# Automated Vehicle Detection and Classification using Acoustic and Seismic Signals

Naoko Evans

Department of Electronics,

University of York,

Submitted for the degree of Doctor of Philosophy

September, 2010

# Abstract

Security threats to important infrastructure cause problems to not only those who live nearby but also in a much wider sense. It is therefore desirable to consider the use of automated systems capable of detection and identification of potential threats. This thesis describes an investigation into acoustic and seismic methods for achieving such a system specifically for commercial road vehicles.

Accurate algorithms have been developed for recognition of moving vehicles using fusion of acoustic and seismic signals. It has been found that seismic signals are less susceptible to interfering signals, making them optimal for detection of vehicles. Their much narrower bandwidth also increases processing efficiency and speed. Thus, the algorithm developed utilises firstly only seismic signals to detect vehicle presence, and then employs both acoustic and seismic signals for classifying type of the vehicle.

The detection algorithm is purely time domain and uses seismic Log Energy together with a modification of Time Domain Signal Coding. The best detection accuracy obtained was 97.71 % with Support Vector Machine and 99.02 % with Learning Vector Quantisation Neural Networks. The classification algorithm to distinguish between trucks and cars utilises three relatively simple time domain methods: Zero-Crossing Rate, Log Energy and Autocorrelation of seismic signals; combined with LPC coefficients collected from acoustic signals. Classification with either SVM or LVQ reached 93.30 % or 80.80 % respectively. This study therefore has demonstrated it is possible to detect an approaching vehicle and classify its type by using acoustic and seismic signal processing.

# Contents

# List of Figures

# List of Tables

# Acknowledgements

I would like to thank first of all to both BAe Systems and the University of York for their funding towards my three-year full-time PhD research. I also would like to thank SENSOR in Netherlands for their generosity, providing us a few geophones utilised during experiment. Secondly I would like to express my appreciation to everyone in the Department of Electronics for their roles in ensuring our less stressfully graduate student lives as well as in offering good working friendship. I would like to thank my thesis advisory panel, Professor Howard for his patience, guidance and inspiration during my PhD study. Above all, I would like to thank my supervisor Dr. Chesmore firstly for giving me this research opportunity, and secondly for his patience, guidance and inspiration throughout the project. Finally, my special thanks goes to all of my family and close friends in the UK and Japan for all of their constant and genuine support.

# Chapter 1

# Introduction

## 1.1 Background and Motivation

Human beings are equipped with inborn capabilities to detect objects that are surrounding, or approaching them, perhaps originally to defend their own safety as well as to acquire food for survival [Yost, 2000]. Since the first recognised invention of motor vehicles around 1800 AD or possibly even earlier, the ability to detect vehicles approaching individuals has relied on human perception, mainly so as to protect their lives. A person's capacity to recognise types of approaching vehicle helps her/him react appropriately, but it depends on more than simple innate capabilities, including experience and learned competence.

Development of machines that are able to imitate human activities has long been pursued. Sometimes it is in order to improve the actual quality of performance itself and/or overall cost effectiveness, on other occasions to free human beings from difficult roles i.e. time consuming, labour intensive work, or tasks potentially linking to high risks on health and safety of personnel [Munich, 2004]. Automated vehicle detection and recognition is one example, and a system that assists the task of vehicle detection and recognition is being considered here; not only for protection of lives but also as a tool to improve infrastructure security. A number of vehicle recognition systems have been published and these will be discussed in Chapter 2. Although monitoring passing vehicles is important for applications such as traffic control or road planning [Nooralahiyan et al., 1997; Toth et al., 2003; Meyer et al., 2006] as well as in recent development of driver assistant systems [Bertozzi and Broggi, 1998; Sun et al., 2006b; Tsai et al., 2007]; needless to say, monitoring unauthorised intruding vehicles for security has been one of the most critical issues for society, governments and industries [Choe et al., 1996; Altmann et al., 2002; Xiao et al., 2006].

Examples of these emerging issues are found particularly in and around the energy industry. A breach of security surrounding pipelines transferring oil or gas could cause serious damage to both the industry and people who live close to the pipelines as well as the livelihood of the consumers. Parfomak [Parfomak, 2004] reported that at the time of his report publication in 2004, there were approximately 756,000 kilometres of gas and oil pipelines in the United States of America. It was pointed out that these pipelines potentially could become a target of terrorist attacks, urging one to consider the possible damage and impact that may be caused by such incidents. Also theft of assets from pipelines is another issue concerning the industry to act against in order to protect both the valuable resources and the residents of the neighbourhood. The media [Unknown, 2006a,b] have reported there have been a significant number of deaths apparently triggered as a result of fuel thieves working on the petrol pipelines. Of course, the area where such concerns apply is not limited only to the United States of America, but also extended widely across the world as there are many important infrastructures, including pipelines [Stafsudd et al., 2008]. Consequently, developing a system that would enhance the security for such essential infrastructures by employing automated moving vehicle recognition is of interest to both the public and the industry [Roper, 2005].

## 1.2 Sensor Technology

As in the case of detecting approaching vehicles by humans, effectively gathering available information of the surroundings, i.e. input signals, is the first significant and crucial task to be completed in recognition processes. According to Arora et al. [Arora et al., 2004], intrusion by a moving vehicle causes certain disturbance in the surrounding environment; namely thermal, seismic, acoustic, electrical, magnetic, chemical, and optical. Therefore, in place of eyes and ears for human perception, a variety of sensing techniques have been adopted in automated vehicle detection to capture such disruption caused by the target vehicles, hence to detect the intrusion.

### 1.2.1 Active Sensing

Generally it is understood that an active sensor needs an external circuit providing a "source of excitation" [Wilson, 2005, p.16], e.g. current or voltage, in order to produce its measurement output whereas a passive sensor does not. The majority of the following four sensing techniques are often seen as "active" sensing although it can actually depend on how the measurement is performed.

#### (a) Radar

Radar has the capacity to monitor across a wide area even in dark conditions [Koch et al., 2006; Suchandt et al., 2006], however it can be expensive to install and perform surveillance compared to others, such as video camera monitoring [Lin et al., 2008].

#### (b) Inductive Loops

Inductive loops have been employed in vehicle detection for traffic monitoring due to their low installation costs for small scale monitoring [Nooralahiyan et al., 1997]. Nevertheless, their installation and maintenance can be expensive for a large scale monitoring [Ding et al., 2004] and more complex than others, for example, laser detection [Cheng et al., 2005]. Another issue with inductive loop sensing for vehicle detection, especially reported in traffic management applications, is that interruption of the traffic is inevitable during installation and maintenance works as the sensors are buried within the ground surface [Ding et al., 2004; Ploetner, 2006].

#### (c) Magnetic Sensing

Magnetic Sensing has a long history especially as a navigation tool. Monitoring earth's magnetic field variation allows detection of vehicle existence as well as estimation of vehicle size and speed [Caruso and Withanawasam, 1999]. However, the susceptibility of magnetic sensing to noise and temperature change can lead to undesirable limitations in applications [Mapps, 1997].

#### (d) Laser

Laser sensing has been employed for traffic monitoring and other applications in the transportation industry as well as for pipeline risk management. Advantages are the capabilities of providing relatively high density and accurate data of a wide area without having to place sensors on the ground, hence no interruption to the objects' movements [Toth et al., 2003]. Moreover, real-time laser data processing can be less computationally intensive compared to inductive loop operations [Cheng et al., 2005]. Nevertheless, installation and maintenance costs are still the main drawbacks of laser sensing [Nooralahiyan et al., 1997] in addition to the disadvantages of being active sensors, which will be recognised once the advantages of passive sensing are covered in the following section.

### 1.2.2 Passive Sensing

The main advantages of using passive sensors in comparison to active sensing technologies can be realised by observing how data collection activities are carried out. Firstly, the data can be collected without interrupting the movements of the target objects, hence useful for practical applications such as traffic monitoring [Nooralahiyan et al., 1997; Sun et al., 2006b]. Secondly, not only data collection can be done without revealing the existence of sensors to strangers [Becker and Gudesen, 2000] but also concealing acoustic signals can be hard once emitted from an intruder's point of view [Srour, 2001], thus suitable for military applications. In addition to the above two, passive sensing is commonly less expensive to install, operate, and maintain than active sensing [Srour, 2001; Lan et al., 2005; Sun et al., 2006a]. Therefore, this section describes the majority of popular passive sensors that have been applied to vehicle detection research.

### (a) Image Processing

Several types of passive image sensing technologies have been employed for target detection, of which three are mentioned below. One of the obvious advantages of these image capturing sensors over other passive sensing is that they support detection of targets that are either still or undertaking only few activities for a long period of time [Hinz and Stilla, 2006]. On the other hand, one of the main drawbacks of any image processing is that obstacles present between the source and sensors, for example tall vegetation or tall buildings or terrain conditions can reduce successful detection results [Nooralahiyan et al., 1997].

#### Image Processing: Video Camera

Nowadays using video cameras to collect data for recognition studies can be achieved at a lower cost compared to other image sensing techniques [Nooralahiyan et al., 1997]. On the other hand, the use of video camera images is not suitable under certain monitoring conditions, for example during the night or foggy conditions [Koch et al., 2006; Cheng et al., 2005]; which may imply the unsuitability of them for the current research.

#### Image Processing: Satellite

Satellite image detection can be suitable when surveillance of a wide area is required although the installation and maintenance costs may be high [Roper, 2005]. The other disadvantages include the low ground resolution and the inability to monitor on a cloudy day.

#### Image Processing: Thermal Infrared

Unlike the other two image processing sensors described so far, thermal infrared image processing allows detection in dark conditions. Particularly aerial infrared image sensing is equipped with an ability to monitor over a wide range. However, treatment of noise can be a very challenging task in some applications [Hinz and Stilla, 2006].

### (b) Acoustic Signal Sensing: Microphone

Acoustic signals can be measured by microphones. The three main microphone types are: dynamic, piezoelectric, and condenser (also called capacitive) microphones. Technical details of these types can also be found in [Wilson, 2005]. The frequency response, directional response, and sensitivity as well as the cost and durability are key features that should be considered carefully so as to select a suitable microphone for a given application.

### (c) Seismic Signal Sensing: Accelerometer /Geophone

An accelerometer measures vibration and shock. Hence, it is often used in geophones to measure seismic signals generated by moving vehicles. The most popular accelerometers are piezoelectric that are robust and able to provide a flat output response over a relatively wide frequency range. Piezoelectric accelerometers are usually made of quartz (natural crystal) or polycrystalline ceramics (man-made ceramics). Other accelerometers may use piezoresistive transducer, capacitive sensor, or servo (or force balance) technology. Details of these accelerometer technologies as well as a self explanatory comparison table of various accelerometer types can be found in [Wilson, 2005].

### 1.2.3 Acoustic and Seismic Sensor Fusion

As described in the previous section, there are a range of sensing techniques that can be selected for acquiring valuable data for the current research. Any of these sensors can be beneficial to the study; however, each one can also bring potentially difficult but unique challenges for algorithm development. Choice of such sensors can have an impact on the sort of challenges one would face during algorithm development and hence requires a good consideration of various factors including the resource and time availability for the task [Arora et al., 2004]. Therefore for the current research, it was firstly decided to use acoustic sensing for the following reasons.

One of the biggest advantages of using acoustic sensors for detection is that they are passive. Also it is possible to constantly gather acoustic data of an area even when visibility is low, for instance due to darkness, fog or when there are obstacles between an intruding vehicle and a sensor. Furthermore, development and operation costs of acoustic systems can be much lower than others, such as satellite or radar surveillance systems [Nooralahiyan et al., 1997; Eom, 1999; Becker and Gudesen, 2000; Srour, 2001; Wu and Mendel, 2004]. In addition, the power consumption of acoustic sensors are generally low [Viangteeravat et al., 2007a]. Due to these benefits, acoustic sensors, which were once in regular use in military applications until the start of the second world war [Dyrdal et al., 2003], are now being revisited for security enhancement.

Perhaps more accurate observation of unauthorised intruding vehicles can be achieved when utilising multiple methods together rather than relying on one type of sensor

[Jacyna et al., 2005]. Therefore, multiple sensor fusion as indicated in Figure 1.1, may be an ideal near future goal for vehicle detection. On the other hand, developing vehicle recognition algorithms using one type of sensors is already a demanding task; thus utilising various sensors simultaneously will require a huge amount of resources such as time, skills and costs [Hsieh et al., 2006] for development and operation. It is possible that even using two kinds of sensors would make the task over complicated for practical use, compared with the case of using acoustic sensors only. Sometimes adopting another sensor can result in compromising the valuable advantages of acoustic sensors; as it can be seen in examples such as, a combination of acoustic and video camera image sensing [Chellappa et al., 2004], or acoustic and magnetic sensing [Ding et al., 2004]. Nevertheless, it may not be the case for adding seismic sensing because of its similarities to acoustic sensing in several aspects such as the relatively low installation and maintenance costs with low power consumption as well as less computational demands. When combined with other sensors, seismic sensing has shown good potential to improve classification performance [Ding et al., 2004]. Combining acoustic and seismic sensing for vehicle detection was also recommended after a primary study using sound "generated by striking a heavy metal plate with a hammer" [Stafsudd et al., 2008, p.85]. Therefore, for the current research, acoustic and seismic sensor fusion was chosen as the data gathering method.

## 1.3 Hypothesis

The hypothesis of this research is;

*"It is possible to detect an unauthorised moving vehicle and classify its type with an automated vehicle recognition system using acoustic and seismic signal processing, so as to enhance protection of the infrastructure."*

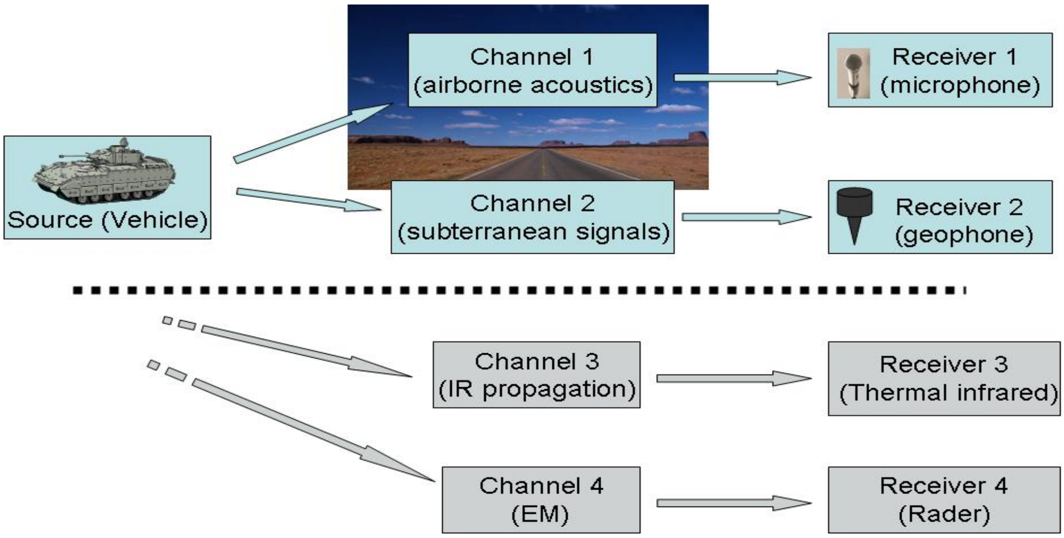


Figure 1.1: Sensor Fusion
(Please note; only the top half is within the scope of this thesis. Pictures are obtained from Microsoft Office 2003 Clip Art)

## 1.4 Contribution to the Field

Based on the literature review presented in the next chapter (Chapter 2), the contribution made by this thesis can be summarised as follows:

(a) Use of seismic signals for road vehicle detection and the development of a novel algorithm using time domain methods.

(b) Development of a new road vehicle classification algorithm using a combination of acoustic and seismic signals.

(c) Collation of a comprehensive literature survey in relation to automated vehicle recognition studies.

## 1.5 Thesis Structure

It is not straightforward to obtain a constructive understanding of various pattern recognition algorithms in a uniform way because researchers often take arbitrary approaches in analysing the same algorithm. However, there have been some suggestions for symbolic representation of standard pattern recognition system structure to help one understand general pattern recognition algorithms [Looney, 1997; Nooralahiyan et al., 1997]. For example, the most common way of depicting such a structure consists of three main stages; "signal pre-processing", "feature extraction" and "classification" [Duda et al., 2001]. There are other variations; for example processing of classification output into a required format can be attached at the end of this model as "post-processing" in some cases, and one may include feature extraction in pre-processing [Bishop, 1995]. As is often the case, these terms used to describe the procedures concerning pattern recognition systems have been employed interchangeably. The structure and particularly the connection between each element may not always be clear nor linear. The variation possibly depends on what strategies one would select for carrying out the development task, but also for presentation of the outcome. Indeed, the latter may have stronger influence than the former, on how such systems are treated in the literature. As a result, it can lead to some misunderstanding during and after the development process if there is no clear indication of what each procedure involves.

The pattern recognition system structure model for this particular study is depicted in Figure 1.2, together with the chapter number of which the explanation of the corresponding part can be found in this thesis.

The first stage, named here as "pre-processing", is composed of "data acquisition" and "noise reduction". Also where applicable, recovery of missing data and either "down sampling" or "up sampling" can be included here. At the data acquisition phase, data of the target source are collected by suitably placed sensors and then fed into the system to be sampled and quantised for conversion to digital signals. Secondly, unwanted signals in the gathered data are removed while signals of interest may be boosted. Next, the treated signals are processed further to separate or extract selected features of signals that represent some distinctive source attributes that

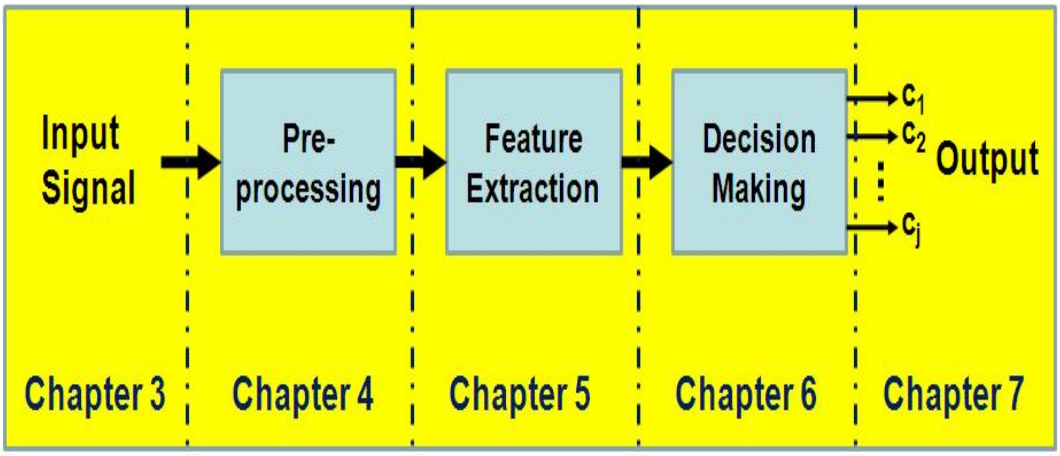


Figure 1.2: General Structure of the Recognition System with Chapter Numbers

may not be obvious at a first analysis, so that accuracy and speed of signal discrimination between different classes can be improved. Once collected, these feature vectors can be normalised particularly if it increases the overall performance. Finally, the category of input signal source is estimated through a decision making process, by using either clustering or classification algorithms.

This thesis documents elements of the development work of the optimum algorithm for automated moving vehicle recognition using acoustic and seismic signal processing as well as machine learning, together with analysis and discussion of the findings. Chapter 2 presents a summary of the literature relevant to this work. Then, Chapter 3 briefly pays attention to the mechanism of how target signals for the study are generated and propagated to the sensors capturing the input data. Next, the three following chapters, 4, 5 and 6 respectively discuss the findings regarding the methods investigated during the project so as to discover more effective signal pre-processing, feature extraction, and decision making algorithms. Each described method is selected based on their relatively high potential illustrated in the literature in relation to the current research. Therefore, theoretical explanation and outcomes of analytical study carried out by software with collected real-life data are included. Chapter 7 firstly discusses the output of various sets of algorithms, secondly introduces the algorithm considered optimum, based on the research performed using a larger set of real data. It also describes the applied evaluation methods and results acquired by using different data sets over the duration of the project, followed by some discussion. Finally the thesis finishes with some concluding remarks on the research, together with suggestions for future development.

# Chapter 2

# Literature Review

A literature survey has been conducted in order to understand the current and past research trends in the area of automated moving vehicle recognition systems, which use acoustic and/or seismic signal processing. Depending on the category of target vehicles, these can be divided into two groups: military vehicle recognition and road vehicle recognition. Research on road vehicle recognition has been so far mainly for developing traffic management or intelligent transport systems including automated driver assistance systems [Nooralahiyan et al., 1997, 1998]. The outcomes of such studies are often published and made widely available in the public domain.

On the other hand, currently there are only few peer-reviewed journal publications aiming specifically for security applications; regardless of the target vehicle groups, i.e. military or road vehicles. However, it is estimated there are quite a few more research findings that have not been made publicly accessible, perhaps because of some constraints in place to reserve sensitive information.

Therefore, this literature survey may only represent some portion, rather than all, of what actually has already been investigated in the related area. Nevertheless, the following should be able to help one grasp the past and current trends as well as the transition, in the research community of automated moving vehicle recognition system development.

## 2.1 Acoustic Vehicle Recognition

In the literature, a particular interest in development of automated moving vehicle detection and recognition systems using acoustic and/or seismic signals started to appear in the 1970s [Thomas and Wilkins, 1970]. Nonetheless, research papers written on development of more or less self-contained algorithms for such tasks, rather than those focusing on only small parts of automated recognition, can only be identified in those published in or after the late 1990s.

**[Thomas and Wilkins, 1970, 1972]**

The first published research on acoustic signals generated by a moving vehicle for recognition purposes was that by Thomas and Wilkins, in which they investigated the feasibility of several feature extraction algorithms. The first of the two papers briefly explained the preliminary results obtained by analysing acoustic signals generated by a moving vehicle. It was argued that acoustic signals of a vehicle travelling at an almost constant speed contain attributes similar to those of human speech signals. This was their reason for choosing autocorrelation and short-term power spectrum, at first, as the methods to extract features from the input signals; allegedly by following the trend seen in speech recognition research at the time. Nonetheless, the outcomes were not satisfactory because the algorithm failed to highlight any distinctive characteristics of different vehicle classes. Instead, they later found that cepstrum analysis, which was reported to be unsuccessful in speech signal processing according to them, produced better results than those mentioned above in displaying the firing rate of the engine to detect the existence of a moving vehicle.

Particularly in relation to the latter paper, which revealed more detailed information about their research, the study by using collected acoustic signals was motivated by the following two objectives; firstly to identify a particular engine contained in various bodyworks as well as running within different environments, and secondly to classify vehicle types. The experimental data were recorded from moving industrial vehicles carrying three types of engines. They chose certain engine types for the investigation based on subjective claims made by some personnel working with industrial vehicles that they were able to distinguish these engines better than others by using human perception only.

### 2.1.1 Research on Military Vehicle Recognition

Early publications regarding automated acoustic vehicle recognition algorithms were focused mainly on military vehicle signals, in order to develop a system that improves surveillance for security.

#### (a) [Choe et al., 1996]

The first notable self-contained automated moving vehicle classification algorithm was developed by Choe and his team, which was published as an invited paper in a conference proceeding. The algorithm only examined acoustic signals of two specific military vehicles. Human perception was utilised to prepare reference data for squeak, engine and exhaust sounds. They applied both distance based and statistical correlation based pattern matching to the feature vectors collected by the Discrete Wavelet Transform (DWT) multiresolution analysis, with Haar and Daubechies-4 (D-4) wavelets. Classification of the signals using this algorithm was performed by finding the best match for the acquired feature vectors from the reference data of those two classes stored in advance in a database. Although the precise detail of the decision making algorithm was not stated in the paper, the final decision was made reportedly as a global decision, accompanying some statistical data indicating the confidence level, which was determined according to the local decision outcomes obtained by each of the two classifiers.

This method did not require training, nonetheless, the preparation of the reference signal database was crucial. This aspect was claimed to be beneficial for military applications since collecting training data set could be a challenging task. For cases of which gathering substantial amount of training data is possible, as a result, the use of Neural Network (NN), in particular for classifiers in non-military applications, was recommended as an alternative decision making method. On the other hand, the necessity of both creating the reference database through human perception and manually analysing acoustic signals of each vehicle type were the drawbacks of their algorithm. It also meant there was no automated vehicle detection algorithm included. Therefore, whilst the study gave some ideas of potential classification algorithms, it was realised that inclusion of an automated vehicle detection algorithm would be necessary for the current study.

#### (b) [Braunling et al., 1997]

It was reported that an adaptive beamforming algorithm was developed for a study of acoustic moving vehicle detection, tracking and classification. They presented some real-time tracking results obtained with a beamforming algorithm that required 8 microphones to collect input signals. It was said to have a relatively wide detection range, although greatly changeable, of up to several kilometres for military ground vehicles. They also showed some classification results; however, no details of the feature extraction nor classification algorithms were described in the paper.

#### (c) [Maciejewski et al., 1997]

By following the study and the associated recommendation by Choe and his team [Choe et al., 1996], the combination of wavelet based feature vector and NN was examined and presented at a conference by Maciejewski and the team. The samples used for this study were acoustic signals generated by four military vehicles, running on several roads with different surface conditions. A feedforward NN with one hidden layer and a Probabilistic Neural Network (PNN) using Radial Basic Function (RBF) were chosen for a comparative study. They introduced a method to reject samples using a range of pre-defined thresholds. The correlation between the classification results and the rejection rates was illustrated. For the both classifiers, the best accuracy obtained in the classification experiments achieved higher than 70%. It was then concluded that the PNN classifier demonstrated slightly better results in terms of the accuracy and learning speed. It was also stated that such methods do not need thorough knowledge regarding “which part of object is responsible for the components of signal” [Maciejewski et al., 1997, p.292].

#### (d) [Jarnicki et al., 1998]

A team of researchers at the same institution as the previous paper, including Maciejewski, published another conference paper on a similar study in the following year. This time, input signals were processed to collect the power spectrum of autocorrelation function and a set of filterbank outputs of Fast Fourier Transform (FFT) as the feature vectors. Although the explanation was not detailed enough, the classification was performed by a distance and adjustable threshold based pattern matching algorithm. The experimental studies were carried out by using acoustic signals of military tanks, military transporters, and “civilian” cars; all running on roads with various surface conditions. The size of samples for the experiment was not declared. The achieved classification accuracies were in the range between 40% and 70% for 3-class. However, once the study was rearranged to carry out 2-class classification between “military” and “civilian” classes; above 84% was observed at the best. Once again they claimed that the knowledge regaridng precise signal origin was unnecessary for the method. It was also stated that the algorithm could be implemented as real-time application although no evidence to support such an argument was included in the paper.

#### (e) [Succi et al., 1999]

Succi and his colleagues published a paper in a conference proceeding. Acoustic signals of moving military vehicles were classified by two kinds of extracted features; the spectral peaks using Short Time Fourier Transform (STFT) above adaptively defined thresholds, and also the Direction of Arrival (DoA) based on phase comparison. Classification was performed by using two sets of multi-layer NN of which each single network was configured to produce one output, corresponding to a class. The first set of NN was to classify signals into 6 groups by estimating the number of cylinders in engines. Its outcome was reported to have achieved 95% accuracy. The second set was for identifying vehicle signals generated by 18 different vehicles; and it consequently accomplished 88%.

The collected samples for their study were recordings of acoustic signals of 18 different vehicles having 6 different engine types in total, of which about half of the samples were used as test samples during the experiments. Each two to three minute long sample was segmented to a set of one second long frames, and each frame of a set was processed for the above two kinds of classification while the accumulative results for each class were obtained. The final decision about each sample set was made by finding the class with the maximum output value accumulated over all processed frames of each two to three minutes long sample set. The training was carried out by Backpropagation algorithm with so-called “early stopping”, which was designed to avoid overfitting.

It was stated that the capability to identify the engine type itself could be valuable even when identification of the vehicle failed, for example because the target vehicle was not included in the training data. Especially for a surveillance system for security, the author agrees with them on that any incomplete information gathered should still be beneficial to make as best decision as possible at each crucial moment, so as to respond appropriately to the appearing situation at any time. Hence, this viewpoint was taken into consideration during the development phase of the present research.

Unfortunately there were some discrepancies found between the results described in the text and those in the tables; consequently, there was some ambiguity regarding the exact accuracy achieved through their experiments. However, roughly speaking, they managed to obtain accurate performance in both engine cylinder type classification and military vehicle identification by analysing the spectral peaks of acoustic signals. Therefore, this could be regarded as an indication of the good potential of the employed methods.

**(f) [Eom, 1999]**

Eom recommended another acoustic vehicle classification algorithm. It consisted of Time-Varying Autoregressive (TVAR) combined with low-order Discrete Cosine Transform (DCT), as well as Maximum Likelihood (ML) decision rules for selecting the optimum model orders for AR and DCT expansion. Principal Components Analysis (PCA) for dimensionality reduction, three-layer feedforward NN trained by Backpropagation algorithm as the classifier were also added. It was suggested that TVAR for feature extraction can be superior to other time-frequency domain approaches in acoustic vehicle recognition due to its capability of representing non-stationary signals with relatively fewer parameters. It was also mentioned, however, that the downside of TVAR was its susceptibility to a wrong model order selection. Nevertheless, they argued that the above issue of finding the optimum number of order for TVAR method was overcome by employing ML decision rules. Although there was not much detailed description regarding their data; it was reported that the algorithm was tested with both synthetic signals as well as some actual acoustic signals of six vehicles, which were most likely military ones because of what the acknowledgement included. It was stated that 88.3% of these actual data were classified correctly.

### (g) [Liu, 1999]

In a master dissertation, a feature extraction module derived from biological hearing model was applied on the US Army Laboratory Acoustic-seismic identification Data Set (ACIDS) database comprises recordings of 9 military vehicles. Learning Vector Quantisation (LVQ) and its descendants were examined, and the original LVQ performed the best reaching 92% although its training time was the longest.

### (h) [Averbuch et al., 2001, 2009]

The outcome of another research work that followed the findings of Choe et al. [Choe et al., 1996] was published [Averbuch et al., 2001] having military applications in mind. In this case, Wavelet Packet (WP) energy, rather than standard Wavelet Transform (WT), was studied; in order to find a narrow dominant frequency band of acoustic signals generated by vehicles. “8-th order spline wavelet packets and the Coiflet5 wavelet packets with 10 vanishing moments” [Choe et al., 1996, p.13] were selected. The paper reported two tasks; the first of which was classification of acoustic signals generated by 3 vehicles, and another was detection of acoustic signals generated by one particular vehicle from a mixture of recordings by consulting the stored model feature vectors. The mixture consisted of signals of the target vehicle, other vehicles and background noise.

For the former, performances of two classifiers; Linear Discriminant Analysis (LDA) and Classification and Recognition Trees (CART) were studied. Although no numerical figure was included, it was reported that CART outperformed LDA. For the latter, only CART was used because of its resultant performance observed for the former. The potential was shown, with appropriate modification, to develop CART classifier to perform such a specific task with very few failures. The latter was expanded in their later publication [Averbuch et al., 2009] to discrimination between acoustic signals generated by vehicles and those by the others. Although neither the precise size of samples for the study nor statistical results regarding performance was clear, the experiment with various acoustic signals recorded under different conditions was illustrated. For this task, 6-th order spline wavelet was employed; and the positive effect of combining some feature dimension reduction algorithms, such as Mahalanobis distance or PCA, was examined.

#### (i) [Srour, 2001]

A paper by Srour of the U.S. Army Research laboratory described the use of spectrum analysis to perform spectral peak detection, a harmonic analysis called Harmonic Line Association (HLA), as well as beamforming for military vehicle detection, localisation and classification. This research was based on the findings published earlier in a technical report [Srour and Robertson, 1995] on real-time detection and tracking algorithm for military vehicles, which also mentioned a great deal about the characteristics and propagation of acoustic signals generated by moving vehicles as well as background noise in a battlefield. It was suggested that when a moving target was further than 20 metres away from the sensor, acoustic signals were influenced by the various environmental factors, causing especially the lower frequency components to be attenuated, thus making the signal processing more challenging. Then, a method developed by the same laboratory in order to provide estimation of the level of the signal attenuation caused during sound propagation was employed to compensate for the challenge. Although only the preliminary results were presented, the paper showed some relatively rare insights into research on vehicle recognition carried out by such an institution.

#### (j) [Wang and Qi, 2002]

In a conference proceedings, a paper was published on acoustic military vehicle recognition using signals collected by distributed acoustic sensor arrays. It was stated that the sample signals for the experiment were from four vehicle classes although there was no indication of how many vehicles involved to form a class.

The features were extracted by both spectrum analysis and Wavelet transform. In the former, FFT (the resolution was not specified) was applied to samples of every one second and then the highest peak was found in the spectrum to estimate the harmonic lines of the signal. Then, once the feature vectors were consolidated, PCA was processed before classifying them by either k-Nearest-Neighbour (kNN) or minimum-distance approach. Unfortunately there were some discrepancies between the figures presented in the table and what was written in the text. Nevertheless, it was understood that kNN outperformed the minimum-distance classification algorithm.

#### (k) [Wu and Mendel, 2003, 2004, 2007]

Wu and Mendel adopted HLA for their algorithm to extract features from acoustic signals of military vehicles using 2nd to 12th harmonics. Their data for the experiments were selected from a dataset ACIDS consisting of "nine kinds of ground vehicles in four environmental conditions" [Wu and Mendel, 2003, p.123], ths ame data set as in above [Liu, 1999]. Performances of Bayesian Classifiers, type-1 and type-2 Fuzzy Logic Rule-Based Classifiers (FL-RBC) were compared, firstly for three sets of binary classification tasks between "tracked" and "wheeled", "heavy-tracked" and "light-tracked", "heavy-wheeled" and "light wheeled" vehicles [Wu and Mendel, 2003], as well as for classification between multiple classes [Wu and Mendel, 2004, 2007]. The classifier selection was emphasised with the argument that a suitable classifier for the task needs to be able to handle "time-variation and uncertainties" [Wu and Mendel, 2003, p.123] because acoustic signals of moving vehicles are much influenced by various unstable factors. It was reported that the type-2 FL-RBC outperformed type-1 FL-RBC, and type-1 FL-RBC produced better results than what achieved by the Bayesian classifier. The training samples were one second blocks Closest Point of Approach (CPA) about the sensors, having the highest Signal to Noise Ratio (SNR) of each recording. It was also added that the classification in so-called “adaptive operational or working mode” showed better performance.

### (l) [Munich, 2004]

These samples from the same database ACIDS as what used in the above studies [Liu, 1999; Wu and Mendel, 2003, 2004, 2007] were utilised in another paper. Without any supporting evidence nor detailed discussion of the argument, Munich stated that vehicle recognition is similar to speech recognition. On such standing, he examined the feasibility of recognition techniques well known within speech recognition on vehicle signals by comparing the performance with that of techniques used in face recognition. The techniques adapted from speech recognition studies were: Mel-Frequency Cepstral Coefficient (MFCC) feature extraction algorithms combined with either Gaussian Mixture Models (GMM) or Hidden Markov Models (HMM). Another algorithm that was used for comparison purpose consisted of STFT, PCA, and Probabilistic Bayesian Subspace models. Classification was processed by ML estimation performed with three different models. Consequently, it was reported that classification accuracies achieved by the algorithms adapted from speech recognition were inferior to the another algorithm developed for face recognition as the latter managed to achieve around 85 to 90%.

### (m) [Qi et al., 2005]

A team of researchers conducted a comparative study regarding recognition of multiple moving vehicles with some Blind Source Separation (BSS) algorithms. The focus was on the front end, the source separation; and corresponding recognition accuracy was reported. The short paper concluded that their algorithm, consisting of BSS with subband filter applied to the STFT output, managed to recognise two signals that were simultaneously present at a time. The outcome was better than other algorithms that first processed the separation of the source before the classification. However, it should be pointed out that little detail information regarding the GMM classification algorithm was included, and more importantly the experimental data were artificially prepared by mixing separate recordings and replaying them through separate speakers.

**(n) [Viangteeravat et al., 2007a,b; Viangteeravat and Shirkhodaie, 2007]**

In the USA Army Laboratory, there have been at least three conference papers published for their research on acoustic vehicle detection and identification performed by the same people; and, some of the results have been released to the public. In general, the three papers discussed the similar experiments carried out, with some discrepancies regarding the details. Without thorough description of the contents, it was stated that the sample data were taken from the US Army Research Laboratory database, which appeared to be the common practice for research in the institution. In addition, all the samples were those selected as CPA; based on the energy levels by a method they called Energy-Level Monitoring (ELM).

Firstly, they introduced an algorithm with experimental results using two military and one civilian vehicles [Viangteeravat et al., 2007a]. In this algorithm, a Hamming window function was applied to the CPA signals before FFT was employed to construct the Harmonic Line Frequency Sets (HLFS) (HLFS) matrices. Then, Singular Value Decomposition (SVD) was utilised to calculate the eigenvectors and eigenvalues of these matrices. The eigenvector, which was associated with the significant eigenvalues, was used as the extracted feature vector. The correlation between feature vectors of the test samples and that of reference data was used for reassuring the CPA detection outcome. The classification was performed with Particle Filtering, achieving on average 80.33% accuracy although the total number of the test samples was not stated in the paper.

On the other hand, in the latter two papers that are very similar to each other [Viangteeravat et al., 2007b; Viangteeravat and Shirkhodaie, 2007]; they used what was called “Low Rank Decomposition based $L_p$ norm” to overcome the problem with frequency shifting caused by the variable moving speed of a vehicle. Although once again there was not enough information regarding precise size of the samples, it was shown that the classification performance was improved by approximately 5% on average for three civilian cars as well as two military and one civilian vehicles, probably by using the same dataset as in the above paper.

### 2.1.2 Research on Road Vehicle Recognition

This section describes research on road vehicles rather than military ones. In general, research on road vehicle recognition is performed for traffic management or automated driver assistance system.

