$ is the magnitude of the *h*th harmonic, and $\mathrm{IHSC}(i)$ is the instantaneous harmonic spectral centroid of the *i*th frame.

The Harmonic Spectral Deviation (HSD) indicates the spectral deviation of the log-amplitude components from a global spectral envelope.

$$\mathrm{HSD}=\frac{\sum_{i=1}^{N_f}\mathrm{IHSD}(i)}{N_f}, \tag{11}$$

$$\mathrm{IHSD}(i)=\frac{\sum_{h=1}^{N_h}\left|\log_{10}(A_i(h))-\log_{10}(SE_i(h))\right|}{\sum_{h=1}^{N_h}\log_{10}(A_i(h))}, \tag{12}$$

$$SE_i(h)=\begin{cases}\dfrac{A_i(h)+A_i(h+1)}{2} & \text{when } h=1\\ \dfrac{\sum_{l=-1}^{1}A_i(h+l)}{3} & \text{when } h\in[2,N_h-1]\\ \dfrac{A_i(h-1)+A_i(h)}{2} & \text{when } h=N_h\end{cases} \tag{13}$$

where $SE_i(h)$ is the *h*th harmonic spectral envelope and $\mathrm{IHSD}(i)$ is the instantaneous harmonic spectral deviation of the *i*th frame.

The Harmonic Spectral Spread (HSS) is the amplitude-weighted standard deviation of the harmonic peaks of the spectrum, normalized by the instantaneous Harmonic Spectral Centroid (IHSC).

$$\mathrm{HSS}=\frac{\sum_{i=1}^{N_f}\mathrm{IHSS}(i)}{N_f}, \tag{14}$$

$$\mathrm{IHSS}(i)=\frac{1}{\mathrm{IHSC}(i)}\sqrt{\frac{\sum_{h=1}^{N_h}A_i^2(h)\cdot(f_i(h)-\mathrm{IHSC}(i))^2}{\sum_{h=1}^{N_h}A_i^2(h)}}, \tag{15}$$

where $\mathrm{IHSS}(i)$ is the instantaneous harmonic spectral spread of the *i*th frame and $\mathrm{IHSC}(i)$ is defined in Eq.(10).

The Harmonic Spectral Variation (HSV) is the normalized correlation between the amplitudes of the harmonic peaks of two adjacent frames.

$$\mathrm{HSV}=\frac{\sum_{i=2}^{N_f}\mathrm{IHSV}(i)}{N_f-1}, \tag{16}$$

$$\mathrm{IHSV}(i)=1-\frac{\sum_{h=1}^{N_h}A_{i-1}(h)\cdot A_i(h)}{\sqrt{\sum_{h=1}^{N_h}A_{i-1}{}^2(h)}\sqrt{\sum_{h=1}^{N_h}A_i^2(h)}}, \tag{17}$$

where $A_{i-1}(h)$ represents the magnitude of the *h*th harmonic peak of the previous frame, and $\mathrm{IHSV}(i)$ is the instantaneous harmonic spectral variation of the *i*th frame.

In addition to the MPEG-7 descriptors, we also use other types of feature including the LPC and MFCC. LPC [21] derives the coefficients of a forward linear predictor using the Levinson-Durbin recursion algorithm:

$$\hat{x}(n)=-a(2)x(n-1)-a(3)x(n-2)-\cdots-a(p+1)x(n-p), \tag{18}$$

where $\hat{x}(n)$ is the estimated signal, p is the order of the prediction filter, $a(2),\cdots,a(p+1)$ are the coefficients of the predictor, and $x(n)$ is the input signal.

MFCC [22] are cepstrum coefficients derived from the discrete cosine transform (DCT) based on the critical bandwidths of human hearing. These critical bandwidths are 100 Hz below 1 kHz, but rise nearly exponentially above 1 kHz. Human perception of sound can be regarded as a nonuniform filter bank, with fine resolution at low frequencies and coarse resolution at high frequencies, arranged according to the *Mel-scale frequency* scale. The relationship between the Mel-scale frequency and linear frequency is given by

$$mel\_f(k)=2595*\log_{10}(1+\frac{f(k)}{700}), \tag{19}$$

where $mel\_f(k)$ is the Mel-scale frequency and $f(k)$ is the linear frequency. The computation of MFCC starts with calculating the energy $E(m)$ of the *m*th band

$$E(m) = 10 \cdot \log\left[\sum_k |X(k)|^2 B_m(k)\right], \quad (20)$$

where $X(k)$ is the FFT of the signal $x(n)$, $B_m(k)$ is the triangular weighting function [22] of the mth band and the summation is taken only for the frequency components in the mth band. After that, the energy $E(m)$ is converted to the Quefrency domain by using discrete cosine transform (DCT) to result in the Mel-scale cepstrum coefficients:

$$\text{MFCC}(n) = \frac{1}{M}\sum_{m=1}^{20} \cos\left[\frac{\pi n(m-0.5)}{M}\right] E(m) \quad (21)$$

Finally, Sound Pressure Level (SPL) in dB is also employed as one of the sound features in the work. Zero-crossing rate measures the rate of sign-changes in a signal. Pitch represents the fundamental frequency of the spectrum. Statistical descriptors including Autocorrelation, Skewness and Kurtosis are also included in the feature set. Skewness is a measure of asymmetry of data about the mean and is defined as

$$SK = \frac{E\{(x-\mu)^3\}}{\sigma^3}, \quad (22)$$

where $E\{\}$ denotes the expectation operation, μ and σ are the mean and the standard deviation, respectively, of the time-domain data x. Kurtosis is a measure of how outlier-prone a distribution is. The kurtosis K of a random variable x is defined as

$$K = \frac{E\{(x-\mu)^4\}}{\sigma^4}. \quad (23)$$

In feature extraction, normalization is usually necessary to ensure numerical performance of the ensuing classification process. The features calculated using the aforementioned procedures are normalized as follows:

Step 1. Divide the features into several parts according to the extraction methods.

