Analysis and Classication of Car Engine Sounds to Diagnose Failures with the Help of Neural Networks

Luis Andrés Méndez-Ortega[1] and Orion Fausto Reyes-Galaviz[2]
[1]Universidad Autonoma de Tlaxcala
Facultad de Ciencias Básicas, Ingenieria y Tecnologia
http://www.ingenieria.uatx.mx/licenciaturas/ic/
[2]Instituto Tecnologico de Apizaco
http://www.itapizaco.edu.mx/
Apizaco, Tlaxcala. México
E-mail: [1]siul222@hotmail.com, [2]orionfrg@ingenieria.uatx.mx
(Paper received on August 12, 2006, accepted on September 25, 2006)

Abstract. The detection of possible car engine failures can be achieved through the identification on the vibration changes that are produced while the motor is operating. A person with enough mechanical experience is able to recognize a flaw in the engine, just by listening to the noise and sensing the vibrations produced. A car owner can notice a flaw in the motor, when a shift on the movement of the engine is felt, or if there's an abnormal noise on it. Nevertheless, it is hard to guess what kind of malfunctions produces certain motors noises and vibrations, making it hard to work directly on the engine's flaw. In this work, a method is proposed to classify, and directly detect, some common car engine failures, just by using the engine's recorded sound. To accomplish this goal, we used Linear Prediction Cepstral Coeficients (LPCC) as the main acoustic characteristics extractor, Principal Component Analysis, and a Time Delay Neural Network for the pattern recognition. We obtained recognition results of up to 95.58% when predicting four common car failures.

1. Introduction

A failure on any engine can produce mild or violent vibrations that can damage the motor; if left ignored. These vibrations can cause fissures, waste and/or overheating of important parts and reduce the machine's performance. Besides, the vibrations are a good indicator of the mechanical performance and they are also very sensitive to the malfunction evolution. In general, these types of flaws are preceded by shifts on the sound conditions, vibration, power loss, etc. These indicators are a sign of some kind of future failure on the engine's performance. Although, not all the sources of sounds or vibrations are inevitable, due to the fact that some are natural to the machine's operation by it self, enhancing with this the importance of identifying which correspond to a possible malfunction. Most of the sounds that an engine produces are generated by rotating mechanisms inside the motor; also, the machines that have an internal combustion system, produce more noise due to the explosions created inside.

© H. Sossa and R. Barrón (Eds.)
Special Issue in Neural Networks and Associative Memories
Research in Computing Science 21, 2006, pp. 75-85

In general, a trained ear can quickly detect the existing differences between the noises that are called normal, and a series of abnormal noises produced, for example, by vibrating gears, poundings on the engine, hisses, etc. If this kind of relevant acoustic information exists, hidden inside the engine's noise, the extraction, recognition, and classification of the acoustic characteristics could be possible, using automatic mechanisms, in order to obtain reliable diagnostics.

The engine noise can be an important clue to the prediction of possible failures, and can be helpful to preserve a good motor performance. The analysis presented in this work, joint with the monitoring systems that actually exist, can produce more tools that can help the motor specialists obtain robust diagnostics when repairing car engines. Taking this into account, it is proposed to capture and analyze the acoustic sound produced by these kind of machinery, and use techniques of acoustic characteristics extraction, and pattern recognition, in order to classify the different types of noises that are produced by engines with good and bad performance. On the next section we will present a review of prior comparative studies on the field. Section 3 details the fundamental basis in the noise recognition process and describes our proposed system. Section 4 deals with acoustic processing and the feature extraction method which uses the Linear Prediction Cepstral Coefficients (LPCCs) method. A fundamental theory on pattern classification, Principal Component Analysis, and Time Delay Feed-Forward Neural Networks, is given in Section 5. The complete system implementation can be found on Section 6. Our experimental results, that go up to 95.58%, are shown in section 7, and the concluding thoughts are presented in Section 8.