#### (a) [Sampan, 1997]

A PhD thesis was published for traffic monitoring by using a circular array, consisting of 152 microphones, but 143 of them were of interest. After the data were pre-processed to only maintain the components between 2700Hz and 5400Hz, the 30-dimension feature vectors were extracted from the energy over each 0.2 seconds in the time domain. PCA was processed to reduce the dimension to 24 before performing classification with either kNN, Multilayer Perceptron (MLP) or Adaptive Fuzzy Logic System (AFLS). Although the exact number of vehicles in each class was not clear, it was stated that they varied between each class; and there were data of 1327 vehicles in total. Problems caused by this inequality of the sizes of training sets was addressed in the thesis by considering the effect on the training as well as introducing the learning factor, a method intended to take care of the issue. Moreover, it attempted to deal with the inequality derived from the fact some classes are easier to be learned than others by allocating misclassification costs for each pattern although there was no logical explanation regarding how they were determined. The obtained classification accuracies were 97.95%, 92.24% and 78.67% for 2-class, 4-class, and 5-class experiments respectively. Althugh the use of such a large number of microphones may not be appropriate for developing a cost effective and compact recognition system, this study played a good role in initiating research on acoustic road vehicle recognition with fairly accurate results.

### (b) [Nooralahiyan et al., 1997, 1998]

Researchers at the University of Leeds published two similar papers on their acoustic road vehicle classification studies for traffic monitoring in the late 1990s. The focus was on classification only thus detection algorithms were not included. The study was motivated by the progress in automated speech recognition research that apparently let them believe feasibility of the short term spectrum, particularly at low frequencies, for the task. Therefore, the resultant algorithm choice was made heuristically but with influence of speech recognition studies.

In the first paper [Nooralahiyan et al., 1997], they first conducted a feasibility study using acoustic signals of four vehicles; such as a small saloon, a medium saloon, a motorcycles, and a light goods vehicle. All of the recordings were collected under relatively controlled conditions, particularly in terms of the level of background noise and vehicle speed. At this stage, the feature extraction methods used were: FFT and autocorrelation method Linear Predictive Coding (LPC) both in MATLAB, and also software for computational modelling of hearing, which was developed by Patterson et al. [Patterson et al., 1995]. For the latter, Equivalent Rectangular Bandwidth (ERB) based gammatone filterbanks, covering between 100Hz and 12kHz, were used to simulate the movement of the basilar membrane in the cochlea. Moreover, the same software platform was also used to simulate auditory nerve activity patterns of the cochlea's inner hair cells. Within these methods, FFT was omitted after the initial stage because its classification outcome was not satisfactory. An unsupervised Kohonen Self-Organising Map (SOM) was used for this first phase of the study.

They then carried out another set of experiments using signals collected at various urban road sites where they had little control compared with the first stage [Nooralahiyan et al., 1998]. In total, over 200 recordings of various vehicles travelling at different speed of no faster than 40 miles per hour (i.e. approximately 64 km per hour), collected at more than 20 sites were utilised. The collected feature vectors were classified by the supervised Time Delay Neural Network (TDNN) with two hidden layers and no feedback, between the following four classes; buses or lorries, saloon cars, motorcycles, and light goods vehicle or vans. The reported classification accuracies achieved, particularly with adaptively changed threshold, were above 84%. These results were seen as good, therefore, it might be a good reference for the the current research.

**(c) [Wu et al., 1998, 1999]**

Another example was reported in two similar papers published firstly in a conference proceedings and then in a journal. The adapted method was called "eigenfaces methods", which had been used in human face image recognition beforehand. It is also known as Karhunen-Loeve expansion or PCA. The feature vectors of acoustic signals were collected with FFT power spectrum, which was then normalised per frame to unit power. Some adjustment was also applied to reduce the impact of insignificant parts before PCA was performed. The class of unknown samples was determined according to the distance of the new feature vector from the reference, which was prepared in advance during training by using only selected data, recorded at the same location under very similar conditions. Recordings of saloon type cars were used as the training samples during the experiment. The distributions of training and test sample sets were exhibited in graphs, but no numerical values for classification accuracy nor sample size was included. It was pointed out that the algorithm was susceptible to the change in recording conditions; furthermore,

more data would be required to study sample distributions that usefully display the signals' distinguishable attributes.

### (d) [Jacyna et al., 2005; Necioglu et al., 2005]

A group of researchers at the MITRE Corporation in the USA have published work on developing acoustic road vehicle classification algorithms for networked sensor systems [Jacyna et al., 2005]. Their approach was based on a hierarchical network topology, and the classification was performed between two road vehicle classes: the "light" and the "heavy". Firstly they looked into algorithm development for a simple classification that can be performed at each sensor node. The feature vectors were extracted by a combination of FFT and 8-band filterbank, and then classified by a so-called "linear-weighted classifier". Although there was not much information about the actual experimental results nor the total number of different vehicles involved in the experiment, other than mentioning the number of 125ms long audio file segments for car and truck signals; it was reported that they managed to achieve the minimum error rate of approximately 13%.

Secondly, according to another paper [Necioglu et al., 2005], a more sophisticated classification algorithm that can be processed on the features collected by the each node described above, but at the next stage of the hierarchical network topology, was examined. Although the feature extraction algorithm was similar to that of the first one, this time they disclosed more information regarding the experiment and the data set, in which the effects of small parameter variation were also examined. The selected classification algorithm for this stage was ML estimation with GMM. It was reported that when the feature vectors, which were the filterbank output of spectral energy, were scaled logarithmically; the minimum error rate was found to be nearly 7%.

**(e) [Sobreira-Seoane et al., 2008]**

An algorithm that has some degree of similarity to the one studied for the current research in a sense as they both consist of two stages, such as vehicle presence detection and classification, was reported at a conference, which some of the findings of this PhD research was also presented at. It was realised that the obvious difference between the two algorithms would be that Sobreira-Seoane purely relied on acoustic signals. For vehicle detection, Short Time Energy (STE) and envelope peaks were used. The chosen feature extraction algorithms for classification stage were; Zero-Crossing Rate (ZCR), “Spectral Centroid (centre gravity of spectral power distribution)”, “Spectral Rolloff Point”, Subband Energy Ratio (SBER), and MFCC. kNN and LDA algorithms were utilised as the classifiers. The primary results showed that the combination of kNN and three feature extraction algorithms; Spectral Rolloff Point, SBER and MFCC, performed comparatively better, and achieved nearly 90% accuracy.

**(f) [Lu et al., 2008a,b]**

A vehicle detection algorithm, based on studies in mammalian perception and neurology, was suggested in articles in IEEE conference proceedings. This was another research initiated from an augment that “there exists similarity between vehicle and speech recognition”[Lu et al., 2008a, p.1336]. However, once again, they included no evidence nor discussion to support such a view point. Their recommendation was to use gammatone filterbank on STFT for feature extraction, dimension reduction and the two-step decision making processes to first detect the presence of vehicles before classifying them into four vehicle groups. They conducted experiments to compare MFCC and gammatone auditory filterbank on STFT. They showed that the gammatone filterbanks outperformed MFCC under noisy environment whereas MFCC was better in less noisy conditions. In addition, a technique called Spectro-Temporal Representation (STR) was described. This enhanced the correlation between neighbouring signal frames, and the use of Nonlinear Hebbian Learning (NHL) for reducing the dimension improved the detection accuracy. In the best case, under less noisy environment where the SNR was 10dB, it achieved about 95% accuracy, on average under two kinds of noise although the performance declined significantly as SNR decreased. They argued that the algorithm could be expanded to classification tasks without difficulty.

#### (g) [Anami and Pagi, 2009]

Recognition algorithms for two-wheel vehicles were investigated. Vectors extracted by three time domain methods, such as ZCR, STE and Root Mean Square(RMS) plus two frequency domain feature extraction methods, "Mean and Standard Deviation of Spectrum Centroid (CMEAN and CSD)" were applied to the NN for classifying between two classes: "bikes" and "scooters". The recorded acoustic signals of 118 vehicles were utilised, of which 58 were for training and the rest was for testing. It achieved 73.33% accuracy between the two classes. The observed effects of vehcie ages and how well they were maintained within the same class were also mentioned. Such information can be useful for certain applications that have only known and limited targets, however, it does not apply to the current project.

## 2.2 Seismic Vehicle Recognition

The potential of exploiting seismic surface waves, in particular Rayleigh waves, for military vehicle and personnel tracking was investigated [Succi et al., 2000], by the same team as discussed in Section 2.1.1 for acoustic military vehicle recognition [Succi et al., 1999]. The outcome was positive regarding military vehicles; however, it was found their method was not suitable for non-military passenger vehicles.

### 2.2.1 Research on Military Vehicle Recognition

#### (a) [Tian et al., 2002]

Classification results of two classifiers, kNN and minimum-distance classifier, processed with seismic signals of two moving vehicles were found in a conference proceeding. Application of a very similar algorithm to acoustic signals gathered by distributed sensors [Wang and Qi, 2002] was also reported in the same conference proceeding, as previously mentioned in Section 2.1.1. The two paper shared two identical authors. The feature extraction algorithm consisted of spectrum analysis, including Power Spectrum Density (PSD), and DWT coefficients, using Daubechies wavelet. Then, PCA was applied to the feature vectors to reduce the dimension before classification was performed. It was reported that kNN outperformed the minimum-distance classification algorithm. Since the experiment was carried out on signals of only two vehicles, the outcome may have been biased, particularly about those after some re-organisation of sample sets were applied; hence, the study may lack any general applicability to others.

#### (b) [Jackowski and Wantoch-Rekowski, 2005]

A conference paper explained some aspects of a seismic military vehicle recognition algorithm suggested by two researchers in Poland. They used LPC and Multi-layer NN with one hidden layer, combined with Backpropagation learning algorithm. Unfortunately, not much detailed information of their study was revealed in the paper, however, the small number of examples shown indicated a good results of the seismic signal moving military vehicle classification algorithm.

### 2.2.2 Research on Road Vehicle Recognition

#### (a) [Lan et al., 2005]

Another research paper discussed moving vehicle recognition using seismic signal processing, especially by focusing on the Rayleigh waves only. Although a moving ground vehicle could generate more than Rayleigh waves and other surface waves, it was suggested that the most significant energy of the impact caused within the ground surface by a moving vehicle is transmitted as Rayleigh waves. Therefore, detection of a moving vehicle was attempted by using vertical axis seismic sensors, trying to capture Rayleigh waves for recognition between three classes: diesel, heavy diesel, and gasoline engine. Unfortunately, the report did not contain some of the details regarding the algorithm, such as the sampling frequency, signal frame size, number of neuron nodes, as well as explanation of how exactly the feature vectors were handled during the computation.

Yet, it was reported that the collected feature vectors consisted of frequency spectrum and amplitude, and they were classified using three-layer multi-layer NN, which included one hidden layer, trained by Backpropagation algorithm. Unlike the ordinal Backpropagation algorithm [Bishop, 1995], the learning rate of this suggested one was not constant. Instead, the learning rate was reduced as the classification error declined. The presented experimental results indicated that the samples were processed with approximately 86% accuracy. Consequently, it was claimed to be possible to achieve a high recognition accuracy by using only vertical axis seismic sensors placed within the ground surface area. Although without much detailed explanation of the idea, it was also suggested that employing more than one types of sensor to collect data would improve the classification accuracy, the processing speed and the cost, for example by combining seismic signals with acoustic or magnetic signals.

## 2.3 Acoustic and Seismic Vehicle Recognition

There is a brief record of a study on automated moving vehicle localisation using acoustic and seismic signals, which was included in a report of the meeting held in the late 1970s [Brooks et al., 1977]. It was stated that they were interested in the velocity change of, and also the Doppler effect observed in a narrow-band spectra of acoustic and seismic signals. Although there are no journal publications found on this study, it indicated that the idea of combining acoustic and seismic signals processing for automated moving target detection had been considered.

### 2.3.1 Research on Military Vehicle Recognition

#### (a) [Altmann et al., 2002; Altmann, 2004]

Research on acoustic and seismic vehicle pattern recognition for security purposes was conducted by Altmann and his team in Germany. They collected a set of acoustic and seismic data, collected from 10 different vehicles running at 7 various speeds on 4 separated lanes from 2 directions in a German armed force site, i.e. 560 data variations in total. They spent one month in 2000 to collect these and meteorological records as well as several recordings of background noise, including non-coloured noise. Their recognition algorithm utilised LVQ performed on the relative power within the first 15 harmonic lines observed in the power spectrum obtained by 2048 points FFT. The maximum recognition accuracy was reported to have reached 98% for four tracked vehicles and 83% for four wheeled vehicles.

It was reported that the classification of the heavy vehicles was relatively more successful; however for lighter vehicles, the algorithm required manual adjustment by the operator. It was also pointed out that the acoustic signal processing algorithm was developed with a focus more on the engine cycle rather than vehicle type. Moreover, in the more recent study, it was found that combining a beamforming algorithm to remove background noise by using microphone arrays can be useful.

However, beamforming was found to be useless for the seismic signal processing. Consequently, it was concluded that a further development of the algorithm is still necessary to improve its performance, especially for vehicle type recognition.

### (b) [Li et al., 2002]

Li and his team published their experimental results of acoustic and seismic 2-class military vehicle type recognition, a classification study between tracked vehicles and wheeled vehicles, in IEEE Signal Processing Magazine. The PSD feature vectors were collected with FFT, from acoustic and seismic signals separately. The classification results by three algorithms, again on acoustic and seismic signals separately, were shown. The employed classifiers were: kNN, ML with GMM, and Support Vector Machines (SVM). It was understood that the SVM classifier achieved on average 94% for acoustic and 96% for seismic classification, and therefore outperformed the other two.

### (c) [Qu et al., 2003]

Qu and others [Qu et al., 2003] published a conference paper on their acoustic and seismic military vehicle classification study. Although the paper explained little about their algorithm and the experiment, it was understood that they used PSD combined with narrow band energy for feature extraction, and fuzzy logic rule based classifier and NN with Backpropagation learning algorithm. It was read that the average classification accuracy achieved during the experiment was approximately 85%.

### (d) [Duarte and Hu, 2004]

Duarte and Hu reported their research on 4-class vehicle recognition using wireless distributed acoustic and seismic sensor network. The following were provided at each network node: a microphone, a geophone, an infrared sensor, together with a processor that performed primarily recognition based on the local data, a transmitter and batteries. The data used in the research were collected at a USA military site over two weeks of November 2001 when they gathered raw data for later evaluation at each sensor node while also locally extracted feature vectors together with class labels that were manually designated by operators using their perception. During their experiment, local classification was performed only on the data segments, of which the presence of vehicles were detected by either the manual analyses or kNN algorithm; by the use of distance information and acoustic energy values. The local classification results of three classifiers; kNN, ML with normal distribution, and SVM were compared. The results showed similar levels of accuracy reached just under 70% by all three classifiers regarding acoustic data, and also slightly inferior performance on seismic classification; in which ML and SVM produced fairly better results than that of kNN.

**(e) [Mazarakis and Avaritsiotis, 2006, 2007]**

Mazarakis and Avaritsiotis attempted a two-class military vehicle classification with acoustic and seismic Wireless Sensor Networks (WSN). The chosen feature extraction algorithm was similar to Time Domain Signal Coding (TDSC) and Time-Encoded Signal Processing And Recognition (TESPAR), which was originally developed by King and Phipps [King and Phipps, 1999] based on the Time-Encoded Speech (TES) research [King and Gosling, 1978]. TDSC and TES are explained later in Section 5.1.5. Results of two classifiers were compared; one was the commercially available Fast Artificial Neural Networks (FANN) that used multi-layer feedforward NN with Backpropagation algorithm, and another was called "Archetype" classifier that measured the distance between the average values of training sample of known class and a test sample regarding matrix representation of time domain signal shapes, developed for TDSC [Chesmore, 2001]. For analysis, they employed signals obtained from 9 runs each of two vehicles: a heavy wheeled truck and a tracked vehicle. The non real-time classification experiment showed results reaching around and above 80%, encouraging for time domain recognition studies, particularly combined with the proven simplicity and inexpensive computaion

**(f) [Xiao et al., 2006]**

Xiao and his team published two articles regarding their military vehicle recognition study using acoustic and seismic signals. It was a comparative study between combinations of two feature extraction, two dimensionality reduction, and two classification algorithms. For feature extraction, they chose STFT and a newly proposed method based on MFCC. Compared with the standard MFCC algorithm, theirs used rectangular filterbank instead of triangular, and there was no DCT included. The dimensions of collected features were reduced by either Genetic Algorithm (GA) or PCA. Firstly, the two classifiers, kNN and SVM with RBF kernel for training, were processed on STFT with no dimensionality reduction algorithm, and it was found that SVM produced better accuracy than kNN.

For SVM and kNN, the effect of five different sizes of windowing function on classification results was also analysed. Interestingly it was reported that the best window size was not the same for each classification algorithm when acoustic signals were used although it turned out to be the same for seismic recognition. This indeed indicated the importance of examining the best window size so as to improve the overall recognition performance. The rest of the experiment was only carried out with SVM algorithm using a fixed window size. It was reported that, on the whole, classification accuracy of acoustic recognition was better than that of seismic one because the acoustic classification achieved above 85% at maximum whereas the maximum accuracy for the seismic classification was around 70% up to 77%. Furthermore, they also mentioned that it would not be trivial to realise a fair comparison regarding results achieved so far in various vehicle recognition studies becasue of the multiple influencial factors involved, including dataset and target class variatoins.

#### (g) [Xiao et al., 2007]

In 2007, two of the three authors of the above published another paper in a collection with some other authors. By using the same acoustic and seismic data sets, this time they examined slightly different algorithms. For the pre-processing and feature extraction, they used above STFT method with a fixed window length. Now as a comparative study, they chose two dimensionality reduction algorithms: PCA as in above and Independent Components Analysis (ICA) (ICA), and also four classifiers: in addition to SVM and kNN as in above, decision tree and PNN. As a result, it was shown that firstly the PCA dimension reduction algorithm performed better than ICA for acoustic signal classifications, however the opposite was true for the seismic signal ones, and secondly the SVM outperformed the other classifier, followed by PNN then kNN. Nevertheless, it was also reported that this order of four classifiers was completely the reverse, in terms of the processing speed. The maximum accuracy achieved for acoustic signal classification was almost 90%, and for seismic, it was 73 %. Overall, acoustic classification produced better results.

## 2.4 Chapter Summary

An extensive literature review in relation to the current study area has been carried out. Research publications on automated vehicle recognition algorithm development have been around since the late 1990s. Among them, however, there are only five studies on road vehicles as seen in Section 2.1.2, and the target signals of the others using acoustic and/or seismic signals processing were mainly focused on military vehicles. Another obvious outcome of the survey is, to the author's knowledge, there is only one published research on automated road vehicle recognition utilising seismic signals; and none on research using acoustic and seismic fusion. Although classification with seismic sensors only has not produced strong results, its feasibility up to a certain level has been demonstrated by others and consequently encouraging the fusion of acoustic and seismic sensors. As stated in Section 1.3, this research is carried out aiming to develop algorithm to protect important infrastructure. For that, the potential threat may not be limited to intruders approaching by military vehicles. Therefore, the rest of the thesis concentrates on road vehicle recognition using both acoustic and seismic signal processing.

Interestingly it was also realised that the most popular approach for the feature extraction algorithm was to use either spectral or harmonic analysis methods based on FFT; including peaks of STFT or HLA or HLFS, filterbank such as MFCC or wavelet based. The relatively minor approach, on the other hand, was to employ either short-time energy, LPC based or AR based methods. Furthermore, over all, algorithms were often combined with PCA for reducing the dimension of the collected vectors.

Once again, because this development aims at security applications, there are several critical aspects that are open for improvement. Performance accuracy and computation speed are possibly two of the most important aspects need to be addressed. In addition, the scope of the research has been defined based on some other points that also require attention due to the project resource limitation; such as development and operational costs as well as time allowance for a PhD study in the United Kingdom.

Anchored in all of these findings, a range of techniques to be studied further in order to develop an optimum vehicle recognition algorithm have been selected. For feature extraction (Chapter 5); first of all, the time domain methods have been investigated owing to their simplicity, speed and inexpensive computation. Then, filterbank is also applied mainly for comparison purpose as a result of its popularity. In terms of classifiers (Chapter 6), statistical basis algorithms including NN have been applied with or without dimensionality reduction. Each of which will be explained together with outcomes of some empirical studies with real vehicle signals collected during the research project.

# Chapter 3

# Signal Generation and Propagation

## 3.1 Signal Generation

Understanding the fundamental characteristics and behaviour of target sources is essential for development of an automated recognition system. One approach to ensuring such comprehension of the attributes of a target signal can be by modelling the characteristic structures of the signals emitted by the source. Modelling of vehicle signals could be advantageous either directly or indirectly to the current research; and such studies of both acoustic [Favre, 1983; Dyrdal et al., 2003; Christou and Jacyna, 2005; Cevher et al., 2007] and seismic [Anderson et al., 2003; Ketcham et al., 2005] signatures of moving vehicles for developing classification systems have been attempted by other researchers. Nevertheless, it is important to remember the purpose of this research is to develop a recognition algorithm that can support distinguishing the types of vehicles, rather than studying details of how such acoustic signatures are generated. Therefore, it was decided to concentrate on identifying distinctive characteristics of the real-life signals, in order to develop appropriate recognition algorithms.

### 3.1.1 Acoustic Signal Sources

The acoustic signals of a moving road vehicle are generated mainly by the movement of various components, such as a power unit and tyres but also others like an exhaust system [Thomas and Wilkins, 1970; Srour and Robertson, 1995; Wu et al., 1999]. Signals of the current research's interest are vehicle exterior signals rather than the interior ones when it is assumed a vehicle is moving within the legal speed range. Note that under such an assumption, consideration of the aerodynamic noise generated around the vehicle body is not usually required [Sandberg and Ejsmont, 2002]. Also the horn and/or burglar alarm sound as well as any additional sound sources that may be equipped in a vehicle (e.g. for warning, entertainment, or marketing purposes), are outside the focus of this research.

With these assumptions, acoustic signals of a moving vehicle are generated by the vibration caused in the engine itself, the resonance of the exhaust pipe as the waste gases pass through, the tyre friction noise as well as vibration of the rest generated both directly and indirectly mainly but not only, by the above three notable sound sources within a vehicle [Thomas and Wilkins, 1970, 1972; Karlsen et al., 1995; Srour and Robertson, 1995; Nooralahiyan et al., 1998; Wu et al., 1999; Wang and Qi, 2002; Munich, 2004; Necioglu et al., 2005]. Since they involve multiple types of motions and materials, the characteristics of the resultant signals are not simple. This consequently suggests some complexity concerned with the current research in understanding the sound source attributes and their modifiers in terms of a recognition target.

#### (a) Power Unit

In the power unit of a common four-stroke petrol engine car, the core movement occurs in the crankshaft, which is connected to the rod bearing, then connecting rod and then the piston that moves up and down; causing the four main steps performed in one cycle of a four-stroke engine. The four steps are "intake", "compression", "power", and "exhaust". The four-stroke is initiated from two cycles of the crankshaft's movement triggering the piston to go down and up twice. During intake some petrol and air mixture is sucked in as the piston goes down. The petrol is brought in to the combustion chamber and gets mixed with the air. Then the mixture is compressed whilst the piston moves up. Next, a spark ignites the fuel and the combustion happens as the piston goes down for the second time, then the exhaust gas is discharged into the exhaust pipe to be thrown out to the exterior while the piston goes up again to complete the four stages and be ready for the next cycle [Ofria, last accessed in 2009]. The noise generated by the power unit has a strong relation to the above operations. It is suggested that its main sources are caused either by the changing pressure of the combustion chamber and the mechanical movement of the various parts [Nelson, 1987] or the recurring ignition [Succi et al., 1999].

### (b) Tyre Friction

In terms of the acoustic signals generated by the tyres, vibration caused by the rotational movement and the contact with the ground surface as well as the signals due to aerodynamic forces are the main source. Through modelling studies [Kujipers and van Blokland, 2001] and their empirical examinations [Kropp et al., 2000; Lui and Li, 2004], it was confirmed that the geometry between a tyre and the ground surface, of which the tyre has contact with, has some notable influence called the "horn effect" [Sandberg and Ejsmont, 2002; Cevher et al., 2007] on amplifying the sound radiated by the tyre. It has been recognised [Sandberg and Ejsmont, 2002] that at a higher travelling speed, sometimes the aerodynamic noise due to wind turbulence becomes more important, but often the distinction between that and the tyre noise is a challenging task.

Sandberg [Sandberg and Ejsmont, 2002] also pointed out that there are some differences between the dominant influential factors for power unit noise and that for tyre noise. The power unit noise is mainly affected by both the power of and the revolving speed of the engine. On the other hand, for the tyre noise, the vehicle's travelling speed as well as the combination of the tyre's properties and the road surface condition are significantly more important. Also the forces caused by acceleration, breaking or cornering influence the latter.

It has been suggested that while the tyre friction noise increases proportionally to the travelling speed, the acoustic signals generated by the power unit are independent of speed variation [Sandberg and Ejsmont, 2002]. Consequently, a combination of the lower vehicle travelling speed and the higher engine repetition speed usually results in the signals generated by the power unit becoming greater than the tyre friction noise. On the other hand, at higher vehicle moving speeds, the lower engine repetition speed and the lower engine power can lead to domination by the friction noise [Nelson, 1987].

In terms of the speed at which the tyre friction noise becomes more dominant in acoustic signals generated by a moving vehicle, Rustighi [Rustighi et al., 2008] stated the threshold as at 40km per hour based on [Sandberg, 2001], whereas some others [Christou and Jacyna, 2005; Jacyna et al., 2005; Cevher et al., 2007] have said it is at around 50km per hour. However, it should be noted that some [Nelson, 1987; Sandberg and Ejsmont, 2002] have also pointed out that it can depend on multiple factors, including the ground surface conditions. Sandberg and Ejsmont showed, in addition, a discussion on how each sound source contributes towards the overall acoustic signals of moving vehicles, and also how that depends on types of vehicles, together with how all these have changed over the last several decades as a result of the progresses seen in the technology, legislation and the market trend.

In terms of the level of noise produced by vehicles in different classes, furthermore, it may be worth noting here that the acoustic signals generated by a heavy vehicle are often louder than that of a small car [Nelson, 1987; Sandberg and Ejsmont, 2002].

### 3.1.2 Seismic Signal Sources

It has been pointed out [Altmann, 2004] the generation mechanism and the propagation of seismic signals of a moving vehicle can be more complex than those of the acoustic signals. The ground surface vibrations that are important for the current application are mainly the seismic surface waves rather than body waves. Although surface waves often consist of various sub-member waves, there are two dominant surface waves, which have been well studied. They are called Rayleigh waves and Love waves, after the pioneer physicist and the geophysicist respectively [Robinson and Coruh, 1988].

As shown in Figure 3.1, Rayleigh waves progress while making ground particles rotate in elliptical circle shapes that have longer axis along the vertical axis, in the opposite direction to that of the transmission of the vibration.

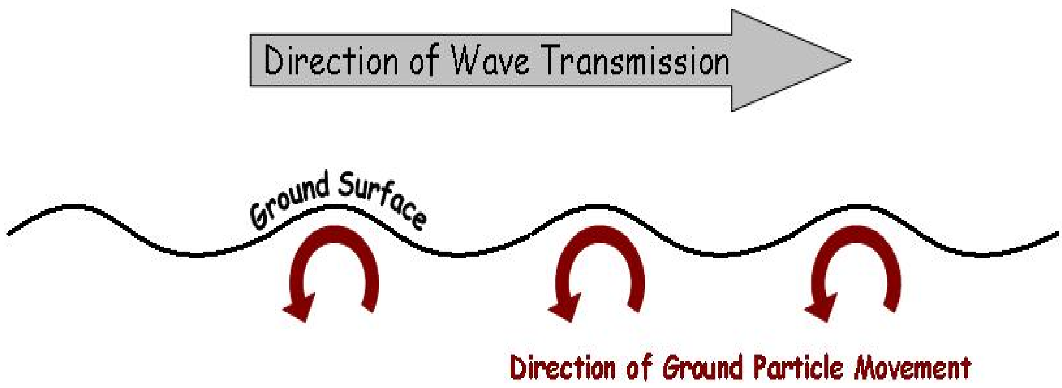


Figure 3.1: Transmission of Rayleigh Waves

On the other hand, Love waves cause sideways movement of the ground surface particles while the vibration progress along them. The unique characteristics, of which can be seen in both Rayleigh and Love waves are that the way the speed of the vibration (i.e. the frequency) starts from a low value and then increases over a period of time. In addition, both Rayleigh and Love waves travel slower than body waves, P-wave and S-wave [Kearey and Brooks, 1984; Robinson and Coruh, 1988; Succi et al., 2000].

It has been shown [Miller and Pursey, 1955; Succi et al., 2000; Tian et al., 2002] around two-thirds of the energy generated by a single impact on a ground surface can be observed in Rayleigh waves. Also Rayleigh waves are the dominant elements of moving vehicles' seismic signals, caused by the tyre treads coming into contact with the ground surface and they decay slower than the body waves [Succi et al., 2000]. Consequently, for the current target detection problem, the important vibrations are mainly Rayleigh waves, which can be captured by a vertical axis geophone placed on the ground surface [Succi et al., 2000; Altmann, 2004].

## 3.2 Signal Attributes and Propagation

With the basic knowledge of the mechanism of how the vehicle signals are generated, as mentioned in the preceding section (Section 3.1), attention is now brought to their attributes and how they may be altered while travelling between the source and the sensors. From a recognition system developer's point of view, such knowledge is equally useful in order to understand data actually collected by the sensors; regarding what characteristics can be expected to appear in the signals, and also about estimating how various factors may have modified the original attributes during propagation. For example, as well as vehicle speed variation, signal characteristics of the same source could be fabricated by uncertainties brought in by source propagation. Although other factors also have large influence, in general, the longer the propagation path the more complicated it could get; this would introduce limitations regarding monitoring range for a sensor to collect useful data. As has been seen in military vehicle recognition [Srour, 2001], integration with sound propagation models could be an interesting approach. However, despite some achievement [Nota et al., 2005], work in this area is also still in progress [Jonasson, 2007]. Therefore, in order to avoid any additional uncertainty for now, it has been determined not to employ such a model for the current study.

### 3.2.1 Acoustic Signal Attributes

In terms of the frequency attributes of the acoustic signals, the majority of moving vehicle signals are observed in the relatively low part of human perception range, i.e. up to a few kHz [Nooralahiyan et al., 1997; Wu et al., 1999; Chen et al., 2002; Altmann, 2004] or even lower [Srour and Robertson, 1995; Srour, 2001]. It is claimed in some papers [Wang and Qi, 2002] some higher frequency components of the sound source, up to several kHz, can also be identified in spectrum analyses due to harmonic components.

Jacyna et al. [Jacyna et al., 2005] reported that acoustic energy emitted by a car passing by a receiver was observed around several hundred Hz range for a relatively short time; on the other hand, a moving truck radiated acoustic energy within several hundred and slightly higher frequency range for a longer time period. It is suggested [Altmann et al., 2002; Altmann, 2004; Cevher et al., 2007] fundamental frequency observed in a spectrum analysis of acoustic signals generated by a moving vehicle indicates the engine's repetition cycle. An overview of the characteristics, including spectrum analyses, of each significant vehicle acoustic signal source can be found in part 3 of a book edited by Nelson, especially in the chapters 6 and 7 [Nelson, 1987].

Acoustic signals are often categorised in two groups, stationary and non-stationary. Stationary signals have fixed frequency regardless of time, hence if they are analysed in the time domain, amplitude changes will only occur at regular intervals. On the other hand, the frequency of non-stationary signals varies over time, thus they are occasionally called time-varying signals [Eom, 1999]. Therefore, the frequency domain signal processing techniques may perform well for stationary signals, whereas techniques that work in the time-frequency domain may be more useful in non-stationary signal processing.

Frequently acoustic signals of a vehicle are defined as non-stationary and it is why some [Eom, 1999; Cowling and Sitte, 2003; Jacyna et al., 2005; Necioglu et al., 2005] claim that dealing with vehicle acoustics with methods generally recognised as suitable for stationary signals will not produce satisfactory outcomes. Nevertheless, there are other suggestions [Wu et al., 1999; Averbuch et al., 2001, 2007; Altmann et al., 2002; Lu et al., 2008a] that acoustic signals generated by moving vehicles can be regarded as stationary when they are analysed over a short period of time since the signals are approximately quasi-periodic. These remarks indicate it is not trivial to state whether vehicle signals are stationary or non-stationary in general terms. Perhaps, it is due to the lack of studies have been carried out regarding vehicle signals, particularly in the case of road vehicles. Or, it is because the answer can vary depending on several factors such as the vehicle's speed of travelling, signal analysis frame length, conditions, in which the vehicle is travelling including the ground surface properties. All these factors can influence the signals collected by the sensors and signal analysis outcomes; hence, it can depend on the actual application of the study. Both points can be true to a certain degree for the current case and probably this is one of the reasons why the current research is challenging. Then, clearly some examination of the real-life signals is required to progress further in this study, and the outcomes of that are reported in Section 4.2.

### 3.2.2 Acoustic Signal Propagation

It is pointed out [Choe et al., 1996; Necioglu et al., 2005; Lu et al., 2008a] there are so many factors that can have significant impact on the characteristics of vehicle acoustics, including: noise and reflected sound caused by road surface conditions, speed of both the vehicle as whole and its engine, Doppler effect, meteorological elements. As a result, the vehicle classification task is challenging.

Acoustic signals that travel near the ground surface are influenced by various factors that cause attenuation, distortion and enforcement due to reflections [Favre and Gras, 1984], particularly when the distance between the sensor and the target is greater than 20m [Srour, 2001]. Ground surface conditions such as porosity and smoothness, terrain structure including vegetation, and also meteorological properties are the main factors that have effects on behaviours of airborne acoustics especially close to the ground [Piercy et al., 1977; Attenborough, 1988, 2002; Becker and Gudesen, 2000].

For example, atmospheric temperature and wind affect the speed, refraction as well as air absorption of sound. It is known [Rossing et al., 2002; Srour, 2001] that sound travels further when a layer of warm air covers colder air near the ground. Also, lower frequency components can typically travel further than the higher frequency ones although appropriate treatment of wind noise would be crucial to intelligibility of some low frequency sources [Becker and Gudesen, 2000]. Obstacles existing in and around the propagation path cause diffraction and scattering [Howard, 2006]. It is reported that a thin layer of water on the road surface increases vehicle signal level, particularly at low speed for cars but at high speed for heavy vehicles [Jonasson, 2003]. On the other hand, acoustic signals travelling over ground covered with snow, is known to be attenuated compared with that over grass [Albert et al., 2008]. Snow also reduces the amplitude of acoustically induced seismic signals, thus alters the characteristics [Albert, 1987]. In addition, the distance between source and sensor yields the sound pressure reduction, known as the “inverse square law” [Howard, 2006, p.28]; causing “reduction of 6dB per distance doubling in all directions” [Attenborough, 2002, p.53] for point source whereas 3dB for line source [Attenborough, 2002].

When working with acoustic signals of moving vehicles, the influence of Doppler effects observed in the collected data can be useful to comprehend the real characteristics of signals generated by the source [Ding et al., 2004]. The speed of a moving object can be determined by examining frequency shifts especially when the source velocity and frequencies are high [Searle, 2005]. It was stated that the influence of Doppler effect on the characteristics observed in the acoustic signals generated by a vehicle moving at about 50km per hour within an environment at normal temperature may be insignificant [Wu et al., 1999]. Additionally, it should be noted that in single element sensing, the speed information obtained only from Doppler calculation can be confused between the case of a slowly moving vehicle close to the sensor and that of a fast moving vehicle at a long distance; hence other information such as SNR may also need to be consulted. Estimation of the travelling speed of a moving vehicle is not particularly included in the important attributes for the current research. As a result, Doppler shift has not been studied separately from any other phenomena.

After their first relatively simple study in noise emitted by a moving vehicle [Favre, 1983], modelling of moving vehicle acoustic signals was attempted by Favre and Gras [Favre and Gras, 1984]. The study focused on analysing, by a comparative study, how the real-life acoustic signals generated by a moving vehicle is influenced by multiple factors. These factors included the interference between direct signal and reflection, Doppler effect, sound absorption characteristics of the air, and also the location and the direction in regards to those of the sensor. It was concluded that the acoustic signals of a vehicle can be seen as monopole source, which is defined as "noise source in which the entire radiating surface vibrates in phase" [Rossing et al., 2002, p.715]. Favre and Gras also reported that the influence of the height of the sensor may be less significant for broadband signals, including vehicle signals, compared with that of narrow band signals.

### 3.2.3 Seismic Signal Attributes and Propagation

The propagation of seismic signals can be complicated due to their dependency on multiple factors including, the geological and topographical conditions of ground surface and shallow sub-surface, such as elasticity, porosity and consistency [Tian et al., 2002; Altmann, 2004; Lan et al., 2005]. As a result, so far, the use of seismic signals for target recognition studies has been limited.

However, compared to acoustic signals, seismic signals are less susceptible to Doppler effects [Tian et al., 2002; Lan et al., 2005]. Moreover, using seismic signals for detection allows the performance to be independent of wind conditions, which can often cause difficulties in acoustic detection. Also, compared to acoustic signals, seismic signals of interest have a much narrower bandwidth but at the lower frequency range. Therefore, firstly a much faster and more efficient signal processing, and secondly a wider detection range can be achieved [Lan et al., 2005].

It has been suggested [Altmann et al., 2002] fundamental frequency observed in a spectrum analysis of the seismic signals generated by a moving vehicle indicates the repetitive cycle of the tyre and tread contact, hence it reveals useful information to calculate the velocity at which the vehicle is travelling at [Succi et al., 2000]. The speed of a moving truck can be estimated fairly accurately by the seismic signals if the size of tread spacing is known, or the accuracy can be reached as high as 30% even if the tread spacing was unknown [Lan et al., 2005].

### 3.2.4 Acoustic and Seismic Signal Coupling

It is also known that the acoustic signals generated by a moving vehicle can also cause the ground to vibrate and this may be captured by a geophone placed within the shallow surface area. Therefore what is often called acoustic to seismic coupling should be considered in order to understand the seismic signals better [Bateman, Jun. 15, 1938; Press and Ewing, 1951; Bass et al., 1980; Sabatier et al., 1986a,b].

For example, the acoustic to seismic signal coupling has been evaluated for estimation of the speed and spatial range of an approaching vehicle, or speculating the ground surface component structure [Moran and Greenfield, 2008]. In general, correlation of acoustic and seismic signals is employed to reduce the effect. It has been pointed out that acoustic and seismic coupling is a “very reliable way of identifying the sources from aircraft and helicopters” [Succi et al., 2000, p.167]. Acoustic and seismic coupling has also been studied for land mine detection research [Sabatier and Xiang, 2001].