Step 2. Find the minimum and the maximum in each data set.

Step 3. Rescale each data so that the maximum of each data is 1 and the minimum of each data is -1. That is, the features fall into the full range within the interval $[-1,1]$.

In order to simplify the processing, it is usually desirable to reduce the feature set to its minimum. To this end, the following Sequential Forward Selection (SFS) procedure can be used:

Step 1. Find the single feature that yields the highest successful detection rate.

Step 2. Keep the already-selected features and try to identify a newly added feature that yields the highest successful detection rate.

Step 3. Repeat step 2 until the desired number of features has been selected or until there is no further improvement in the successful detection rate.

3. CLASSIFICATION METHODS

In this section, a review of several algorithms used for the present noise classification problem is given. Only the concepts relevant to the current discussion are addressed.

A. Nearest Neighbor Rule

NNR [12] is a straightforward approach to the problem of classification. In this method, we simply find the closest object from the training set. The thus found object is associated with the same class as that training data. The Euclidean distance is used as the distance measure:

$$dist(x, x_{train}) = |x - x_{train}|^2, \quad (24)$$

where x_{train} is the training data and x is the object being classified. The flowchart of NNR is shown in Fig. (**3**).

B. Artificial Neural Networks

ANN [13] is an important technique in artificial intelligence that mimics human's learning process. Biological learning involves adjustments to the synaptic connections that exist between the neurons, as shown in Fig. (**4**). The relationship of the input and output of a neuron can be represented by the following equation:

$$y = f(net) = f\left(\sum_{n=1}^{N_x} w_n x_n + \theta\right), \quad (25)$$

where y is the output, x_n is the input, w_n is the weighting function, N_x is the number of input nodes, θ is the bias, and $f(\bullet)$ is the activation function which is usually a nonlinear function. In this paper, the following hyperbolic tangent function is used:

$$f(x) = \tanh(x). \quad (26)$$

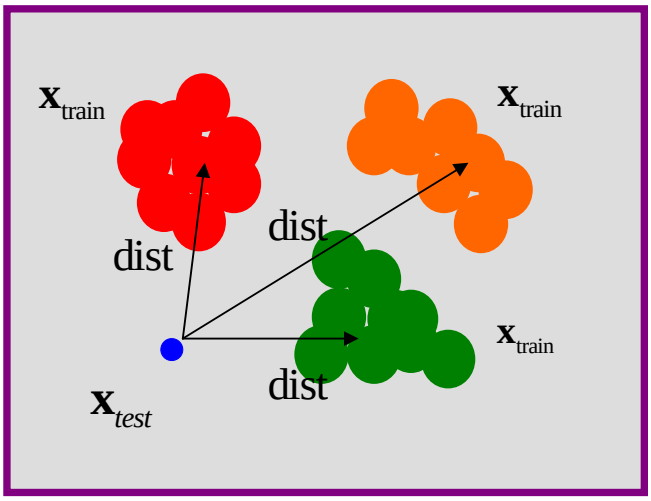


Fig. (3). The idea underlying the NNR algorithm. The distance measure is the Euclidean norm.

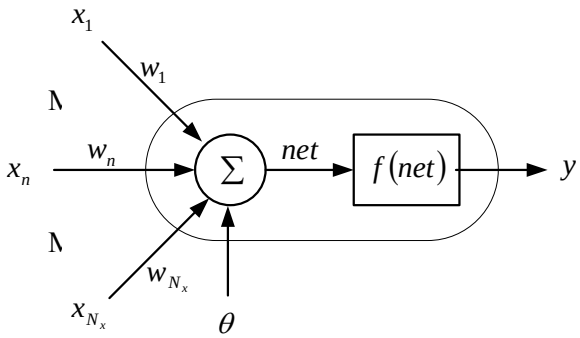


Fig. (4). The schematic structure of a neuron. Learning process in ANN involves adjustments of weights assigned to the connections between the neurons.

Multilayer feedforward networks are widely used which can be shown as Fig. (**5**). Extending the above input and output relationship of the neuron to the multilayer feedforward network leads to the following expression:

$$y_o = f_o\left[\sum_{h=1}^{N_z} w_{oh} \cdot g_h\left(\sum_{i=1}^{N_x} v_{hi} x_i + \theta_{vh}\right) + \theta_{wo}\right], \quad (27)$$

where x_i is the input to the input layer, y_o is the output from the output layer, v_{hi} is the weighting function between the input layer and the hidden layer, w_{oh} is the weighting function between the hidden layer and the output layer, N_x is the number of neurons in the input layer, N_Z is the number of neurons in the hidden layer, N_y is the number of neurons in the output layer, θ_{vh} is the bias of the hidden layer, θ_{wo} is the bias of the output layer, $g_h(\bullet)$ is the activation function of the hidden layer, and $f_o(\bullet)$ is the activation function of the output layer.

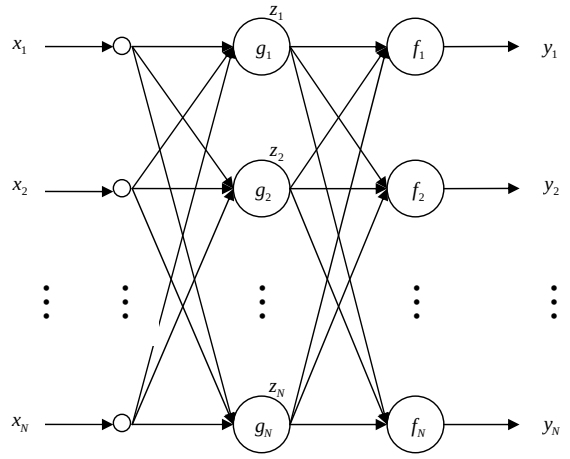


Fig. (5). The structure of a multilayer feedforward ANN. The ANN used in this paper has three layers including the input layer, the hidden layer and the output layer.