2. State of the Art

Many systems have been proposed on the acoustic pattern recognition field in the past few years. Depending on the problem that has to be solved, a common point of some approaches is focused on the detailed spectral analysis of the acoustic signal. In [1] Mario E. Munich uses Mel Frequency Cepstral Coefficients (MFCC) for the acoustic analysis, Gaussian Mixture Models (GMM), and Hidden Markov Models, combined with Bayesian Subspace Methods, applied to the automatic recognition of acoustic characteristics from vehicles; this is used on military operations, for surveillance purposes, achieving a precision of 83%. Huadong Wu, Et Al [2], also worked on the vehicle recognition field. They proposed that each vehicle model has the same kinds of noises, vibrations, hops, and tire friction. They used a method called eigenfaces, used most commonly on the face recognition field, to characterize the noise patterns and use them to recognize the vehicle; this method is also known as the Korhunen-Loeve expansion or as the Principal Component Analysis (PCA), the results on this work seem promising. In [3], Edgar A. Estupiñan, Et Al use an analysis of mechanical vibrations, as a part of their predictive maintenance, to establish the mechanical health of their machines, preventing with this, future flaws. They propose techniques based on the Fast Fourier Transform to analyze the noise produced by the vibration of slow rotatory machinery. Pedro N. Saavedra [3], [4], presents some techniques applied to the vibration analysis; he particularly analyzed the use of fissure detection on machinery axis and rafters. For the theoretical study, the Finite Element was used, and the fissure was modeled using lineal fractomechanical theory.

It is difficult to compare these systems; because the data bases, experimental conditions, number of samples per second, frame size, and type of recognitions, are different. Although, the analysis of different techniques, suggests interesting possibilities of combining them in order to obtain better recognition results.

3. Car Engine Noise Recognition Process

The car engine noise recognition process is basically a pattern recognition problem, and it is similar to speech recognition. The goal is to take the motor's sound wave as an input, and at the end recognize the engine's malfunction. Generally, the engine noise recognition process is done in two steps; the first step is the acoustic processing, or features extraction, while the second is known as pattern processing or classification. The proposed system can be seen in Figure 1. For this case, in the acoustic analysis, the engine's signal is processed to extract relevant features in function of time. The feature set obtained from each noise sample is represented by a vector, and each vector is taken as a pattern. As for the pattern recognition methods, four main approaches have been traditionally used: pattern comparison, statistical models, knowledge based systems, and connectionist models. We focus on the use of the last one.

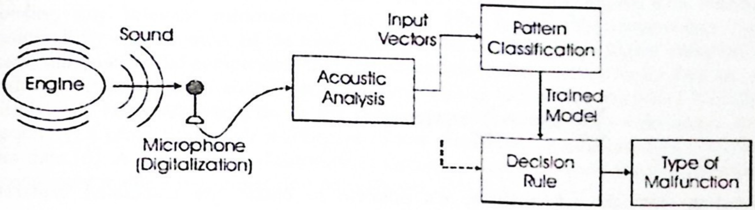

Fig. 1 Engine Noise Recognition Process.

4. Acoustic Processing

The acoustic analysis implies the application and selection of filter techniques, feature extraction, signal segmentation, and normalization. With the application of these techniques the signal is described in terms of its fundamental components. An engine signal is too complex and codifies more information than the one needed to be analyzed and processed in real time applications. For this reason, in the engine noise recognition process we use a feature extraction function as the first plane processor. Its input is the engine sound signal, and its output is a vector of features that characterizes key elements of the sound wave. In this work we used Linear Prediction Cepstral Coefficients (LPCC) [5] as the feature extraction method.

4.1. Linear Prediction Cepstral Coefficients

The Lineal Prediction (LP) method is historically one of the most important methods used for the voice analysis [5]. Its fundamental basis is to establish a filter model for the sound source. With enough number of parameters, the LP model can establish a suitable approximation to the spectral structure of any kind of sound. This is why this technique is used to analyze our acoustic samples. The LP method gets this name because it pretends to extrapolate the value of the sound's next sample $x(n)$ as the weighted sum of the previous samples $x(n-1), x(n-2), \ldots, x(n-k)$:

$$x(n) = \sum_{i=1}^{K} a_i x(n-1) \quad (1)$$

To do so, a coefficient computation must be made, by minimizing an error function E, specifically the least square, over a window of size N.

$$E = \sum_{n=0}^{N-1} e^2(n) \quad (2)$$

$$E = \sum_{n=0}^{N-1} \left(x(n) - \sum_{i=1}^{K} a_i x(n-1) \right)^2 \qquad 0 \le n \le N-1$$

Departing from the LP analysis, it's possible to obtain the associated cepstral coefficients expression (LPCC):

$$c(0) = \log(1) = 0 \quad (3)$$

$$c(i) = -a(i) - \sum_{j=1}^{i-1} \left(1 - \frac{j}{i} \right) a(j) c(i-j) \quad 1 \le i \le N_c \quad (4)$$