Nevertheless, so far, it has not been recognised as important in vehicle type identification studies. For the current study, no particular algorithm to treat any effect of the acoustic to seismic coupling has been introduced because the purpose of the research is not to gain a detailed understanding of the signals' origins but to extract useful features from the signals to be processed by classifiers. Therefore, the first attempt has been not to include any treatment to cancel such coupling contained in the recorded seismic signals unless proved to be otherwise. Equally, this coupling has not been investigated when processing acoustic signals.

### 3.2.5 Road Vehicle vs Military Vehicle

As noted in Chapter 2, there are currently only a few publications on road vehicle recognition studies using acoustic signal processing, and none was found for a study using both acoustic and seismic signals. It is worth noting here that filling this identified gap of knowledge within the topic of moving vehicle recognition is one of the contributions of the research being presented in this thesis.

For military vehicles, on the other hand, some relatively good classification results have been been obtained (Chapter 2). However, it was often pointed out that compared with road vehicles, signals generated by military are louder and more distinguishable [Choe et al., 1996; Succi et al., 1999]. In addition to the vehicle mass, data used in military vehicles are often collected when they are travelling on untreated surfaces, which may contribute to the louder signals. Moreover, it is often required to perform classification between tracked and wheeled vehicles. It has been mentioned that "some tracked vehicle may have a high-frequency peak associated with the sprocket and track system" [Srour and Robertson, 1995, p.7]; these characteristics could be utilised for military vehicle recognition. On the other hand, the current study attempts classification of road vehicles that are all with tyres.

For acoustic military vehicle recognition, it is suggested that the source of lower frequency components with greatest emission are due to engine and ground contact of either tracks or wheels. Although sound modifiers exist including exhaust and muffler, the repetitive characteristics of sources may result in harmonic structure observed in received signals [Srour and Robertson, 1995; Succi et al., 1999; Viangteeravat and Shirkhodaie, 2007]. On the other hand, it is pointed out that such harmonic structure could be missing in commercial vehicle signals as a result of design aimed to reduce noise generated by them [Jacyna et al., 2005]. Evidently, much research has been carried out that often led to legislation being created for environmental noise control in relation to moving road vehicle signals [Nelson, 1987; Nota et al., 2005].

## 3.3 Chapter Summary

This chapter has provided a general background regarding the target signals for the current study, acoustic and seismic signals generated by moving vehicles. It has revealed how the target signals are expected to consist of multiple sources, as well as how the characteristics of the signals collected by sensors may have received influence of many factors. As a consequence, some of the challenges regarding the task have been exposed. Familiarisation with the information presented in this chapter should help one to identify both signal processing and pattern recognition methods that are likely to be more appropriate than others as discussed in the following chapters.

# Chapter 4

# Signal Pre-processing

## 4.1 Data Collection

For reasons explained in Section 1.2, it was decided to utilise acoustic and seismic sensors for obtaining input signals for this vehicle recognition system development. Now, the chosen data collection method is discussed further in this section.

### 4.1.1 Sensor Layout

A system can have single or multiple input channel(s) to collect data. Multiple input channels can be achieved in various ways for example by using sensor arrays [Czyzewski, 2003; Ciosek et al., 2006] or distributed sensor networks [Agre and Clare, 2000; Li et al., 2002; Duarte and Hu, 2004; Jacyna et al., 2005; Kotecha et al., 2005; Mazarakis and Avaritsiotis, 2006]. Table 4.1 lists the advantages and disadvantages of three major sensor configurations.

Two additional points to be considered regarding layout are that firstly the position of the sensor in relation to the source; such as the height, direction and distance, would affect the received signal's characteristics [Nota et al., 2005; Jonasson, 2007]. Secondly, the ideal sensor position can also depend on other factors, including the

| | Advantages | Disadvantages |
|---|---|---|
| Mono (Single) | 1. Low cost<br>2. Input signal processing is less complex.<br>3. Low power and low space requirement.<br>4. Data can be collected at one location (not require network communication, and easier maintenance). | 1. Direction of the signal source cannot be localised.<br>2. There is no backup to ensure the detection or support when a sensor fails. |
| Sensor Arrays | 1. Able to localise the direction of the signal source.<br>2. Still low power and small space required (although maybe more than mono).<br>3. Data can be collected at one location (not require network communication, and easier maintenance). | 1. More expensive than having just a single input.<br>2. More complex maintenance than mono.<br>3. More complex computation is required. |
| Distributed Sensor Networks | 1. It can provide more accurate localisation of a target.<br>2. Faults occurring at a sensor have little effect on the whole system performance, and possible to replace small sensor part without halting the whole system.<br>3. Size and shape of coverage can be flexibly determined according to the needs and restrictions.<br>4. It receives little influence from terrain conditions and environmental ambient by placing nodes at close distance from the objects of interest.<br>5. It can be low cost for the scale of coverage it makes it possible. | 1. It requires data transmission network between multiple points.<br>2. It either needs power supply chain reaching to every node, or portable power supply equipped at each node.<br>3. Locating and replacing faulty nodes may not be a simple job, especially when multiple nodes fail.<br>4. Limited data transmission bandwidth within a network.<br>5. Processing input data collected at multiple nodes may cause computational complexity.<br>6. It can be expensive. |

Table 4.1: Variations of Sensor Placement

sensor characteristics and accessibility of the monitored area.

### 4.1.2 Microphone

As previously noted, the frequency response, directional response and the sensitivity regarding the extent, where a microphone can capture the acoustic signals generated by a source existing in the surrounding area, are the three crucial specifications. In addition to the robustness and the expense, they are to be looked at in selecting a suitable microphone for a task. In terms of frequency response, it is desired to have a microphone with a flat frequency response over the frequency range where the signal of interest dominates. This allows collection of data with little modification due to the sensor's characteristics, i.e. more authentic target signals. Regarding the suitable directional response for the current research, on the other hand, real-life acoustic signals of a moving vehicle need to be collected outdoors; generally in some open field near the ground surface with some but not too strong wind present. Data collection was performed by placing microphones on the side of a single road without other road crossing. Thus, having a directional microphone is beneficial to collecting signals directly emitted from the moving vehicles whilst minimising wind and other unwanted noise signals; in other words, maximising the SNR. The issue of sensitivity should also be considered in relation to the requirement of the final product's surveillance range.

### 4.1.3 Geophone

Maxwell and Faber and others at Sensor in Netherlands [Maxwell and Faber, 1994; Faber and Maxwell, 1997] suggested a set of geophone attributes that require particular attention because of their potential in affecting the characteristics of the seismic signals to be collected around the ground surface. Many geophone specifications are fairly similar to those found in that of a microphone. However, in addition to what has already been mentioned regarding the microphone specification to collect acoustic signals, the following are terminologies that are typically used to describe the sensor attributes for a geophone. The first is the natural frequency, which is the resonant frequency of the spring responding to the movement of the shallow ground along its working axis. It determines the lowest frequency that can be measured by the sensor. The second is referred to as the spurious frequency, which is the resonant frequency along the axis perpendicular to that of the natural frequency. Spurious frequency, sometimes referred to as the lowest of spurious resonances, determines the highest measurable frequency by using the particular geophone. A set of instructions for testing a geophone to ensure the reliability of the data collected during the experiment was also given [Hagedoom et al., 1988]. By following this, some of the main attributes of the sensor adopted for the present study were evaluated, and the outcome is included in Appendix A.

### 4.1.4 Data Acquisition

In terms of the strategy to meet the research hypothesis (Section 1.3), there may be two ways that can be taken in relation to data collection. One of which is by collecting data of the whole population that is every single vehicle that might ever turn up in a monitored area, and performing analysis on all of them with all the possible combinations of various methods available as well as every single parameter setting pattern that can be employed in order to obtain the optimum method. Of course, this approach allows one to obtain a very good understanding of the concerned signals characteristics, but only realistically achievable when the whole population of related data is relatively small. However, the population of interest for the present study are too large to adopt such tactics. Then, an alternative is to obtain data that can be utilised as representative of the whole population. When the latter approach is taken, the aim for a research such as this one is to develop an algorithm that has ability to classify samples as correct as possible when a new unseen set of data is presented. In machine learning, this ability of an algorithm is often referred to as "generalisation", and building a moving vehicle pattern recognition systems that generalises well may be seen as the current goal. Unfortunately though, it is not always easy to obtain such useful data at the first instance of a study, especially when there are so many factors that influence characteristics of data like the case of moving vehicle signals. Consequently, the following approach was employed as a starting point.

Firstly, based on knowledge obtained from the literature review, a relatively smaller set of primary data were collected under several different conditions as seen in Figure 4.1, in order to identify an appropriate data collection method, including sensor choice and layout.

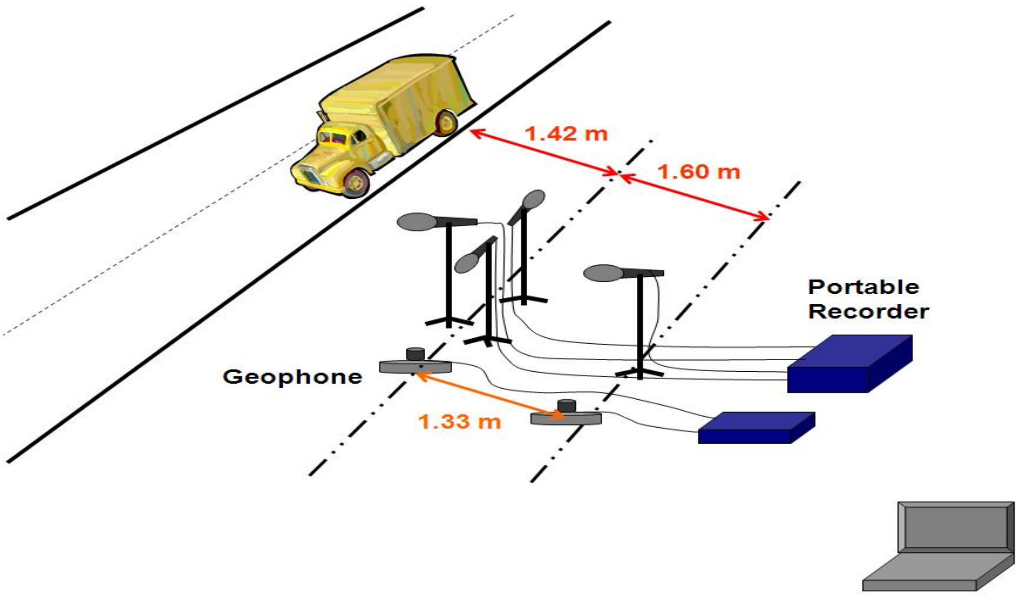


Figure 4.1: Data Collection Set-up for the Initial Phase

The below is the list of apparatus employed for the initial stage.

- Microphone: two types of condenser microphones, Rode NT5 (cardioid) and AKG CK93 (hypercardioid).

- Geophone: Sensor SM-4 vertical basic geophone unit.
- Recorder: Edirol portable 4-channel recorder and Marantz portable 2-channel recorder.

Secondly, by using these data, the typical characteristics of vehicle signals such as amplitude, waveform variation and dominant frequency components were studied by using various software packages including WaveLab, Cool Edit and MATLAB. Some of the feature extraction methods that were already selected for the investigation by then were utilised to make some judgements regarding the acquisition method. Thirdly another set of data were collected by using the established approach but under different conditions in terms of the distance between the sensor and vehicles as well as the locations. The newly collected signals were processed and analysed by the same approach to evaluate the data collection method. The second and third procedures above were performed repeatedly until some degree of discrepancy in the observed characteristics of moving vehicles belonging to different types were noticed, and these were more or less present in majority of the gathered samples.

Figure 4.2 depicts the set-up for collecting data that was eventually standardised to obtain both training and test samples throughout the project. The recording location and the associated meteorological data about each recording session carried out using this standard set-up can be found in Appendix A.

The apparatus for the standard setting is:

- Microphone: Rode NT5 condenser cardioid microphone.
- Geophone: Sensor SM-4 vertical basic geophone unit.
- Recorder: Edirol portable 4-channel recorder or Marantz portable 2-channel recorder.

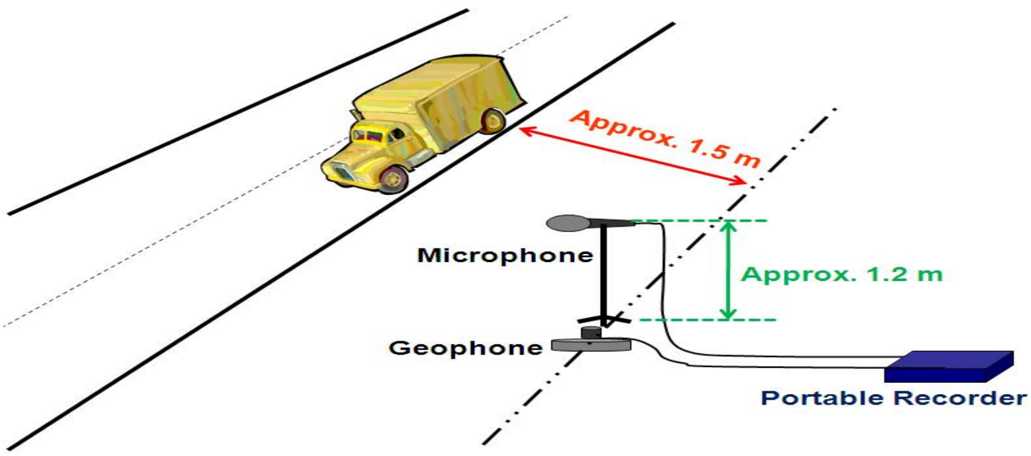


Figure 4.2: Standard Data Collection Set-up for the Study

Note, two other microphones and a portable recorder were also used during some sessions, and their details are in Appendix A. Throughout the project, all the data were digitised with a sampling frequency at 44.1kHz and quantised to 16 bits.

During the formation of the standard set-up, the following assumptions were made:

(a) The vehicles are travelling on a dry road with metalled surface.

(b) The vehicles are travelling at a speed approximately between 30km per hour and 100km per hour.

(c) Only one vehicle passes directly in front of the sensors at any one time, i.e. only single target recognition is considered.

(d) The SNR is at least 7dB for acoustic and 15dB for seismic signals.

## 4.2 Noise Reduction

The collected real-life input signals inevitably contain unwanted signals, including wind noise, noise generated by the contact between the receiver and the system, mechanical noise, sampling and quantisation noise. Keeping these unwanted signals could result in unsatisfactory classification performance, hence applying some kind of noise reduction algorithm to the digitised signals so as to improve recognition results is important [Atal and Rabiner, 1976].

To develop a pre-processing algorithm, first of all, the collected sample data were examined in the time domain as shown by the examples in Figure 4.3. This figure also illustrates how the signal characteristics, including duration, differ both between vehicle types and between sensor types. With regard to the assumption made earlier about SNR, it was noticed that sometimes seismic signals only last for 1 second although acoustic signals of cars tended to last longer (typically between 1 and 4 seconds). Both signals of trucks were typically longer than 2 seconds. It should be noted that multiple factors influence the duration, including the sensitivity of sensors and vehicle travelling speed. It was also observed that the amplitude of seismic signals of cars were smaller than those of trucks.

The collected data were then normalised in WaveLab version 4.0b then analysed with MATLAB. Waveform, spectrum and spectrogram analyses of 44100 samples are depicted in Figures between 4.4 and 4.7 for acoustic and seismic signals of a truck, a car, and background noise.

Note: due to the natural frequency and spurious frequency of the seismic sensor, as stated in its datasheet, frequency components of interest are those between 10 Hz and 180 Hz, whereas the frequency range for analysing acoustic signals was determined according to the sampling frequency of 44.1 kHz that should be enough to cover the range mentioned in Section 3.2.

It was found that generally the main components of the acoustic signals of vehicles are no higher than 16 kHz, and the dominant frequency components can be seen in under 8 kHz range. On the other hand, it was noticed that the frequency components of acoustic background noise were more explicit in the lower band, especially below several hundred Hz region. Although seismic signals were also collected with

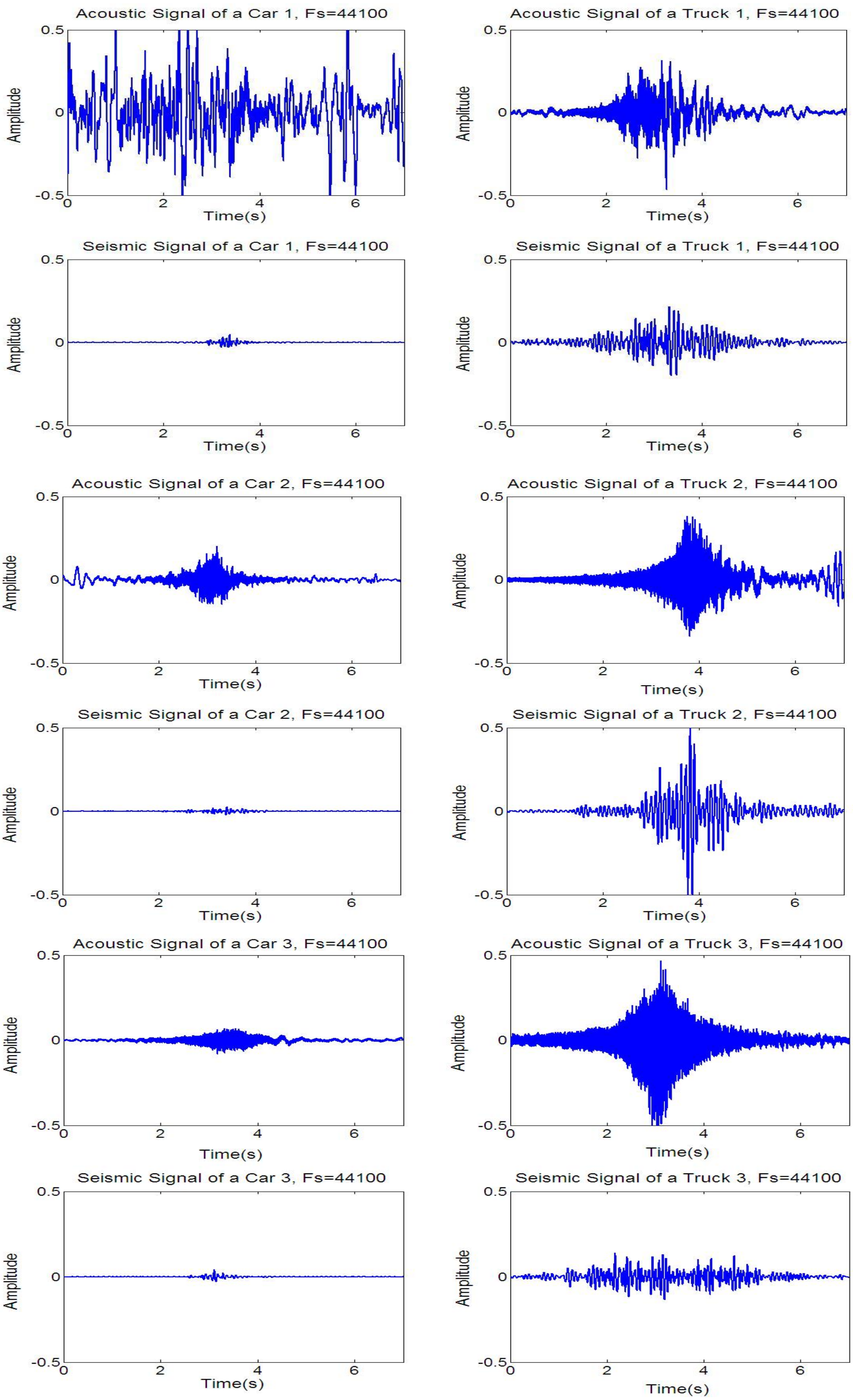


Figure 4.3: Waveforms of Acoustic and Seismic Signals of Cars (left column) and Trucks (right column) over 7 seconds, Fs=44.1kHz

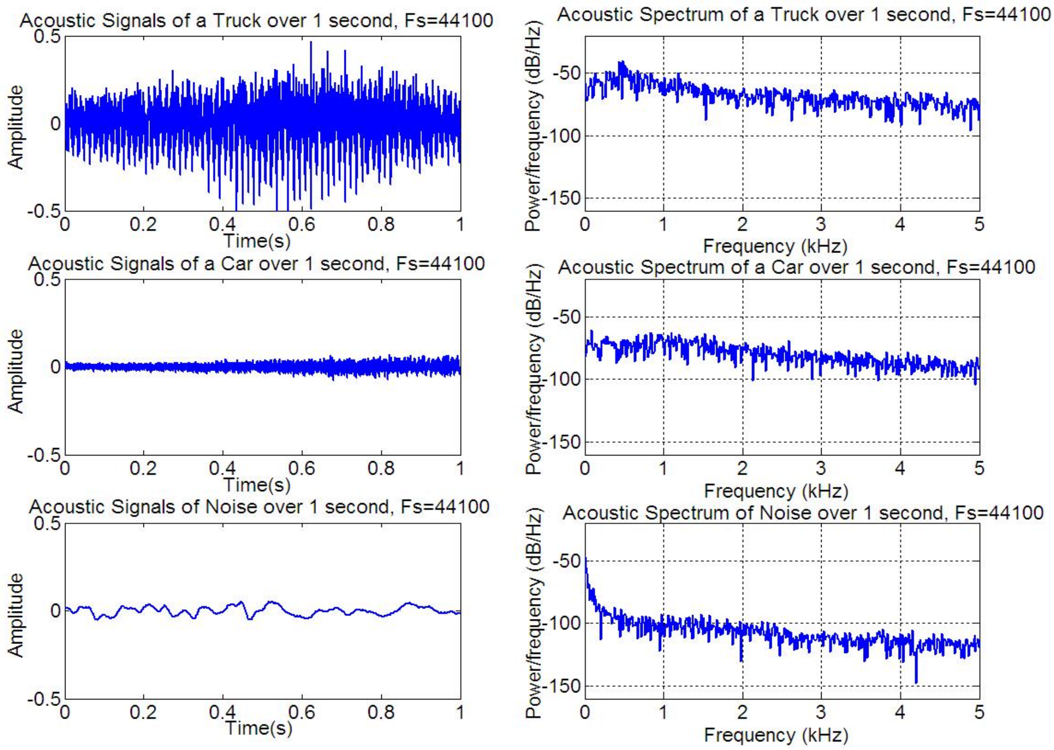


Figure 4.4: Waveform and Spectrum of Acoustic Signals (44100 samples, frequency range: 0Hz-5kHz, Fs=44.1kHz)

a sampling frequency at 44.1kHz, any components outside the sensor's capturing range are not within the interest as they do not necessarily represent the target signal attributes but that of the sensor; thus they should be removed at a relatively early stage of processing as unwanted signals. Keeping them may cause inefficiency or decline in overall performance.

For a closer look, the same samples as in Figure 4.4 and Figure 4.5, but either only for 4096 samples or over narrower frequency band is shown in Figures 4.8 and 4.9 respectively.

Interestingly, in those figures of truck signals appear to show existence of stationary components in the characteristics, which could be due to the repetitive engine spikes, modified by the exhaust with some resonance frequencies. Such quality could be useful in acoustic classification particularly if this is the case for all trucks.

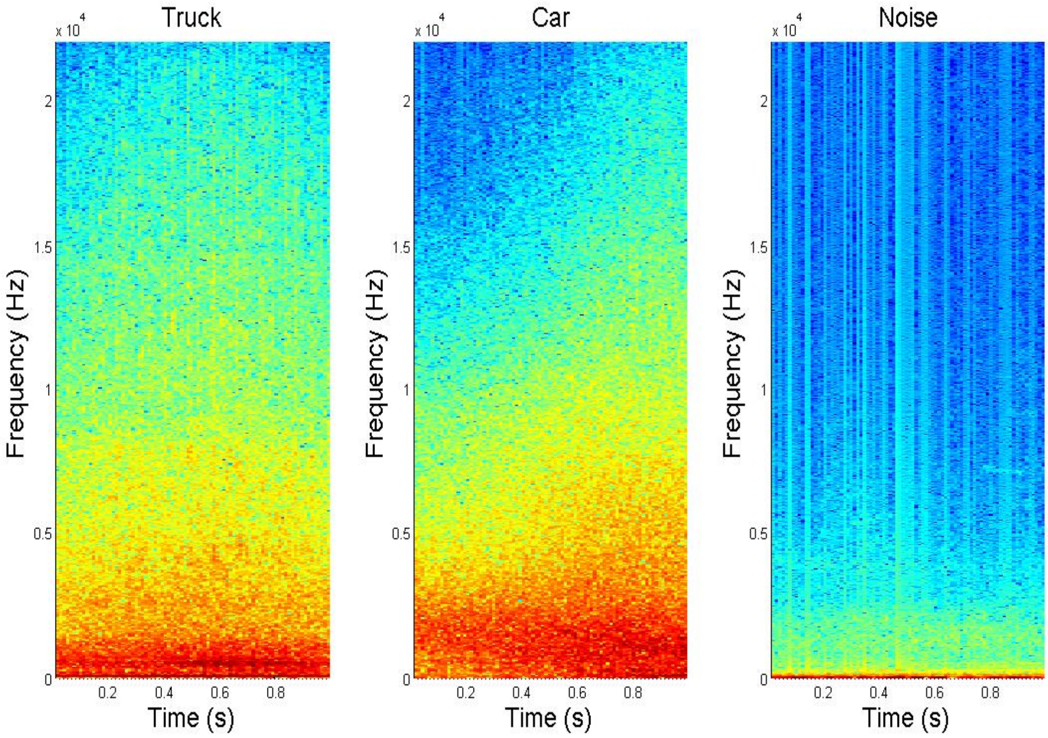


Figure 4.5: Spectrogram of Acoustic Signals (44100 samples, frequency range: 0Hz-22.05kHz, Fs=44.1kHz)

However, it was found that this could depend on the samples as shown in Figure 4.10. Similarly, although they are less apparent than those of the trucks, some car samples also show stationary elements in the low frequency region (Figure 4.11). Effects of Doppler shift can also be observed in some of these figures.

### 4.2.1 FIR Filtering

Digital filters can be categorised into two types; Infinite Impulse Response (IIR) and Finite Impulse Response (FIR) filter. There are known advantages and disadvantages in both types of filters. Due to the following advantages, FIR filter was chosen although it requires more memory and processing speed to obtain a filter with narrower transition width [Ifeachor and Jervis, 2002].

- A linear phase response is achievable.
- Filter performance is stable.

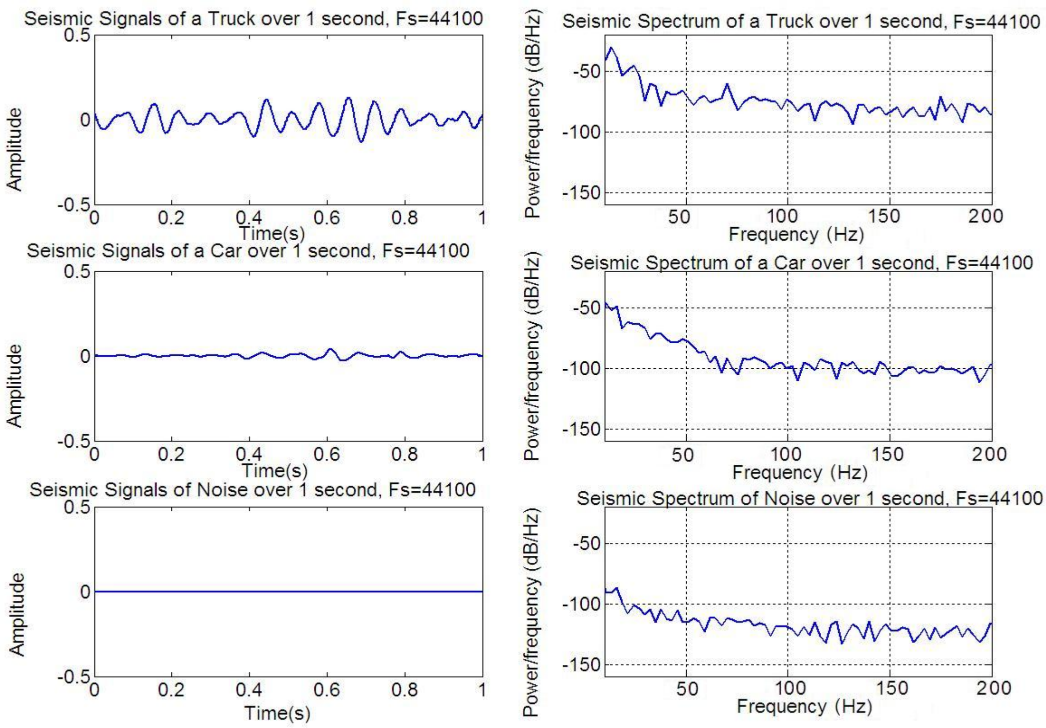


Figure 4.6: Waveform and Spectrum of Seismic Signals (44100 samples, frequency range:10Hz-200Hz, Fs=44.1kHz)

- Lower round off noise or coefficient quantisation noise is generated.

The basic characteristics of FIR filter response can be described by the following two equations.

$$y[n] = \sum_{k=0}^{N-1} h[k]x[n-k]; \tag{4.1}$$

$$y[n] = \sum_{k=0}^{N-1} h[k]z^{-k}; \tag{4.2}$$

where: $y[n]$ is the filtered output signal in time domain,

$x[n]$ is the input signal in time domain

$h[k], (k = 0, 1, 2, ..., N)$ is the impulse response coefficient

$N$ is the filter length or sometimes called number of taps

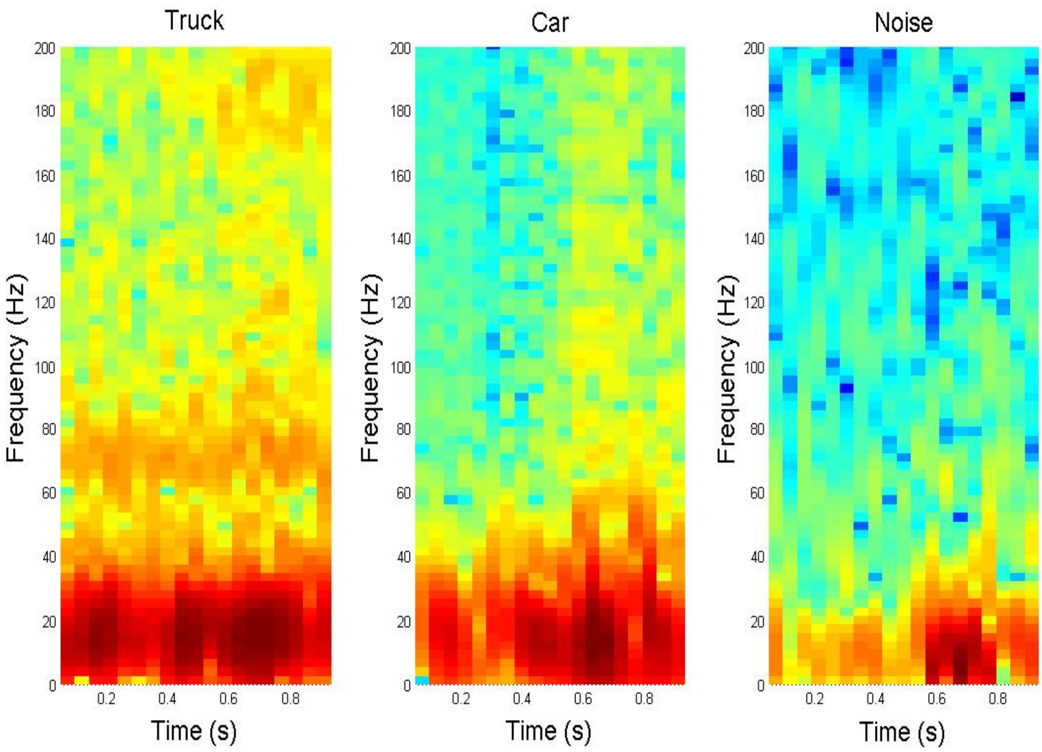


Figure 4.7: Spectrogram of Seismic Signals (44100 samples, frequency range: 0Hz-200Hz, Fs=44.1kHz)

Equation 4.1 is the difference equation and Equation 4.2 is the transfer function, which is in the Z domain [Proakis and Manolakis, 1996; Oppenheim et al., 1999; Ifeachor and Jervis, 2002].

For FIR filter design, there are several main design methods to be chosen from: window method, frequency sampling method [Rabiner and Schafer, 1971], and an optimum method that is known as "Equi-ripple" method or "Parks-McClellan" algorithm [Bateman and Paterson-Stephens, 2002, pp.348-352,604-610]. For simplicity in terms of the algorithm and computation [Oppenheim et al., 1999; Ifeachor and Jervis, 2002], the window method is employed.

The impulse response of an ideal filter that has a "brick wall" like roll-off can result in infinite number of coefficients. Nonetheless, a digital FIR filter can only manage implementation of a truncated version since it is not capable of reproducing an

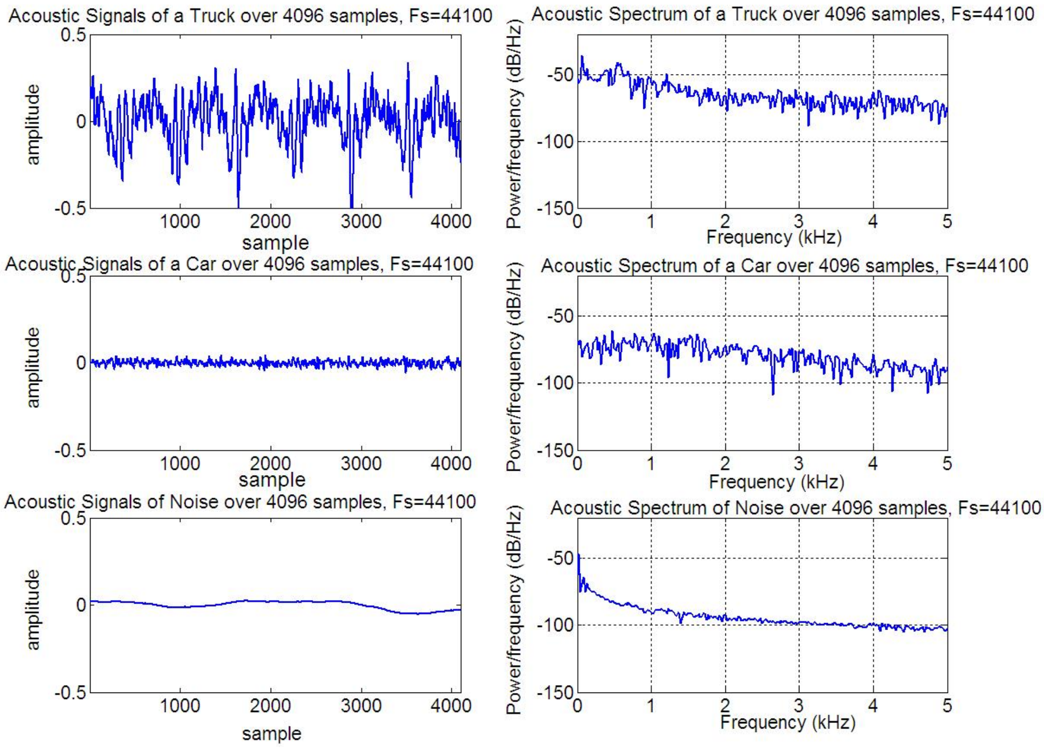


Figure 4.8: Waveform and Spectrum of Acoustic Signals as in Figure 4.4 (4096 samples (approx. 93ms), frequency range:0Hz-5kHz, Fs=44.1kHz)

infinite length impulse response. The use of a windowing function supports this truncation of impulse response. As illustrated in both Equation 4.3 and Figure 4.12, the actual FIR coefficients can be calculated by multiplying the ideal impulse response by the window function when this window method is applied.

$$h[n] = h_D[n] * w[n] \tag{4.3}$$

Different window functions bring unique effects on the resultant impulse response [Bateman and Paterson-Stephens, 2002]. For example a direct truncation, which is equivalent to using a rectangular window, would lead to unwanted ripples and overshoot, the so called Gibbs phenomenon, especially with fewer coefficients [Ifeachor and Jervis, 2002]. Therefore selecting a suitable window function is important for designing an efficient FIR filter.