The features of signals are used as the inputs to the network. While training the network, the outputs are compared to the targets of the training set. The error between the outputs and the targets is "back propagated" to the network and updates the weightings functions of the nodes in the hidden and output layers. The procedure of back propagation algorithm is given as follows:

Step 1. Decide the network structure and number of neurons.

Step 2. Initialize the network weighting functions.

Step 3. Provide input targets of training set.

Step 4. Calculate the network outputs.

Step 5. Calculate the cost function based on the current weighting functions.

Step 6. Update the weighting functions by using the gradient descent method.

Step 7. Repeat step 3 to step 6 until the network converges.

C. Fuzzy Neural Networks

FNN [17] is a technique that combines fuzzy reasoning with ANN. With reference to Fig. (**6**), the FNN adopted in this paper has five layers: the input layer, the membership layer, the rule layer, the hidden layer, and the output layer.

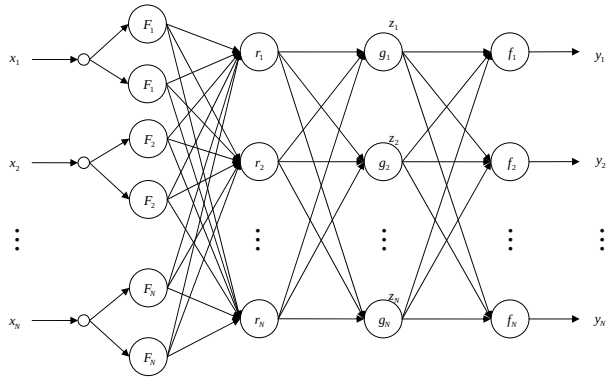


Fig. (6). The structure of the FNN. The FNN used in this paper has five layers including the input layer, the membership layer, the rule layer, the hidden layer and the output layer.

The features of signals are used as the inputs to the network. The second layer is the membership function that connects the fuzzy logics with the neural networks. In the paper, the Gaussian function is selected as the membership function:

$$F_i = e^{\frac{-(x_i-\mu_0)^2}{2\sigma_0^2}}, \quad (28)$$

where μ_0 and σ_0 are the mean and the standard deviation, respectively, of the Gaussian function, and x is the feature of the signal. The fuzzy reasoning rule is established with the following format:

If x_1 is F_1 and …and x_{Nx} is F_{Nx} then y_1 is w_{11} and …and y_{Ny} is w_{NyNz}, (29)

where $x_1,\ldots,x_{Nx}$ are the features of the signal, $F_1,\ldots,F_{Nx}$ are the associated membership functions, $y_1,\ldots,y_{Ny}$ are the outputs of the FNN, and $w_{11},\ldots,w_{NyNz}$ are the weights of the neural network.

The third layer is the rule layer and can be expressed as follows:

$$r=\prod F_i, \tag{30}$$

where r is the truth value of the rule.

The third layer of the FNN is the input layer of the neural network. We use the preceding ANN to train the FNN. The Centroid method is used to obtain the output of the FNN as follows:

$$y_o=f_o\left[\sum_{h=1}^{N_z} w_{oh}\cdot g_h\left(\sum_{i=1}^{N_x} v_{hi}r_i+\theta_{vh}\right)+\theta_{wo}\right] \tag{31}$$

D. Hidden Markov Models

HMM [18] is a very powerful technique for modeling the speech signals and therefore is widely used in speech recognition. An HMM is characterized by five parameters. The first and the second parameters are the number of states in the model (N) and the number of distinct observation symbols per state (M). The third parameter is the state transition probability distribution (**A**):

$$\begin{cases}\mathbf{A}=\{a_{ij}\}\\ a_{ij}=P(q_{t+1}=S_j\mid q_t=S_i),\quad 1\le i,j\le N\end{cases}, \tag{32}$$

where $S=\{S_1,S_2,...,S_N\}$ denotes the set of the states and q_t denotes the state at time t.

The fourth one is the observation symbol probability distribution (**B**):

$$\begin{cases}\mathbf{B}=\left\{b_j(k)\right\}\\ b_j(k)=P(V_k\ at\ t\mid q_t=S_j),\quad 1\le j\le N\\ \qquad\qquad\qquad\qquad 1\le k\le M\end{cases}. \tag{33}$$

where $V=\{V_1,V_2,...,V_M\}$ denotes the set of the symbols.

The fifth one is the initial state distribution (π):

$$\begin{cases}\pi=\{\pi_i\}\\ \pi_i=P(q_1=S_i),\quad 1\le i\le N\end{cases}. \tag{34}$$

An HMM is in general denoted as $\lambda=(A,B,\pi)$ for convenience. Three fundamental problems for HMM design can be stated as follows [18]:

Problem 1. Given the observation sequence $O=o_1,o_2,\cdots,o_T$ and a model $\lambda=(A,B,\pi)$, how do we efficiently compute $P(O\mid\lambda)$, the probability of the observation sequence?

Problem 2. Given the observation sequence $O=o_1,o_2,\cdots,o_T$ and a model $\lambda=(A,B,\pi)$, how do we choose an optimal state sequence $Q=q_1,q_2,\cdots,q_T$ that best explains the observations?

Problem 3: How do we adjust the model parameters $\lambda=(A,B,\pi)$ to maximize the probability of the observation sequence $P(O\mid\lambda)$?

The solution of the problem 1 can be obtained by using a forward-backward procedure. The solution of the problem 2 can be obtained by using the Viterbi algorithm. The solution of the problem 3 can be obtained by using the Baum-Welch algorithm. The details of these HMM solution algorithms can be found in Reference [18].

In this paper, we use HMM to classify noises due to isolated scooter faults. The block diagram of an HMM-based classifier is shown in Fig. (**7**). It is assumed that we have V types of noise to classify and each type is modeled by a distinct HMM. The observations are sound features extracted from the noises corresponding to each fault type. The states are variations in spectral composition of the noise.

Step 1. For each fault type V, we construct a model λ^v by estimating the model parameters that optimize the likelihood of observation vectors in the training set.

