Acceptance and Classification of Gears Based on Sound Signals Spectrum Analysis

Isaack Adidas Kamanga[1] and Peter Mwita Joseph [2]

Electronic Engineering Department

[1-2]Tianjin University of Technology and Education

Tianjin

China, P.R

ABSTRACT

This paper work presents our proposed systematic approach in solving the industrial problem on identifying the good and bad gears by analyzing the spectrum of sound signals produced by the gears; In this proposed approach we treat this problem as a pure machine learning and classification problem having two classes (good gears and bad gears), whereby we have analyzed the spectrum of several sound signal samples from good and bad gears then extracted five audio features from their spectrum (Short time energy, zero crossing rate, Spectral entropy, pitch and block energy entropy) which after investigation, we found a significant difference between the two classes. We formed a 5D features vector; we used 10 features vectors from good gears and 10 from bad gears as our training samples for finding the discriminating point. After features extraction, we apply Support Vector Machine (SVM) learning classification method to classify the new features vector extracted from the unknown sound from the gear.
Key Words: Gear, Short-Time Energy, Zero crossing rates, Energy Entropy, Spectral entropy.

1. INTRODUCTION

The problem of classifying good and bad gears can be solved by many ways, including by image pattern recognition, ANN and machine learning and Acoustic signal spectrum analysis. This report presents the approach of using spectrum analysis of sound wave produced by the gears to separate the bad gear from good one, This problem was inspired to us by our course master, Prof Lil, of advanced digital signal processing who also arranged the factory visit for us to experience the challenges faced by the factory in sorting the two gears (the gears are quiet small so wastage of time and accuracy in sorting are among the big challenges), this work is primarily aimed at eliminating these challenges. In this work we treat this problem as pure machine learning and classification problem having two classes (good gears and bad gears), so the whole report is about given a gear sound, how actually we classify it as being good or bad. So finally we input sound wave produced by a gear (audio acquisition) extract features, then classify

2. METHODOLOGIES

The first step in any audio classification system is to extract features i.e. identify the components of the audio signal that are good for classification and discarding all the other stuffs which carries information like background noise, emotion etc In this approach number of features have been extracted, those features that showed no significant difference between good gears and bad gears during training, like Mel frequency cepstral coefficients (MFCC) feature, where not considered. The features extracted and used for training in this work are short-time energy, block entropy of the signal, spectral entropy, zero crossing rate and pitch. Each of these features is discussed shortly in this chapter. After features extraction and training, a clear line has been drawn to approximate the discriminating point between the two classes and, an appropriate method for classifying the new features vector extracted from the sound produced by the gear whose status (bad or good is to be determined).For testing purpose, several sound were collected from ten good gears and ten bad gears then classified them by using our approach, figure 1 below shows the steps followed in identifying the status of a given gear, finally we came up with good results and the conclusion and recommendations are drawn.
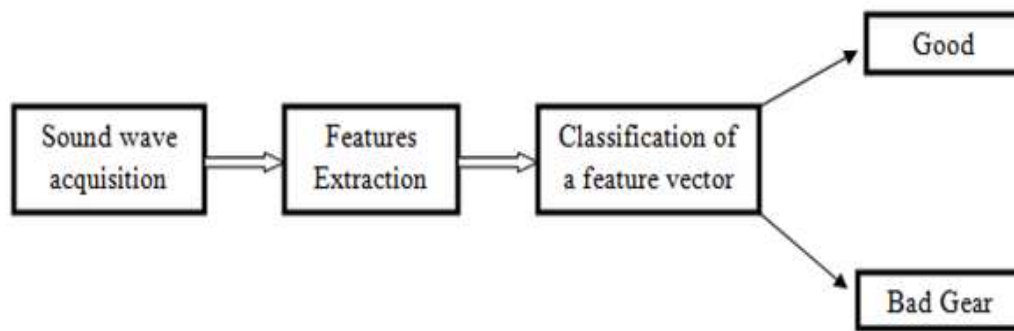
Figure 1: Procedures in identifying the status of a given gear