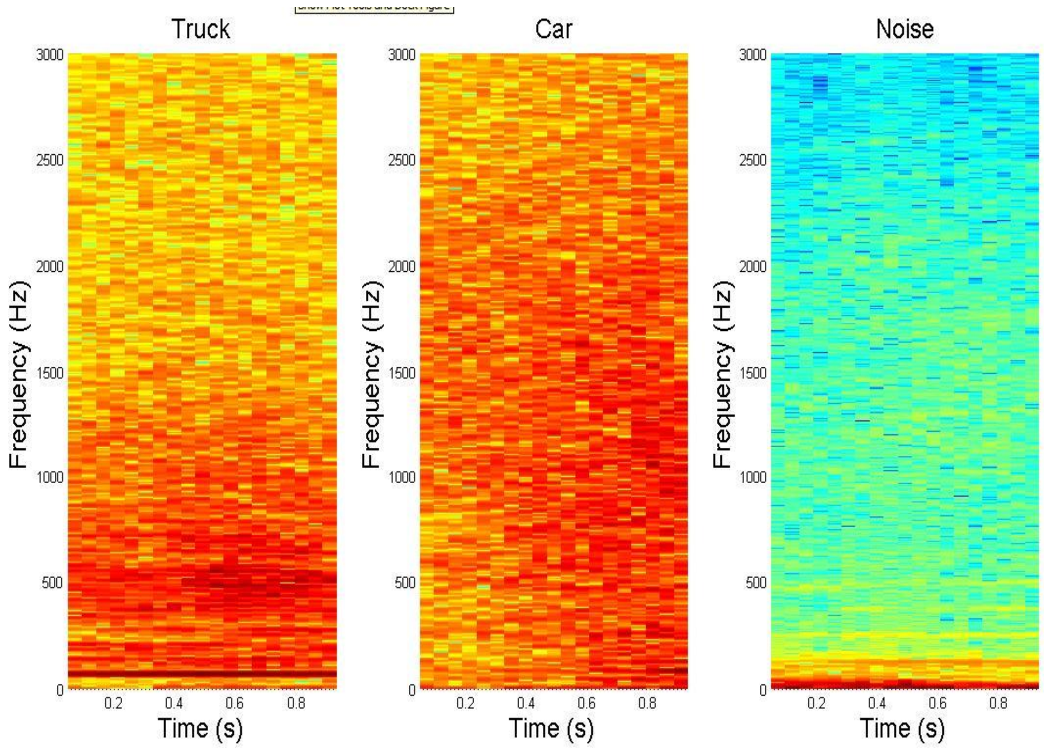


Figure 4.9: Spectrogram of Acoustic Signals as in Figure 4.5 (44100 samples, frequency range: 0Hz-2kHz, Fs=44.1kHz)

Mainly due to their computation and stopband ripple, Blackman window and Hamming window functions were selected for the current study. After examining the real-life data of moving vehicles in the frequency domain, the cut-off frequency and transition width ($Df$) were also determined to calculate each set of coefficients. The narrower the transition width, the more coefficients required; their relationship is inverse proportion. Thus it was realised FIR is unsuitable by itself for seismic signals as the signals of interest are in such a narrow band. Therefore a simple application of FIR has been considered only for acoustic signals of moving vehicles. A comparative study of the effects due to the window function characteristics as well as the data frame size was conducted. Figures 4.13 to 4.16 depict the differences observed in the filter response according to the window function and transition width when bandpass filters with cut-off frequencies set at 180 Hz and 8 kHz. Table 4.2, on the other hand, shows the number of filter coefficients required for each setting,

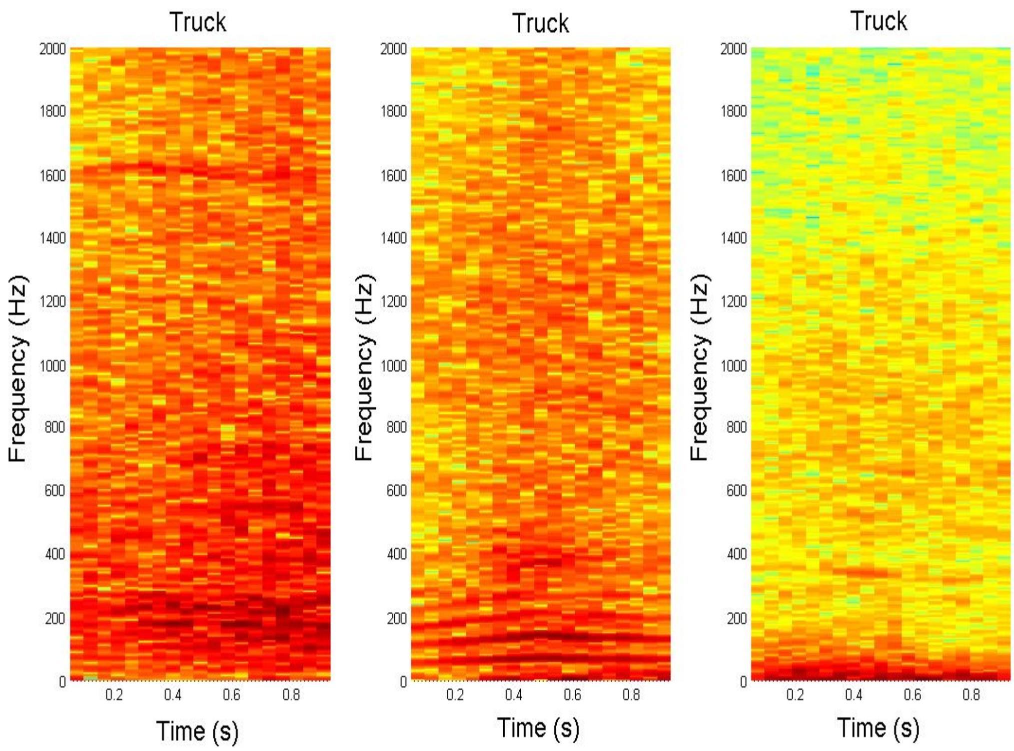


Figure 4.10: Spectrogram of Acoustic Signals of Three More Trucks (44100 samples, frequency range: 0Hz-2kHz, Fs=44.1kHz)

all of which were designed with sampling frequency ($Fs$) at 44.1 kHz.

| Transition Width | Hamming Window | Blackman Window |
|---|---|---|
| 500 Hz | 291 | 485 |
| 150 Hz | 970 | 1617 |

Table 4.2: Number of Filter Coefficients Required

The trade-off between computation and effectiveness of a filter should be considered, especially for practical use of the algorithm. Nonetheless, during the search for the optimum feature extraction and classification methods, the signals have been processed with the filter designed by using Blackman window with transition width of 150 Hz, since its ability to suppress frequency components outside the range of interest while maintain the others is seen as the greatest. Figure 4.17 shows an example of how the filter performs on the actual collected signals of a vehicle, examined both in the time and the frequency domain.

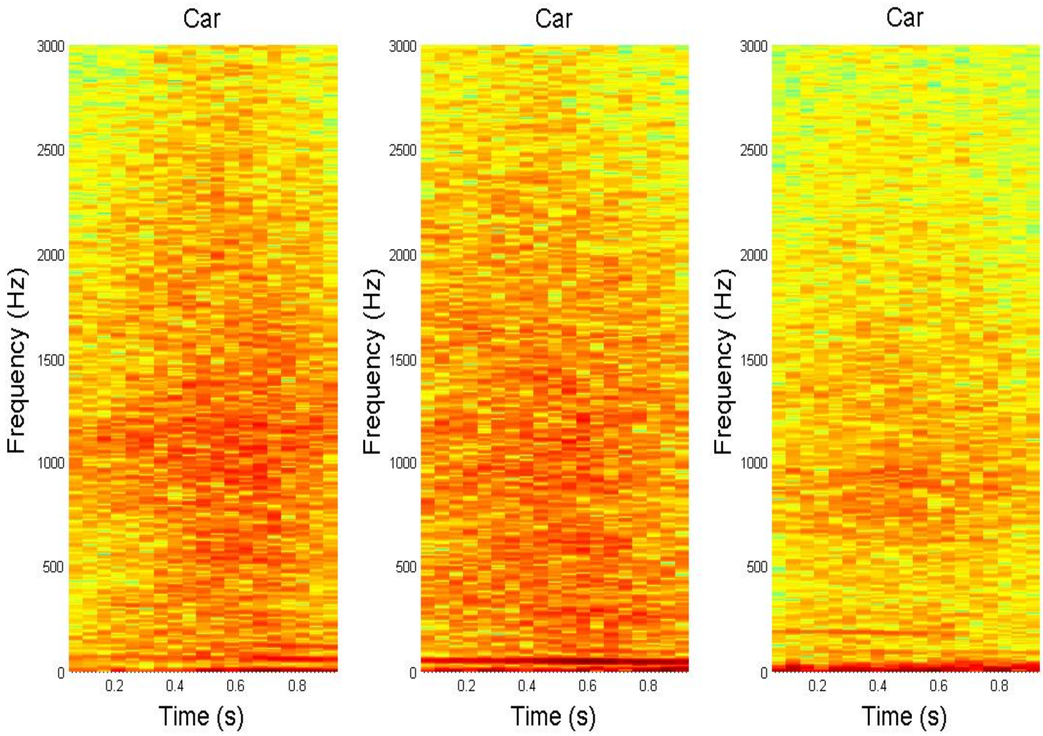


Figure 4.11: Spectrogram of Acoustic Signals of Three More Cars (44100 samples, frequency range: 0Hz-3kHz, Fs=44.1kHz)

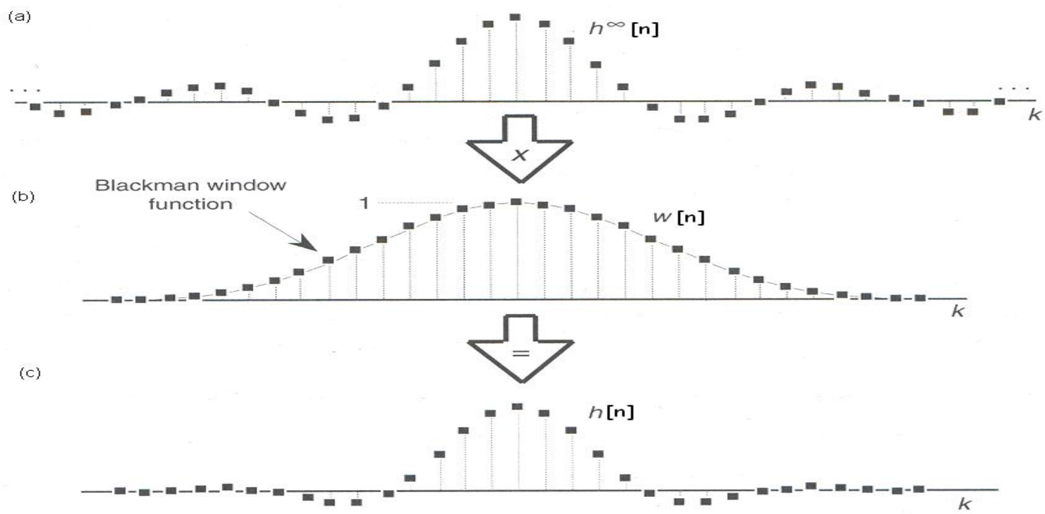


Figure 4.12: Impulse Response of a FIR Filter Designed with Window Methods: (a) Ideal Impulse Response, (b) Blackman Window Function, (c) Actual Impulse Response.(Obtained from Lyons [Lyons, 1997, p.185].)

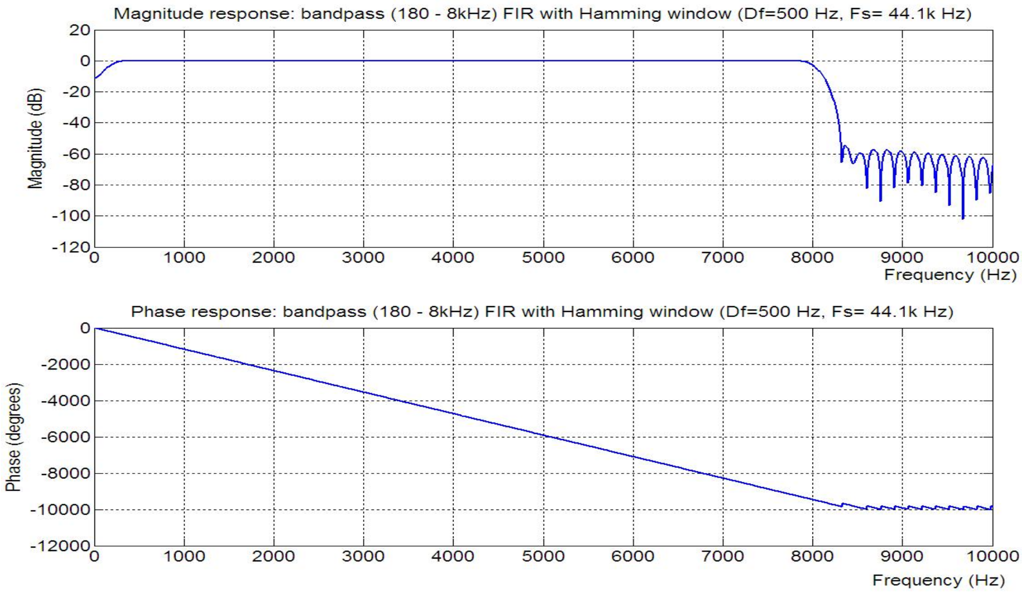


Figure 4.13: Bandpass Filter Response (designed with Hamming window, Df=500Hz, Fs=44.1kHz)

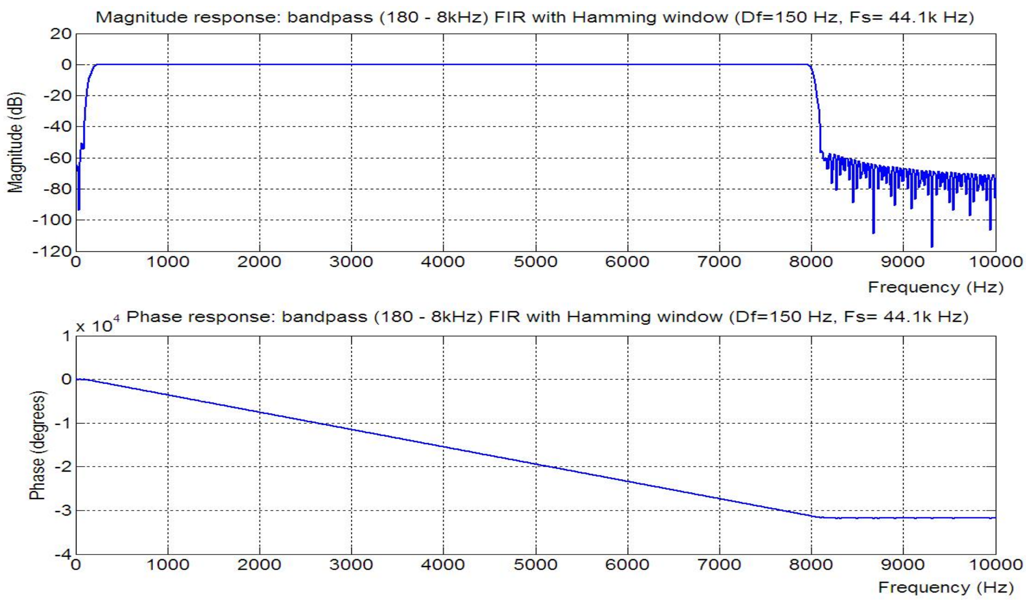


Figure 4.14: Bandpass Filter Response (designed with Hamming window, Df=150Hz, Fs=44.1kHz)

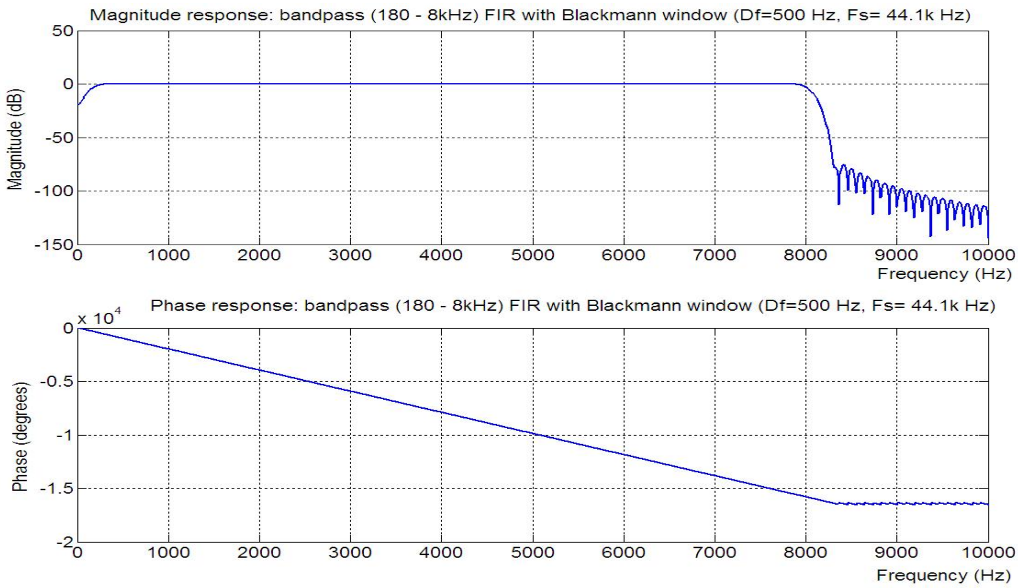


Figure 4.15: Bandpass Filter Response (designed with Blackman window, Df=500Hz, Fs=44.1kHz)

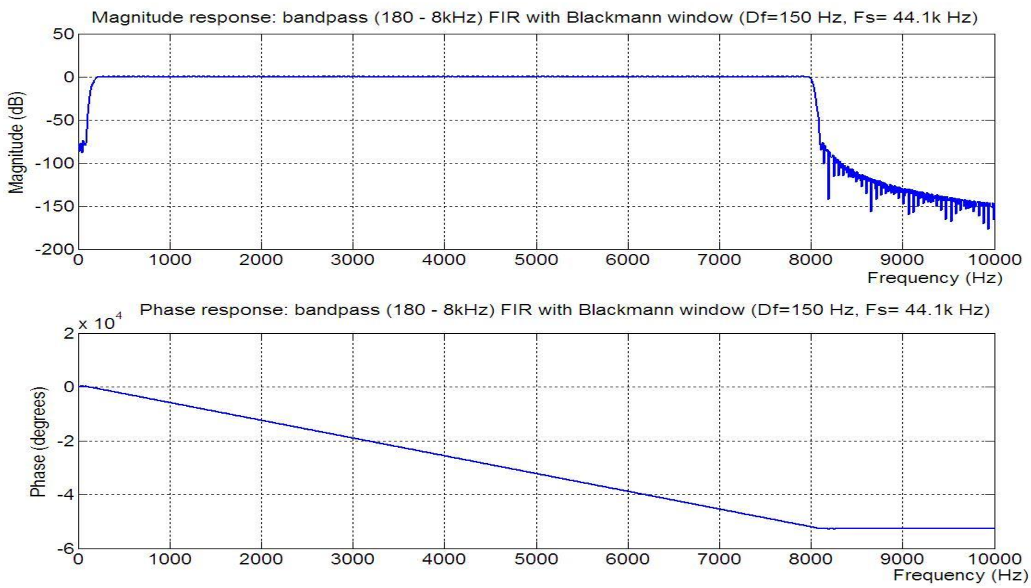


Figure 4.16: Bandpass Filter Response (designed with Blackman window, Df=150Hz, Fs=44.1kHz)

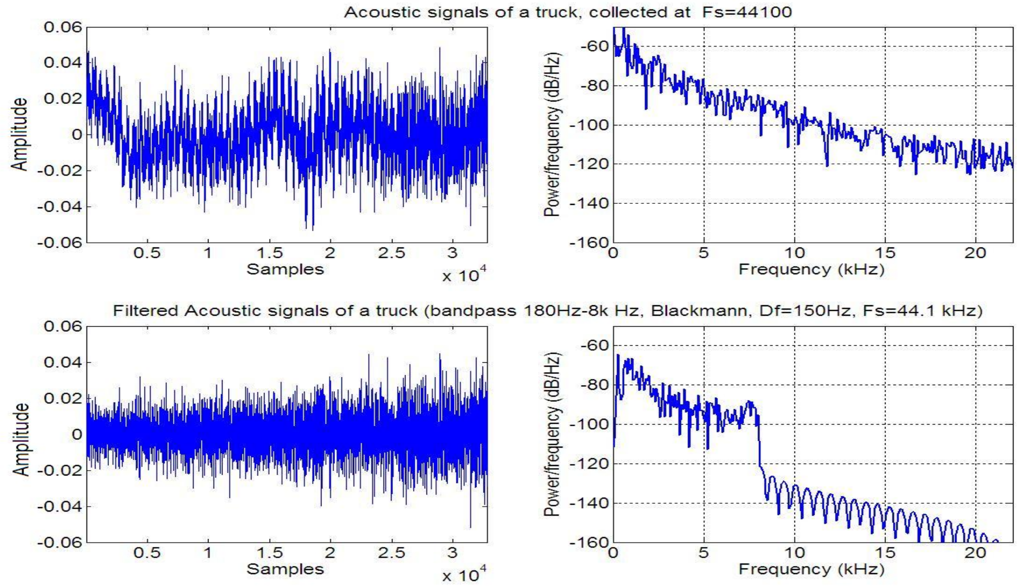


Figure 4.17: Frequency Response of Filtered Acoustic Signal of a Moving Truck (Filter designed with Blackman window, Df=150Hz, Fs=44.1kHz)

For seismic signal processing, a FIR filter can be applied so as to suppress the frequency components above the spurious frequency of the sensor, which is 180Hz.

However, aiming to reduce those below the sensor's natural frequency at 10 Hz by only a FIR filter may not be realistic with the current sampling rate. Although it can be designed in MATLAB as shown in Figure 4.19, it requires 24255 coefficients if the transition width was set to 10Hz for example, hence the computation becomes more expensive and processing slows down.

### 4.2.2 Spectral Subtraction of Estimated Noise

As a simple method to reduce the amount of noise mixed in signals, "spectral subtraction" was introduced to speech recognition study by Boll [Boll, 1979]. It is achieved by subtracting the estimated power spectrum of noise from that of the input signal. In the original method, the spectral power of noise is estimated by heuristically calculating the average spectrum over some number of frames which

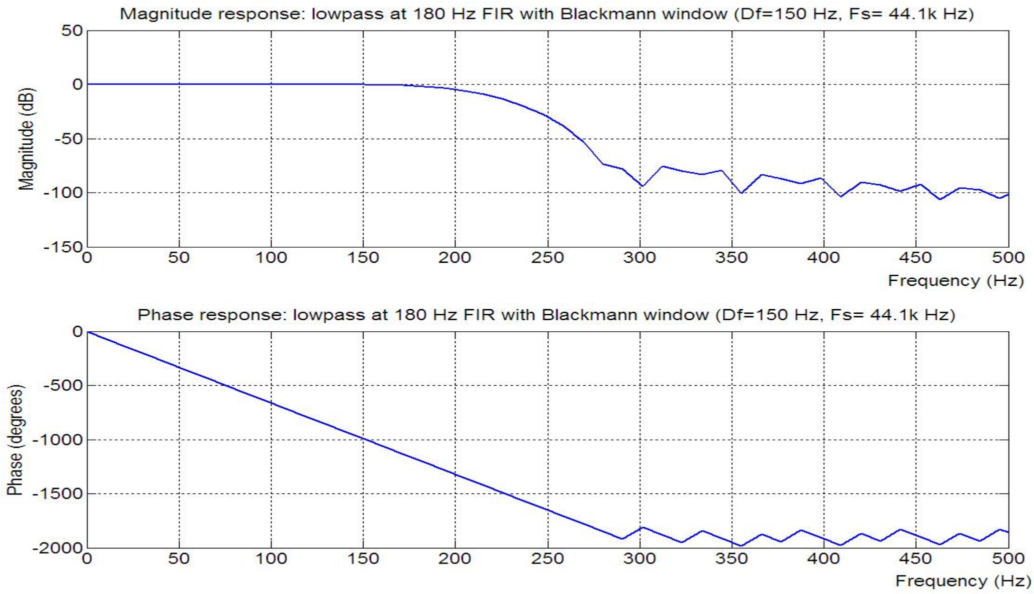


Figure 4.18: Lowpass Filter Response (designed with Blackman window, Df=150Hz, Fs=44.1kHz)

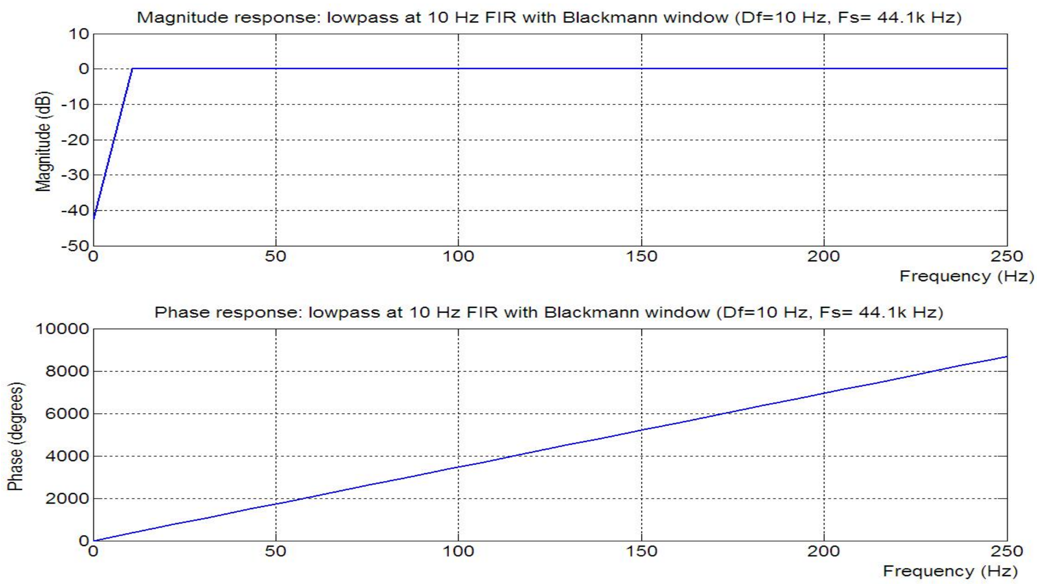


Figure 4.19: Lowpass Filter Response (designed with Blackman window, Df=10Hz, Fs=44.1kHz)

are known to contain only noise. The phases of the instantaneous input signals are maintained and added to the subtraction results. In order to avoid having negative values in the frequency range, where the magnitude of estimated noise spectrum is greater than that of the instantaneous input signal, the resultant spectral components are replaced by zeros for these specific frequency bands. Other researchers also have suggested variations of the algorithms often in relation to speech intelligibility. For example, it was pointed out that the above original algorithm sometimes results in rapid change between frequency bands of spectrum causing another kind of noise, often named "musical noise", which can interrupt the intelligibility of the processed signals. Therefore, some algorithms include additional procedures to reduce the effect of musical noise [Berouti et al., 1979].

For the current research, however, the intelligibility of the signals is less important. Instead, the ultimate purpose of employing any noise reduction algorithm is to improve either the overall recognition performance or computation efficiency. Therefore, if spectral subtraction is to be applied, the original algorithm may be sufficient depending on the outcomes of empirical studies using the collected data. During the experiment, it was assumed that the signals of interest are much greater level than that of noise, due to the assumption made earlier in Section 4.1.4. Since it is important to be able to reconstruct the signals once the subtraction is performed in the frequency domain, the estimated noise spectrum is found as the inter-quartile mean of standard DFT over some numbers of frames that are known to contain noise, rather than using the spectral power. This estimated noise is simply subtracted form the standard DFT of the input signal frame by frame, with 50% overlap between the consecutive frames to compensate for any discontinuities caused by the segmentation. Then the time domain signals are reconstructed from the resultant DFT. Note, both the amplitude and the phase of signals and the noise are treated together during the whole process. Figure 4.20 depicts the spectrum analysis of the original input signal of a truck passing by a microphone (blue, top line), the resultant signal after subtracting the estimated noise (red, middle), and estimated noise (black, bottom). It shows how the frequency components corresponding to the range, where the dominating part of the estimated noise appear, were attenuated after subtraction.

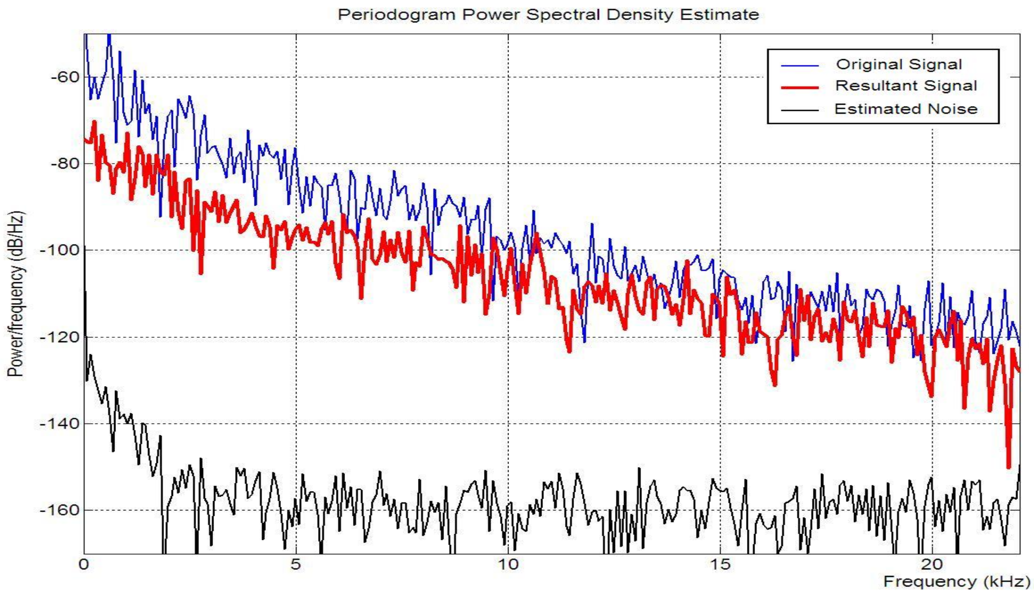


Figure 4.20: Effect of Spectral Subtraction on Acoustic Signal of a Truck

This outcome was compared with the windowed original signals, as shown in Figure 4.21; whereas the difference between the reconstruction (the subtraction results added to the estimated noise) and the original for the case is in Figure 4.22. As the scale of the y-axis of both graphs in Figure 4.22 indicates, the difference is fairly small. By observation of the graphs, it was decided to accept the algorithm as it was for the rest of the project as the reconstruction was satisfactory in a subjective judgement.

### 4.2.3 Wavelet Packet (WP) Noise Reduction

Wavelet transform is a powerful time-frequency signal analysis tool employed in many practical applications, particularly in the last two decades [Bentley and Mc-

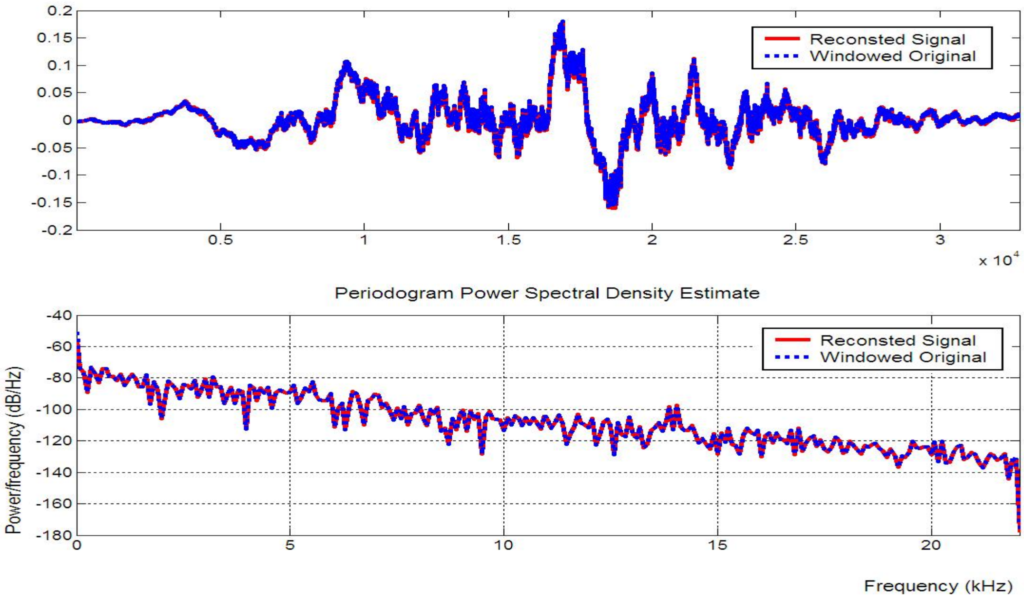


Figure 4.21: Effect of Spectral Subtraction on Acoustic Signal of a Truck

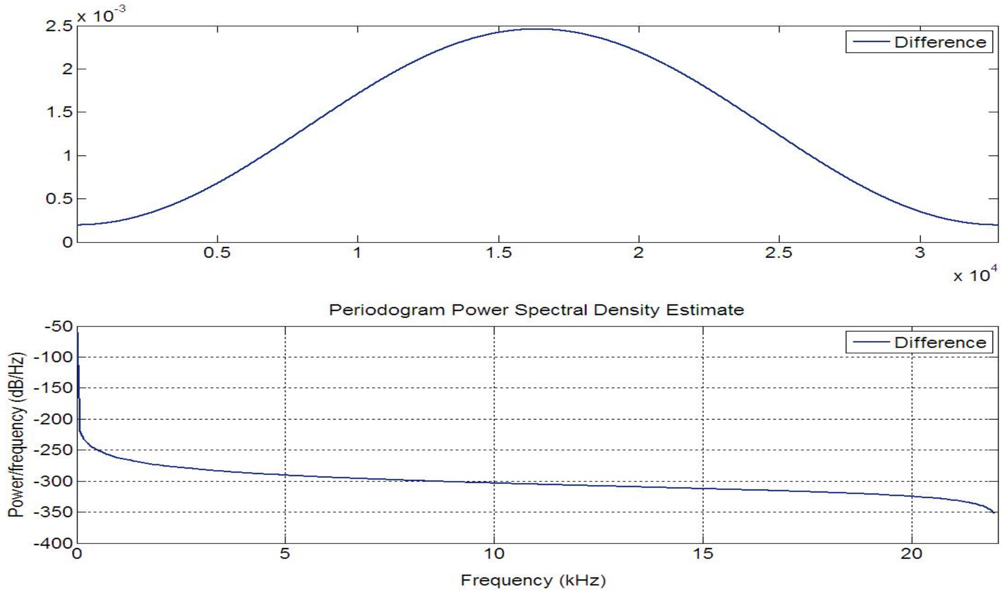


Figure 4.22: Difference Between the Original and Reconstructed Signal After Noise Subtraction

Donnell, 1994]. Compared to STFT, wavelet allows signal analysis localised in both time and frequency domains. The use of wavelets for reducing noise can be traced back to Grossmann and Morlet [Grossmann and Morlet, 1984], proving firstly the wavelet transform can be invertible as long as certain conditions are met and secondly that small variation made on the wavelet coefficients results in mathematically related variation of the signals reconstructed by the inverse transform; hence they have isomorphic property [Ramchandran and Vetterli, 1993]. Wavelet transform is discussed in more detail in Section 5.3.2.

WP analysis utilises a Quadrature Mirror Filter (QMF), which is a pair of lowpass and highpass filters constructed with “the same set of wavelet filter, but with altering signs and in reversed order” [Walczak and Massart, 1997, p.82]. Combined with down sampling, the signals are decomposed into high and low frequency components at each step. It is known to be effective in segmenting high frequency transient signals and low frequency more continuous ones. [Walczak and Massart, 1997] gives a comprehensive explanation of WP decomposition and reconstruction, particularly regarding noise reduction.

Noise reduction with WP has been studied in many applications including speech enhancement [Ghanbari and Karami-Mollaei, 2006]. In addition, it has been applied to vehicle signal enhancement by [Lopez et al., 1999], who provided a comprehensive explanation of the mechanism involved in both decomposition and reconstruction, together with a comparative list of different wavelet that may be suitable for the purpose. The general rules to be followed in applying such methods are that when the signals are decomposed by WP, the parts of interest are seen in only some of the WP components having much higher coefficients. Thus, either eliminating (hard thresholding) or reducing the magnitude of coefficients according to a pre-defined rule (soft thresholding) those below the threshold results in removal of insignificant wavelet components in general term. For example, [Gurley and Kareem, 1999] provided a simple demonstration for both hard and soft thresholding.

It can be achieved by using WP decomposition of signal with using orthogonal or biorthogonal wavelet, and the hard thresholding procedures are summarised below.

(a) Perform wavelet decomposition of input signal using a QMF.

(b) Eliminate minor coefficients below the threshold by setting them to zeros.

(c) reconstruct the signal from the only maintained coefficients by using conjugates of QMF.

Finding the optimum threshold, thresholding method and indeed the wavelet for the application can be an exhausting task. For the current study, Donoho's [Donoho, 1995] "universal threshold" (Equation 4.4), soft thresholding (Equation 4.5), with "db4" wavelet as well as "sln" rescaling in MATLAB were selected after some empirical study. The de-noising processing was performed with the MATLAB wavelet toolbox. Both equations were taken from [p.88][Walczak and Massart, 1997].

$$thr = \sqrt{2\log_2(N * \log_2(N)} \tag{4.4}$$

where, $N$ is the signal length.

$$\alpha_j = \begin{cases} 0 & if \mid \alpha_j \mid < thr \\ sign(\alpha_j)(\mid \alpha_j \mid - thr) & if \mid \alpha_j \mid \geq thr \end{cases} \tag{4.5}$$

where, $\alpha_j$ are the soft thresholding coefficients.

Needless to say, the results were widely influenced by the parameter settings. Figure 4.23 shows an example result obtained for a truck signal, in which some components were reduced from the signal after processing. Figure 4.24 is the same signal but segmented by Blackman window function instead of Hamming window.

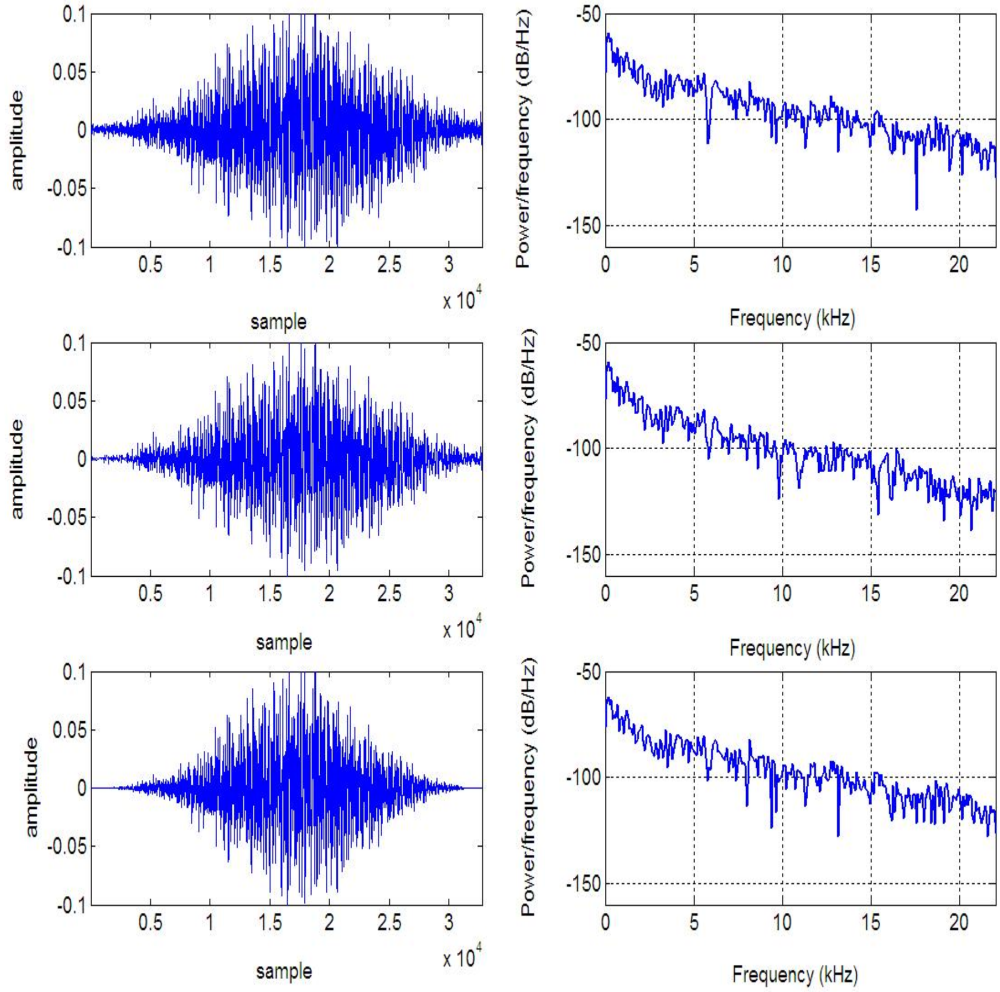


Figure 4.23: Example of WP Noise Reduction with Acoustic Signals of a Truck (top: original signal framed by a Hamming window, middle: WP noise reduction result, bottom: the difference. Frame Size=32768, Fs=44.1 kHz.)