Step 2. For each unknown fault type to be classified, calculate the features of the noise data and form the observation sequence $O=\{o_1\ \ o_2\ \ \cdots\ \ o_T\}$. Compute the likelihoods for all possible models $P(O\mid\lambda^v),\ \ 1\le v\le V$ by solving problems 1 and 2 using the Viterbi algorithm. Select the model that yields maximum likelihood.

$$v^*=\arg\max_{1\le v\le V}[P(O\mid\lambda^v)]. \tag{35}$$

4. EXPERIMENTAL INVESTIGATION

In order to validate the proposed noise classification techniques, a series of experiments are performed for scooters. Various noise types were created purposely on the platform of an engine rig, as shown in Fig. (**8**). A scooter with an electronic fuel injection system, single-cylinder, four-stroke, 0.125-liter internal combustion engine was employed in this application. A 1/2-inch condenser microphone was used to measure the noise produced by the scooter. The experiments were undertaken in a semi-anechoic room. Three practical cases were investigated in the experiments.

Case 1. Three kinds of noise including a muffler expansion noise, a one way clutch noise, and a belt damage noise were examined. These noises were measured while the engine was running. The spectrogram of each noise data is shown in Fig. (**9a-c**).

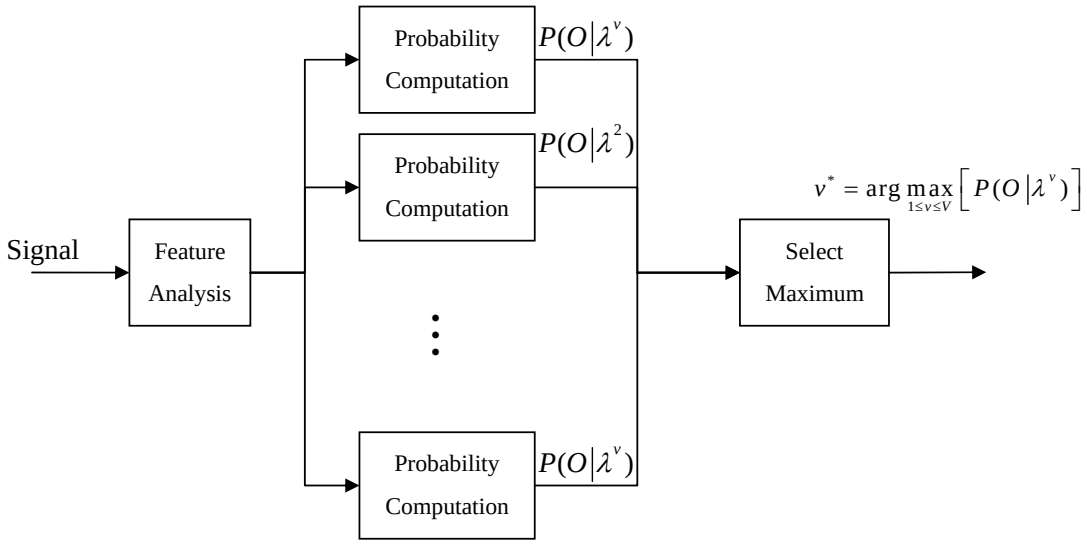


Fig. (7). The block diagram of an HMM-based classifier. In the paper, the computation of probability $P(O|\lambda^v)$ is carried out by using the Viterbi algorithm.

The time-domain waveform of each noise data is shown in Fig. (**10a-c**). The associated features are calculated and shown in Fig. (**11**).

Case 2. Seven kinds of noise including an ac generator noise, a one way clutch noise, a noise of the engine R-cover, a knocking noise of the engine L-cover, a knocking noise of the engine R-cover, a knocking noise of the engine back-cover, and a knocking noise of the engine top-cover were examined. The first three kinds of noise were measured while the engine was running, whereas the others were measured while the engine is idle.

Case 3. Five kinds of noise including the sound from the engine under the normal condition, the noise due to intake manifold leakage, the clutch damage noise, the pulley damage noise, and the belt damage noise were examined. The noises were measured while the engine was running.

In the experiments, noise data were measured and stored under the sampling rate 44.1 KHz with 16 bit resolution. In a total, 600 frames were employed in each case, where 5/6 of the data was used for training, while 1/6 for testing. The signal processing parameters are: hop size = 441, window/frame size = 1323, and FFT size = 2048.

Feature extraction and noise classification were then performed according to the procedure is depicted in Fig. (**2**). Features of the noise were extracted and normalized into the range [-1, 1]. The dimension of the feature space template was 75. With the input features and the target fault types, we started to train the diagnostic system. The previously mentioned classification techniques, NNR, ANN, FNN, and HMM were applied to the scooter noise data. After training, we entered the testing phase to identify the noise types.

The performance of the classification methods is compared in terms of the successful detection rate which is