A Gear

A gear is a rotating machine part having cut *teeth*, or cogs, which mesh with another toothed part to transmit torque [3]. A gear to be a good has to meet all the necessary features after fabrications, good cut teeth, correct thickness and diameter, for gears in our case the number of cut teeth has to be 50 on a reducer and 11 cut teeth for inner section. Any gear that will have the features opposite to those enlightened above shall be treated as a bad one. Figure 2 below illustrates by pictures how the good gear and bad one look like by differentiating their cut teeth arrangement
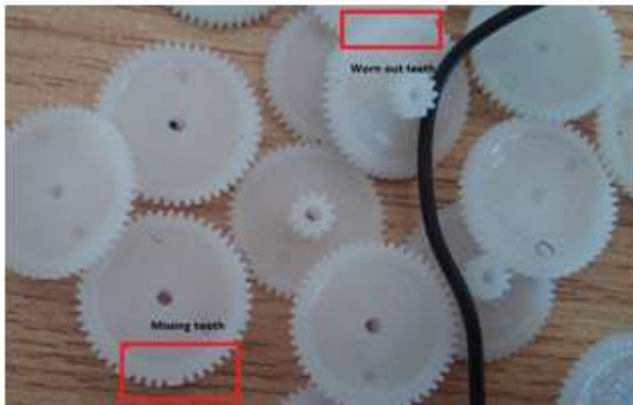

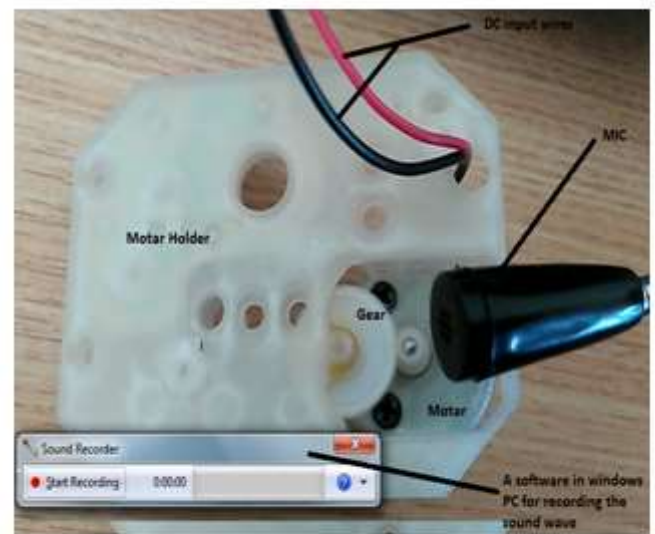Figure 2: Image illustrating good and bad gears by cut teeth arrangement and shape