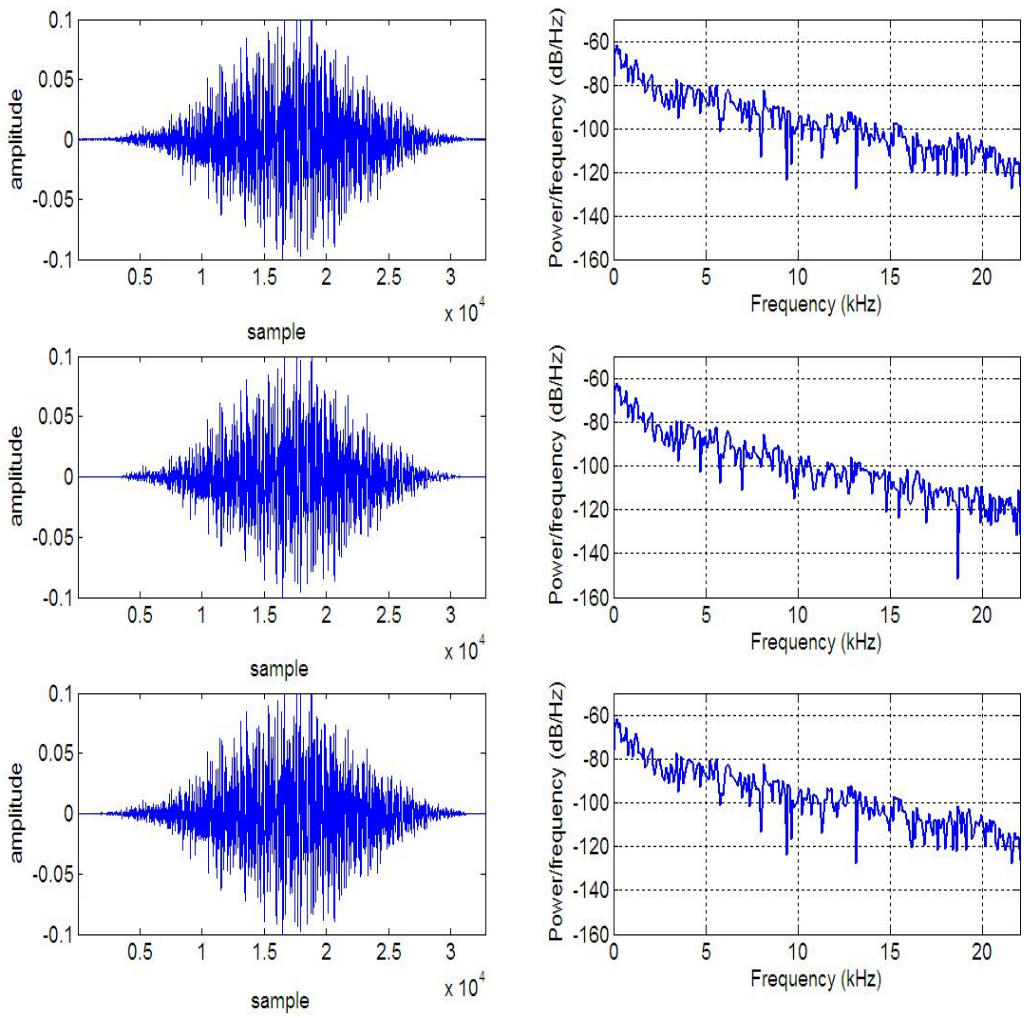


Figure 4.24: Example of WP Noise Reduction with Acoustic Signals of a Truck (top: original signal framed by a Blackman window, middle: WP noise reduction result, bottom: the difference. Frame Size=32768, Fs=44.1 kHz.)

The effect of WP de-noising is perhaps too subtle to be evaluated by itself. Unfortunately, for pattern recognition study, the effect of noise reduction is not easily assessed without continuing further processing. The evaluation will be explained in Chapter 7.

## 4.3 Chapter Summary

The data collection method adopted for the current study has been introduced and some findings regarding examined signal pre-processing methods have been reported. Chapter 7 reports some of the findings of a comparative analysis of three noise reduction methods explained in this chapter; FIR Filtering, Spectral Subtraction and Wavelet Packet (Section 4.2). The purpose of this pre-processing is either to remove unwanted signals or to enhance the components of interest, or indeed both so as to improve the resultant performance accuracy plus computational cost and efficiency. In order for the research to be moved forward, some of the decisions for these parts needed to be made at a relatively early phase. However, the influence of these early decisions had to be re-considered repeatedly throughout the research although there was some limit such as time and resource in terms of what can be achieved. Some of that is reported in the rest of the thesis. Yet, it is also realised there are much more can be and need to be done about these aspects of pattern recognition system development. Perhaps, fully analysing the overall effects of each data collection or pre-processing method can become an exhaustive but very interesting research in itself.

# Chapter 5

# Feature Extraction

As described in Section 1.5, the purpose of including feature extraction algorithms in a pattern recognition systems is to gather a set of information that efficiently and optimally represents significant characteristics from the input data. A good feature extraction algorithm can reduce the overall data amount while retaining or enhancing signal characteristics inherent to each class in order to improve classification performance [Fukunaga, 1990]. Therefore, for a successful system, it is essential to design an optimum algorithm, often unique to the requirements of each application, so that a specific target object can be detected and identified accordingly. It is a challenging but crucial task for this research to empirically understand which features are more suitable than others in order to develop an accurate vehicle recognition system.

There are a wide range of techniques that can be applied to the current study. Nonetheless, this chapter explains a selection of various feature extraction techniques recommended by other researchers, particularly with regard to their feasibility for the current acoustic and seismic vehicle recognition study. An ideal algorithm has a quality that allows feature vectors of one class to converge well while the distances between the vectors of different classes remain large in the feature space. Additionally, it is suggested that large dimension data could result in poor generalisation [Hand, 1997]. If the algorithm can also perform the task with low cost and inexpensive computation, it can make it an excellent choice of feature extraction for the study. Hence, the analysis regarding various methods have been conducted from such a view point, and the outcome is discussed in Chapter 7.

Feature extraction techniques are often grouped into three categories in terms of the focused signal processing domain; such as the time, the frequency, and the time-frequency domains [Wang and Qi, 2002]. The rest of this chapter is arranged to describe studies of the selected feature extraction techniques in these domains. However, it should be noted that in processing of real-life vehicle signals collected for the research, features extracted by using a method categorised in the time domain here can maintain some frequency information, and vice versa.

## 5.1 Time Domain Techniques

Time domain signal processing techniques analyse input signals according to their waveforms, observed over a certain period of time; hence the focus is on the amplitude variation during a selected time period. Such techniques have been applied in various field including speech processing [Benesty et al., 2008]. As discussed in Chapter 2, some vehicle classification studies [Mazarakis and Avaritsiotis, 2007; Sobreira-Seoane et al., 2008] indicated the positive potential of some of time domain techniques. Nevertheless, it should be noted that there are some suggestions [Wang and Qi, 2002; Ding et al., 2006] that acoustic signal processing in the time domain may not be appropriate due to their susceptibility to environmental noise, meteorological noise, and Doppler effect. This section lists the time domain signal processing techniques, of which the author selected and studied for the current research, based on the findings of the literature survey.

### 5.1.1 Envelope Analysis

Envelope analysis allows capturing the time domain amplitude variation over a period of time; hence, it is occasionally called “waveform detection”. It can be achieved by applying a lowpass filter with pre-defined cut off frequency [Bateman and Paterson-Stephens, 2002, p.317-319]. Envelope analysis can be useful at an early stage of algorithm development to grasp the approximate time domain attribute, rather than as a feature extraction method to be included in the final algorithm.

Some examples of envelope analysis are shown in Figure 5.1, which depicts the characteristic difference in acoustic signals of vehicles of different types. The graphs show the fairly repetitive but not exactly stationary quality of vehicle signals. Envelope analysis is a simple method whose observation of the outcome is done in the time domain; however, it can also illustrate some information regarding the frequency components. This truck signal seems to contain more low frequency components compared with that of the car. Although it is far from a detailed analysis, this may portray how some time domain techniques can also offer some useful facts about the frequency components without adding too much computational demand. This may encourage the use of a time domain feature extraction technique for the present study.

### 5.1.2 Zero-Crossing Rate (ZCR)

The techniques that count the number of the point at which the signal amplitude polarity alters during a pre-fixed length of time are usually referred to as “Zero-Crossing Count”. Usually a slightly more sophisticated variation ZCR or “Short-Time Average Zero-Crossing Rate” [Rabiner and Schafer, 1978, p.127], which is found by normalising the Zero-Crossing Count with the signal length along the time line, is employed in recognition studies [Sobreira-Seoane et al., 2008], because of

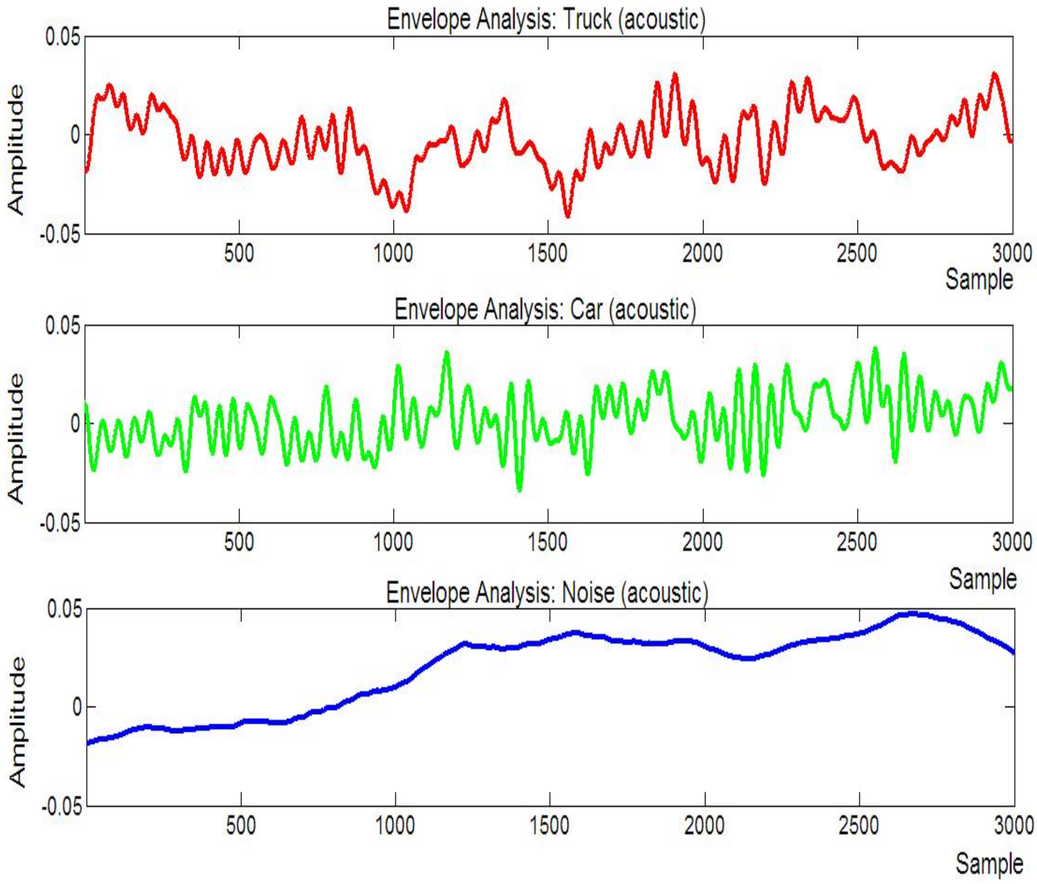


Figure 5.1: Example of Acoustic Envelope Analysis of a Truck (top), a Car (middle) and Noise (bottom), Frame Size=32768, Fs=44.1kHz, Processed with a FIR Lowpass Filter with cut-off Frequency at 1 kHz, Df=400Hz.

the capability to adaptively examine signals of different lengths. This can be used as a compact indicator of the frequency components, particularly in case of treating signals with a narrow bandwidth [Rabiner and Schafer, 1978].

Figure 5.2 is examples of plotted graph showing ZCR collected from acoustic signals of just under 1 second long. Each ZCR value was calculated per a signal frame of 1024 sample long, segmented with 50% overlap.

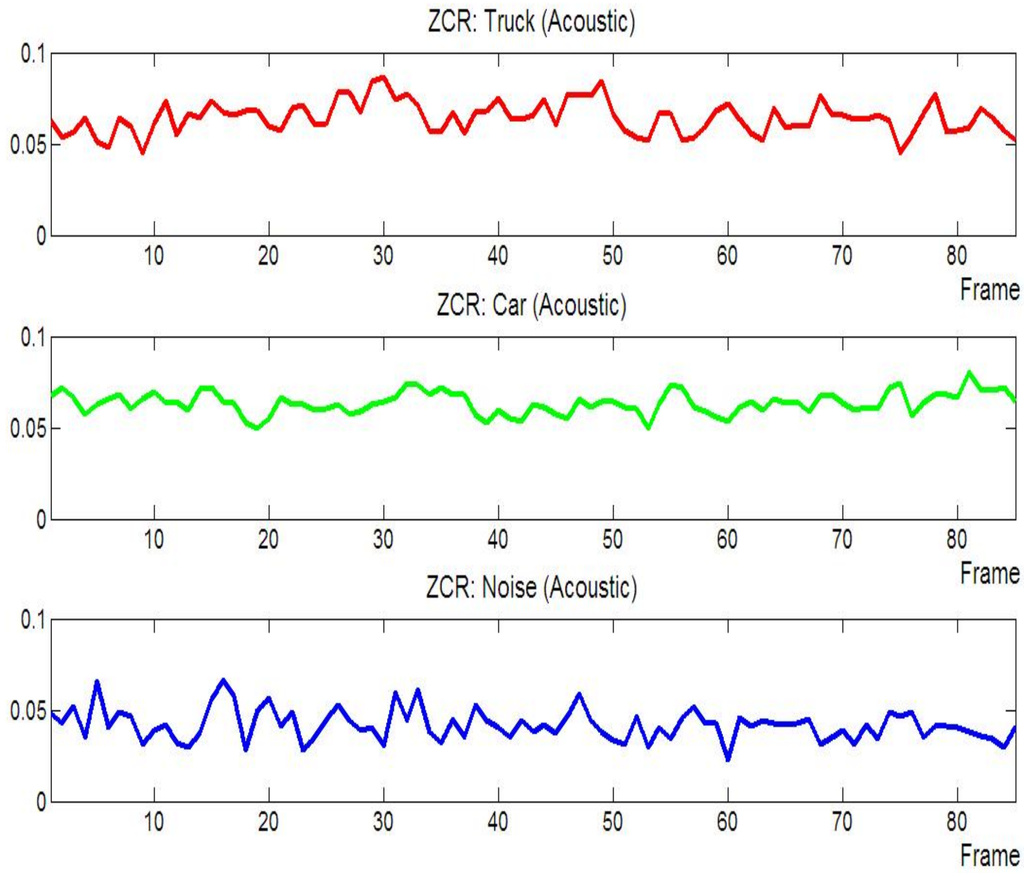


Figure 5.2: Example of Acoustic ZCR of a Truck (top), a Car (middle) and Noise (bottom), Frame Size=1024 with 50% Overlap, Fs=44.1kHz.

### 5.1.3 Energy / Log Energy

When the amplitude at a sampling instant $n$ is described as $x[n]$, the energy of that signal can be found from the squared value of the amplitude, i.e. $x^2[n]$. In practice the summed energy value of signals of a fixed finite length frame may be employed to perform such analysis by using what Owens called "short-time energy function" $E_m$ [Owens, 1993, p.70]; this can be described by Equation 5.1, Note the segmentation of the signals into a pre-fixed frame length is expressed as application of a rectangular windowing function $W[m]$. Similarly, in order to obtain the average energy value over the frame [Atal and Rabiner, 1976], the short-time energy function may be divided by the number of samples per each frame, $N$; as shown in Equation 5.2 [Owens, 1993, p.70].

$$E_m = \sum_{n=1}^{N}(x[n] \cdot W[n-m])^2; \tag{5.1}$$

$$E_m = \frac{1}{N}\sum_{n=1}^{N}(x[n] \cdot W[n-m])^2; \tag{5.2}$$

Atal and Rabiner [Atal and Rabiner, 1976] utilised the logarithmic value of the above energy in their speech recognition algorithm. To avoid having computation of logarithm of zero, a small constant value $\epsilon$ (the example value used on their paper was $10^{-5}$) was added to the summed energy value (Equation 5.3 [Atal and Rabiner, 1976, p.202]).

$$E_s = 10 \cdot log\{\epsilon + \frac{1}{N}\sum_{n=1}^{N}(x[n] \cdot W[n-m])^2\}; \tag{5.3}$$

Figure 5.3 shows example graphs of collected logarithmic energy (Log Energy) of acoustic signals of just under 1 second long for each 1024 samples with 50% overlap. The logarithmic energy value has been applied to the vehicle detection algorithm of the present research, using seismic signals, and its discussion is included in Section 7.2.1.

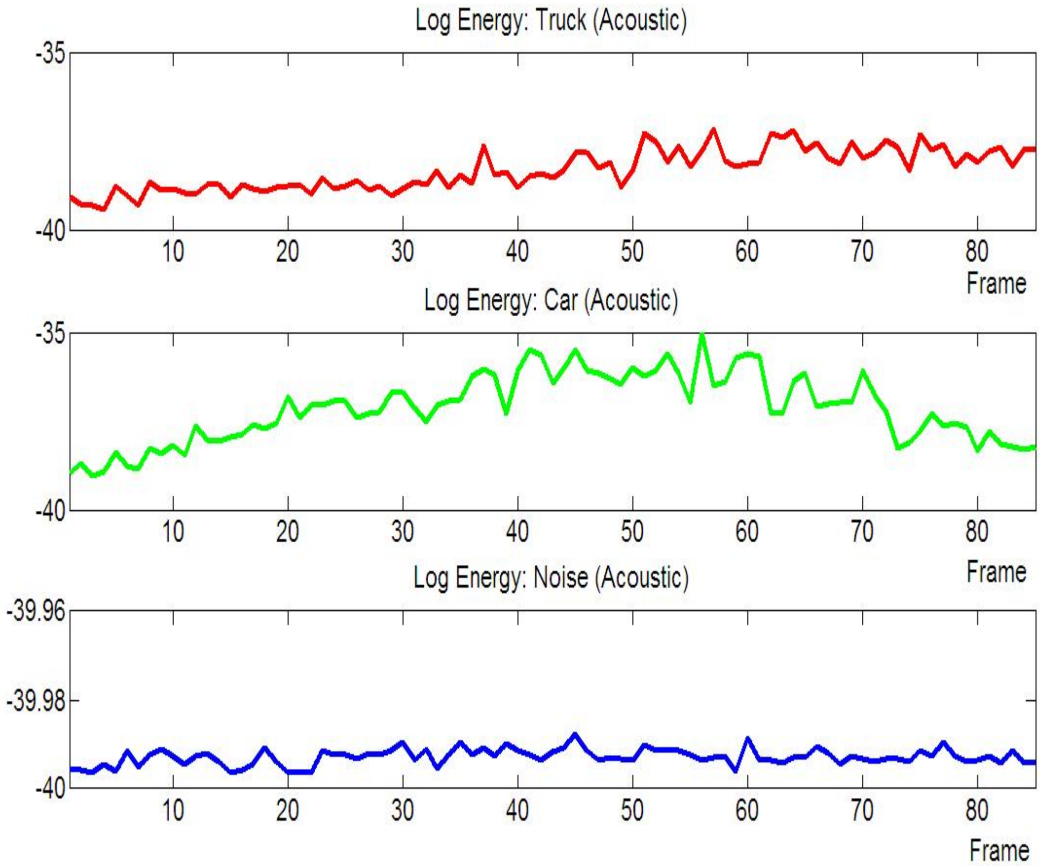


Figure 5.3: Example of Acoustic Log Energy of a Truck (top), a Car (middle) and Noise (bottom), Frame Size=1024 with 50% Overlap, Fs=44.1kHz.

### 5.1.4 Time Domain Correlation

Crosscorrelation of signals is one of the simplest but most widely used time domain signal processing techniques that can be employed to measure the similarity between two sets of signals. For example in radar target detection, correlation between transmitted signal and received signal may reveal target presence if a delayed transmitted signal (i.e. reflected signal) was captured within the received signal [Proakis and Manolakis, 1996; Ifeachor and Jervis, 2002; Kumar et al., 2005]. Crosscorrelation can be calculated by;

$$R_x y = \sum_{n=-\infty}^{\infty} (x[n] \cdot y[n-k]); \quad for \;\; k = 0 \pm 1 \pm 2, ... \tag{5.4}$$

or equally,

$$R_x y = \sum_{n=-\infty}^{\infty} (x[n+k] \cdot y[n]); \quad for \;\; k = 0 \pm 1 \pm 2, ... \tag{5.5}$$

where; $k$ is delay (or the size of shifting)

$x[n]$ and $y[n]$ indicate samples at instant of each signal set.

Both of above were obtained from [Proakis and Manolakis, 1996, p.120].

Similarly autocorrelation, which is crosscorrelation between the current sample and a shifted sample of the same signal, is often used to estimate the periodicity of the signal that may be hidden due to noise, for example. Autocorrelation has often been employed in speech pitch detection algorithms [Atal and Rabiner, 1976], but also the use of autocorrelation in estimation of engine firing rate for vehicle type recognition has been proposed [Thomas and Wilkins, 1970]. Instead of dealing with signals of infinite lengths as in the above Equations 5.4 and 5.5; usually short-time correlation, hence correlation of windowed signals, is calculated in practical applications. A short-time autocorrelation can be expressed as;

$$R_x x = \sum_{n=1}^{N} (x[n] \cdot w[n])(x[n+k] \cdot y[n+k]); \quad for \;\; k = 0, 1, 2, ... \tag{5.6}$$

where; $k$ is delay (or the size of shifting)

$w[n]$ is a windowing function

$N$ is the size of a windowing function $w[n]$

(Obtained from Owens [Owens, 1993, p.58])

Figure 5.4 contains examples of autocorrelation collected from acoustic signals of just under 1 second long, normalised per each frame. Again, each calculation was performed over a frame of 1024 sample long, segmented with 50% overlap.

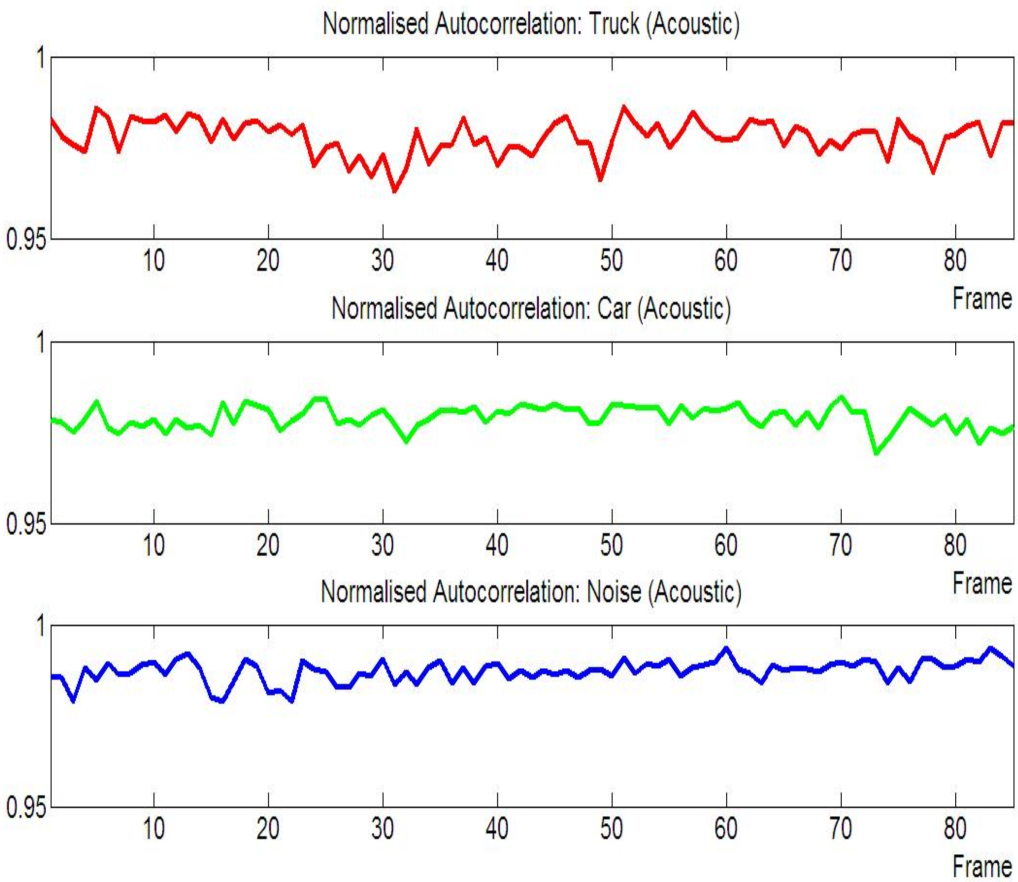


Figure 5.4: Example of Normalised Acoustic Autocorrelation of a Truck (top), a Car (middle) and Noise (bottom), Frame Size=1024 with 50% Overlap, Fs=44.1kHz.

### 5.1.5 Time Domain Signal Coding (TDSC)

TDSC has been developed with diverse applications, such as diagnosis of heart condition [Swarbrick, 2001], machinery maintenance [Lucking, 1997] and animal species identifications [Chesmore, 2001, 2004, 2007; Farr, 2007; Farr and Chesmore, 2007]. It has also been applied to a military vehicle recognition research using acoustic and seismic sensors [Mazarakis and Avaritsiotis, 2006], which demonstrated the optimistic outcome for those types of vehicles. TDSC was derived from King and Goslings' research [King and Gosling, 1978], in which a time domain speech encoding technique was suggested. The original TES algorithm consisted of segmentation of speech signals into samples between two consecutive zero crossing in the time domain. Then, the duration ($D$), i.e. the number of samples per segment, and shape ($S$) information based on the number of complex zeros per segment were found. Finally, the TES algorithm was completed by determining the closest match from the two-dimension array codebook, according to the duration and the shape.

The standard TDSC [Chesmore, 2001] is also a time domain signal processing technique that focuses on waveform descriptors, generated purely in the time domain. However, unlike the TES, it no longer works with complex zeros, since it was found challenging but unnecessary [Chesmore, 2001]. Instead, the shape information of TDSC is represented by the number of minima per each segment. As there is some positive potential recognised in developing the TDSC algorithms further, research has continuously been carried out by either exploring variations of shape descriptors [Chesmore, 2001], constructing automated codebook generation algorithms [Swarbrick, 2001], or building algorithms that do not require a codebook [Farr, 2007].

For the current study TDSC was modified slightly simply to make later handling of the features easier for the author's program. The shape ($S$) information was taken as number of positive maxima and negative minima instead of number of positive minima and negative maxima used in the original one so that there is no epoch with

($S$) of 0.

During the preliminary study with TDSC, it was found that the values observed for both the durations ($D$) and shapes ($S$) collected from each epoch of seismic vehicle signals were relatively higher than those of background noise (Figure 5.5).

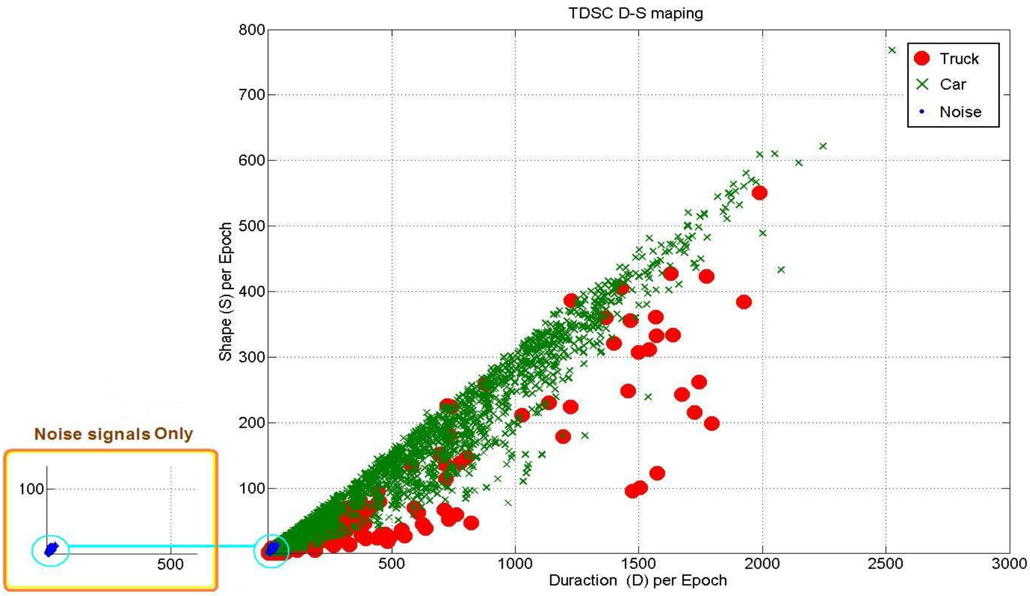


Figure 5.5: TDSC D-S Mapping of Seismic Signals from a Training Set, Frame Size=32768, Fs=44.1kHz.

However, after collecting more data, it was realised such a tendency was not always true as shown in Figure 5.6, which depicts the mapping of $D$ and $S$ values obtained from the training set of two more recording sessions.

Then, as an example is shown in Figure 5.7, the following method was developed [Evans and Chesmore, 2008].

(a) Extract duration per epoch ($D$) and number of either positive maxima or negative minima ($S$) over a frame of a fixed length.

(b) Multiply the $D$ value by $S$ value about each epoch.

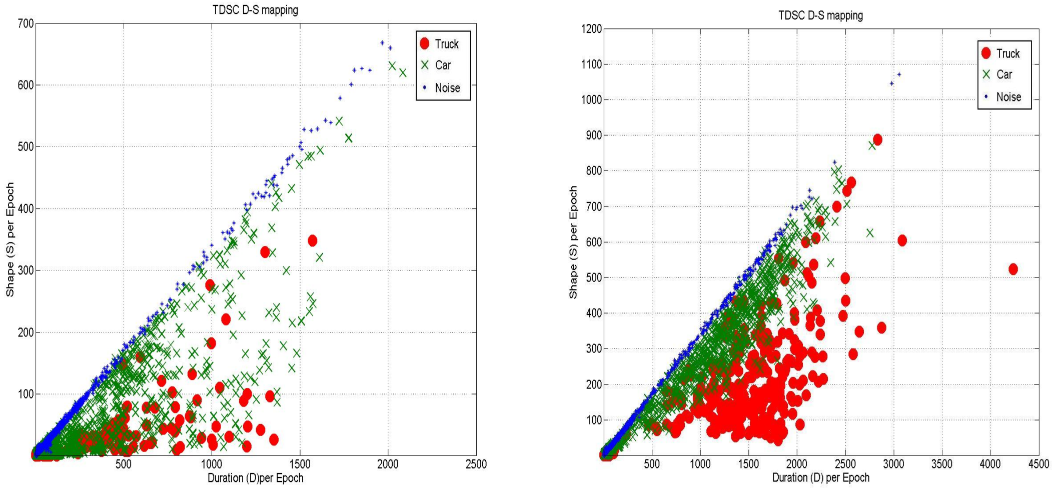


Figure 5.6: TDSC D-S Mapping of Seismic Signals from a Training Set, Frame Size=32768, Fs=44.1kHz.

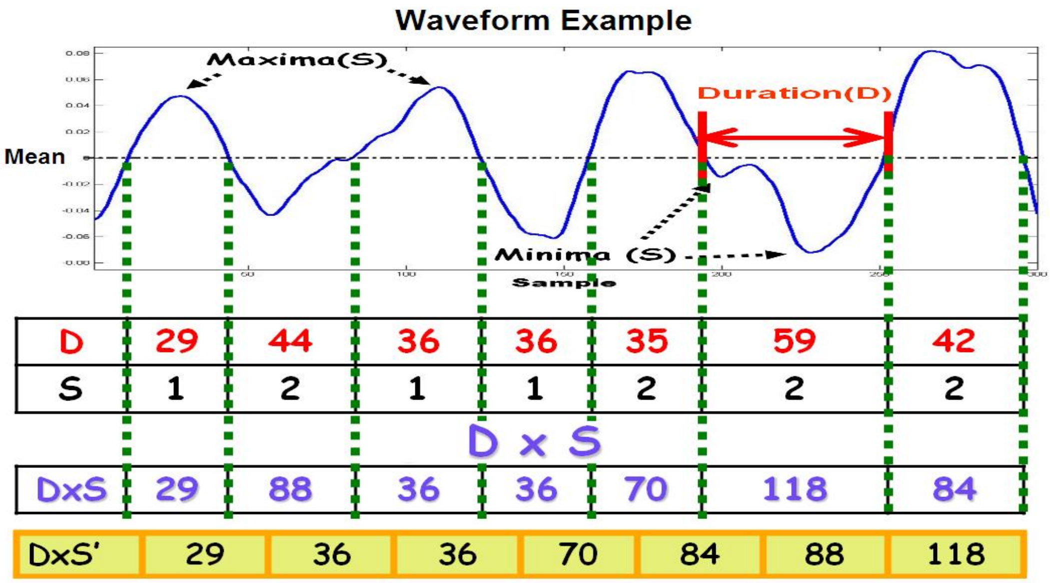


Figure 5.7: TDSC Feature Extraction Example, Frame Size=32768, Fs=44.1kHz.

(c) Sort the $DS$ products per each epoch obtained by the multiplication in ascending order within a frame.

As seen in Figure 5.8, a plotted graph of this sorted $DS$ product tends to reach higher along the x-axis if greater number of epochs was found in a frame. On the other hand, having greater $S$ values results in extension of the graph along the y-axis. This allows the use of information regarding the number of epoch found over a frame as well as some relativity between $D$ and $S$ values without requiring preparation of a codebook, which was necessary but sometimes seen as the disadvantage of the original TDSC.

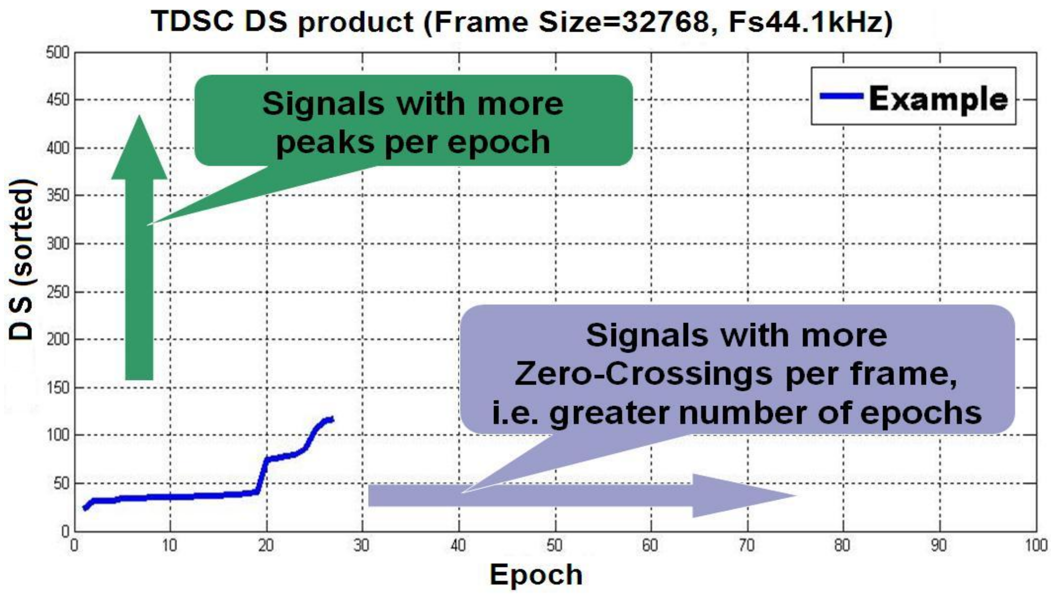


Figure 5.8: TDSC Sorted DS Products Example, Frame Size=32768, Fs=44.1kHz.

Once $DS$ products are collected from multiple sample signals, it emerged that these properties may help distinguishing signals between vehicles and noise, i.e. non-vehicles as in Figure 5.9. Note, the black dotted line was inserted only to help visually assigning the likely point for segmentation.