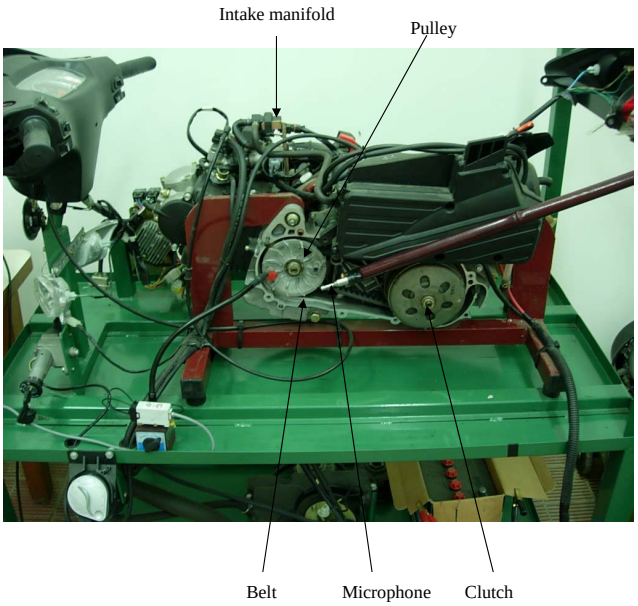
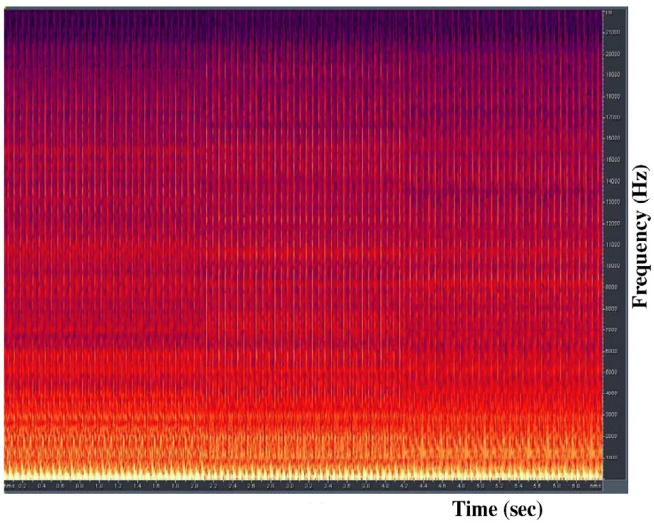


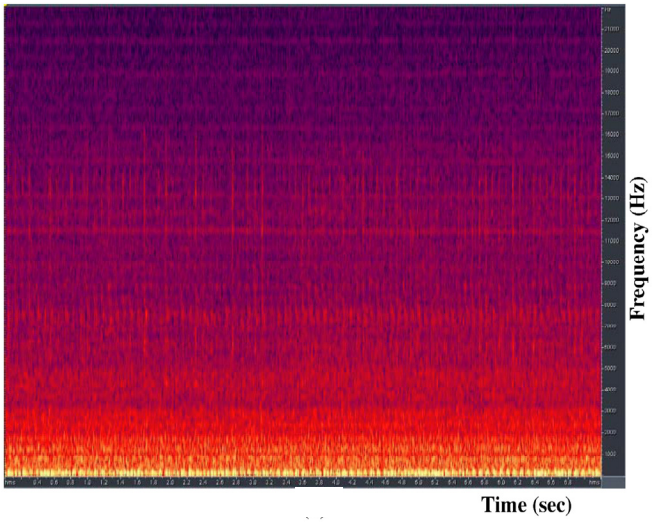
Fig. (8). The Photo of the scooter test rig for the experiments on which various modes of faults can be created.

defined as the ratio of the number of successful detection and the total number of the test frames. For clarity, Tables **1-3** summarize the experimental results of cases 1-3, respectively. The dimension of the feature space is 75. The number of training data is 500 frames and the number of testing data is 100 frames. The performance of classification methods are compared in terms of successful detection rate which is the ratio of the number of frames of correct identification and the total number of frames. The successful detection rate of case 1 was all very high (100%) because the difference of the three noises was clearly audible, even with human hearing. For a more difficult situation of case 2, the successful detection rate was still high (from 86% of NNR to 90% of HMM), but slightly varying with different fault types. Among which,

(**a**)



(**b**)



(**c**)

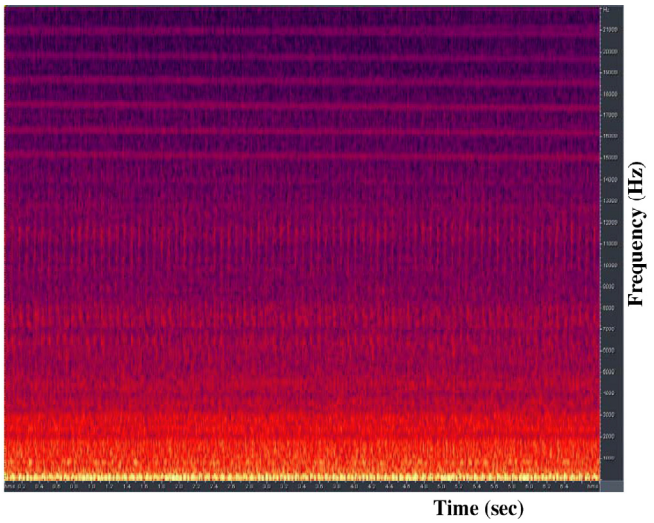
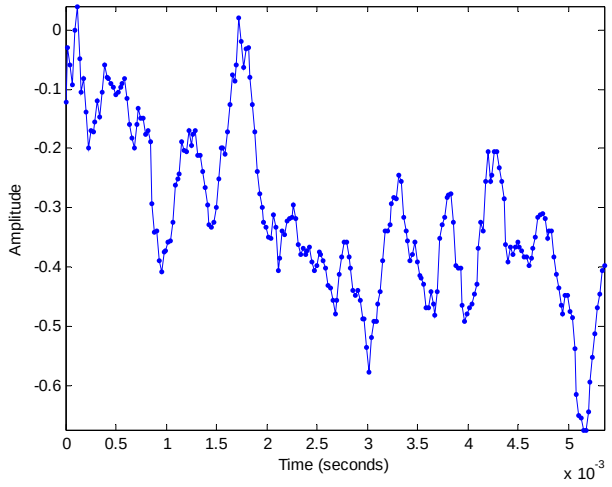
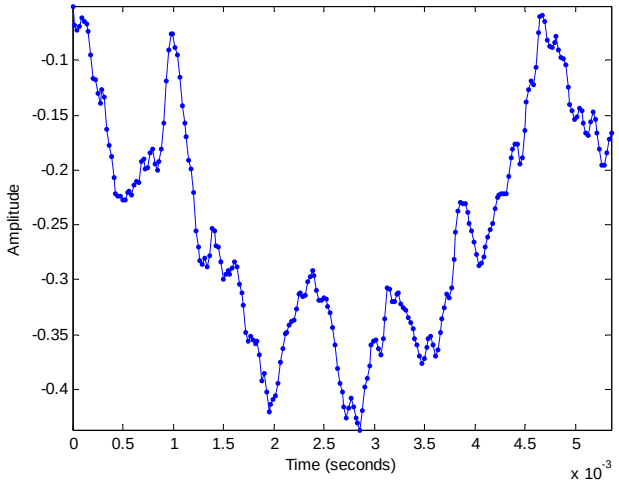