Figure 3: Arrangement used to input sound waves generated from several gears

3. SOUND WAVE ACQUISITION

Arrangement used to record the audio is shown in figure 4 below following the principle illustrated by the figure 3 below, the MIC was connected to PC and the sound wave was recorded for 10 good gears and then repeated the same process for bad gears, because of our old version of MATLAB software, we converted the wave from .wma to .wav file for smooth analysis. We then acquired the recorded sound from working directory for analysis, Initial samples up to 250000 are on flat spectrum so are of less use to our research, these samples could be due to time lag between starting recording software and start of the sound (ringing) from the gear and should not be considered as a general case. Since the recorded stereo is double channelled, we take only a left part for our analysis, in many cases the left channel is used because you may experience strange effects when mixing stereo channels. The phase-shift differs for each frequency and some frequency may get lost when averaging. Especially when analyzing sound properties, the use of only one channel is encouraged.
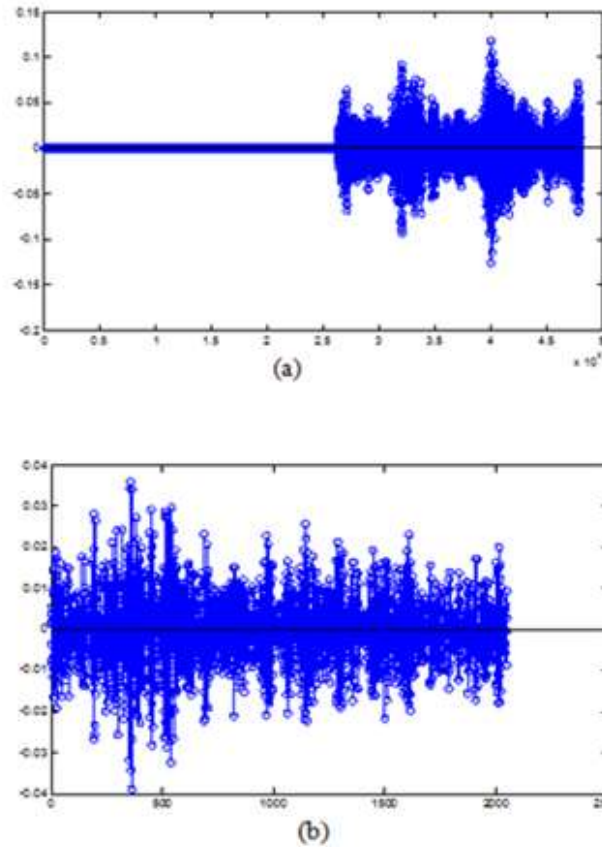
Figure 4: (a) Input sound wave and (b) the cut version of it

We then further reduced the number of samples before extracting features to avoid unnecessary wastage of memory, figure 5 below illustrate the final interval of input signal we considered for analysis.
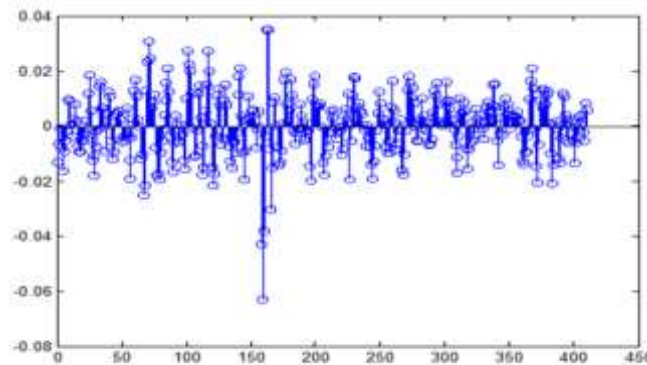

Figure 5: Samples considered for features extraction

4. FEATURES EXTRACTION

4.1. Definition Of Features Used
4.1.1 Short-Time Energy

The energy E of a discrete time signal $x(n)$ is defined by the equation 1 below. For many audio signals such a measurement is of less importance, since it gives little information about time dependent characteristics of such signals.

$$E = \sum_{n=-\infty}^{\infty} x^2(n) \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \text{(1)}$$

Instead of considering the energy for the entire duration which does not make sense for the short duration signal, we consider the short

time energy which shows a significant and notable value. Also the amplitude of an audio signal varies with time. A convenient representation that reflects these amplitude variations is the short time energy of the signal. In general, the short time energy is defined by equation 2 bellow

$$E_m = \sum_n \left[x(n)w(m-n) \right]^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

The above expression can be rewritten as

$$E_m = \sum_n (x(n))^2 h(m-n) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(3)$$

Where by $h(m) = w^2(m)$

In the above expression the term $h(m)$ is interpreted as the impulse response of a linear filter. The nature of short time energy representation is determined by the choice of the impulse response. The bandwidth of the hamming window is twice the bandwidth of a rectangular window of the same length. Moreover the hamming window results in a much higher attenuation outside the bandwidth when compared to the rectangular window. However, in both cases, increasing the length of the window decreases the bandwidth. Short term energy is used in different audio classification problems. In speech signals, it provides a basis for distinguishing voiced speech segments from unvoiced ones [4]. In the case of a very high quality speech, the short term energy features are used to distinguish speech from silence

4.1.2 Implementation

A MATLAB function has been programmed, it takes in the short piece of signal acquired from the gear, then with user defined length of the window and sample step (same step must be used for all samples), it returns the matrix of short time energy of length L, the maximum value of this is taken as a feature as it showed a difference for good and bad gear. The calculated values for all samples are shown in table 1