More investigation on this interesting aspects of TDSC features has been carried

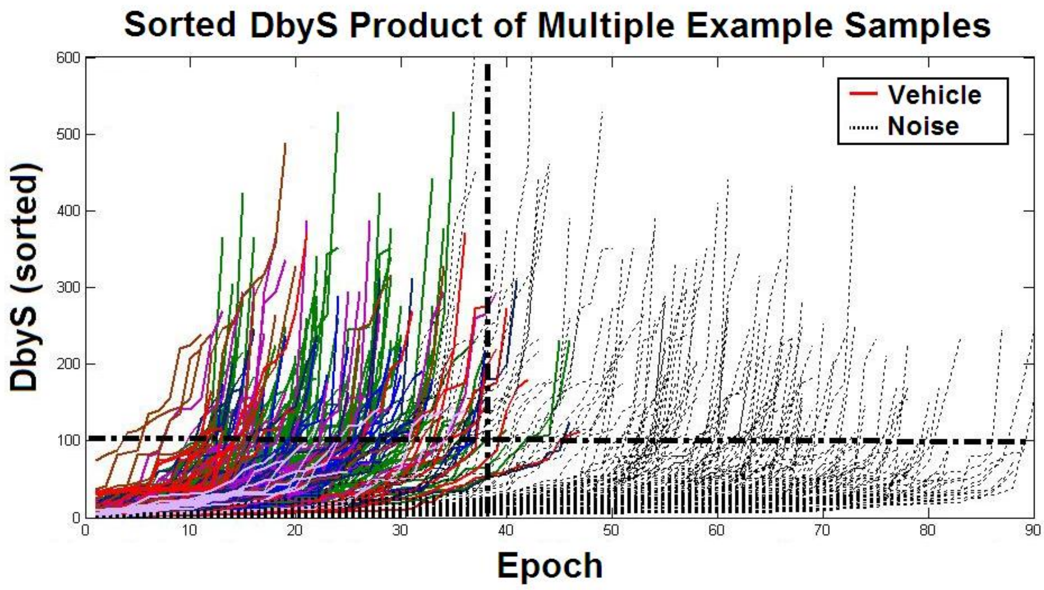


Figure 5.9: TDSC Sorted DS Products Example, Frame Size=32768, Fs=44.1kHz.

out, and the findings are discussed in Chapter 7.

### 5.1.6 Linear Predictive Coding (LPC)

LPC is a feature extraction technique, which has often been applied to speech recognition studies. In the literature, several acoustic vehicle recognition papers listed LPC as their choice of technique. Although usually without enough evidence to support their arguments, these researchers stated the reasons why they made such choices as either that acoustic signals of a moving vehicle travelling at an almost constant speed are similar to those of human speech signals [Thomas and Wilkins, 1970], or that they were simply encouraged by the popularity and success of LPC proved particularly in unvoiced speech recognition [Nooralahiyan et al., 1997]. For this reason, feasibility of LPC for the current study was investigated. Section 7.3.4 discusses the findings, but first the theoretical background is introduced here.

It is widely recognised that LPC has become a powerful tool in speech processing since Atal and Hanauer showed its applicability in 1971 [Markel and Gray, 1976; Owens, 1993], although there are some other historical records regarding the beginning of the LPC development. LPC is based on the hypothesis that it is possible to predict the current signal sample value from a certain number of the preceding samples when the signal is more or less stationary. The basic principle of LPC is to find and update the weighting coefficients or also called (linear) predictor coefficients (e.g. typically at every 10ms to 30ms for speech coding), so that the difference between the actual sample value and the predicted sample value at an instant is minimised. It is realised that the function to achieve the above by minimising the squared differences between the predicted and the actual values is, after all, practically a set of filter coefficients [Makhoul, 1975b; Holmes and Holmes, 2001; Owens, 1993; Rabiner and Juang, 1993]. The advantages of LPC are described in reference such as [Rabiner and Juang, 1993]. Within those, the simple and relatively fast computation would be especially beneficial to this study.

There are two well-known methods used to calculate the coefficients generating the minimum prediction error: “autocorrelation method” and “covariance method”. These names were derived due to the fact that one of the LPC equations contains the same features as the autocorrelation formula, and another uses a covariance matrix in the calculation. Occasionally the covariance method is also called all-pole system because it has no zeros but only poles [Makhoul, 1975a; Rabiner and Schafer, 1978]. A comprehensive tutorial review [Makhoul, 1975b] explained the difference between the two methods; stating that the autocorrelation methods are suitable for stationary signals whereas the covariance methods are for non-stationary signals. However, it was also added that the former can be applied to signals that can be seen as stationary when analysed over a short time period. As mentioned before (Section 3.1.1), vehicle signals can be included in this category. In addition, the former is always stable unlike the latter. Hence, the autocorrelation method based on [Rabiner and Schafer, 1978; Rabiner and Juang, 1993] was adopted to write a MATLAB function for the current research.

Parameter settings for the LPC order $p$ and the signal frame size strongly influence the outcome. Although much research regarding these have already been performed for speech signals processing, to the author's knowledge, there is no established standard practice for the use in vehicle signal processing. Therefore, an empirical search for the best parameters based on the knowledge gained from the literature was conducted. The following is a list of some aspects considered while selecting the LPC order $p$.

- In speech recognition, $p$ is known to depend on the sampling frequency.
- In speech recognition, having one pole per each 1kHz plus additional 3 to 4 is common, representing formant frequencies.
- In speech recognition, increasing $p$ more than a certain level does not result in much of improvement but it only increases the computational demand.
- In speech recognition, $p$ is usually set a value between 8 and 12.
- The signal frequency range of interest is approximately between 180 and 8kHz, but the sampling frequency is 44.1kHz.

### 5.1.7 Co-Occurrence Matrix

The Co-Occurrence Matrix, which had been used actively in image processing particularly for texture analysis [Haralick et al., 1973], was first introduced to speech signal processing by Terzopoulos [Terzopoulos, 1985], who suggested that the Co-Occurrence Matrix could be exploited to advance what had already been known as "voiced/unvoiced/silence classification and pitch detection" algorithm [Atal and Rabiner, 1976] in speech signal processing research.

A Co-Occurrence matrix consists of a two-dimensional matrix that embodies frequencies of occurrence of particular amplitude pairs at a pre-defined "temporal lag" $k$ within a frame of $M$ samples, segmented by a windowing function $w[n]$.

$$w[n] = \begin{cases} 1 & 0 \leq n \leq M-1 \\ 0 & otherwise \end{cases} \quad (5.7)$$

Once the co-occurrence matrix is generated, the suggested five descriptors can be obtained to illustrate signal attributes;

f1: nonuniformity

$$f_1[\Phi(S_n, k)] = \sum_{i=0}^{Q-1} \sum_{j=0}^{Q-1} \phi_{ij}(S_n, k)^2; \quad (5.8)$$

f2: entropy

$$f_2[\Phi(S_n, k)] = \sum_{i=0}^{Q-1} \sum_{j=0}^{Q-1} -\phi_{ij}(S_n, k) \cdot \log \phi_{ij}(S_n, k); \quad (5.9)$$

f3: energy

$$f_3[\Phi(S_n, k)] = \sum_{i=0}^{Q-1} i^2 \sum_{j=0}^{Q-1} \phi_{ij}(S_n, k); \quad (5.10)$$

f4: variation

$$f_4[\Phi(S_n, k)] = \sum_{i=0}^{Q-1} \sum_{j=0}^{Q-1} (i-j)^2 \phi_{ij}(S_n, k); \quad (5.11)$$

f5: correlation

$$f_5[\Phi(S_n, k)] = \sum_{i=0}^{Q-1} \sum_{j=0}^{Q-1} \frac{ij\phi_{ij}(S_n, k) - \mu_x \mu_y}{\sigma_x \sigma_y}; \quad (5.12)$$

where: $\phi_{ij}(S_n, k)$ represents each cell of a co-occurrence matrix with lag $k$,

$\mu_x$ is mean of row sums,

$\mu_y$ is mean of columns sums,

$\sigma_x$ is standard deviation of row sums,

$\sigma_y$ is standard deviation of column sums,

$Q$ is quantisation level.

(Equations from 5.7 to 5.12 are obtained from [Terzopoulos, 1985, p.5-7])

A MATLAB function has been written to generate a co-occurrence matrix of input signals and a plotted examples of collected co-occurrence matrix descriptors are shown in Figure 5.10. Finding the optimum parameters, such as the windowing function length $M$, the lag $k$ and the quantisation level $Q$, can be a long task. Based on the previous studies using the method, the initial settings were chosen, but their variations and the consequence were examined. After some preliminary investigation, more parameter variations were tested, and the program was improved a little to increase speed of processing and memory efficiency. Results are discussed in Section 7.3.5.

## 5.2 Frequency Domain Techniques

Unlike the above time domain techniques, there are others suitable for examining the signal characteristics and their variations in the frequency domain. They are likely to be either derived from, or inspired by, the studies of Fourier Transform and/or auditory perception. For the latter, particularly regarding the frequency selectivity of mammalian hearing function studies led to the development of feature extraction techniques that can be beneficial to this research. It needs to be noted, however, that the use of frequency domain techniques for the current study can be discouraged by suggestions such as that acoustic signals generated by non military vehicles may not have distinctive narrow band harmonic structure, perhaps as

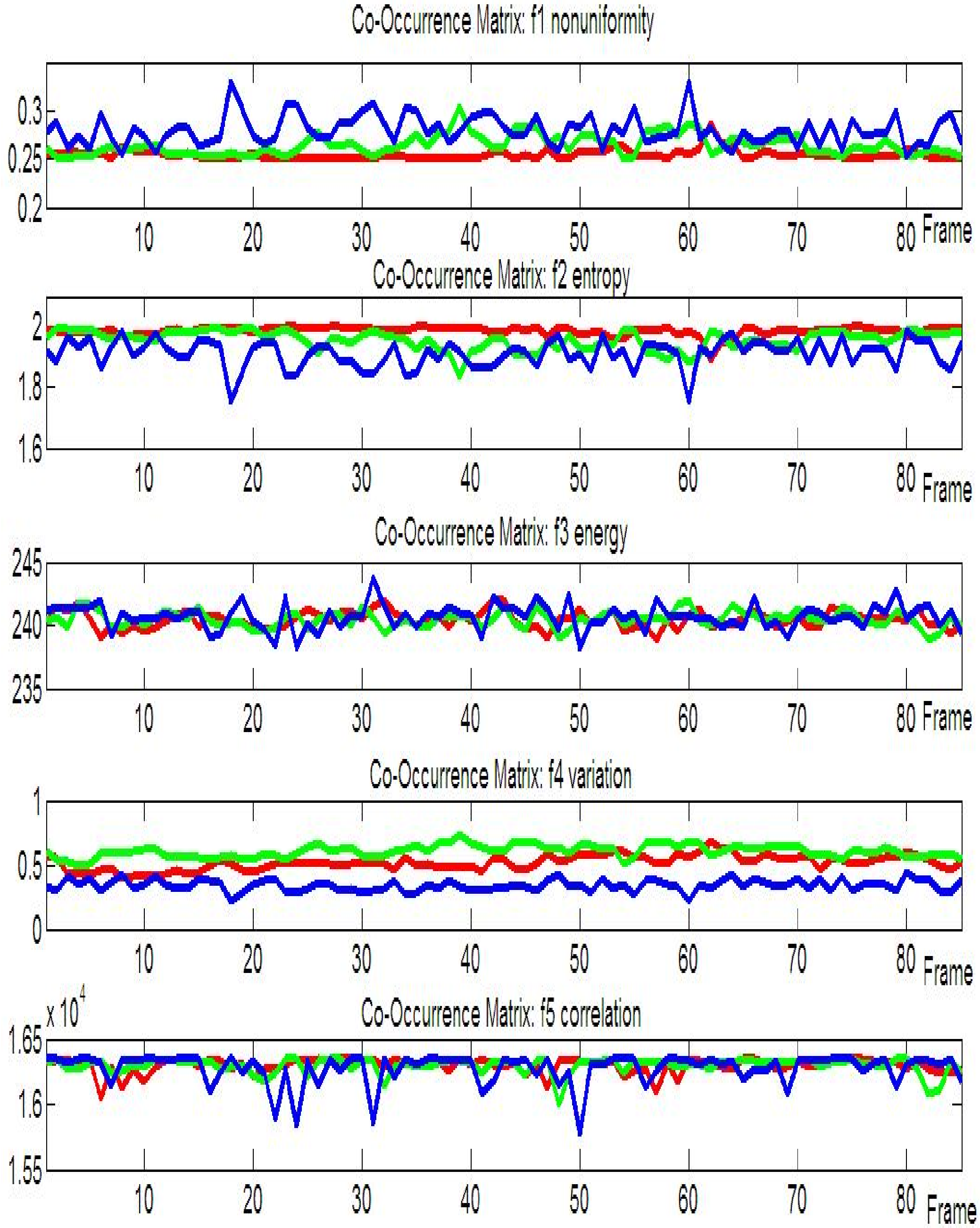


Figure 5.10: Example of Co-occurrence Matrix Descriptors Extracted from Acoustic Signal of a Truck (red), a Car (green) and Noise (blue), Frame Size=32768, Fs=44.1kHz.

a result of the effort to reduce noise emission by the car manufacturers [Jacyna et al., 2005]. Nevertheless, this section describes the frequency domain techniques that have been applied to the similar research whose outcomes were encouraging (Chapter 2), together with some of their variations.

### 5.2.1 Spectum Estimation

Spectrum estimation techniques are divided into two groups; non-parametric and parametric. In non-parametric analysis, firstly autocorrelation of a given sample data is calculated; then, where applicable, the harmonic structure of signals can be estimated by using transform functions. These transform functions are derived from the Fourier Series.

One of the most used transform methods in vehicle acoustic signal processing is the Discrete Fourier Transform (DFT) (Chapter 2), which was developed to treat discrete signals. In fact, the DFT is also one of the most popular and well studied choices for many applications [Jan, 2000]. Although this area of signal processing has been studied extensively by many researchers, perhaps not enough has been reported in relation to acoustic and/or seismic signals of moving vehicles, particularly for road vehicle recognition purposes. Therefore, this section describes a summary of findings obtained by examining usability of non-parametric spectrum estimation with DFT for the current study.

When performing non-parametric spectrum estimation and analysis, several issues need to be addressed and dealt with care, so as to obtain useful outcomes. Good explanation of such issues and their suggested solutions in general terms can be found in many signal processing text books [Brigham, 1974; Jan, 2000; Ifeachor and Jervis, 2002; Mulgrew et al., 2003]. Therefore, this is merely a summary of them with a specific focus on certain aspects which have particular influence on the present research. The first of which is the speed and efficiency of processing DFT.

It is widely known that the FFT algorithm, developed by Cooley and Tukey in 1965 [Cooley and Tukey, 1965], can be employed to overcome this problem; as long as the signal length is a power of 2. Since the original proposal, the algorithm has been modified to achieve even better efficiency and speed in calculating. However, this is not within the scope of this thesis hence will not be expanded here any further. The second is the problem occurring due to the necessity of processing segmented signals rather than that of infinite length in a real-life situation. The consequence is a problem known as spectral leakage, caused by a poor frequency resolution or termed as “coarse spectrum”, and it leads to inaccurate harmonic estimation.

In practice, signals are divided into fixed-length segments (or frames) before performing the transform. This frame length, if signals were to be processed as they were, determines both time resolution and frequency resolution of the analysis. However, it is known that the frequency resolution can be increased without distorting anything else by adding zeros on the signals, i.e. zero padding. Perhaps a more serious problem can be caused by the fact that DFT works on some assumptions; such as the target signals are circular signals. If signals were simply divided into a set of fixed length segments, it is practically the same as applying a rectangular window to the signal.

As it has been demonstrated in standard text books, this might lead to having discontinuity between the beginning and the end of a signal frame, and it will result in inaccurate spectral estimation, phenomena known as “spurious peaks” (exhibition of false peaks) and “spectral leakage” (disguised true peaks). Therefore, it is often the case that a more suitable windowing function should be applied before calculating DFT. In addition, applying a windowing function to overcome the spectral leakage can lead to another problem, known as “smearing”, which is the phenomenon caused by the main lobe widened while minimising the side lobe levels. Therefore, appropriately choosing a windowing function and its length is crucial to accomplishing a good spectrum estimation. Previous studies on the effect of some of well-known windowing functions [Harris, 1978] were consulted to identify the optimum windowing function and signal frame length.

Another problem in using DFT for feature extraction is that the resultant feature vectors have relatively large dimensions. Then, an algorithm developed for handwritten signature image recognition [Lam and Kamins, 1989] contains some interesting suggestions in dealing with DFT values for recognition. They firstly eliminated part of the resultant feature vectors and kept only that within the band range of interest after processing DFT. Secondly the resultant DFT values were normalised by "sample variance" [Lam and Kamins, 1989, p.40], which was the variance of each DFT value corresponding to the same frequency components over the one whole signature. The collected feature vectors were also accompanied by their "signal-to-noise ratio", which was found as "the square of the mean divided by the standard deviation" [Lam and Kamins, 1989, p.42]. Perhaps, these ideas can be adopted for this acoustic and seismic vehicle signal recognition algorithm development in employing DFT. Nevertheless, due to the limitation of time, it will be left for the future study.

### 5.2.2 Filterbank or Subband Analysis

As previously mentioned in Section 1.1, human hearing has some ability to recognise types of approaching vehicles. A healthy adult with standard or acute auditory perception can often achieve high accuracy in recognising or estimating the approaching vehicle type by listening only. Therefore it was assumed that feature extraction techniques that have been developed based on the knowledge obtained in studies of human hearing should be beneficial to the current research. In this section, those techniques particularly relating to the "frequency selectivity", or sometimes referred to as "frequency resolution" or "frequency analysis" [Moore, 2003, p.65], of human hearing are described. Within these terms, "frequency selectivity" is employed.

There have been many studies on the frequency selectivity of human hearing reported in literature. For example; in the late 19th century, Mayer [Mayer, 1876] published his findings in relation to the phenomenon, now known as “masking”. With such findings, Mayer stated that only sound with a lower pitch with some reasonable relative intensity can mask the other sound, and not in cases with higher pitch sound having the equally great relative intensity. This report by Mayer was rewarded with the best credit in early studies on masking by Wegel and Lane [Wegel and Lane, 1924], who studied the masking in terms of the intensities and frequencies of pure tones. Particularly, it was pointed out that masking was observed when the two tones are closer in frequency range.

In one of his publication; Fletcher [Fletcher, 1940] described his findings on auditory perception dynamic range. He secondly described the behaviour of the inner ear and auditory nerve in response to pure tones, followed by explanation of pure tone masking by using a “noise audiogram” [Fletcher, 1940, p.51]. Thirdly, together with some experimental results, he explained the relationship between the frequency of sound and the position at which the excitation occurs in the Basilar Membrane from the observation of wide band noise masking as well as the perceived pitch change based on the earlier studies [Wegel and Lane, 1924; Shower and Biddulph, 1931]. Moreover, also by following previous findings including his own [Fletcher, 1938a,b], he established the relationship between frequency and intensity of a tone as well as frequency, intensity and band width of masking noise. This finding of the rectangular filtering width “critical band width” [Fletcher, 1940, p.55] has led to much study in masking and frequency selectivity of human perception, including the development of “auditory filter” modelling. Detailed explanation about critical band and auditory filter, as well as their historical development can be found in [Greenwood, 1961a,b; Scharf, 1970; Moore and Glasberg, 1983; Patterson and Moore, 1986; Moore, 2003; Fastl and Zwicker, 2007].

Many acoustic signal recognition researchers, particularly in speech or speaker recognition studies, have been developing feature extraction algorithms based on these findings. The use of a filterbank, or similarly subband signal analysis, in feature extraction were developed based on the study of the frequency selectivity of the basilar membrane in human inner ear. The algorithms often consist of filterbanks, in order to account for the auditory filter. Although it is sometimes seen as techniques in the time-frequency domain, filterbanks are included in this section because of this connection to the frequency selectivity.

There are variations in terms of structure (or scaling) of bandpass filtering; i.e. what frequency bands are covered by each set of filterbanks. It is possible to utilise a much simpler filterbank such as a uniform filterbank, consisting of a set of band-pass filters whose frequency range is equally spaced over the whole analysis band. On the other hand, the ranges or the positions of the centre frequencies of band pass filters in non-uniform filterbanks are not equal. Some algorithms use a logarithmic scale, others may be a combination of liner and logarithmic scaling. As previously described in Chapter 2, these feature extraction algorithms were suggested to benefit the current research.

MFCC analysis uses a mel-scale filterbank, whose centre frequency of each band is commonly set fixed at 100Hz up to 1000Hz and then increases logarithmically above the 1000Hz transition point. Frequently a triangular filterbank with total of 20 band pass filters is used. MFCC feature vectors in digital signal processing can be collected by the following procedures.

(a) The sampled and quantised input signals are segmented into frames of fixed length by a winding function.

(b) DFT is performed on each frame.

(c) The log of the sum of the energy of DFT multiplied by the mel-scale filterbank filtering coefficient is calculated for each passband.

(d) The collected logarithmic values are often processed further to reduce dimension.

On the other hand, ERB filterbanks can be built by Equation 5.13 [Howard, 2006, p.76].

$$ERB = 24.7 \times [(4.37 \times f_c) + 1]Hz \tag{5.13}$$

where; $f_c$ is the filter centre frequency in kHz.

It is stated that the effects caused due to the scaling variations of critical band are "insignificant with regard to design of filterbanks for speech recognition purpose" [Rabiner and Juang, 1993, p.78]. In addition, ERB appeared to depict more distinguishable features than MFCC when they were studied with some randomly selected data of moving vehicles. Based on these as well as its simplicity in terms of calculation, it was decided to focus on ERB filterbanks.

Figure 5.11 shows graphs to compare some of the filterbank outputs collected from the real-life signals. Visually, it seems to depict some ability of ERB filterbanks to separate between truck and car signals, as well as the background noise. Nonetheless, filterbank method's feature vectors have relatively high dimension; thus, it is assumed the use of it reduces the computation efficiency and processing speed for an automated recognition system.

## 5.3 Time-Frequency Domain Techniques

Techniques in the time-frequency domain can be advantageous, in particular, for analysing signals described as time-varying or non-stationary. As it has been pointed

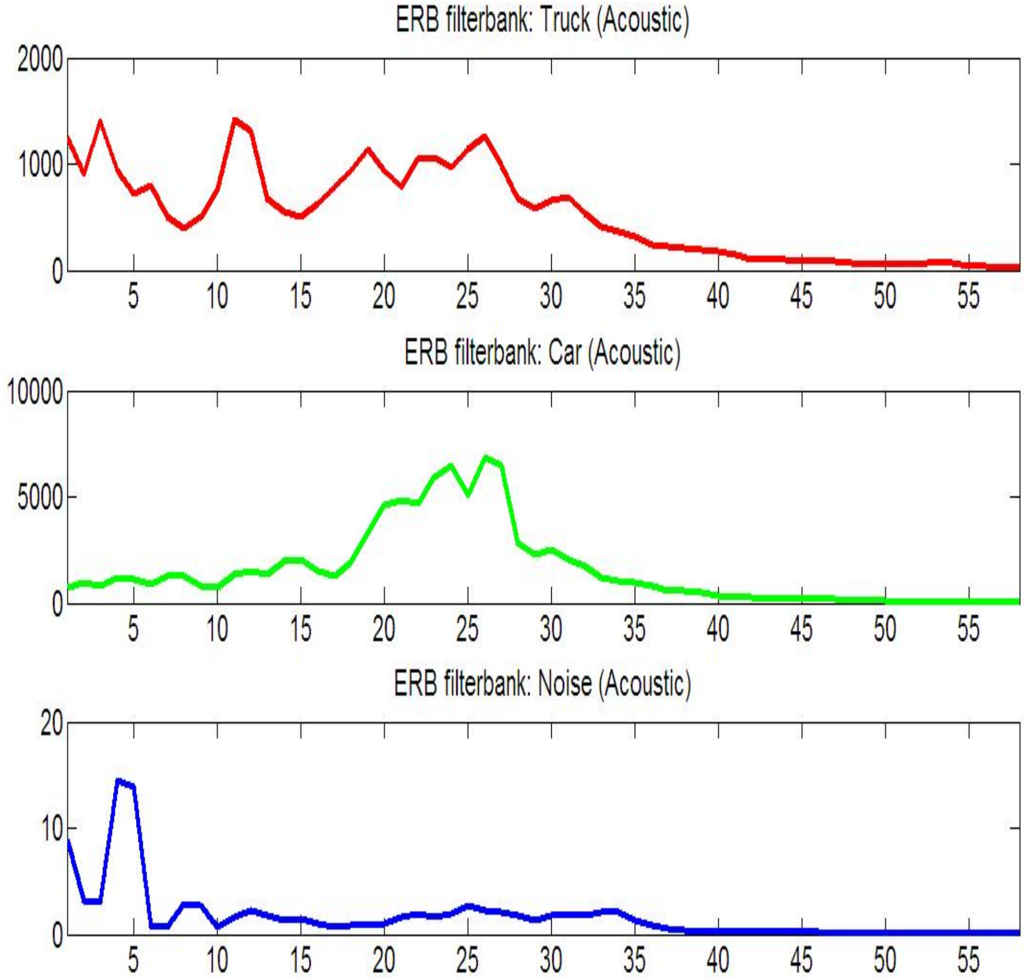


Figure 5.11: Examples of ERB Filterbank Output Extracted from Acoustic Signal of a Truck (top), a Car (middle) and Noise (bottom), Frame Size=32768, Fs=44.1kHz. (The x-axis corresponds to each filterbank.)

out (Section 3.2); signals generated by moving vehicles can be included in this category. However as Heisenberg's "inequality" [Wickerhauser, 1994; Gargour et al., 2009] based on the well-established "Uncertainty principle"[Cohen, 1995, p.44] indicates, there are trade-offs exist between the time resolution and the frequency resolution. If the time resolution is expressed as $\Delta t$, and the frequency resolution is noted as $\Delta f$, then the Equation 5.14 represents the relationship between the two, based on the uncertainty principle.

$$\Delta t \Delta f \geq \frac{1}{4\pi} \tag{5.14}$$

Nevertheless, there have been much research carried out to establish such signal analysis techniques, which may be applicable to the current research. Many text books on time-frequency signals analysis have been published [Cohen, 1995; Debnath, 2001; Qian, 2002] for those wanting to know about both theoretical development and practical applications of the topic in general. In fact, it has grown to be a very large research area where numerous progresses are being reported. Nevertheless, readers should be noted that the following sections only cover the most relevant parts of the research so as to stay focused. For the present research, it is important to identify the techniques with a high potential in delivering accurate classification performance. Therefore, first of all, the literature was consulted (Chapter 2), and then secondly, several techniques were particularly selected for the current heuristic study based on the findings.

### 5.3.1 Short Time Fourier Transform (STFT)

STFT was introduced to improve the already existing standard DFT, so that a better timing localisation of spectral variation can be achieved while keeping the frequency resolution level. Cohen states STFT is "the most widely used" [Cohen, 1995, p.93] non-stationary signal analysis technique. STFT employs a fixed-length windowing function to split the time domain signals into smaller frames prior to processing DFT. Similar to the spectral estimation case (above 5.2.1), this signal segmentation is treated by applying a pre-defined window function to the signals. Once again, a care must be taken regarding the choice of windowing function and its length on such signal segmentation, in order to avoid introducing too many artefacts that could potentially interfere the resultant analysis. In terms of the frame length, there is a minimum limit for achieving meaningful spectral estimation, depending the signal property [Cohen, 1995].

Although the timing information, which is at least better than DFT, is useful; the lack of flexibility regarding the window size is problematic [Gargour et al., 2009]. Particularly in non-stationary signal processing, STFT may be outperformed by other time-frequency methods such as wavelets transforms. Nevertheless, the literature review in Chapter 2 did not make it clear whether STFT could bring valuable benefit to the current research or not, hence some experimental study was performed to clarify the point. Consequently it was eliminated from the options for further investigation.

### 5.3.2 Discrete Wavelet Transform (DWT)

It is said "Time-frequency wavelets are suited most specifically, to the analysis of quasi-stationary signals" [Jaffard et al., 2001, p.7].

Readers interested in about the historical development of the theory should refer to many available books and articles on wavelet and related subjects; such as that by Jaffard et al. [Jaffard et al., 2001, chap.2]. One thing, which is clear to many is that the wavelet theory has been developed over so many decades in diverse fields; including mathematics, physics, geophysics, astronomy, and signal processing. Perhaps this infers its flexibility, which will be mentioned next. Similarly, applications for wavelet transform techniques can also be found in a wide range of studies; including audio processing, image processing (e.g., for classification and edge detection), seismology study, astronomy, biology and finance. In addition, the use of wavelet transform is not limited to feature extraction, or sometimes called "time-frequency spectral estimation" [Bentley and McDonnell, 1994, p.175] for classification, but also used as compression or noise reduction tools as introduced earlier in Section 4.2.3.

Compared to other methods in the time-frequency domain, such as STFT, wavelet transforms are known to achieve "a better trade-off between time and frequency resolution" [Gargour et al., 2009, p.60] due to the dilation and translation functions. Signals are analysed according to a set of (basis) functions that are either scaled or shifted versions of what often referred to as a "mother wavelet", rather than performing analysis by a unified windowed sinusoid, as previously seen with the STFT (above 5.3.1). This additional freedom earned for a more sophisticated analysis has provided the now much appreciated progress in signal processing. As a result, high frequency components over a short time period and low frequency components over a longer time period can be captured simultaneously.

In fact, this ability to capture high frequency transient signals buried within continuous low frequency signals is perhaps the best selling point of the wavelet transform, compared to the conventional Fourier transform. For the current research, the signals of interest can be what are called pseudo-stationary signals; in other words, it is not very clear how useful it is for feature extraction purpose. Preliminary studies showed less positive outcomes compared to the others introduced here so far while the features have high dimension, and this is not ideal. However, for noise reduction at least, it may be useful to discriminate unwanted signals other than vehicles' signals, such as wind noise and some transient noise that may be captured during recording. This is why noise reduction by wavelet transform, particularly wavelet packet algorithm, is mentioned in Section 4.2.3 of this thesis.

## 5.4 Chapter Summary

Based on the literature review in Chapter 2, various feature extraction techniques that may be beneficial to the present research have been discussed. These techniques are categorised according to the signal processing domains, in which each technique tends to have more impact on the outcomes. However, it should be noted here that in practical digital signal processing of moving vehicles, generally signals of a finite length are analysed, rather than that of infinite length. Hence the outcomes of the signal examination using a frequency domain method can contain some timing information regarding the particular signal segment. Also, the findings of a time domain method may have frequency information of the target signals that can be retrieved. Therefore, in selecting the optimum method, it is less important in which domain the techniques are categorised here. Instead, the practical feasibility of them, including how well the distinguishable features are extracted, computation speed and efficiency, are more crucial points to be considered. In such aspects, each technique has been analysed with real-life data collected. As a result, ZCR, Log Energy, Time Domain Correlation, LPC, TDSC, Co-Occurrence Matrix and ERB filterbanks are the selected options for further examination (Chapter 7).

# Chapter 6

# Decision Making

## 6.1 Dimensionality Reduction

Increasing the number of extracted features used to represent input data attributes, more than certain dimensions, can actually lead to poorer classification results, the phenomenon known as "the curse of dimensionality" [Duda and Hart, 1973, p.95]. As many pattern recognition textbooks suggest [Bishop, 1995; Webb, 1999; Marsland, 2009], inserting an appropriate dimension reduction algorithm to filter out the less salient elements of the collected feature vectors before processing the final decision making algorithms could improve overall recognition performance, in terms of accuracy and /or computational efficiency. Benefits of using such algorithms can be, for example, due to reduction of data amount or noise [Jain and Zongker, 1997; Marsland, 2009].

Although there are different views on grouping of such methods, as can be seen in [Koren and Carmel, 2004] for example, that roughly two approaches can be taken to reduce the dimension of the feature vectors for such purposes. One is keeping a selected subset of the features by following some rules whilst discarding the rest before performing classification. Particularly when the extracted multiple features are correlated to each other, there are likely to be some redundant data included;

such a simple selection procedure can be beneficial to improving the operation outcomes. The reduction rules to be followed can be arbitrary, perhaps simply by a subjective choice of a developer; however, more sophisticated tactics are likely to bring more benefit. Needless to say, determining the optimum selection criteria is an important task for the resultant recognition performance as explained by Bishop [Bishop, 1995].

Another approach is the use of more complex dimensionality reduction algorithms, in terms of making the most of the available information. This entails applying some kind of transform algorithms in an attempt to integrate either the characteristics of the target data or the training data or the both, in order to acquire a better representation of the distinguishable, but sometimes hidden, signal attributes. Occasionally, such algorithms are seen as classifiers, rather than part of feature extraction, depending on the perception of or approach taken by the developer. In this chapter, as long as the focus is on reducing the data dimension for optimising the decision making performance, they are treated as independent dimension reduction algorithms.

Overall the common methods for dimensionality reduction are: unsupervised learning or clustering, PCA, LDA, NN, GA. This section explains these listed methods in terms of dimension reduction only. Please note; some of them reappear in the next section (Section 6.2) as clustering or classification algorithm, again with emphasis on those aspects.

### 6.1.1 Simple Selection

Simply selecting features can sometimes be valuable; for example when the feature vectors are too large, or available computation abilities do not match with the demand. In selecting the features to maintain for the succeeding data processing, algorithm developers can either follow some existing criteria or make their own. Although it depends on the complexity of the data in the spotlight as well as constraints due to other requirement, such as application and resource; ideally one should construct own criteria through experimental investigation.

### 6.1.2 Principal Components Analysis (PCA)

Vehicle acoustic signals and their features may contain a variety of unnecessary information for detection and recognition processes. PCA is also frequently referred to as Karhumen-Love (K-L) transform [Yen and Lin, 2000; Wu et al., 1999]. It is a linear method that maintains more significant parameters that vary more radically than others and are uncorrelated to each other, while reducing the dimension of the generated feature vector, hence reducing the data size [Wang and Qi, 2002].

The procedures for calculating PCA are explained next. Firstly, the mean values of all entries in each dimension of input data need to be calculated. Secondly, a "zero-mean matrix" of the input is generated by shifting values of the data matrix according to the mean values. Thirdly, a covariance matrix of the zero-mean matrix is acquired, then fourthly, its eigenvectors are determined. Next, these eigenvectors are sorted in a hierarchical order concerning magnitude of matching eigenvalues. Finally, to obtain the principal component of input data, the multiple dimension input data matrix of interest is multiplied by the eigenvector that corresponds with the largest eigenvalue. The number of eigenvectors to be employed at the final multiplication step can be increased depending on how much of original data should be maintained or in other words how many dimensions of data should be reduced [Bishop, 1995; Jolliffe, 2002; Wang and Qi, 2002; Singhal and Seborg, 2005].

Since the magnitude range of each descriptor varies, some scaling on the collected feature vectors may be required before any automated dimensionality reduction is processed. Perhaps normalising the values within the feature vector is necessary unless a better solution is found. These are examined in Chapter 7.

### 6.1.3 Linear Discriminant Analysis (LDA)

The use of LDA for reducing the dimension of the feature vectors can be seen as similar in a way to that of PCA, although LDA employs supervised learning instead of unsupervised learning [Raymer et al., 2000; Tmer and Demir, 2005]; hence it has been called "adaptive dimensionality reduction" [Ding and Li, 2007]. The aim of LDA supervised learning algorithm is to obtain a transform function that maximises the class separability, or more precisely, the ratio between inter-class variability and intra-class variability, both accompanied by the probability distribution about all classes. It has been claimed that LDA may be more appropriate than above PCA for classification tasks [Yen and Lin, 2000; Yu et al., 2007]. LDA is also discussed in 6.2.1.

### 6.1.4 Neural Networks (NN)

NN, inspired by functions of biological neurons, have been employed widely in non-linear dimension reduction of feature vectors [Bishop, 1995; Lerner et al., 1999; Ture et al., 2007] since their first introduction to chemical engineering by Kramer [Kramer, 1991; Hsieh, 2001]. According to Ture et al. [Ture et al., 2007], it has been proved that PCA equivalence can be achieved by a one-layered linear NN. NN is also mentioned in 6.2.5.

### 6.1.5 Genetic Algorithm (GA)

Data dimension reduction with GA, an inspiration by the biological gene selection and evolution mechanisms seen in natural world, has often been called "feature selection" process [Raymer et al., 2000; Tmer and Demir, 2005] because GA selects a subset of feature vectors that maximally retains the quality of the original feature set with less dimensions unlike PCA, which generates a new set of feature vectors with less dimensions from the original feature sets. Although the simplicity of GA algorithm is advantageous, the uncertainty caused due to the involvement of possibly many either random or pseudo-random procedures is sometimes not desirable. This study has only limited time available for investigation thus such challenges by the uncertainty may not be solved appropriately. Therefore GA has been regarded as unsuitable for this study.

## 6.2 Clustering and Classification

In automated pattern recognition, the class to which a new observation belongs is determined by following pre-defined rules. There are a wide range of decision making techniques that have been developed and utilised for various pattern recognition applications in diverse fields; including finance, marketing, and healthcare as well as engineering. For a set of samples, automated pattern identification or recognition can be performed either by clustering or classification. Converging acquired input signals (or extracted features) according to the similarities or dissimilarities between them determined in proportion to a measurement, is called clustering. On the other hand, sorting new input signals one after another into finite number of groups by applying pre-established methods is called classification.

In order for such conversion or class separation to be achieved, often observation of training samples is required during development and implementation of a pattern recognition algorithm. Training samples are sets of data whose classes are already known; therefore can be used to modify an algorithm so as to improve recognition performance in terms of accuracy and processing efficiency. This algorithm modification process is called “learning”.

Machine learning can be categorised into three forms: supervised learning, unsupervised learning and reinforcement learning. Supervised learning incorporates target data, i.e. the ideal output for each training sample is known beforehand. Also, usually there is an associated error function or cost function that determines the modification rules so that ideally the performance of the algorithm is improved progressively. Unsupervised learning does not make use of target data, instead the objectives may be to determine similarities and dissimilarities between available data, often in an attempt to estimate their distribution, and sometimes to predict the likelihood of which a new data belong to each of the finite categories formed in advance by using the training samples. Finally, reinforcement learning does not require target data, either; its focus is on whether the output is right or wrong regardless of how right or wrong it may be [Bishop, 1995; Duda et al., 2001].

Typically, clustering algorithms perform unsupervised learning; whereas both supervised and unsupervised learning can be seen in classification algorithms. For the decision making procedure in a pattern recognition algorithm such as the current research, including a classification method may be more appropriate. Nevertheless, clustering techniques can be beneficial during development processes in order to measure the class separability, or the usefulness in other words, of each method for pre-processing or feature extraction. The term that is commonly used to express how well an algorithm perform either classification or clustering tasks on a new unknown data is generalisation [Bishop, 1995]. Thus, in this section, techniques in both clustering and classification employed for the current study are to be explored together with their generalisation ability observed during the examination with some of the collected real-life data.