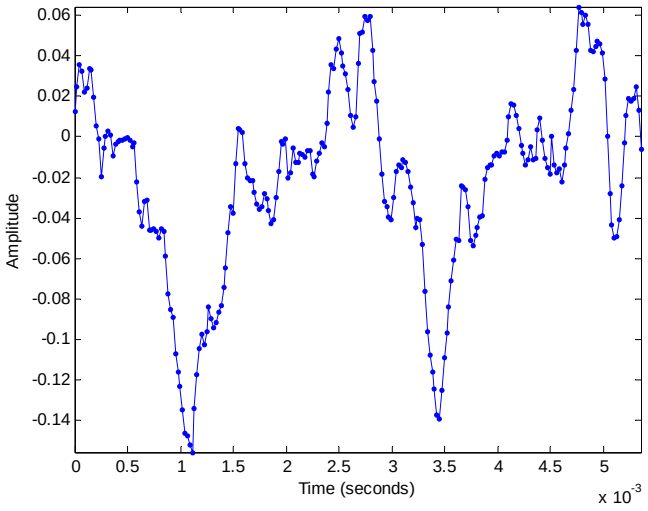
Fig. (9). The spectrogram of the three kinds of noise. (**a**) a muffler expansion noise, (**b**) a one way clutch noise, and (**c**) a belt damage noise.



(a)



(b)



(c)

Fig. (10). The time-domain waveform of three kinds of noise. (**a**) Muffler expansion noise, (**b**) one way clutch noise, and (**c**) belt damage noise.

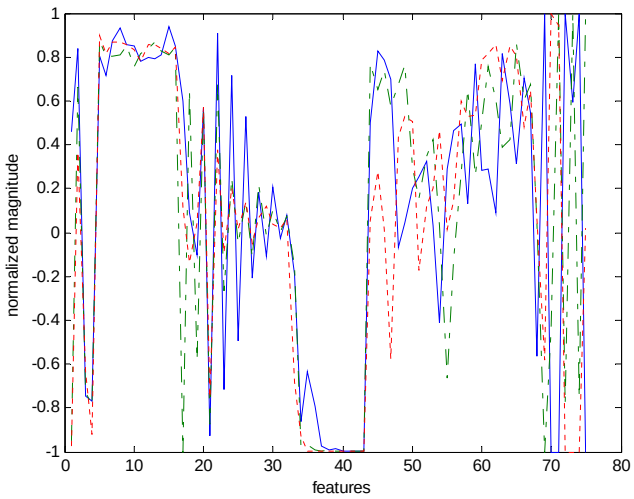


Fig. (11). The features of three kind of noise including a muffler expansion noise (-), a one way clutch noise (-.), and a belt damage noise (:). The abscissa shows the sequence of the features according to APPENDIX.

the ac generator noise and the knocking noise of the engine L-cover seemed to be more difficult to identify than the other fault types, regardless of the classification algorithm used. The successful detection rate of case 3 was also very high (from 96% of HMM to 100% of NNR). A closer inspection of the result reveals that the belt damage was the most difficult fault to identify. In contrast, the engine under normal condition and the air leakage in intake manifold were two easiest fault types for our system. The experimental results suggest that the four classification techniques are effective in identifying fault type correctly using noise data. In particular, FNN has achieved the best performance of fault identification in the test. Introduction of fuzzy logic seemed to have enhanced the performance of plain ANN.

CONCLUSIONS

A scooter fault diagnostic system that makes use of feature extraction and intelligent classification algorithms has been developed in this study. Nineteen sound features including the MPEG-7 descriptors and several other features in the time and frequency domains are extracted from noise data. The extracted features are normalized prior to classification. Classification algorithms including the NNR, the ANN, the FNN, and the HMM are exploited to identify and classify the scooter noise. The proposed diagnostic system was validated by means of practical noise measurement for various fault types. Experimental results revealed that these four classification techniques had attained high successful detection rate in identifying faults of the scooter. Nevertheless, the identification performance may vary slightly with the algorithm and the type of noise used in the tests. Overall, the classification system based on FNN has achieved the best performance in the test. Introduction of fuzzy logic seemed to have enhanced the performance of plain ANN.

As a limitation of the present research, the proposed technique has been verified only by using internal tests. If external tests are used, where the training data and the testing

Table 1. Identification Results of case 1 Summarized for Four Classification Algorithms. Three Kinds of Scooter Noise Including a Muffler Expansion Noise, a One Way Clutch Noise, and a Belt Damage Noise were Examined

(A)

Nearest Neighbor Rule	Muffler Expansion	One Way Clutch	Belt Damage
Muffler Expansion	100	0	0
One Way Clutch	0	100	0
Belt Damage	0	0	100
Successful Detection Rate	100%		

(B)

Artificial Neural Network	Muffler Expansion	One Way Clutch	Belt Damage
Muffler Expansion	**100**	0	0
One Way Clutch	**0**	100	0
Belt Damage	**0**	0	100
Successful Detection Rate	100%		

(C)

Fuzzy Neural Network	Muffler Expansion	One Way Clutch	Belt Damage
Muffler Expansion	100	0	0
One Way Clutch	0	100	0
Belt Damage	0	0	100
Successful Detection Rate	100%		

(D)

Hidden Markov Model	Muffler Expansion	One way Clutch	Belt Damage
Muffler Expansion	100	0	0
One Way Clutch	0	100	0
Belt Damage	0	0	100
Successful Detection Rate	100%		

(E)