4.1.2 Zero Crossing Rates

The zero-crossing rate is the rate of sign-changes along a signal [2], i.e. the rate at which the signal changes from positive to negative or back. This feature has been used heavily in both speech recognition and music information retrieval, being a key feature to classify percussive sounds. In the case of discrete-time signals, a zero crossing is said to occur if successive samples have different algebraic signs. The rate at which zero crossings occur is a simple measure of the frequency content of a signal. Zero-crossing rate is a measure of number of times in a given time interval/frame that the amplitude of the gear sound signals passes through a value of zero. Rough estimates of spectral properties can be obtained using a representation based on the short time average zero-crossing rate [2].
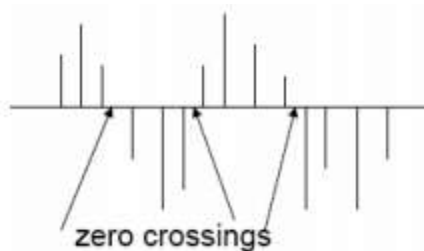

Figure 6: Definition of zero crossing

Mathematically zero crossing rates are defined by equation 4 below

$$z_n = \sum_{m=-\infty}^{\infty} \left| \text{sgn}[x(m)] - \text{sgn}[x(m-1)] \right| w(n-m) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. \quad (4)$$

Where

$$\text{sgn}[x(m)] = \begin{cases} 1, x(m) \geq 0 \\ -1, x(m) < 0 \end{cases}$$

And

$$w(n) = \begin{cases} 0, otherwise \\ \dfrac{1}{2N} \; for, 0 \leq n \leq N-1 \end{cases}$$

Implementation

A MATLAB function has been programmed to calculate the zero crossing rates of a particular gear signal, Zr the mean value from Zr matrix is computed and taken as a feature; the results are included in table 1 of results.

4.1.3 Energy Entropy

Entropy as a thermodynamic state variable was introduced into physics by German physicist Rudolf Claudius in second half of 18th century [7]. It was originally defined as a reversibly received elementary heat over absolute temperature [7]. Of course such a definition has no sense for signal processing. However, it started a diffusion of entropy as a term into the other areas. The entropy as a measure of system disorganization appeared for the first time in connection with the First postulate of thermodynamics: "Any macroscopic system which is in time t0 in given time-invariant outer conditions will reach after a relaxation time the so-called thermodynamic equilibrium. It is a state in which no macroscopic processes proceed and the state variables of the system gains constant time-invariant values." The entropy of a system is maximal when the system has reached the thermodynamic equilibrium. The above depicted key idea promoted entropy to a generic measure of system disorganization. Entropy (or an entropy-based feature) can be computed from any finite set of values, e.g. a parametric vector, a discrete spectral density estimate, or directly from a segment of a digital signal [7] .Another definition of entropy (equation 5 below) was later proposed for use in mathematics, especially statistics:

$$dS = -\sum_{k=1}^{N} p(x_k) \log_2 p(x_k) \; \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \; (5)$$

Where $\{x_1, x_2 \dots\dots x_N\}$ is a set of random phenomena, and $p(x_k)$ is a probability of a random phenomenon x_k

Relation between entropy and sound from gears is based on our finding that the disorganization of the sound wave from bad gears has the highest entropy value while the good gears showed significantly lower entropy value as it is more organized and required an extra energy to be produced in such an organized form. According to the above findings, the entropy can be used in signal processing for separating the useful signal from a background noise also

The probability $p(x_k)$ is approximated by the difference of the spectrum component and the mean value

$$p(x_k) = \left| s^-[k] - s[k] \right| \; \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \; (6)$$

