# Analysis and Classification of Car Engine Sounds to Diagnose Failures

Luis Andrés Méndez-Ortega[1] and Orion Fausto Reyes-Galaviz[2]

Universidad Autónoma de Tlaxcala
Facultad de Ciencias Básicas, Ingeniería y Tecnología
Apizaco. Tlaxcala. México
[1]siul222@hotmail.com,
[2]orionfrg@ingenieria.uatx.mx
http://ingenieria.uatx.mx/

**Abstract.** The detection of a possible car engine failure 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 capable of recognizing a flaw in an 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 being produced by the same. Nevertheless, it's hard to guess what kind of malfunctions produce certain motors noises and vibrations, making it hard to work directly on the engine's flaw. In this work, an interface is designed to classify, and directly detect, four common car engine failures, just by using the engine's recorded sound. To accomplish this goal, we use Linear Prediction Cepstral Coefficients (LPCC) and Mel Frequency Cepstral Coefficients (MFCC) as the main acoustic characteristics extractors, and a Time Delay Neural Network for the pattern recognition. Some results are shown of up to 95.55% when predicting common car failures.

**Key words:** Engine noise, LPCC, MFCC, Time Delay Neural Networks, Expert System

## 1 Introduction

A failure on any engine can produce mild or violent vibrations that can damage the motor, if left unattended. 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 sound/vibration corresponds to a possible malfunction. Most of the sounds that an engine produces are generated by

© A. Gelbukh, Á. Kuri (Eds.)
*Advances in Artificial Intelligence and Applications*
*Research in Computer Science 32, 2007, pp. 131–140*

*Received 13/06/07*
*Accepted 31/08/07*
*Final version 28/09/07*

rotating mechanisms inside the motor; also, the machines that have an internal combustion system, produce more noise due to the explosions created inside.

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 by vibrating gears, pounding 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 could help the motor specialists obtain robust diagnostics when repairing car engines. Taking this into account, an interface is designed to capture and analyze the acoustic sound produced by these kind of machinery, some acoustic characteristics extraction techniques are used, 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) and Mel Frequency Cepstral Coefficients (MFCCs) methods. A fundamental theory on pattern classification, and Time Delay Feed-Forward Neural Networks, is given in Section 5. The complete system implementation, including the user interface, can be found on Section 6. Our experimental results, that go up to 95.58%, are shown on section 7, and the concluding thoughts are presented on 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), te results on this work seem promising. In [3], Edgar A. Estipiñán, Et Al use an analysis of mechanical

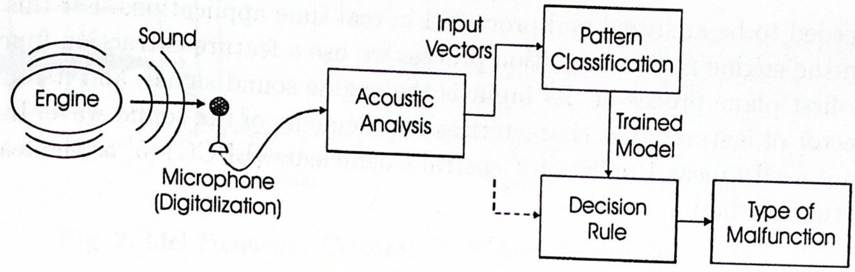


**Fig. 1.** Engine Noise Recognition Process

vibrations, as a part of their predictive maintenance, to establish the mechanical health of their machines, preventing future flaws with this technique. 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 fractomecanical 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 Fig. 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 connectionists models. We focus on the use of the last one.