Classification Method	Successful Detection Rate
Nearest Neighbor Rule	100%
Artificial Neural Network	100%
Fuzzy Neural Network	100%
Hidden Markov Model	100%

Table 2. Identification Results of Case 2 Summarized for Four Classification Algorithms. Seven Kinds of Scooter Noise Including the AC Generator Noise, a One Way Clutch Noise, a Noise of the Engine R-Cover, a Knocking Noise of the Engine L-Cover, a Knocking Noise of the Engine R-Cover, a Knocking Noise of the Engine Back-Cover, and a Knocking Noise of the Engine Top-Cover

(**A**)

Nearest Neighbor Rule	AC Generator	Knocking Noise of L-Cover	One-Way Clutch	Noise-R-Cover	Knocking Noise of R-Cover	Knocking Noise of Back-Cover	Knocking Noise of Top-Cover
AC Generator	76	0	0	24	0	0	0
Knocking Noise of L-Cover	0	89	0	0	11	0	0
Noise-One-Way Clutch	12	0	84	6	0	0	0
Noise-R-cover	2	0	0	98	0	0	0
Knocking Noise of R-Cover	0	12	0	0	88	0	0
Knocking Noise of Back-Cover	0	2	0	0	12	86	0
Knocking Noise of Top-Cover	0	10	0	0	0	4	86
Successful Detection Rate	86%						

(**B**)

Artificial Neural Network	AC Generator	Knocking Noise of L-Cover	One-Way Clutch	Noise-R-Cover	Knocking Noise of R-Cover	Knocking Noise of Back-Cover	Knocking Noise of Top-Cover
AC Generator	85	0	0	15	0	0	0
Knocking Noise of L-cover	0	84	0	0	16	0	0
Noise-One-Way Clutch	0	0	98	2	0	0	0
Noise-R-cover	0	0	1	99	0	0	0
Knocking Noise of R-Cover	0	10	0	0	90	0	0
Knocking Noise of Back-Cover	0	2	0	0	0	95	3
Knocking Noise of Top-Cover	0	0	0	0	0	3	97
Successful Detection Rate	92%						

(**C**)

Fuzzy Neural Network	AC Generator	Knocking Noise of L-Cover	One-Way Clutch	Noise-R-Cover	Knocking Noise of R-Cover	Knocking Noise of Back-Cover	Knocking Noise of Top-Cover
AC Generator	96	0	0	4	0	0	0
Knocking Noise of L-cover	0	82	0	0	10	5	3
Noise-One-Way Clutch	0	0	99	1	0	0	0
Noise-R-cover	0	0	2	98	0	0	0
Knocking Noise of R-Cover	0	0	0	0	98	0	2
Knocking Noise of Back-Cover	0	4	0	0	0	96	0
Knocking Noise of Top-Cover	0	0	0	0	1	0	99
Successful Detection Rate	95%						

(Table 2) contd…..

(**D**)

Hidden Markov Model	AC Generator	Knocking Noise of L-cover	One-Way Clutch	Noise- R-Cover	Knocking Noise of R-cover	Knocking Noise of Back-Cover	Knocking Noise of Top-Cover
AC Generator	80	0	0	20	0	0	0
Knocking Noise of L-cover	0	84	0	0	16	0	0
Noise-One-Way Clutch	0	0	90	10	0	0	0
Noise-R-cover	0	0	10	90	0	0	0
Knocking Noise of R-Cover	0	0	0	0	95	5	0
Knocking Noise of Back-Cover	0	0	0	0	4	96	0
Knocking Noise of Top-Cover	0	0	0	0	1	0	99
Successful Detection Rate	90%						

(**E**)

Classification Method	Successful Detection Rate
Nearest Neighbor Rule	86%
Artificial Neural Network	92%
Fuzzy Neural Network	95%
Hidden Markov Model	90%

Table 3. Identification Results of Case 3 Summarized for Four Classification Algorithms. Five Kinds of Scooter Noise Including the Engine Under The Normal Condition, the Noise Due to Intake Manifold Leakage, the Clutch Damage Noise, the Pulley Damage Noise, and the Belt Damage Noise

(**A**)

Nearest Neighbor Rule	Normal	Air Leakage	Clutch	Pulley	Belt
Normal	100	0	0	0	0
Air Leakage	0	100	0	0	0
Clutch	0	0	100	0	0
Pulley	0	0	0	100	0
Belt	0	0	0	0	100
Successful Detection Rate	100%				

(**B**)

Artificial Neural Network	Normal	Air Leakage	Clutch	Pulley	Belt
Normal	100	0	0	0	0
Air Leakage	0	100	0	0	0
Clutch	0	0	98	2	0
Pulley	0	0	2	98	0
Belt	0	0	2	3	95
Successful Detection Rate	98%				

(Table 3) contd…..

(**C**)

Fuzzy Neural Network	Normal	Air Leakage	Clutch	Pulley	Belt
Normal	100	0	0	0	0
Air Leakage	0	100	0	0	0
Clutch	0	0	98	2	0
Pulley	0	0	2	98	0
Belt	0	0	3	3	94
Successful Detection Rate	98%				

(**D**)

Hidden Markov Model	Normal	Air Leakage	Clutch	Pulley	Belt
Normal	100	0	0	0	0
Air Leakage	0	99	1	0	0
Clutch	0	3	96	1	0
Pulley	0	1	0	97	2
Belt	0	4	0	6	90
Successful Detection Rate	96%				

(**E**)

Classification Method	Successful Detection Rate
Nearest Neighbor Rule	100%
Artificial Neural Network	98%
Fuzzy Neural Network	98%
Hidden Markov Model	96%

APPENDIX

Sound features generally fall into. Nineteen features in three categories, spectral features, temporal features and statistical features, are used in this paper. For instance, the spectral features are descriptors calculated using the Short Time Fourier Transform (STFT) or other model-based methods such as Audio Spectrum Centroid, Audio Spectrum Flatness, Linear Predictive Coding (LPC), Mel-scale Frequency Cepstrum Coefficients (MFCC), etc. The temporal features are descriptors calculated using the time-domain signal such as Zero-crossing Rate, Temporal Centroid and Log Attack Time. The statistical features are descriptors calculated based on statistical analysis such as Skewness and Kurtosis. The descriptors from MPEG-7 are marked with "※".