4.1.4 Spectral Entropy

Spectral flatness often referred to as Wiener entropy is a measure used in digital signal processing to characterize an audio spectrum [8]. Spectral flatness is typically measured in decibels, and provides a way to quantify how noise-like a sound. The spectral flatness for a sampled digital signal can be computed by implementing equation 7 below

$$W.E = \frac{\exp\left(\dfrac{1}{N}\sum_{n=1}^{N} \ln x(n)\right)}{\dfrac{1}{N}\sum_{n=1}^{N} x(n)} \; \dots\dots\dots\dots\dots\dots\dots\dots\dots \; (7)$$

Where $x(n)$ is a sequence and N is the number of samples used, we used $N = 400$ at a step of 5 for the first 2000 samples of 2048 samples for the computation. A function for the computation of spectral entropy has been included in appendix of this report. Hence flatness is calculated by dividing the geometric mean of the power spectrum by the arithmetic mean of the power spectrum equation 8 above. Generally after our investigation we found that the spectral entropy for good gears is low and bad gears have higher values.

With the pre assumption that a bad cut teethed gear will make louder noise at bad teeth location suggest that the spectral entropy for bad gears shall have higher values than for good gears. The results are summarized in table 1 in chapter 2 of this report. Wiener entropy is a pure number, that is, it does not have units. On a scale of 0-1, white noise has an entropy value of 1 and complete order, and a pure tone has an entropy value of 0.
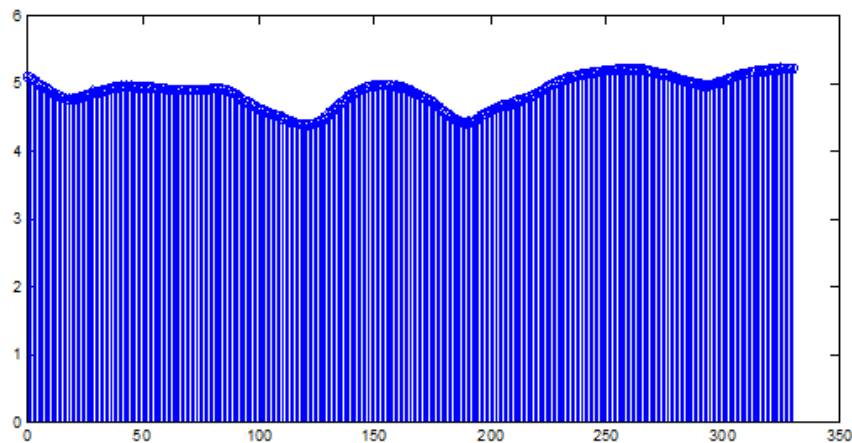
(a)
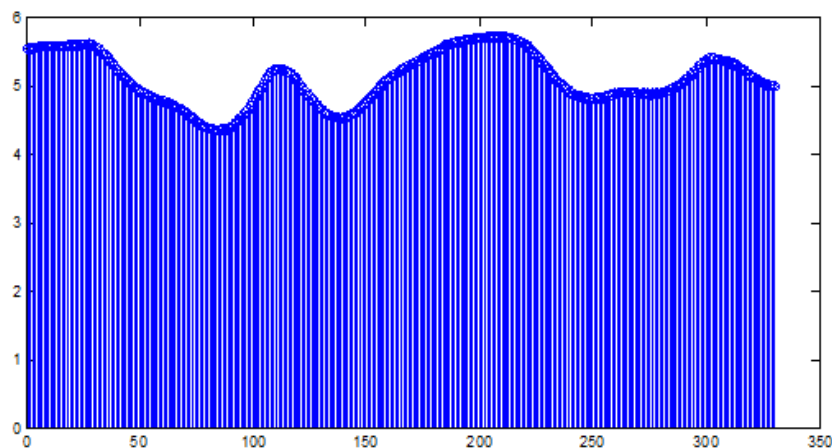
(b)