A common transformation, over this kind of coefficients, is known as the delta cepstral coefficients or delta cepstrum coefficients. We can obtain these by applying the next expression:

$$\Delta c_j(i) = \frac{1}{2T+1}\sum_{k=-T}^{T} k \cdot c_{j+k}(i) \tag{5}$$

5. Engine Noise Pattern Classification

After extracting the acoustic features of each noise sample, the feature vectors are obtained; each one of these vectors represents a pattern. These vectors are later used for the classification process. For the present work, we focused on connectionist models, also known as neural networks (NN), to classify these vectors (patterns). The NN are combined with Principal Component Analysis to reduce the vector's dimensionality; in order to improve the training/testing processing time, obtaining with this, a more efficient classification system.

5.1. Principal Component Analysis

The Principal Component Analysis (PCA), is commonly used in signal processing, statistics, and neural computation. The objective of PCA is to reduce the dimensionality of a variable set p to a set m with less variables to improve the readability of the data. The variable set, or original characteristics, occasionally are redundant, that is why it's useful to reduce the dimension of the original data, without loosing any relevant information. The PCA tries to find the components that successfully explain most of the total variation; the data that has a higher variation is kept, eliminating the components that contribute less to the variation in the data set. A reduced set is easier to analyze and interpret. This reduction has important benefits; firstly, the computational cost in the subsequent processing stages is reduced, and secondly, a projection inside a subspace of low dimensions is very useful to visualize the data [6]. A simple PCA illustration is shown in Figure 2, where the first principal component inside a two-dimension set is shown.

Fig. 1. Principal Component Analysis of a bidimensional data cloud. The line shows the direction of the first principal component.

5.2. Neuronal Networks

Artificial neural networks (ANN) are widely used on pattern classification tasks, showing good performance and high accuracy results. In general, an artificial neural network is represented by a set of nodes and connections (weights). The nodes are a simple representation of a natural neural network while the connections represent the data flow between the neurons. These weights are dynamically updated during the network's training. In this work, we will use the Feed Forward Time Delay Neural Network model; because this model has shown good results in voice [7] and non-voice [8] recognition tasks.

5.3. Feed-Forward Time Delay Neuronal Network

Input Delay refers to a delay in time, in other words, if we delay the input signal by one time unit and let the neural network receive both the original and the delayed signals, we have a simple time delay neural network. This neural network was developed to classify phonemes in 1987 by Weibel and Hanazawa [9].

The Feed-Forward Time Delay neural network doesn't fluctuate with changes; the features inside the sound signal can be detected no matter in which position they are in. The time delays let the neural network find a temporal relation directly in the input signal and in a more abstract representation of the hidden layer. It does this by using the same weights for each step in time [7].

5.4. Scaled conjugate gradient back-propagation

The neural network training can be done through a technique known as Backpropagation. The scaled conjugate methods are based on the general optimization strategy. We use scaled conjugate gradient back-propagation (SCGBP) to train the neural networks; because this algorithm shows a lineal convergence on most of the problems. It uses a mechanism to decide how far it will go on a specific direction, and avoids the time consumption on the linear search by a learning iteration, making it a fast second order algorithm [6].

6. System Implementation

In most of the towns and cities in the Mexican territory, the use of public transportation is very common. One car model, that is commonly used for this purpose, is the *Combi Van* from Volkswagen, models '82, '83, '86, '88, and '90.

These cars work almost every day from 6:00 a.m. to 21:00, and there are some hours that they carry up to 18 persons and travel several kilometers with this weight. Some car shops, in cities where these vehicles are commonly used as public transportation, specialize on maintaining and repairing only these kind of engines (Figure 3), giving the engine specialists, the ability to quickly diagnose the most common motor flaws, and directly work on the problem of these specific engines. For these reasons we choose this kind of engines. The conditions to select and record an engine were:

- Old models or engines previously repaired.
- Similar engine characteristics between engines.
- Several similar car models, or with the same engine.

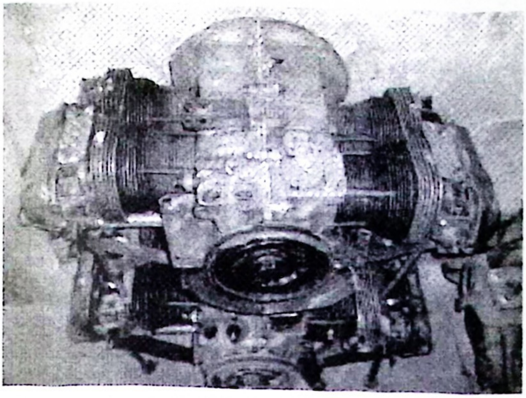

Figure. 2. Volkswagen *Combi* engine with 4 cylinders.