## 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 $x(n-1), x(n-2), ..., x(n-k)$:

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

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

$$\begin{aligned} E &= \sum_{n=0}^{N-1} e^2(n) \\ &= \sum_{n=0}^{N-1} \left( x(n) - \sum_{i=1}^{K} a_i x(n-1) \right)^2 \quad 0 \le n \le N-1 \end{aligned} \tag{2}$$

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

$$c(0) = \log(1) = 0 \tag{3}$$

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

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:

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

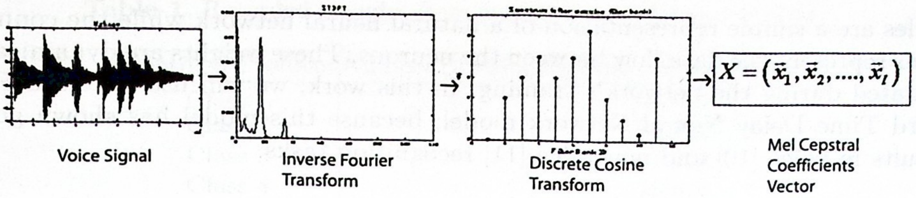


**Fig. 2.** Mel Frequency Cepstral Coefficients extraction method.

### 4.2 Mel Frequency Cepstral Coefficients

The first step of the sound processing, once we have the noise samples, is to obtain (from these samples) the spectral characteristics. This step is necessary because the important information of the sample is codified in the frequency domain, and the engine noise samples are recorded by means of electronic devices in the time domain. When the time domain is converted to the frequency domain we obtain the parameters which indicate the occurrence of each frequency.

There is a wide variety of ways to represent the noise samples in their parametric form. One of the most commonly used on speaker recognition tasks are MFCCs. The human ear decomposes the received sound signals in its fundamental frequencies. Located in the inner ear we find the cochlea which has a conic spiral form. This is one of the three cavities that form the physical structure of the ear [6]. This cochlea filters the frequencies in a natural way. The sound waves are introduced inside this structure bouncing on its walls and getting inside the spiral with low or high frequency, taking into account each frequency's wave length [7]. MFCCs are based on the frequency response the human ear perceives. This method behaves as a filter bank linearly distributed in low frequencies and with logarithmic spacing on the higher frequencies. This is called the Mel Frequency Scale, which is linear below 1000 Hz, and logarithmic above 1000 Hz (Fig. 2) [8].

## 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 artificial neural networks (ANN), to classify these vectors (patterns).

### 5.1 Neural 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 [10] and non-voice [11] recognition tasks.

### 5.2 Feed-Forward Time Delay Neural 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 [12].

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 [10].

### 5.3 Scaled conjugate gradient back-propagation

The neural network training can be done through a technique known as back-propagation. 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 [9].

## 6 System Implementation

In most towns and cities in the Mexican territory, the use of public transportation is very common. A 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 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 used as public transportation, specialize on maintaining and repairing only these kind of engines, giving the engine specialists, the ability to quickly diagnose most of the 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.

**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 |

We recorded the engine's 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 (no acceleration). Also, some noises were recorded from tuned engines in both phases.

## 6.1 Noise Samples

To record the engine noise, we visited car shops recommended by the 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 audio recorder, and the output files were saved in WAV format with the following settings; 8 Khz, 16 bits, and monaural channel. Each of these files has a time period of 3 to 5 seconds,and a total of 26 different engines were recorded. 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.

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, noise contamination, 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, four common flaws were considered, but two of them directly related to a single common flaw, leaving us with only four different classes.

**Table 2.** Recorded 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 |

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 different configurations of LPCC and MFCC (as shown on Tables 3 and 4). The LPCC and MFCC configurations were obtained heuristically, and were the ones that gave the best overall results. 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 7.0, and the characteristic extraction was done with the help of the Auditory Tool-box for Matlab [13].

# 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 | Configuration | Precision |
|---|---|---|---|
| LPCC 38c 50ms | 80.77% | MFCC 38c 50ms | 80.00% |
| LPCC 36c 100ms | 95.41% | MFCC 36c 100ms | 85.60% |
| LPCC 26c 100ms | 95.22% | MFCC 26c 100ms | 82.97% |
| LPCC 20c 100ms | 89.74% | MFCC 20c 100ms | 75.79% |

The results on both tables show that the best precision was given when using 36 LPCCs for each 100ms time frame. 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.