Figure 7: Typical spectral flatness for (a) Bad gear and (b) Good gear

4.1.5 Pitch

Pitch is an auditory sensation in which a listener assigns musical tones to relative positions on a musical scale based primarily on their perception of the frequency of vibration [1][2], Pitch is closely related to frequency, but the two are not equivalent. Frequency is an objective, scientific attribute that can be measured. Pitch is each person's subjective perception of a sound wave, which cannot be directly measured. However, this does not necessarily mean that most people won't agree on which notes are higher and lower. Since it's difficult to measure pitch of a sound wave, we then calculated the maximum of the absolute value of the spectrum of a sound wave. The findings are summarized in table 1 for good and bad gears. Generally we found that the value is significantly higher for the bad gear sound than for good gear sound. A fast Fourier transform is used to obtain the spectrum of a piece of sound wave (not less than one revolution of a gear), then the maximum value of its absolute value is taken for the training. Fourier transform is defined by equation 8 below, so instead of considering pitch which is technically not possible to approximate, we considered the maximum value of the spectrum

$$X(k) = \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N}, 0 \le k \le N-1 \dots\dots\dots\dots\dots\dots\dots\dots\dots (8)$$

Which are complex exponential sequences, as a result, the DFT coefficients, $X(k)$ are complex numbers even if $x[n]$ are real, that is why for training we computed the maximum of the absolute value

4.2 Results of Features Extraction

Below is a table (table 1) of the values obtained after features extraction, b stands for bad gear and g stands for good gear.

Table1: Table of results of features extracted

Signal	Block Energy Entropy	Short Time Energy(max)	Zero crossing rate(mean)	Spectral Entropy (max)	Max of spectrum
b1	9.732	0.1155	0.111	0.5234	6.717
b2	9.63	0.0807	0.124	0.4892	8.454
b3	9.82	0.1074	0.096	0.4123	3.081
b4	8.89	0.0758	0.132	0.38	3.876
b5	9.821	0.0943	0.118	0.5127	6.516
b6	8.89	0.0942	0.112	0.4653	18.29
b7	9.821	0.0628	0.098	0.4652	10.71
b8	9.679	0.07	0.121	0.4653	6.23
b9	9.693	0.101	0.109	0.4123	5.037
b10	9.798	0.1038	0.122	0.55	6.456
g1	10.05	0.1874	0.181	0.3281	2.67
g2	10.12	0.1739	0.211	0.2331	2.291
g3	10.04	0.1802	0.196	0.1822	2.015
g4	10.1	0.1689	0.187	0.2221	2.367
g5	9.997	0.1787	0.221	0.3101	2.12
g6	10.1	0.191	0.218	0.1837	1.893
g7	10.13	0.172	0.187	0.1789	2.312
g8	10.06	0.1598	0.204	0.2362	2.891
g9	10.1	0.1766	0.198	0.1827	1.968
g10	10.1	0.1824	0.214	0.2168	2.541

4.2.1 Training

We used artificial neural network (ANN) to train our data

The structure of neural network is shown below in figure 8 below
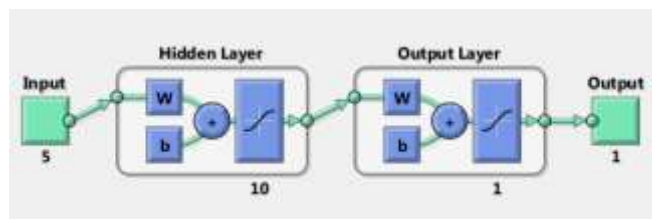
Figure8: Structure of neural network used for training data in table 1

4.2.1.1 Training Inputs

Input to the neural network consists of five inputs (we have five features) for twenty samples which makes up a 5x20 matrix. Five columns of feature values in table 1, becomes the five rows for the input, each row representing one feature

4.2.1.2 Training Target

Training target consists up of a 1x20 matrix (table 2) whereby each column representing the known class of an input sample, we used 0 to represent the class bad gear and 1 to represent class good gear