We recorded the engine sound in two phases; when the car first arrived to the shop and after the car was diagnosed and repaired. This method helped us label the recordings and compare the noise produced by the engine before and after the service. The two phases consisted where; accelerating the engine, and in *Ralenti* state (minimum speed). Also, some noises were recorded from tuned engines in both phases.

6.1. Noise Samples

To record the engine's noise, we visited car shops recommended by public transportation drivers. The recorded samples were obtained at five different car shops, three of them specialized only on Combi Vans. Each recorded sample was obtained with the help of a ZicPlay digital recorder, and the output files were saved in WAV format with the following settings; 8 Khz, 16 bits, and monaural. Each of these files has a time period of 3 to 5 seconds, and we recorded a total of 26 different engines. These samples were classified in four categories, tuned engines (Class 1), crank related problems (Class 2), piston related problems (Class 3), and valves related problems (Class 4), see Table 1. On Table 2 we can see the number of samples obtained from engines on Ralenti state.

Table 1. Recorded samples from accelerating engines, per class.

Identified Problem	Number of Samples	Time (in seconds)
Class 1	7	18.036
Class 2	6	18.502
Class 3	6	17.664
Class 4	7	18.006
Total	26	72.208

Table 2. Recorder samples from engines in Ralenti state, per class.

Identified Problem	Number of Samples	Time (in seconds)
Class 1	7	16.396
Class 2	6	15.430
Class 3	6	15.744
Class 4	7	16.590
Total	26	64.208

It has to be taken into account that these samples were difficult to obtain; in some cases some samples had to be rejected, due to the poor sound quality, minor flaws, and other issues that made the sample unusable for the system's goals. 3 different flaws were used, plus one from tuned/fixed motors, because according to the experts, these were the most common flaws. At first, five common flaws were considered, but two of them directly related to the three flaws used for this work, ending with only 3.

The sound recordings, obtained from engines in both, accelerating and Ralenti states, where segmented in 1 second samples and labeled from classes 1 thru 4; for each segmented sample we extracted a different LPCCs number (as shown on Tables 3 and 4). On one experiment, the vectors were used without any reduction, and in another experiment, the resulting vectors were reduced to 52% its original size by using PCA. The LPCC configurations and PCA reduction were obtained heuristically, and were the ones that gave the best overall results. The reduced vectors are then used to train/test the TDNN; we used 80% of the segmented samples to train the NN and 20% to test it. The network architecture consists on; an input layer with the number of nodes corresponding to the input vectors size, a hidden layer with 60% less nodes than the input layer, and an output layer corresponding to the number of classes to predict, in this case 4. The NN was implemented with the help of the Neural Network Tool-box, in Matlab v7.0, and the characteristic extraction was done with the help of the Auditory Tool-box for Matlab [10].

7. Experimental Results

On Table 3 we can see the results obtained by using the samples from engines in Ralenti state. Table 4 shows the results obtained from samples recorded from accelerating engines. Each result is the best overall result from 10 experiments. On both tables, the first column shows the number of characteristics extracted (c) per each time frame (ms).

Table 3. Results obtained with samples from engines in ralenti state.

Configuration	Precision (without PCA)	Precision (without PCA)
LPCC 36c 100ms	95.22%	94.37%
LPCC 26c 100ms	95.41%	91.62%
LPCC 36c 100ms	89.74%	78.55%

Table 4. Results obtained with samples from accelerating engines.

Configuration	Precision (without PCA)	Precision (without PCA)
LPCC 36c 100ms	95.55%	94.27%
LPCC 26c 100ms	93.01%	95.58%
LPCC 36c 100ms	86.47%	82.58%

The results on both tables show that the best precision was given when using 36 and 26 LPCCs for each 100ms time frame, when using the samples from accelerating engines; here we can say with 36 coefficients the training time was longer, and there is not so much difference on the recognition results, showing with this that 26 LPCCs give the best overall results. Taking this into account, we randomly selected 12 WAV files from the training samples, previously separated, and fed them to the trained NN, calculating with this the precision percentage given by each output node. The results obtained are shown in Table 5.