### 6.2.1 Linear Discriminant Analysis (LDA)

For a two-class example, LDA makes classification decision by comparing the probability of which the sample of interest belongs to one class against a threshold. If the probability is greater than the threshold, the sample is assigned to the class, otherwise it is classified to another class as expressed by Equations 6.1 to 6.3 [Bishop, 1995, p.38, eq.(2.13-2.15)].

$$y_k(\mathbf{x}) = \mathbf{W}_k^T \mathbf{x} + w_{k0}, \tag{6.1}$$

where

$$\mathbf{W}_k^T = \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}^{-1}, \tag{6.2}$$

$$w_{k0} = -\frac{1}{2} \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k + lnP(C_k) \tag{6.3}$$

The above $k$ is the number of classes, $\mathbf{W}$ and $w_{k0}$ are usually called "weight vector" [Nagy, 1968, p.204] and "threshold weight" [Duda and Hart, 1973, p.131] respectively. $C_k$ indicates the $k$-th class, and $P(C_k)$ is a priori probability for class k. It is based on the assumption that the each class distribution is an independent normal distribution, and the covariance of each class is equal.

The algorithm may be useful for assessing the feasibility of each feature extraction algorithm by using the training samples. However, it has limited practical use for automated classification of unknown data due to the relatively intensive computation requirement [Jing et al., 2004] and its dependency upon samples to belong to the linearly separable distribution that can be indicated after training for each class $k$ [Zhu and Martinez, 2006]. As a result, LDA requires a large training samples that comprise the a priori distributions for the algorithm to produce good outcomes [Peng et al., 2008].

More advanced discriminant analysis methods have been suggested by many researchers [Jing et al., 2004; Koren and Carmel, 2004; Zhu and Martinez, 2006; Peng et al., 2008]. Consequently, it is no longer distinguishable, in many respects, from some of other classification methods that also use statistical prior knowledge learned with training samples. Therefore its further explanation is omitted here.

### 6.2.2 K-means Algorithms

The simplicity of K-means clustering algorithm may be one of the biggest advantages; however its weakness is that the results can be easily influenced by the number of cluster centres and their initial positions [Anderberg, 1973; Hartigan, 1975; Duda et al., 2001]. In other words, finding both the best number of $K$ (number of clusters) and appropriate initial positions of these cluster centres is critical. Ray and Turi [Ray and Turi, 1999] have suggested a method to automatically determine the best number of $K$, combined with rules in positioning the new cluster centres when $K$ is increased. This proposal by Ray and Turi may inspire a better clustering result; especially when the samples are linearly separable in the sample space, but the ideal number of clusters is unknown because there are perhaps multiple cases that can be adopted. In such cases, the suggested measurement rules can be useful to determine the optimum $K$. Especially during the initial period of the project, it was utilised to analyse how good each feature extraction technique option can be by observing the best number of $K$ outputs by the algorithm. It was sometime useful but often the resultant values were too big to be adopted for developing an algorithm that best suits the study. Also, there are still some room for improvement. Perhaps more knowledge regarding sample distribution can be accompanied, if available, in deciding the positions of new clustering centres when the number of clusters is increased. Rather than simply splitting the cluster with maximum variance into two, thresholding based on intra-cluster distance could be included in such splitting.

### 6.2.3 k-Nearest-Neighbours (kNN)

kNN is a classification method, but uses similar idea as the above two. Its fundamental rule is that $k$ samples that are close to each other in the Euclidean space should belong to the same class. It is said to be accurate, however, it requires a high number of calculations to be processed with a large number of training samples and large feature vectors [Wang and Qi, 2002; Duarte and Hu, 2004; Xiao et al., 2006].

### 6.2.4 Gaussian Mixture Model (GMM)

Mixture Models have been popular in speaker/speech recognition [Reynolds, 1995]. The commonest algorithm combines ML function with Gaussian distribution, hence GMM, and its key advantage is its capability of modelling random distributions [Necioglu et al., 2005] for multiple-class classification. Nonetheless, because of such ability, GMM typically require large training samples [Temko and Nadeu, 2006].

The fundamental concept of GMM with ML in pattern recognition is that the feature vectors collected from all the possible samples are assumed to construct a normal distribution. In terms of training process, the developer firstly estimates distributions that are Gaussian for each class by using the samples available for training. Then, when processing the feature vectors collected from new unknown samples, the probabilities for the vectors to belong to each distribution are calculated. The class accompanying the highest probability to contain each of the new samples is predicted. Such a method is known as Expectation-Maximisation (EM) algorithm; formed by Dempster, Laird and Rubin [Dempster et al., 1977].

This algorithm can be very effective if one has access to every single sample that can exist as the distributions can be estimated accurately in such cases. Nevertheless, in practical vehicle recognition research, it is highly unlikely to be the case since there are so many factors that cause variations to the input characteristics (Chapter 3).

### 6.2.5 Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANN), or simply NN is an extension of statistical pattern recognition [Bishop, 1995] inspired by the behaviour of biological neurons. The study of NN can be understood as what derived from Rosenblatt's Perceptron [Vapnik, 2000]. There exist both supervised and unsupervised learning NN algorithms, and the algorithms usually employ some kind of error functions so as to improve their performance [Bishop, 1995; Looney, 1997; Haykin, 1999]. NNs can be categorised by the number of NN layers or by the training methodologies. From the wide range of NNs, LVQ has been selected for further investigation, based on the findings of previous research on acoustic vehicle recognition (Chapter 2) as well as other recognition studies.

#### (a) Learning Vector Quantisation (LVQ)

Self-Organising Featuring Maps (SOFM), or often simply SOM, employs unsupervised, competitive learning. Once a winner neuron is decided, not only the weights of the winner but also that of the neighbouring neurons are updated according to the distance between the winning neuron and the rest; neurons that are close to the winner receive positive gain on their weights and weights of those that are far from the winner are decreased [Looney, 1997; Haykin, 1999]. LVQ is similar to the above SOFM except that LVQ uses target data during training procedures, in other words it stays in supervised mode while being trained. Although the original LVQ algorithm only updates the winner's codebook vectors during training [Kohonen, 2001], many LVQ variations have been proposed since its first introduction, including those that update codebook vectors of multiple neurons with relatively good performance as well as the winner's [Kohonen, 1990], or others that combine algorithms to optimise initialisation [Merelo et al., 1997]. Nevertheless, an interesting mathematical analysis of five variations of LVQ reported that the original LVQ performed better than the four others in terms of susceptibility to different initialisations and stability [Ghosh et al., 2006].

A previous study on acoustic recognition of insects using NN [Farr, 2007] showed a very good results by LVQ. Thus, after some experiments, LVQ became the choice for the current study.

### 6.2.6 Support Vector Machine (SVM)

SVM is a supervised learning classifier which aims to identify "the particular hyperplane for which the margin of separation is maximised" [Haykin, 1999, p.320]. For a standard two-class linear classification algorithm, it performs under the following assumptions:

(a) Samples from two classes are linearly separable in their feature space.

(b) There exists a hyperplane that maximises the "margin", or the sum of shortest distances between each sample point in each class and a linear line that separates two classes.

However, SVM can perform classification of non-linearly separable samples by applying a method known as "kernel mapping"[Burges, 1998]. Kernel mapping technique transfers the feature vectors in the feature space into a higher dimension space, in which the samples can be linearly separable [Gunn, 1997].

The possible disadvantages of SVM are that its training can be time consuming and also its high demand in memory [Umer and Khiyal, 2007]. SVM has been applied to a range of applications, including classification of indoor acoustic event [Temko and Nadeu, 2006] and automated acoustic wood condition monitoring in the transportation industry [Yella et al., 2007].

## 6.3 Chapter Summary

A selection of clustering and classification algorithms have been discussed together with some dimensionality reduction techniques that can be combined so as to improve the decision making performance for the system. Of course, there are so many others that have not been mentioned here but could be considered for the current study. Nonetheless, as it has been the case so far, these options were determined based on the literature review outcomes and also findings of pilot studies carried out during the relatively early stages of the project. Like many other algorithms discussed so far, each of clustering, classification and indeed the method to process collected feature vectors before performing decision making procedures has various parameters that can be set and altered depending on the application. Unfortunately for a developer, each parameter influences the results, therefore the more parameters that need to be appropriately defined, the more challenges one may face in order to find the best ones. Also similar to pre-processing and feature extraction techniques, it is not easy to determine the feasibility of a decision making technique on its own. Hence, in the next chapter, classification methods of choice; LVQ and SVM are utilised with feature vectors extracted from the real-life signals.

# Chapter 7

# Results and Discussion

An optimum system can be built by developing an algorithm that consists of some or perhaps all of those methods from each of the three steps that can be found in a pattern recognition system (Figure 1.2). In order to achieve that, one needs to establish a way of measuring the performance of each method and various combinations of multiple methods. It is crucial to find as optimum as possible parameter settings for each method. Some studies [Duin, 1996; Dietterich, 1998; Webb, 1999] have been carried out on the search for effective performance comparison approaches. Within these, one common ground accepted by researchers seems that there is no universal single approach to perfectly evaluating classification performance. It depends on the application, dataset, and other factors; even including the level of expertises of the developer. Consequently, the process introduces another level of uncertainty, which also needs to be addressed carefully during a research study like this one.

So far, many methods have been introduced for each of three main steps. If the purpose of this research was simply to create machines that can almost mimic human activities of classifying vehicle types, methods that were developed based on the knowledge regarding human activities would probably appear to be the best. However, the aim is to develop a highly accurate and efficient pattern recognition algorithm independent of human capabilities. Therefore, one should not be limited to any categories, but to be open to all the techniques that may have a positive potential when being integrated in an optimum vehicle recognition algorithm. Methods described so far in Chapter 4, Chapter 5, and Chapter 6 regarding pre-processing, feature extraction, and decision making methods respectively, attracted higher interest of the author because of their potential qualities that were supported by recommendations made in the literature, but also what was discovered during experimental analyses.

This chapter briefly introduces the evaluation methods utilised during the study, and then discusses what has been decided to be the optimum system as a result while explaining how it was determined.

## 7.1 Evaluation Methods

When evaluating the performance of an automated recognition system, it is important to clarify how it is measured. Depending on the measurement, or the interpretation of the outcome, the results may imply different meanings. In terms of examining hypothesis in machine learning, there are some known and widely accepted performance measurements. Therefore this section introduces some of them, especially those seen to be appropriate for the current hypothesis testing.

### 7.1.1 Classification Error

The most common measurement used widely in machine learning is the "error rate" or "misclassification rate" [Hand, 1997, p.97] that indicates the ratio of how much of the sample data have been misclassified by an algorithm. Although it is simple and easy to use, sometimes the information given out by the error rate may not be enough for a developer to identify how to improve the performance. Also each misclassification is treated as equal, hence no consideration regarding the real risk is included. For example, for the current study, recognising background noise as an unauthorised moving vehicle (false alarm) and trigger some action to respond as a result, is seen as equally problematic as a case when the system recognise a vehicle actually intruding as background noise. Clearly depending on the application and purpose of implementing such a system, a suitable risk assessment regarding each misclassification case should be performed. Nonetheless, error rate is a useful tool during this development especially because not enough detail about application as well as precise objectives to deploy the system is determined yet.

During the evaluation, in most cases, collected data were split in two and they were referred to as training and validation sets [Bishop, 1995]. In order to avoid allowing the results to be too biased by the samples used for training, the validation set is only utilised to evaluate how well the classifier generalises with the previously unknown data. The data for the present study were collected at various locations at different recording sessions. As mentioned in Chapter 3, there are so many factors that could influence the signals characteristics, therefore, it was ensured the both training and validation set have almost equal balance in terms of the number of samples taken from each recording sessions for each class. This chapter presents the misclassification observed mainly for the validation set, but some times that for the training set is also indicated.

### 7.1.2 Cross-Validation

With cross-validation, the data set is divided into a finite number of sub groups, and a classifier is trained by almost all subset but one, and the one remaining subset plays the role of validation set. The same process is repeated until all the subsets take the validation part, then the average of each trial can be calculated. When there are few sample data available, cross-validation can be a valuable approach although a concern regarding over-fit has been raised for such methods [Webb, 1999].

### 7.1.3 Receiver Operation Characteristic (ROC) curve and Confusion Matrix

Receiver Operation Characteristic (ROC) curve and confusion matrix are both common methods utilised to visualise the classification results [Schalkoff, 1992; Fawcett, 2006]. With these, the true positive, true negative, false positive and false negative rate can be depicted. This information can be very useful especially when the risk for each case is known so that the performance can be assessed in relation to the risk.

### 7.1.4 Parameter Setting

For each of the methods investigated, findings in relation to variations of parameter settings were studied during the research. The frame length for the signal processing was also altered so that its effect could be inspected. Although experimental analysis was carried out with many other frame lengths as well, sometimes only those powers of 2 were preferred for comparison between methods, particularly those making use of an FFT algorithm. As discussed in the next section, a vehicle detection algorithm using seismic signals has been developed in the research. Therefore, the rest of the time was spent on developing a classification algorithm that could follow the detection since the overall structure shown in Figure 7.1 was roughly established. This subsequently placed an upper limit of the frame length for processing acoustic signals at 1 second; thus, investigation with longer frames were not considered.

The frequency components of interest are in the range between around a few hundred Hz and 8 kHz for acoustic whereas between 10Hz and 180 Hz for seismic signals. With a sampling frequency of 44.1 kHz, the lower end for the seismic signals, 10 Hz corresponds to 4410 samples per cycle whereas 245 samples per cycle for acoustic signal case, 180 Hz. Consequently, to achieve an organised illustration

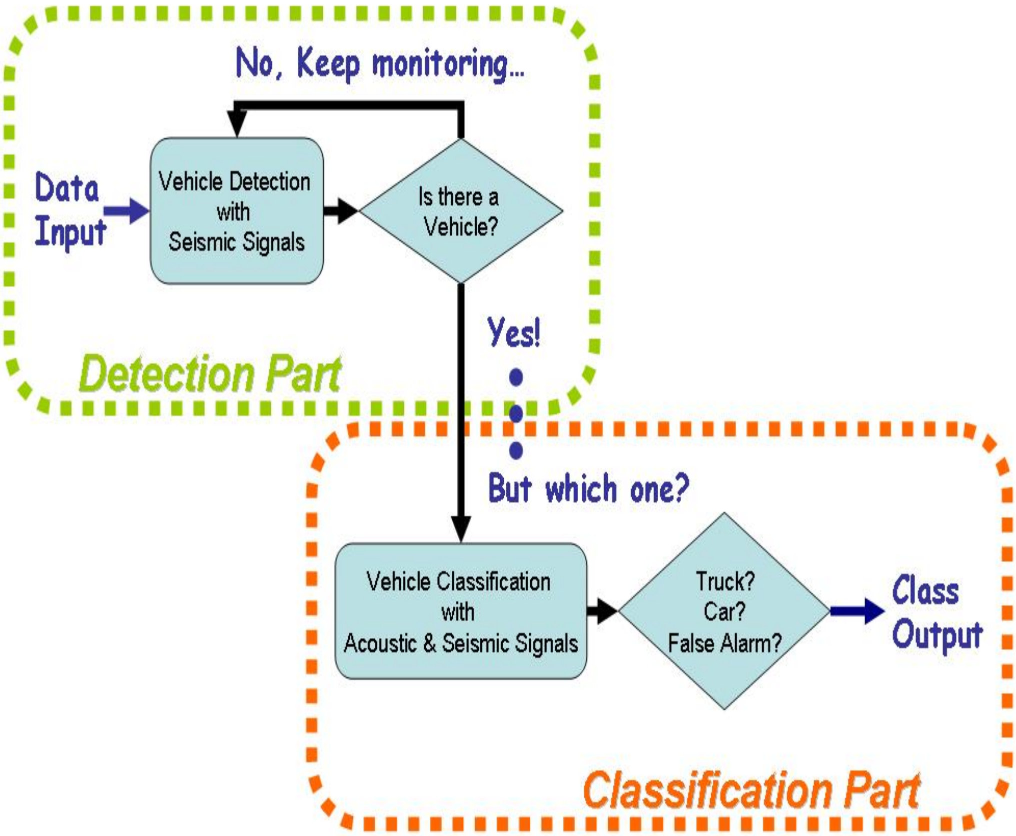


Figure 7.1: Overall Structure of the Optimum System

of the results, the frame length variations for inspection whose results are presented in this chapter are usually: 32768, 16384, 8192 for seismic and 32768, 16384, 8192, 4069, 2048, and 1024 for acoustic signals. Two more options for both acoustic and seismic signals, such as 512 and 256 for the former and 4096 and 2048 for the latter, were initially employed as well, however, they were later eliminated as these did not appear to be significant.

### 7.1.5 Experimental Data

For evaluation, real-life acoustic and seismic data collected by using the method described in Section 4.1 were used. In order to identify the optimum methods for the current study, a range of training and validation sample sets were prepared. These samples were collected at 5 different locations over 6 recording sessions. As the author did not have access to the knowledge regarding vehicle models, engine types or size for almost all the samples, each passing by vehicle was observed manually and subjectively labelled during the recording; initially between saloon car, van, four-wheel drive, pick-up truck and bigger truck classes. Nonetheless, a concern was raised regarding the reliability of this labelling, especially about the three mid-range classes as they could be too dependent on the subjective judgement of personnel who is no expert of vehicles. Thus, two of the relatively distinguishable classes, saloon car (called cars from now on) and bigger truck (truck as the former) were chosen for the rest of this study.

Data collected at each recording session were analysed manually with WaveLab in order to prepare sample blocks of CPA; each having one second long, i.e. 44100 samples, based on listening and observation of signal amplitude. Sample sets for background noise were also prepared.

There were fewer trucks than cars recorded during all the sessions. As Sampan [Sampan, 1997] pointed out, having unequal sample size for each class may result in a biased analysis and indeed influences the measurement of the generalisation ability of a classifier. Although this bias can be useful to have when the expected size of each class population is already known for an application, this is not the case for the current study.

Therefore, some data sets were prepared with sample sets comprising an equal size for each class although it means many data were wasted during the development. Regarding car and noise classes for such cases, the same number of samples found for truck class at one recording session were randomly selected to match the sample size for all three categories. About half from each recording were used as training set. Tables 7.1 to 7.3 list the data set contents. Note, data set B was prepared by using part of those in set A. This is because some results of failure was discovered in the seismic data of recording session V, probably caused by travelling. Therefore results presented in this chapter for any classification examination involving seismic signal processing would be that of either data set B or C.

| | Training Set A | | | Validation Set A | | |
|---|---|---|---|---|---|---|
| Recording | Truck | Car | Noise | Truck | Car | Noise |
| I | 2 | 2 | 2 | 1 | 1 | 1 |
| II | 1 | 1 | 1 | 1 | 1 | 1 |
| III | 8 | 8 | 8 | 7 | 7 | 7 |
| IV | 1 | 1 | 1 | 0 | 0 | 0 |
| V | 19 | 19 | 19 | 19 | 19 | 19 |
| VI | 23 | 23 | 23 | 23 | 23 | 23 |
| Total | 54 | 54 | 54 | 51 | 51 | 51 |

Table 7.1: Data Set A

| | Training Set B | | | Validation Set B | | |
|---|---|---|---|---|---|---|
| Recording | Truck | Car | Noise | Truck | Car | Noise |
| I | 2 | 2 | 2 | 1 | 1 | 1 |
| II | 1 | 1 | 1 | 1 | 1 | 1 |
| III | 8 | 8 | 8 | 7 | 7 | 7 |
| IV | 1 | 1 | 1 | 0 | 0 | 0 |
| V | 19 | 19 | 19 | 19 | 19 | 19 |
| Total | 31 | 31 | 31 | 28 | 28 | 28 |

Table 7.2: Data Set B

| | Training Set C | | | Validation Set C | | |
|---|---|---|---|---|---|---|
| Recording | Truck | Car | Noise | Truck | Car | Noise |
| I | 2 | 23 | 19 | 1 | 22 | 19 |
| II | 1 | 7 | 33 | 1 | 7 | 32 |
| III | 8 | 16 | 54 | 7 | 16 | 54 |
| IV | 1 | 7 | 75 | 0 | 7 | 74 |
| V | 19 | 19 | 28 | 19 | 19 | 28 |
| Total | 31 | 72 | 209 | 28 | 71 | 207 |

Table 7.3: Data Set C

When a range of frame size ere used, each frame was taken with 50% overlap from each of one second block CPA files segmented in advance with WaveLab. Table 7.4 lists how this affects the number of frames that were constructed of one CPA block, depending on the selected frame size.

| Frame Size | 32768 | 1024 | 8192 | 4096 | 2048 | 1024 |
|---|---|---|---|---|---|---|
| #Frame(s) per File | 1 | 4 | 9 | 20 | 42 | 85 |

Table 7.4: Frame Size and Number of Frames Per File (#Frames(s) per File) with 50% overlap

### 7.1.6 Implementation

Algorithms have been implemented in MATLAB. Where an original function was written by the author, it was evaluated in either or both of the following ways;

(a) Comparison with a hand calculation.

(b) Comparison with the MATLAB built-in function.

## 7.2 Detection Algorithms

The reason why the fusion of acoustic and seismic signal processing became the option for the current research was, based on collected evidence, the use of seismic sensor was believed to enhance the performance of acoustic vehicle recognition system while adding little costs in terms of finance, time and computation (Chapters 1 to 3). In order to implement this in practice, collected seismic signals were examined, and features extracted from them were obtained. As a consequence, a vehicle detection algorithm that utilises seismic signals only was developed during the project, and this section reports about the findings.

The optimum system can have a structure such as in Figure 7.1, consisting of two steps, first for detecting vehicle existence and second to classify the vehicle. During the preliminary study, the high potential of an event detection algorithm consisting of seismic features obtained with Log Energy (Section 5.1.3) and TDSC (Section 5.1.5) was discovered. As well as their accuracy, these methods were especially preferred because of their advantages of simplicity and inexpensive computation.

### 7.2.1 Seismic Log Energy

If the energy function of acoustic signals is used, like others attempted [Sobreira-Seoane et al., 2008], the measurement of background noise level needs to be carried out under various conditions. Since acoustic signals can be more susceptible to many factors that could alter the characteristics; including the weather and the variation of human activities in the surrounding area during a day or over different seasons. The use of seismic signals on the other hand, may reduce the amount of measurement required to find a suitable threshold at a location although they are still influenced by some factors, such as humidity or dampness of the ground surface property (Section 3.2.3).

Figures 7.2 to 7.4 and Table 7.5 show Log Energy distribution and the statistical data (mean $\mu_{En}$, median $\tilde{x}_{En}$ and standard deviation $\sigma_{En}$) collected from the training set A for each of the 3 frame length variations with 50% overlap.

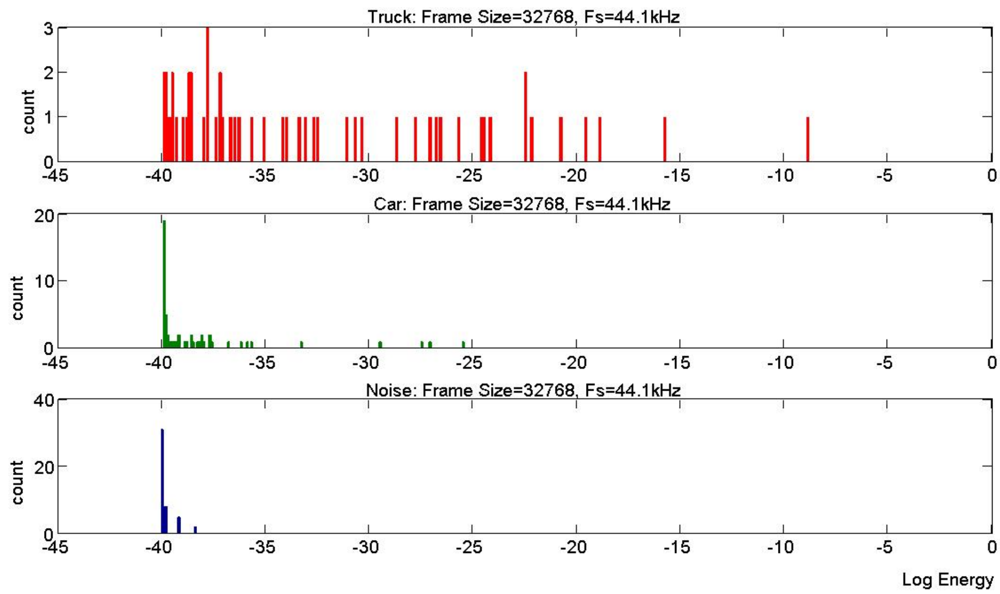


Figure 7.2: Distribution of Seismic Log Energy of Truck (top), Car (middle) and Noise (bottom), Frame Size=32768, Fs=44.1kHz.

Figures 7.2 to 7.4 illustrate how useful the seismic Log Energy could be, for both vehicle detection and classification. For detection, SVMs were implemented using a SVM toolbox called LIBSVM [Chang and Lin, 2001] with RBF kernel on data set B. Note, this toolbox with the kernel was utilised for all the SVM analyses presented in the thesis. The results for each frame size is shown in Table 7.6.

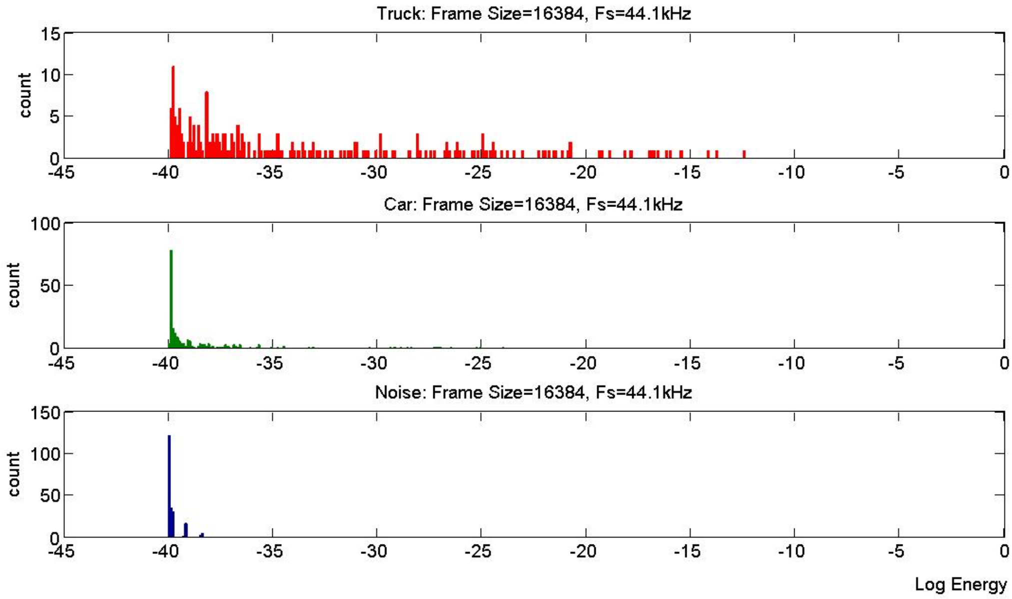


Figure 7.3: Distribution of Seismic Log Energy of Truck (top), Car (middle) and Noise (bottom), Frame Size=16384, Fs=44.1kHz.

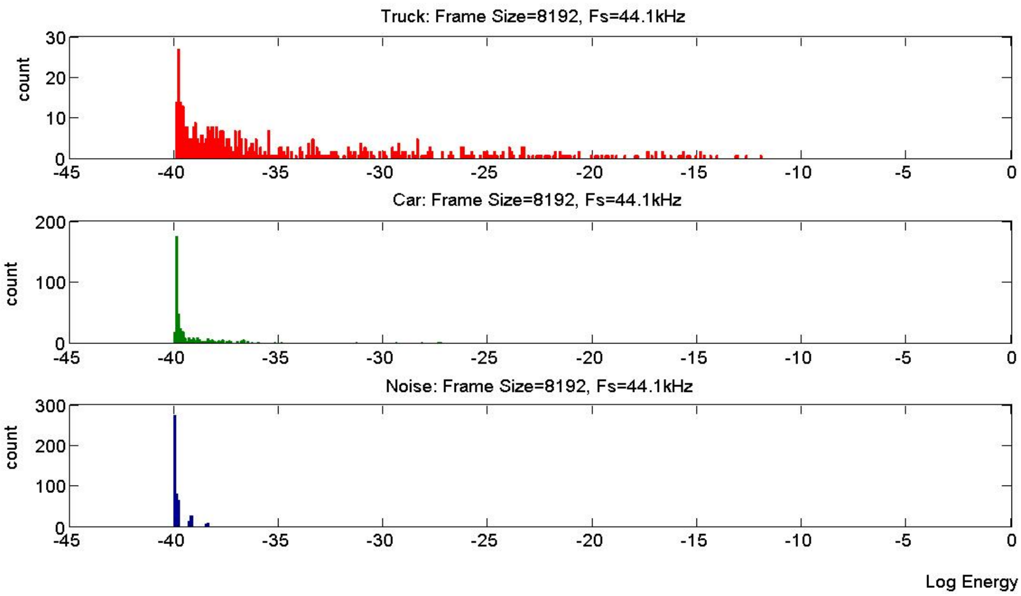


Figure 7.4: Distribution of Seismic Log Energy of Truck (top), Car (middle) and Noise (bottom), Frame Size=8192, Fs=44.1kHz.

| | Truck (seismic) | | | Car (seismic)) | | | Noise (seismic)) | | |
|---|---|---|---|---|---|---|---|---|---|
| Frame | $\mu_{En}$ | $\tilde{x}_{En}$ | $\sigma_{En}$ | $\mu_{En}$ | $\tilde{x}_{En}$ | $\sigma_{En}$ | $\mu_{En}$ | $\tilde{x}_{En}$ | $\sigma_{En}$ |
| 32768 | -32.67 | -35.47 | 7.59 | -38.34 | -39.59 | 3.05 | -39.95 | -39.98 | 0.11 |
| 16384 | -33.17 | -36.24 | 7.08 | -38.51 | -39.69 | 2.95 | -39.94 | -39.98 | 0.10 |
| 8192 | -33.42 | -36.55 | 7.11 | -38.57 | -39.81 | 2.93 | -39.95 | -39.98 | 0.10 |
| 4096 | -33.64 | -36.92 | 7.13 | -38.63 | -39.83 | 2.92 | -39.94 | -39.98 | 0.11 |
| 2048 | -33.87 | -37.12 | 7.10 | -38.71 | -39.87 | 2.84 | -39.95 | -39.99 | 0.11 |

Table 7.5: Statistical Data of Seismic Log Energy: Truck, Car and Noise Samples in Training Set A.

| Frame Size | Training Set B | Validation Set B |
|---|---|---|
| 32768 | 93.55 % (87/93) | 92.86 % (78/84) |
| 16384 | 93.82 % (349/372) | 91.37 % (307/336) |
| 8192 | 83.99 % (703/837) | 82.14 % (621/756) |

Table 7.6: Detection Results by Seismic Log Energy Only

With the bigger dataset C, slightly higher values were observed for all frame sizes (Table 7.7).

| Frame Size | Training Set C | Validation Set C |
|---|---|---|
| 32768 | 98.08 % (306/312) | 97.71 % (299/306) |
| 16384 | 97.76 % (1220/1248) | 97.39 % (1192/1224) |
| 8192 | 97.79 % (2746/2808) | 97.28 % (2679/2754) |

Table 7.7: Detection Results by Seismic Log Energy Only

Although it is not 100 % accuracy that may be preferred for security appllications, the results have demonstrated the positive outcomes of vehicle detection by seismic Log Energy only.

### 7.2.2 Seismic TDSC Modification

For TDSC, graphs (Figures 7.5 to 7.7) were compared by visual observation, so as to identify the best frame size setting for feature extraction.

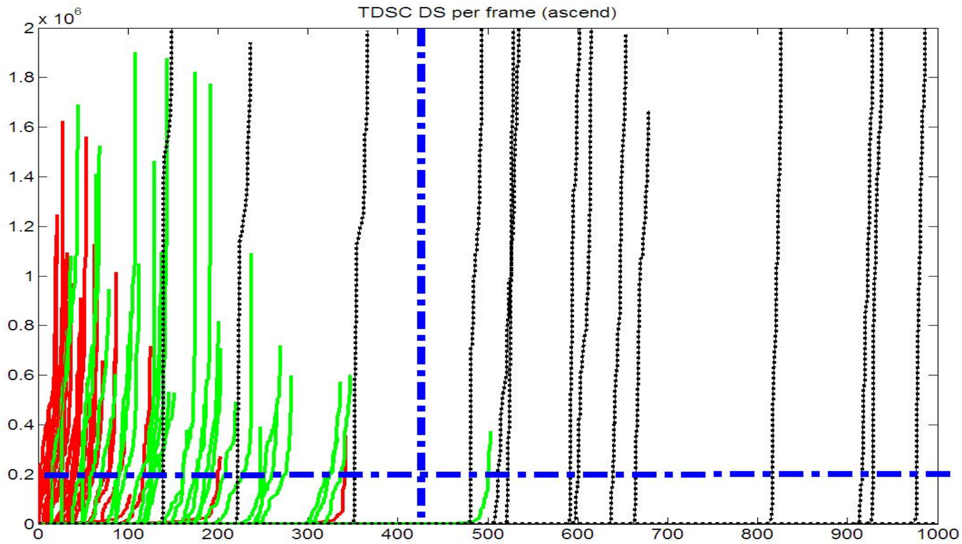


Figure 7.5: TDSC Sorted DS Products of Truck (red), Car (green) and Noise (black dotted line) of Seismic Signals from Training Set B. Frame Size=32768, Fs=44.1kHz.

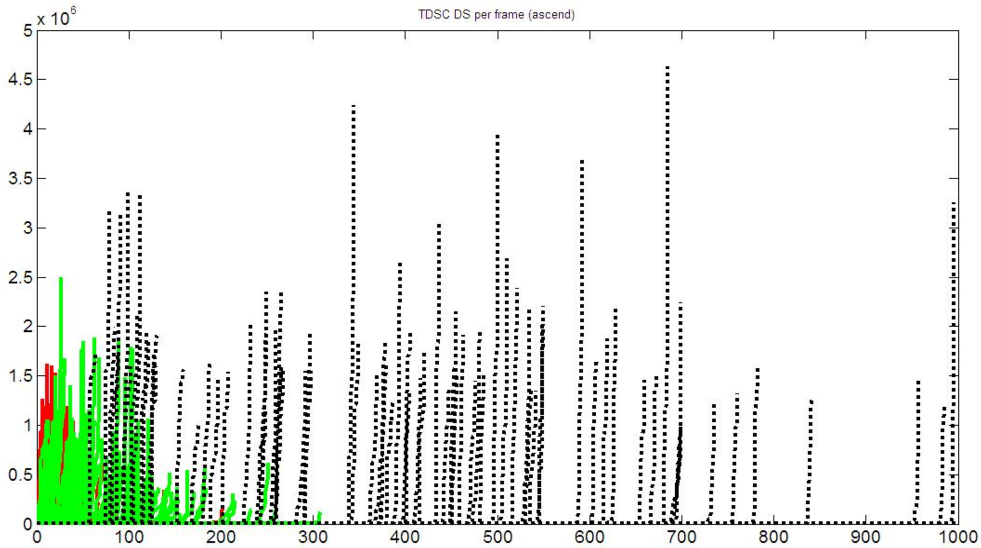


Figure 7.6: TDSC sorted DS Products of Truck (red), Car (green) and Noise (black dotted line) of Seismic Signals from Training Set B. Frame Size=16384, Fs=44.1kHz.

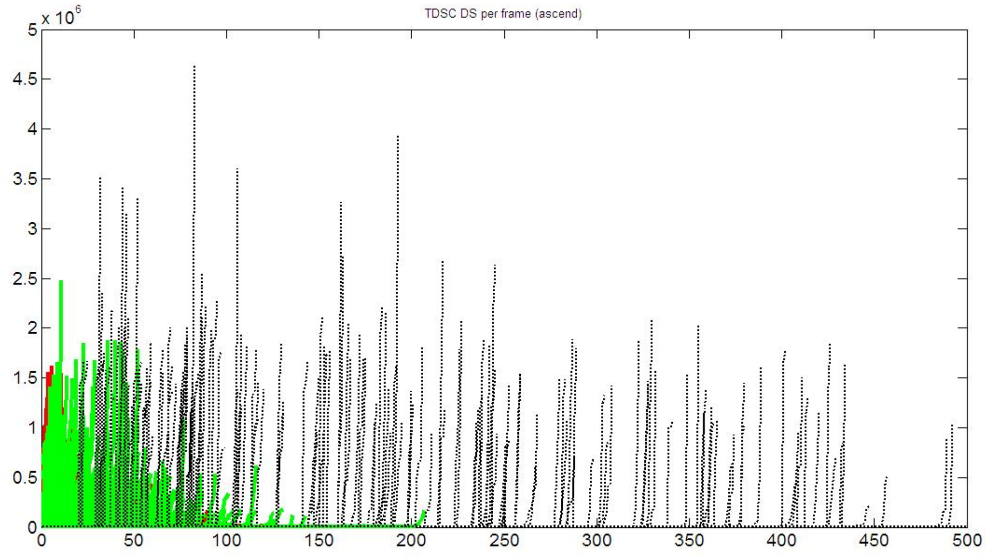


Figure 7.7: TDSC Sorted DS Products of Truck (red), Car (green) and Noise (black dotted line) of Seismic Signals from Training Set B. Frame Size=8192, Fs=44.1kHz.

The blue dashed line in Figures 7.5 indicates the position of a suggested threshold for detecting vehicle signals. Although this empirical process to find the optimum threshold for detecting the presence of vehicle at an arbitrary location can be seen time consuming, it should be much easier than other detection algorithms using acoustic signals. Nonetheless, the smaller the frame size, the less apparent the optimum threshold point becomes. Another problem addressed was how to implement automated classification with this method as TDSC sorted $DS$ products vary in length and range for each processed frame. Then, the following approach was established.

Once sorted $DS$ product of each epoch is collected for a frame, it is used to generate a $DS$ detection vector according to a set of pre-defined thresholds.

(1 ) Determine a segmentation point in a plotted graph by observation as indicated earlier by the lines in Figure 7.5.

(2 ) Collect the coordinate of the position against epoch (x-axis) and $DS$ values (y-axis); as $(\Xi, \Upsilon)$.

(3 ) With the $\Xi$, firstly check whether the number of total epochs found in a frame is less than the threshold. If so, the frame is classed as vehicle.

(4 ) Secondly, see if the $\Xi th$ entry of the sorted $DS$ values is greater than the $\Upsilon$. If so, the frame is classed as vehicle.