Table 2: Training target

0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1

4.2.1.3 Training Results

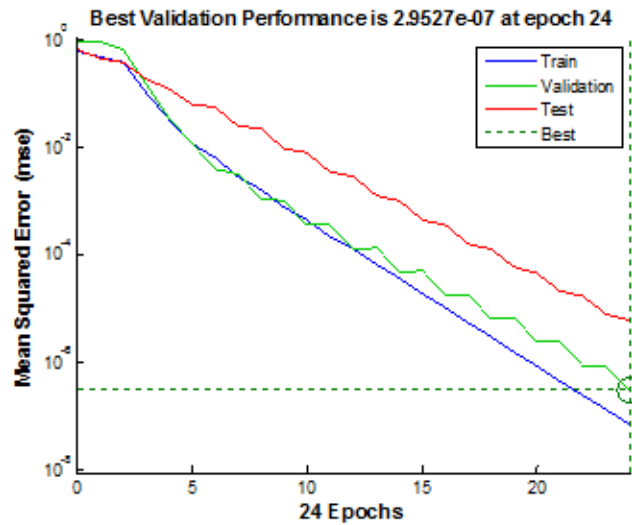Training performance plot shows very good results, with little error for testing samples
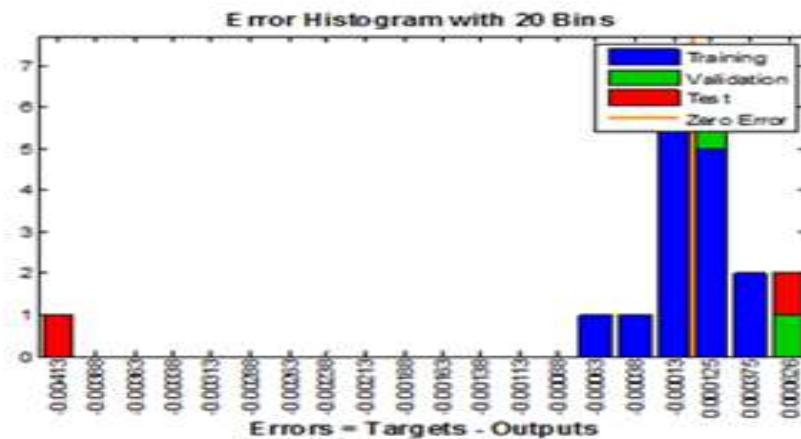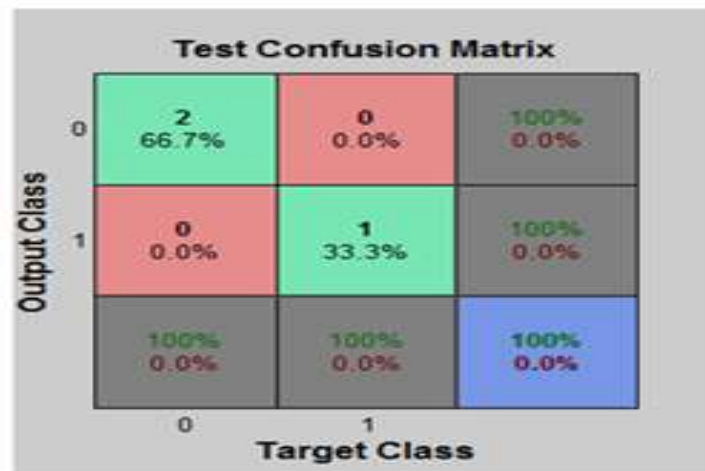

Figure 9: Training performance plot of the table 1

Errors histogram and test confusion matrix below suggest that the training and testing results are very good

(a)

(b)

Figure 10: Training error histogram and the test confusion matrix of the table 1

5. CLASSIFICATION

5.1 Definition

It is the process by which we automatically assign an individual item to one of a number of categories or classes, based on its characteristics [7]. The items are gear's produced sounds; their characteristics are the features we extract from them (Short time energy, zero crossing rate, Spectral entropy pitch and block energy entropy); the classes (good or bad gear) fit the problem definition. The complexity lies in finding an appropriate relationship between features and classes [7]. Class models must be learned on many sound samples to properly account for between/within class variations. The natural range of features must be well represented on the sample population Failure to do so leads to over fitting: training data only covers a sub-region of its natural range and class models are inadequate for new data [7].