**Table 4.** Results obtained with samples from accelerating engines

| Configuration | Precision | Configuration | Precision |
|---|---|---|---|
| LPCC 36c 100ms | 95.55% | MFCC 36c 100ms | 84.85% |
| LPCC 26c 100ms | 93.01% | MFCC 26c 100ms | 88.35% |
| LPCC 20c 100ms | 86.48% | MFCC 20c 100ms | 83.81% |

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

| Test | Class 1 | Class 2 | Class 3 | Class 4 | Real Class |
|---|---|---|---|---|---|
| 1 | 94.59 | 0.0 | 5.40 | 0.0 | Yes |
| 2 | 100.0 | 0.0 | 0.0 | 0.0 | Yes |
| 3 | 89.18 | 0.0 | 0.0 | 10.81 | Yes |
| 4 | 0.0 | 75.67 | 21.62 | 2.70 | Yes |
| 5 | 5.40 | 94.59 | 0.0 | 0.0 | Yes |
| 6 | 2.70 | 67.56 | 8.10 | 21.62 | Yes |
| 7 | 0.0 | 0.0 | 97.29 | 2.70 | Yes |
| 8 | 0.0 | 5.40 | 86.48 | 8.10 | Yes |
| 9 | 0.0 | 0.0 | 2.70 | 97.29 | No (C3) |
| 10 | 0.0 | 0.0 | 0.0 | 100.0 | Yes |
| 11 | 0.0 | 0.0 | 8.10 | 91.89 | Yes |
| 12 | 0.0 | 0.0 | 8.10 | 91.89 | Yes |

## 8 Conclusions and Future Work

The results show that an engine in bad conditions does in fact emit different sounds than a tuned motor, also that they can be recognized by the TDNN, depending on the type of flaw that's affecting the motor efficiency. This information can help a motor specialist give fast diagnosis, prices estimates for the reparation, and directly work 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 extractions such as Wavelets and LPC, to compare the shown results with these analysis, and make a trustworthy system. We are also collecting more sound samples mainly from Combi Van, Chevy motors, and from other car brands, to build a system that is able to recognize other engines and other failures. We also want to design a Matlab-independent user-interface that can be used on car shops, and predict engine failures, only by feeding directly the motor noise, or a pre-recorded sample.

## 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. Analisis De Vibraciones Aplicado A Las Maquinas Rotatorias De Baja Velocidad. 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 Caracerísticas. Facultad de Informática. Universidad Politécnica de Madrid (UPM). November 22, 2004.
6. A. J. Balagué, Diccionario Enciclopédico Baber, Editorial Baber, 1119 E. Broadway Street Glendale, California 91205. (1991).
7. Bernal J., Bobadilla J., Gómez P., *Reconocimiento de voz y fonética acústica*, Ed. Alfaomega Ra-Ma , Madrid, España. (2000).
8. Miramontes de León G., De la Rosa Vargas J. I. and García E., 'Application of an Annular/Sphere Search Algorithm for Speaker Recognition', Proceedings of the 15th International Conference on Electronics, Communications and Computers, 0-7695-2283-1/05 IEEE. CONIELECOMP, (2005).
9. Orozco-Garca, Jos. Extraccin y Anlisis de Caractersticas Acsticas del Llanto de Bebés para su Reconocimiento Automático Basado en Redes Neuronales. Tesis. INAOE, México. February 2004.
10. Hilera J., Martínez V. Redes Neuronales Artificiales Fundamentos, modelos y aplicaciones. Alfaomega, Madrid, España. (2000).
11. Reyes-Galaviz O. F., Reyes-Garca, 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.
12. 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).
13. Auditory Toolbox. Malcolm Slaney, Interval Research Corporation http://rvl4.ecn.purdue.edu/ malcolm/interval/1998-010/AuditoryToolbox.zip