8. Conclusions and Future Work

The results show that a motor in bad conditions does in fact emit different sounds than a tuned motor, also that the sounds are different, or at least are recognized by the TDNN, depending on the type of flaw that's affecting the motor efficiency. This information can help the motor specialist give fast diagnosis, prices estimates for the reparation, and work directly on the motor's problem. We also concluded that is viable to work on this problem and build a cheap system that can be used by motor manufacturers and car shop owners.

We are working with other acoustic characteristic extraction such as MFCC and LPC, and noise reductions methods, by using wavelets filters, in order to compare the shown results with these methods, and make a trustworthy system. We are also collecting more sound samples from Combi Van motors and from other car brands, to build a system that is able to recognize other motors and other failures. For the final tests on Table 5 we worked on a user interface, designed in Matlab, that accepts WAV files as an input, automatically extracts the acoustic features and, with a trained NN previously loaded, emits the predicted failure. But, we want to design an independent user interface that can be used on car shops and predict engines failures, only by feeding directly the motor noise, by microphone, or pre recorded engine noise samples.

Table 5. Comparison between the predicted class and the real class.

Test	Class 1	Class 2	Class 3	Class 4	Real Class
1	94.5945	0.0	5.4054	0.0	yes
2	100.0	0.0	0.0	0.0	yes
3	89.1891	0.0	0.0	10.8108	yes
4	0.0	75.6756	21.6216	2.7027	yes
5	5.4054	94.5945	0.0	0.0	yes
6	2.7027	67.5675	8.1081	21.6216	yes
7	0.0	0.0	97.2972	2.7027	yes
8	0.0	5.4054	86.4864	8.1081	yes
9	0.0	0.0	2.7027	97.2972	no
10	0.0	0.0	0.0	100.0	yes
11	0.0	0.0	8.1081	91.8918	yes
12	0.0	0.0	8.0181	91.891892	yes

References

[1] Mario E. Munich. Bayesian Subspace Methods for Acoustic Signature Recognition of Vehicles. Published at the 12*th* European Signal Processing Conference (EUSIPCO-2004), Vienna, Austria. September 6-10, 2004.

[2] Huadong Wu, Mel Siegel, and Pradeep Khosla. Vehicle Sound Signature Recognition by Frequency Principle Component Analysis. The Proceedings of IMTC'1998, selected in the IEEE Transaction on Instrumentation and Measurement Vol. 48, No. 5, page 1005 - 1009, (ISSN 0018-9456), October 1999.

[3] Pedro N. Saavedra and Edgar A. Estupiñan. Análisis De Vibraciones Aplicado A Las Máquinas Rotatorias De Baja Velocidad. http://internal.dstm.com.ar/sites/mmnew/bib/notas/Estupi1.pdf. Chile 2000.

[4] Pedro Nelson Saavedra. La Medición y Análisis de las Vibraciones como Técnica de Inspección de Equipos Y Componentes, Aplicaciones, Normativas Y Certificación. Facultad de Ingeniería - Universidad de Concepción, Casilla 160 - Concepción -Chile. http://www.aaende.org.ar/sitio/biblioteca/material/CONFCHILE.pdf. 2001.

[5] Agustín Álvarez Marquina. Fundamentos del Reconocimiento Automático de la Voz, Algoritmos de Extracción de Características. Facultad de Informática. Universidad Politécnica de Madrid (UPM). November 22, 2004.

[6] Orozco-Garca, Jos. Extracción y Análisis de Características Acústicas del Llanto de Bebés para su Reconocimiento Automático Basado en Redes Neuronales. Tesis. INAOE, México. February 2004.

[7] Hilera J., Martínez V. Redes Neuronales Artificiales Fundamentos, modelos y aplicaciones. Alfaomega, Madrid, España. (2000).

[8] Reyes-Galaviz O. F., Reyes-García, C.A., Infant Cry Classification to Identify Hypoacoustics and Asphyxia with Neural Networks, in LNCS 2972: Advances in Artificial Intelligence. Springer, Berlin, 2004, pp 69-78, ISBN 3-540-21459-3, ISSN 0302-9743.

[9] A. Weibel, T. Hanazawa, G. Hinton, K. Shikano, and K.J. Lang, "Phoneme Recognition Using Time Delay Neural Networks," IEEE Trans. Acoustics, Speech, Signal Proc., ASSP-37: 32'339, (1989).

[10] Auditory Toolbox. Malcolm Slaney, Interval Research Corporation http://rvl4.ecn.purdue.edu/ malcolm/interval/1998-010/AuditoryToolbox.zip