5.2 Classification Methods Categories

Classification algorithms are divided into supervised and unsupervised algorithms. In a supervised classification, a labelled set of training samples is used to "train" the algorithm whereas in the case of an unsupervised classification the data is grouped into some clusters without the use of labelled training set. Parametric and nonparametric classification is another way of categorizing classification algorithms. The functional form of the probability density of the feature vectors of each class is known in parametric methods. In nonparametric methods on the other hand, no specific functional form is assumed in advance, instead the probability density is rather approximated locally based on the training data [6].

5.3 Support Vector Machine

Support vector machine is a supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier [6][7], we used this method for our experimentation, implemented by MATLAB and the input consisted of a 5-D feature vector and the result were the percentage of match, the class that won the higher percentage was considered a winner class

6. EXPERIMENT AND RESULTS

6.1 Experiment Set Up

We recorded the sounds from 5 good gears and 5 sounds from bad gears from the samples given from the factory whose actual status were known (figure 11), for experiment only, we used SVM classification method by MATLAB to classify each of the input signal from the table 2 below. The status of the gear from SVM classier was noted as well as the actual class of the gear

Figure 11: Good and Bad gears used to extract sounds for testing

6.2 Result Discussion

From the results in table 3 and our training results, it is clear that the approach gives more convincing results which can be improved by increasing the size of training samples and use as many features as possible.

Table 3: Table of experiment results

Input signal	Actual status	Experiment results
s1	bad	bad
s2	bad	bad
s3	bad	bad
s4	bad	bad
s5	bad	bad
s6	good	good
s7	good	bad
s8	good	good
s9	good	good
s10	good	good

Efficient of our approach is given by equation 6 below

$$efficiency = \frac{Total_{gears} - Wrongly_{classified\_gears}}{Total_{gears}} x100\% \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (9)$$

$$efficiency = \frac{10-1}{10} x100\%$$

$$efficiency = 90\%$$

7. CONCLUSION AND RECOMMENDATIONS

The number of samples used in our research is very small but the results obtained is very good, the efficiency could be boosted by using a large size of training input and increased number of feature vector. According to Bayes classification ideas, he suggested for any classification to approach optimal (Bayes method is an optimal one demanding the statistics of the event/events to be totally known) the size of straining samples should be very large and the dimension of the training vector to be very large. Also since this is a two class classification problem, one could use a likely hood ratio test to minimize the average risk in classification.

ACKNOWLEDGEMENT

REFERENCES

[1] Sanjit K.M, (2006), Digital Signal Processing: A Computer Based Approach, Third Edition. Publishing House of Electronics Industry Beijing

[2] Theodoros, G and Aggelos, P (2014), Introduction to Audio Analysis: A MATLAB Approach, Academic Press Elsevier Ltd , San Francisco

[3] Erling Wold, Thom Blum, Douglas Keislar, James Wheaton, "Content-Based Classification Search and Retrieval of Audio", *IEEE Multimedia Mag.*, vol. 3, pp. 27-36, 1996

[4] H. Misra, S. Ikbal, H. Bourlard, and H. Herman sky, "Spectral entropy based feature for robust asr," in Proc. ICASSP, May 2004, pp. 193–196.

[5] https://en.wikipedia.org/wiki/Spectral_flatness

[6] Abdillahi Hussein Omar; Audio Segmentation and Classification
[7]Juan Pablo Bello, Sound Classification

AUTHOR PROFILE

Isaack Adidas Kamanga received the B.S.in Telecommunications Engineering from University of Dar Es Salaam in 2012, Tanzania, worked as Engineer in fields of RF, VSAT, Optic fiber network for 2 years, currently working as Tutorial assistant at Dar Es Salaam Institute of Technology (DIT), and doing masters degree in Signal and Information processing Engineering at Tianjin University of Technology and Education, P.R China, with research base in computer vision.

Peter Mwita Joseph: Telecom Engineer at the Ministry of Work, Transport and Communication Tanzania and PG Student at the Tianjin University of Technology and Education. His research field is Security Systems, Machine Learning, Signal and Image Processing. He obtained his B.Sc. in Electronics and Communication at the University of Dar es Salaam and CCNA certified.