Spectral Features	Dimension	Sequence	Spectral Features	Dimension	Sequence
※ Audio Spectrum Centroid	1	33	※ Harmonic Spectral Centroid	1	69
※ Audio Spectrum Flatness	24	44~67	※ Harmonic Spectral Deviation	1	70
※ Audio Spectrum Envelope	10	33~43	※ Harmonic Spectral Spread	1	71
※ Audio Spectrum Spread	1	68	※ Harmonic Spectral Variation	1	72
※ Spectral Centroid	1	74	Sound Pressure Level	1	19
LPC	13	20~32	MFCC	13	4~16
Pitch	1	17	Autocorrelation	1	2
Temporal Features	**Dimension**	**Sequence**	**Temporal Features**	**Dimension**	**Sequence**
※ Log Attack Time	1	73	※ Temporal Centroid	1	75
Zero-crossing rate	1	1			
Statistical Features	**Dimension**	**Sequence**	**Statistical Feature**	**Dimension**	**Sequence**
Skewness	1	18	Kurtosis	1	3

data belong to different set, it is anticipated that not only the detection rate will drop but also the difference in robustness of each classification algorithm against uncertainties and variations of the data may become clear. This conjecture should be examined *via* external tests in the future research. Extension of the present system to accommodate more fault types by using more features and classification algorithms is currently on the way.

ACKNOWLEDGEMENTS

The work was supported by the National Science Council in Taiwan, under the project number NSC94-2212-E009-018.

REFERENCES

[1] Bai MR, Hsiao IL, Tsai HM, Lin CT. Development of an on-line diagnosis system for rotor vibration via model-based intelligent inference. J Acoust Soc Am 2000; 107: 315-323.

[2] Bai MR, Huang JM, Hong MH, Su FC. Fault diagnosis of rotating machinery using an intelligent order tracking system. J Sound Vibrat 2005; 280 (3-5): 699-718.

[3] Jia J, Kong F, Liu Y, *et al*. Noise diagnosis research based on wavelet packet and fuzzy C-clusters about connecting rod bearing fault. Trans Chin Soc Agric Mach 2005; 36: 87-91.

[4] Wang HM, Chen X, An G, Fan XH. I.C. Engine misfire fault diagnosis based on singular value theory and neural networks. Trans Chin Soc Intern Combust Engines 2003; 21(6): 449-452.

[5] Martinez JM. MPEG-7 Overview (version 9),"ISO/IEC JTC1/SC29/WG11 N5525, Pattaya, Thailand, March, 2003.

[6] Crysandt H, Wellhausen J. Music classification with MPEG-7," Proc. SPIE Storage and Retrieval for Media Databases, Santa Clara, January, 2003.

[7] Peeters G.A. Large set of audio features for sound description (similarity and classification) in the CUIDADO project, Technical report, IRCAM, Paris, France, 2004.

[8] Peeters G, Rodet X. Automatically Selecting Signal Descriptors for Sound Classification, Proc. International Computer Music Conference, Goteborg, Sweden, September, 2002; pp. 455-458.

[9] Arie A, Livshin, Rodet X. Musical Instrument Identification in Continuous Recordings," Proc. the 7th Int Conf on Digital Audio Effects, Naples, Italy, October, 2004.

[10] Arie A, Livshin, Peeters G, Rodet X. Studies and improvements in automatic classification of musical sound samples," Proc. the International Computer Music Conference, September, 2003.

[11] Zongker D, Jain A. Algorithms for feature selection: An evaluation," Proc. the 13th International Conference on Pattern Recognition, Vienna, Austria, August, 1996.

[12] Cover T, Hart P. Nearest Neighbor pattern classification. IEEE Trans Inform Theor 1967; 13(1): 21-27.

[13] Demuth H, Beale M. Neural Network Toolbox User's Guide, MathWorks, 1998.

[14] Scott P. Music Classification using Neural Networks. Project of EE37B, Stanford university 2001.

[15] Wang CH, Liu HL, Lin CT. Dynamic optimal learning rates of a certain class of fuzzy neural networks and its applications with genetic algorithm. IEEE Trans Syst Man Cybern Part B: Cybernetics 2001; 31(3): 467-475.

[16] Li CJ, Lee C, Wang S. Diagnosis and Diagnostic Rule Extraction Using Fuzzy Neural Network, Symposium on Intelligent Systems, International Mechanical Engineering Congress and Exposition, ASME, November 11-16, 2001.

[17] Lin CT, Lee CSJ. Neural Fuzzy Systems, Prentice-Hall, 1996.

[18] Rabiner L. A tutorial on hidden markov models and selected applications in speech recognition. Proc IEEE 1989; 77(2): 257-286.

[19] Chang WS. Development of PSHMM Algorithm and Its Application to Continuous Mandarin Digits Recognition, Master Thesis and Technique Report, Multi-media Information Lab., Institute of Electrical and Control Engineering, National Chiao Tung University, 2001.

[20] Chai W, Vercoe B. Folk Music Classification Using Hidden Markov Models, Proc. Internation Conference on Artificial Intelligence, Las Vegas, June 2001; pp. 4.3.4/1-5.

[21] Harma A. Linear predictive coding with modified filter structures. IEEE Trans. Speech Audio Proc 2001; 9(8): 769-777.

[22] Logan B. Mel Frequency Cepstral Coefficients for Music Modeling, proceedings of the Int. Symposium on Music Information Retrieval (ISMIR), Plymouth, USA, October 2000.

[23] Oppenheim AV, Schafer RW. Buck JR. Discrete-Time Signal Processing, Prentice Hall, 1999.

Received: January 5, 2008 Revised: February 12, 2008 Accepted: February 14, 2008