(5 ) Finally store a value in a $DS$ detection vector as below, based on the decisions made.

$$DSvector = \begin{cases} 1 & vehicle \\ -1 & nonvehicle \end{cases} \tag{7.1}$$

$DS$ detection vector can be put into a classifier. Table 7.8 shows the SVM classification performance achieved by this seismic TDSC method only. The decline of the accuracy for the frame size 8192 could largely depend on the difficulty in setting the threshold values. Therefore, the use of larger frames would be recommended although it may depend on other conditions related to an application, such as type and speed of target vehicles since seismic signals of smaller (lighter) and fast moving vehicles tend not to last long, thus shorter frames can become an inevitable choice.

| Frame Size | Training Set B | Validation Set B |
|---|---|---|
| 32768 | 93.55 % (87/93) | 92.86 % (78/84) |
| 16384 | 93.82 % (349/372) | 91.37 % (307/336) |
| 8192 | 83.99 % (703/837) | 82.14 % (621/756) |

Table 7.8: Detection Results by Seismic TDSC Only

The results were improved when the total amount of data was increased as shown in Table 7.9.

| Frame Size | Training Set C | Validation Set C |
|---|---|---|
| 32768 | 98.08 % (306/312) | 96.73 % (296/306) |
| 16384 | 94.79 % (1183/1248) | 94.44 % (1156/1224) |
| 8192 | 88.68 % (2490/2808) | 88.60 % (2440/2754) |

Table 7.9: Detection Results by Seismic TDSC Only

### 7.2.3 Combination of TDSC and Log Energy

It was sometimes found that the signals that were incorrectly classified by one method tended to be classified correctly by another. Hence, it was anticipated that combining these two methods together could yield a fairly accurate detection performance, hopefully better than either of the single method cases above. However, it is not trivial to identify the best weighting one should allocate to each method. Therefore, features collected from the two methods were passed onto a classifier as two-dimensional input as they were, one dimension for each.

Table 7.10 and 7.11 show the performance by SVM for data sets B and C, whereas Table 7.12 and 7.13 are those processed by LVQ with various node settings.

| Frame Size | Training Set B | Validation Set B |
|---|---|---|
| 32768 | 95.70 % (89/93) | 92.86 % (78/84) |
| 16384 | 93.28 % (347/372) | 91.37 % (307/336) |
| 8192 | 90.44 % (757/837) | 88.89 % (672/756) |

Table 7.10: SVM Detection Results by Combination of Seismic Log Energy and TDSC. Data Set B.

| Frame Size | Training Set C | Validation Set C |
|---|---|---|
| 32768 | 98.72 % (308/312) | 97.71 % (299/306) |
| 16384 | 97.76 % (1220/1248) | 97.47 % (1193/1224) |
| 8192 | 97.83 % (2747/2808) | 97.28 % (2679/2754) |

Table 7.11: SVM Detection Results by Combination of Seismic Log Energy and TDSC. Data Set C.

| Node | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|---|
| Training | 100% | 66.67% | 66.67% | 66.67% | 66.67% | 66.67% | 66.67% | 66.67% |
| Validation | 96.43% | 66.67% | 66.67% | 66.67% | 66.67% | 66.67% | 66.67% | 66.67% |

Table 7.12: LVQ Detection Results by Combination of Seismic Log Energy and TDSC. Data Set B, Frame Size=32768.

| Node | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|---|
| Training | 94.55% | 98.72% | 98.72% | 98.72% | 98.72% | 99.68% | 97.75% | 98.72% |
| Validation | 95.76% | 97.71% | 97.71% | 97.71% | 97.71% | 99.02% | 96.41% | 97.71% |

Table 7.13: LVQ Detection Results by Combination of Seismic Log Energy and TDSC. Data Set C, Frame Size=32768.

By comparing these tables, first of all, a part from Table 7.12, generally high accuracies were attained. At the same time, Table 7.12 perhaps indicates susceptibility of LVQ to the influence by the dataset inequality when the data size is smaller.

Another experiment was also designed and carried out with SVM to find how the recording conditions influence these results about signals recorded during a single session, regarding 4 different sessions. Since there were not many samples available for each recording, 4-fold cross-validation was conducted. Table 7.14 shows the construction of the data set for each recording as well as the average achieved. The results seem to show little difference between signals collected during at least three out of four sessions although the discrepancy observed particularly for session II could be due to the smaller data size.

| Recording | #Vehicles | #Noise | Training Average | Validation Average |
|---|---|---|---|---|
| I | 135 | 38 | 99.61 % | 99.40 % |
| II | 29 | 65 | 79.00 % | 71.29 % |
| III | 113 | 116 | 94.76 % | 88.72 % |
| IV | 23 | 155 | 98.69 % | 95.39 % |

Table 7.14: SVM Detection Results by Combination of Seismic Log Energy and TDSC. 4-fold Cross-Validation for Each Recording Session. Data Set B, Frame Size=32768.

As introduced earlier in Section 4.2, a noise reduction algorithm can improve the results by reducing unwanted signals varied within the seismic data. For these methods, performance with spectral subtraction (Section 4.2.2) was compared with that of the combination of spectral subtraction plus FIR lowpass filtering (Section 4.2.1). It was found using only the spectral subtraction of estimated noise was more effective in enhancing the distinguishable characteristics of those collected features.

## 7.3 Classification Algorithms

### 7.3.1 Zero-Crossing Rate (ZCR)

The feasibility of ZCR (Section 5.1.2) was investigated with the collected data, and Figure 7.8 depicts the distribution taken from the acoustic training data set A with frames of 32768 samples.

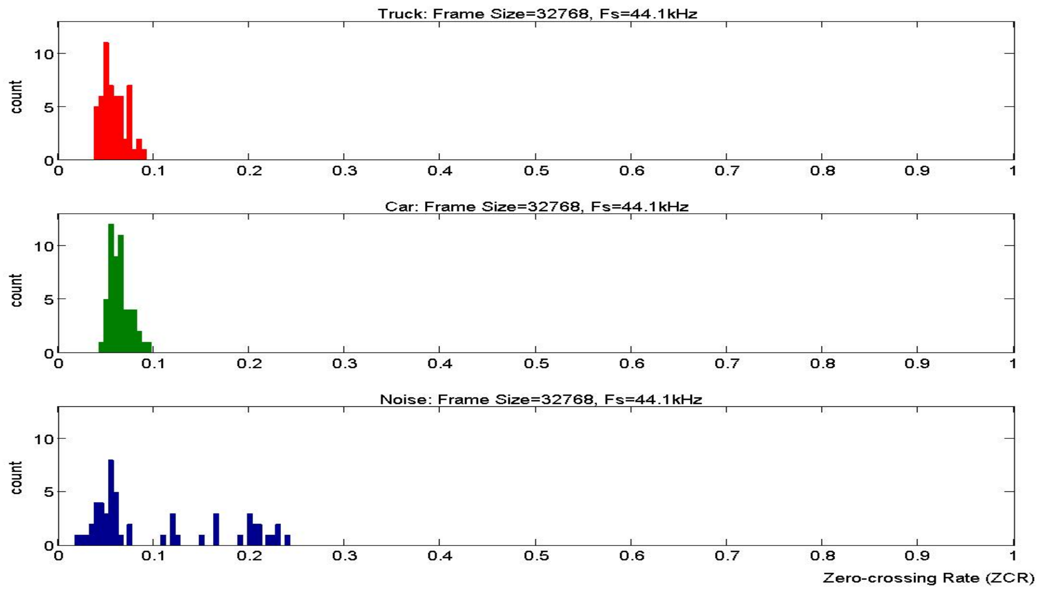


Figure 7.8: Distribution of Acoustic ZCR of Truck (top), Car (middle) and Noise (bottom), Frame Size=32768, Fs=44.1kHz.

From an observation of the Figure 7.8, it was noticed ZCR of noise appears to be more wide spread over the range than either truck or car signal. However, both truck and car seem to occupy similar regions, which may imply little possibility to segment signals of these two classes one from another by using only ZCR. Such distributions were examined for various frame sizes and with a range of pre-processing methods. It was found they all tend to show similar results on graphs (Appendix B.1). Then, some statistical data such as the mean $\mu_{ZCR}$, median $\tilde{x}_{ZCR}$, standard deviation $\sigma_{ZCR}$ about ZCR collected from the training set are listed in Table 7.15.

| | Truck (acoustic) | | | Car (acoustic) | | | Noise (acoustic) | | |
|---|---|---|---|---|---|---|---|---|---|
| Frame | $\mu_{ZCR}$ | $\tilde{x}_{ZCR}$ | $\sigma_{ZCR}$ | $\mu_{ZCR}$ | $\tilde{x}_{ZCR}$ | $\sigma_{ZCR}$ | $\mu_{ZCR}$ | $\tilde{x}_{ZCR}$ | $\sigma_{ZCR}$ |
| 32768 | 0.0586 | 0.0566 | 0.01 | 0.0643 | 0.0624 | 0.01 | 0.1031 | 0.0608 | 0.07 |
| 16384 | 0.0577 | 0.0550 | 0.01 | 0.0638 | 0.0618 | 0.01 | 0.1025 | 0.0595 | 0.07 |
| 8192 | 0.0575 | 0.0547 | 0.01 | 0.0638 | 0.0614 | 0.01 | 0.1018 | 0.0586 | 0.07 |
| 4096 | 0.0572 | 0.0542 | 0.02 | 0.0638 | 0.0615 | 0.01 | 0.1004 | 0.0571 | 0.07 |
| 2048 | 0.0566 | 0.0542 | 0.02 | 0.0635 | 0.0615 | 0.01 | 0.0992 | 0.0547 | 0.07 |
| 1024 | 0.0552 | 0.0527 | 0.02 | 0.0629 | 0.0615 | 0.01 | 0.0966 | 0.0527 | 0.07 |
| 512 | 0.0529 | 0.0508 | 0.02 | 0.0618 | 0.0605 | 0.01 | 0.0917 | 0.0488 | 0.08 |
| 256 | 0.0458 | 0.0430 | 0.02 | 0.0551 | 0.0547 | 0.02 | 0.0762 | 0.0234 | 0.08 |

Table 7.15: Statistical Data of Acoustic ZCR: Truck, Car and Noise Samples in Training Set A.

These data indicate how ZCR of acoustic signals may be beneficial to detecting vehicle presence before performing classification. At the same time, the ability to separate between trucks and cars with ZCR may not be good. It is also hard to identify the best frame size for this feature by these data only. Figure 7.9 and Table 7.16 present the same type of information for those extracted from the seismic signals. Graphs for other frame size variations can be found in Appendix B.1. Although it is not so clear in the distribution figure, the statistical data may imply some good potential of seismic ZCR for classification.

### 7.3.2 Energy / Log Energy

Figure 7.10 and Table 7.17 show graphical examples of the acoustic Log Energy distribution and their statistical data collected from the training set A. Again, other graphs can be found in Appendix B.2.

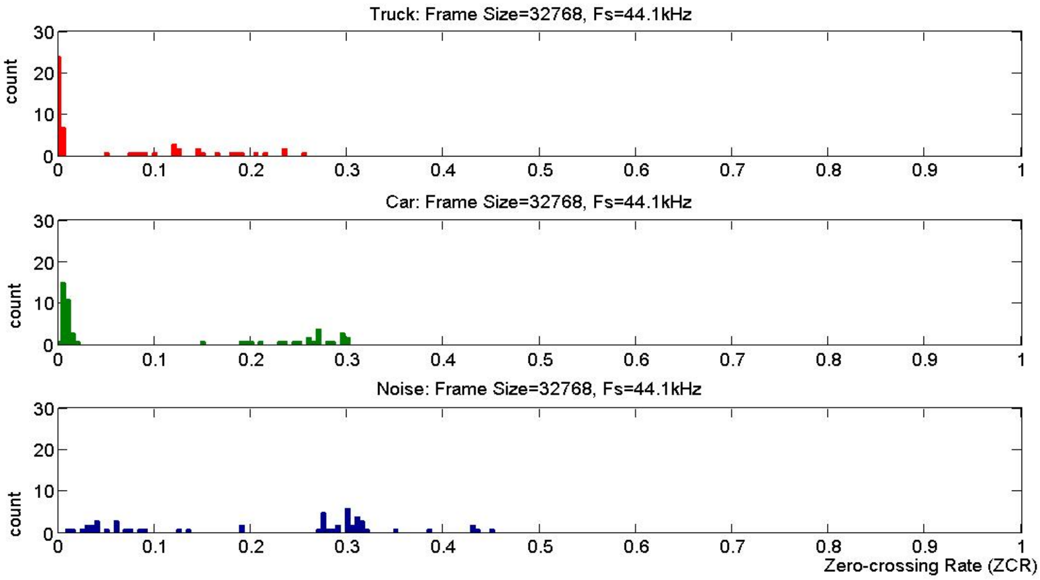


Figure 7.9: Distribution of Seismic ZCR of Truck (top), Car (middle) and Noise (bottom), Frame Size=32768, Fs=44.1kHz.

| | Truck (seismic) | | | Car (seismic)) | | | Noise (seismic)) | | |
|---|---|---|---|---|---|---|---|---|---|
| Frame | $\mu_{ZCR}$ | $\tilde{x}_{ZCR}$ | $\sigma_{ZCR}$ | $\mu_{ZCR}$ | $\tilde{x}_{ZCR}$ | $\sigma_{ZCR}$ | $\mu_{ZCR}$ | $\tilde{x}_{ZCR}$ | $\sigma_{ZCR}$ |
| 32768 | 0.064 | 0.003 | 0.08 | 0.112 | 0.013 | 0.12 | 0.216 | 0.277 | 0.13 |
| 16384 | 0.062 | 0.004 | 0.08 | 0.110 | 0.012 | 0.12 | 0.216 | 0.276 | 0.14 |
| 8192 | 0.063 | 0.04 | 0.08 | 0.111 | 0.013 | 0.12 | 0.217 | 0.277 | 0.14 |
| 4096 | 0.063 | 0.004 | 0.08 | 0.112 | 0.014 | 0.13 | 0.218 | 0.274 | 0.14 |
| 2048 | 0.064 | 0.004 | 0.08 | 0.112 | 0.015 | 0.13 | 0.218 | 0.269 | 0.14 |

Table 7.16: Statistical Data of Seismic ZCR: Truck, Car and Noise Samples in Training Set A.

| | Truck (acoustic) | | | Car (acoustic) | | | Noise (acoustic) | | |
|---|---|---|---|---|---|---|---|---|---|
| Frame Size | $\mu_{En}$ | $\tilde{x}_{En}$ | $\sigma_{En}$ | $\mu_{En}$ | $\tilde{x}_{En}$ | $\sigma_{En}$ | $\mu_{En}$ | $\tilde{x}_{En}$ | $\sigma_{En}$ |
| 32768 | -31.49 | -32.31 | 5.38 | -34.14 | -34.92 | 4.23 | -39.97 | -39.99 | 0.05 |
| 16384 | -31.53 | -32.89 | 5.67 | -34.31 | -35.48 | 4.34 | -39.97 | -39.99 | 0.05 |
| 8192 | -31.73 | -33.25 | 5.70 | -34.48 | -35.70 | 4.33 | -39.97 | -39.99 | 0.05 |
| 4096 | -31.92 | -33.47 | 5.71 | -34.61 | -35.91 | 4.32 | -39.97 | -39.99 | 0.05 |
| 2048 | -32.03 | -33.62 | 5.72 | -34.68 | -35.99 | 4.31 | -39.97 | -39.99 | 0.06 |
| 1024 | -32.10 | -33.65 | 5.72 | -34.72 | -36.04 | 4.32 | -39.97 | -39.99 | 0.06 |
| 512 | -32.16 | -33.75 | 5.74 | -34.77 | -36.10 | 4.32 | -39.97 | -39.99 | 0.07 |
| 256 | -31.51 | -33.16 | 5.99 | -34.49 | -35.76 | 4.38 | -39.97 | -39.99 | 1.51 |

Table 7.17: Statistical Data of Acoustic Log Energy: Truck, Car and Noise Samples in Training Set A.

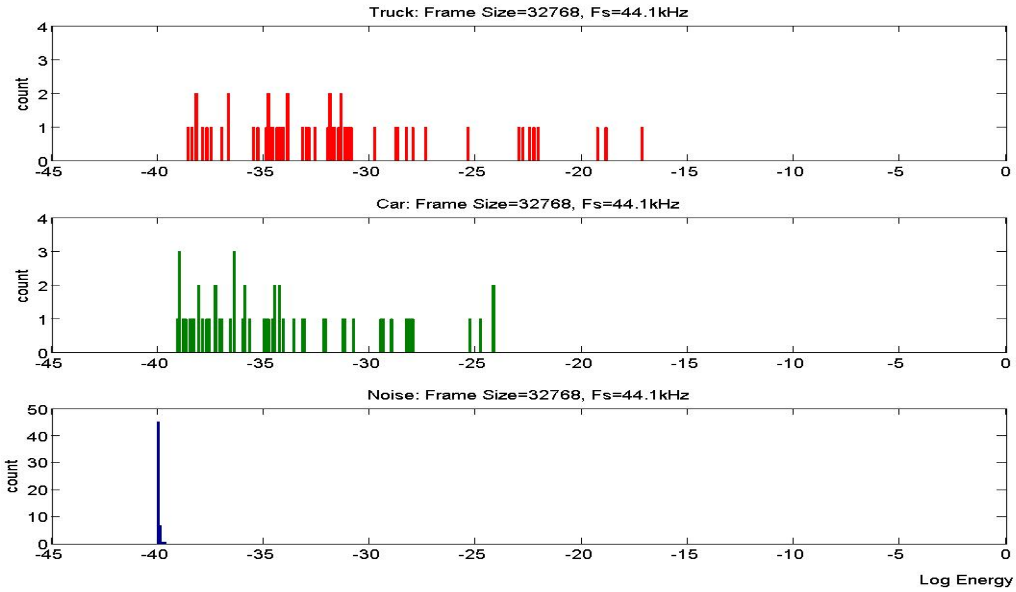


Figure 7.10: Distribution of Acoustic Log Energy of Truck (top), Car (middle) and Noise (bottom), Frame Size=32768, Fs=44.1kHz.

The figure shows that it may be useful to separate between vehicles and noise, secondly Log Energy of some trucks reach higher values than that of most cars, as it can be read from the difference in the statistical data. However, Log Energy of truck and cars are not linearly separable with all examined frame size variations.

### 7.3.3 Time Domain Correlation

The distribution of normalised autocorrelation is seen in Figure 7.11, and Table 7.18 shows statistical data (mean $\mu_{Cor}$, median $\tilde{x}_{Cor}$ and standard deviation $\sigma_{Cor}$) for normalised autocorrelation, both extracted from the acoustic signals in training set A.

The collected features do not seem to exhibit any distinguishable characteristics between different classes. Instead, they show how most of the signals possess high autocorrelation property, apart from some noise samples. This suggest the correlation method can bring little benefit to the present research.

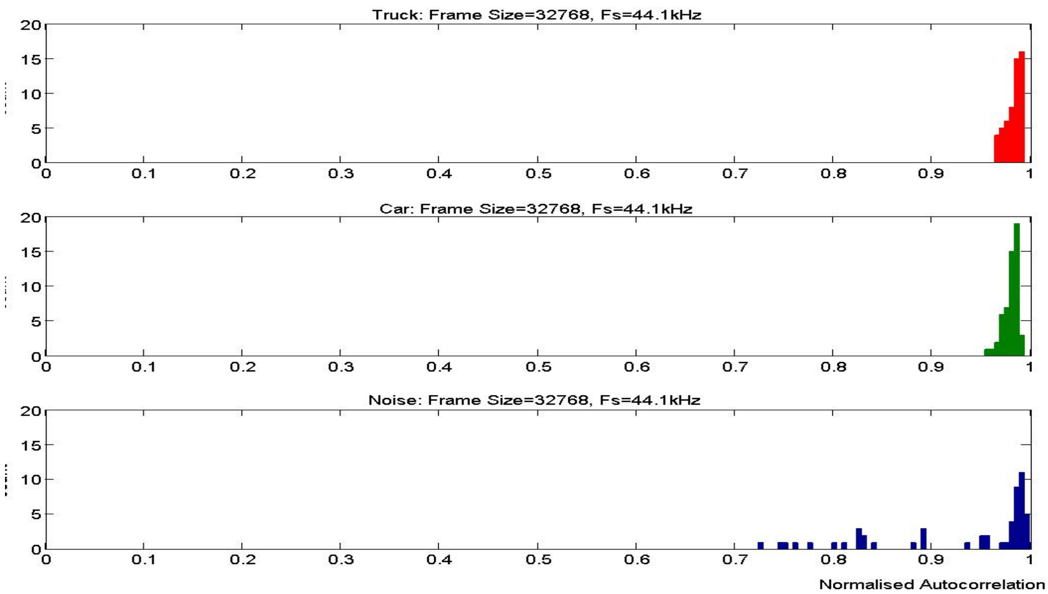


Figure 7.11: Distribution of Acoustic Normalised Autocorrelation of Truck (top), Car (middle) and Noise (bottom), Frame Size=32768, Fs=44.1kHz.

| | Truck (acoustic) | | | Car (acoustic) | | | Noise (acoustic) | | |
|---|---|---|---|---|---|---|---|---|---|
| Frame Size | $\mu_{Cor}$ | $\tilde{x}_{Cor}$ | $\sigma_{Cor}$ | $\mu_{Cor}$ | $\tilde{x}_{Cor}$ | $\sigma_{Cor}$ | $\mu_{Cor}$ | $\tilde{x}_{Cor}$ | $\sigma_{Cor}$ |
| 32768 | 0.982 | 0.983 | 0.01 | 0.979 | 0.981 | 0.01 | 0.930 | 0.982 | 0.08 |
| 16384 | 0.982 | 0.985 | 0.01 | 0.979 | 0.981 | 0.01 | 0.930 | 0.982 | 0.08 |
| 8192 | 0.982 | 0.985 | 0.01 | 0.979 | 0.981 | 0.01 | 0.930 | 0.982 | 0.08 |
| 4096 | 0.982 | 0.985 | 0.01 | 0.979 | 0.981 | 0.01 | 0.930 | 0.983 | 0.08 |
| 2048 | 0.982 | 0.985 | 0.01 | 0.979 | 0.981 | 0.01 | 0.930 | 0.984 | 0.08 |
| 1024 | 0.982 | 0.985 | 0.01 | 0.979 | 0.981 | 0.01 | 0.929 | 0.984 | 0.09 |
| 512 | 0.981 | 0.985 | 0.01 | 0.978 | 0.980 | 0.01 | 0.928 | 0.982 | 0.09 |
| 256 | 0.982 | 0.986 | 0.01 | 0.979 | 0.981 | 0.01 | 0.935 | 0.984 | 0.09 |

Table 7.18: Statistical Data of Acoustic Autocorrelation: Truck, Car and Noise Samples in Training Set A.

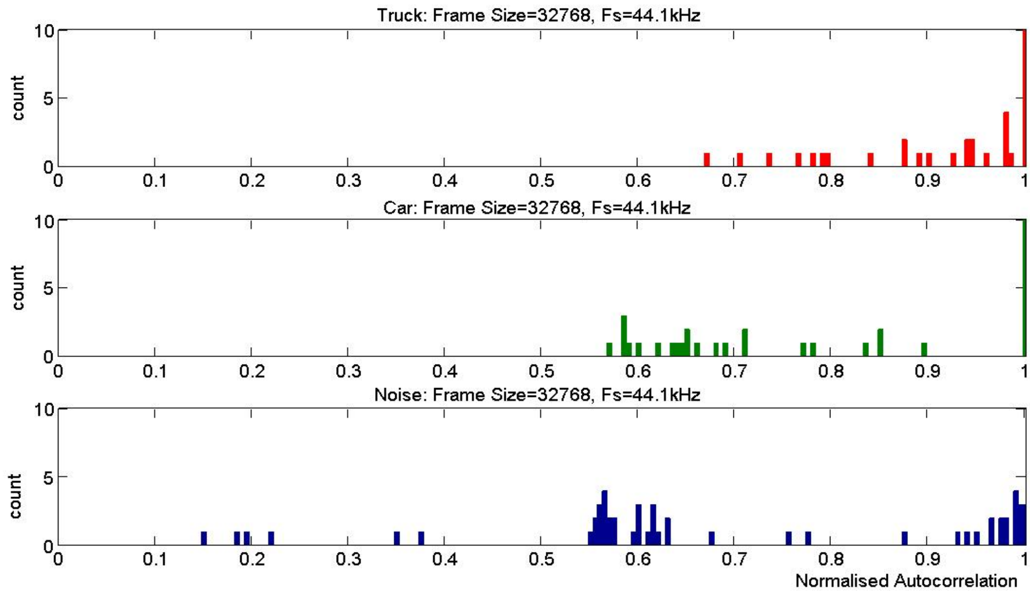


Figure 7.12: Distribution of Seismic Normalised Autocorrelation of Truck (top), Car (middle) and Noise (bottom), Frame Size=32768, Fs=44.1kHz.

| | Truck (seismic) | | | Car (seismic) | | | Noise (seismic) | | |
|---|---|---|---|---|---|---|---|---|---|
| Frame Size | $\mu_{Cor}$ | $\tilde{x}_{Cor}$ | $\sigma_{Cor}$ | $\mu_{Cor}$ | $\tilde{x}_{Cor}$ | $\sigma_{Cor}$ | $\mu_{Cor}$ | $\tilde{x}_{Cor}$ | $\sigma_{Cor}$ |
| 32768 | 0.948 | 1.000 | 0.09 | 0.866 | 0.999 | 0.17 | 0.700 | 0.615 | 0.24 |
| 16384 | 0.947 | 1.000 | 0.09 | 0.865 | 0.999 | 0.17 | 0.694 | 0.617 | 0.26 |
| 8192 | 0.944 | 1.000 | 0.09 | 0.861 | 0.999 | 0.17 | 0.688 | 0.617 | 0.26 |
| 4096 | 0.942 | 1.000 | 0.09 | 0.859 | 0.999 | 0.18 | 0.679 | 0.619 | 0.26 |
| 2048 | 0.939 | 1.000 | 0.10 | 0.856 | 0.998 | 0.18 | 0.666 | 0.614 | 0.27 |

Table 7.19: Statistical Data of Seismic Autocorrelation: Truck, Car and Noise Samples in Training Set A.

### 7.3.4 Linear Predictive Coding (LPC)

A comparative study on 2-class (between trucks and cars) classification ability of LPC coefficients and the minimum error with different order $p$; such as 12, 24, 36, and 48, combined with various pre-processing methods and frame sizes, as listed in Table 7.20 was conducted.

This experiment was carried out with SVM using RBF kernel on data set A. Once

| Pre-processing Method | Frame Size |
|---|---|
| Window Framing | 32768 |
| FIR filter | 16384 |
| Spectral Subtraction | 8192 |
| Wavelet Packet | 4096 |
| Spectral Subtraction & FIR filter | 2048 |
| n/a | 1024 |

Table 7.20: List of Pre-Processing and Frame Size Variations

collected, the feature vectors were either, input into the classifier as they were, normalised to -1 and 1 range, or reduced to 3-dimension by PCA before classification. Consequently, the following was found to be better than the others.

- LPC order $p$: either 12 or 24.
- Pre-processing method: spectral subtraction of estimated noise.
- Frame Size: either 32768, 16384, or 8192; with subtle difference with others.
- Pre-classification treatment: either no processing or normalised to -1 and +1.

Table 7.21 and 7.22 show the achieved validation accuracy for the best and worst cases respectively. Since the higher the order $p$, the greater the computational demand, the best and worst choices of $p$ were determined as 12 and 48 correspondingly.

| Frame Size | 32768 | 16384 | 8192 | 4096 | 2048 | 1024 |
|---|---|---|---|---|---|---|
| no-processing | 78.70% | 73.28% | 73.64% | 72.35% | 71.24% | 70.35% |
| normalised | 75.93% | 72.06% | 67.32% | 65.74% | 65.27% | 67.88% |
| PCA(3) | 52.94% | 73.28% | 70.48% | 43.87% | 70.59% | 69.23% |
| normalised & PCA(3) | 74.07% | 71.08% | 70.81% | 69.71% | 70.05% | 51.22% |

Table 7.21: LPC (p=12) Classification (2-class) Results with Spectral Subtraction Only. (The numbers in brackets for PCA indicate the number of principle components kept.)

| Frame Size | 32768 | 16384 | 8192 | 4096 | 2048 | 1024 |
|---|---|---|---|---|---|---|
| no-processing | 50.00% | 48.28% | 47.39% | 47.50% | 48.55% | 51.28% |
| normalised | 50.98% | 55.15% | 54.14% | 55.74% | 53.64% | 53.82% |
| PCA(3) | 52.94% | 51.47% | 51.63% | 52.01% | 54.30% | 55.67% |
| normalised & PCA(3) | 54.90% | 56.37% | 52.51% | 52.01% | 48.95% | 48.09% |

Table 7.22: LPC (p=48) Classification (2-class) Results with Spectral Subtraction plus FIR Filtering. (The numbers in brackets for PCA indicate the number of principle components kept.)

LPC coefficients with $p = 12$ were also classified by LVQ using mean squared error and three variations of learning rate (in the "Rate" column), as shown in in Table 7.23 (for frame size 32768) and 7.24 (for 16384).

| Rate | 10 nodes | 20 nodes | 30 nodes | 40 nodes | 50 nodes | 60 nodes | 70 nodes |
|---|---|---|---|---|---|---|---|
| 0.005 | 69.61% | 71.57% | 67.65% | 69.61% | 57.84% | 73.53% | 61.76% |
| 0.01 | 70.59% | 67.65% | 67.65% | 73.53% | 71.57% | 68.63% | 69.61% |
| 0.02 | 70.59% | 73.53% | 73.53% | 74.51% | 66.67% | 73.53% | 69.61% |

Table 7.23: LPC (p=12) LVQ Classification (2-class) Results with Spectral Subtraction Pre-Processing, Frame Size=32768.

| Rate | 10 nodes | 20 nodes | 30 nodes | 40 nodes | 50 nodes | 60 nodes | 70 nodes |
|---|---|---|---|---|---|---|---|
| 0.005 | 66.91% | 69.85% | 68.63% | 70.10% | 69.85% | 69.36% | 68.38% |
| 0.01 | 69.17% | 68.14% | 67.65% | 69.61% | 64.65% | 66.42% | 67.89% |
| 0.02 | 67.40% | 66.18% | 64.46% | 66.91% | 66.67% | 65.93% | 67.16% |

Table 7.24: LPC (p=12) LVQ Classification (2-class) Results with Spectral Subtraction Pre-Processing, Frame Size=16384.

Example confusion matrix and ROC are shown in Table 7.25 and Figure 7.13 respectively for LVQ classification. They both indicate how the classifier's output was biased towards one class.

It is known that LPC feature extraction algorithm is suitable for "estimating the basic speech parameters, e.g. pitch, formants, spectra, vocal tract area functions" [Rabiner and Schafer, 1978, p.396]. As it has been mentioned in Chapter 2, LPC was

| | Target Class: Truck | Target Class: Car |
|---|---|---|
| Output Class: Truck | 30 | 8 |
| Output Class: Car | 24 | 46 |

Table 7.25: Example Confusion Matrix of LPC LVQ Classification (2-class) with Spectral Subtraction Pre-Processing, Node=40, Learning Rate=0.01, Frame Size=32768.

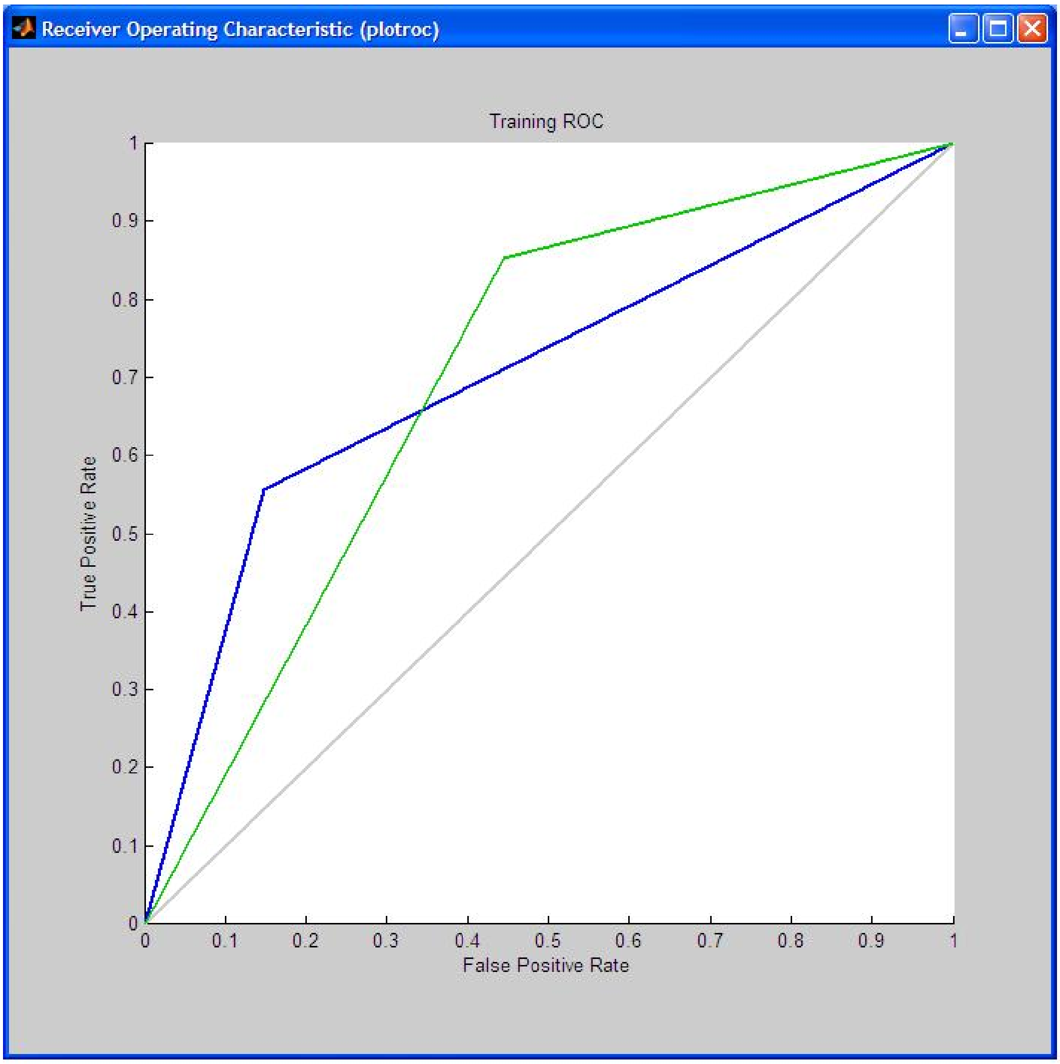


Figure 7.13: Example ROC curve of LPC LVQ Classification (2-class) with Spectral Subtraction Pre-Processing, Node=40, Learning Rate=0.01, Frame Size=32768.

used in other vehicle recognition studies. Some researchers stated acoustic signals of vehicles are similar to that of speech. Nevertheless, to the author's knowledge, no scientific justification for such argument was offered so far, and many simply mentioned the popularity and positive prospects demonstrated in speech recognition. The relatively good results shown above with LPC, however, may relate to what has been shown in Section 4.2 regarding the observation of resonance in some truck and car signals.

### 7.3.5 Co-Occurrence Matrix

Three of the five descriptor functions, in particular, showed some positive outcome of the co-occurrence matrix during the preliminary study. However, with more data, inferior performance was observed with the method.

Using the variations listed in Table 7.20, a search for the optimum pre-processing and frame size was carried out regarding Co-Occurrence Matrix descriptors, for three lag $k$ values: 1, 20 and 256. Differences between such variations were relatively smaller compared to the cases with other feature extraction methods. None of the achieved accuracy was above 70%. The discrepancies due to the setting of lag $k$ was subtle as well; the best and worst cases are in Table 7.26.

| | Best | Worst |
|---|---|---|
| Pre-processing Method | Spectral Subtraction & FIR filter | Spectral Subtraction |
| Frame Size | 32768 | 1024 |
| pre-classification processing | none | normalised & PCA |
| Lag k | 20 | 20 |
| Results (2-class) | 65.69 % | 44.82 % |

Table 7.26: Best and Worst for Co-Occurrence Matrix Descriptors.

The best case was also examined with LVQ classifier as shown in Table 7.27. Overall, these are slightly worse than that of SVM.

| Rate | 10 nodes | 20 nodes | 30 nodes | 40 nodes | 50 nodes | 60 nodes | 70 nodes |
|---|---|---|---|---|---|---|---|
| 0.005 | 58.82% | 57.84% | 57.84% | 58.82% | 57.84% | 58.82% | 54.90% |
| 0.01 | 58.82% | 54.90% | 58.82% | 57.84% | 57.84% | 58.82% | 57.84% |
| 0.02 | 55.88% | 58.82% | 55.88% | 54.90% | 57.84% | 57.84% | 58.82% |

Table 7.27: Co-Occurrence Matrix LVQ Classification (2-class) Results with Spectral Subtraction plus FIR Bandpass Filter Pre-Processing, Lag k=20, Frame Size=32768.

Table 7.28 and Figure 7.14 depict the confusion matrix and ROC for the best case achieved with LVQ.

| | Target Class: Truck | Target Class: Car |
|---|---|---|
| Output Class: Truck | 8 | 0 |
| Output Class: Car | 46 | 54 |

Table 7.28: Example Confusion Matrix of Co-Occurrence Matrix LVQ Classification (2-class) with Spectral Subtraction plus FIR Pre-Processing, Node=140, Learning Rate=0.01, Frame Size=32768.

### 7.3.6 ERB Filterbank

With the same set of variations regarding pre-processing methods and frame size as in Table 7.20, the best and worst cases respectively for the filterbank method in terms of the achieved accuracy are shown in Table 7.29 and 7.30.

| Frame Size | 32768 | 16384 | 8192 | 4096 | 2048 | 1024 |
|---|---|---|---|---|---|---|
| no-processing | 50.00% | 50.00% | 50.76% | 54.56% | 60.60% | 70.65% |
| normalised | 56.86% | 58.33% | 58.14% | 57.84% | 58.52% | 58.43% |
| PCA(3) | 74.07% | 71.08% | 70.81% | 69.71% | 70.05% | 49.80% |
| normalised & PCA(3) | 54.90% | 50.49% | 62.64% | 59.80% | 57.84% | 57.83% |

Table 7.29: Best: ERB Filterbank Classification (2-class) Results with Wavelet Packet Pre-Processing. (The numbers in brackets for PCA indicate the number of principle components kept.)

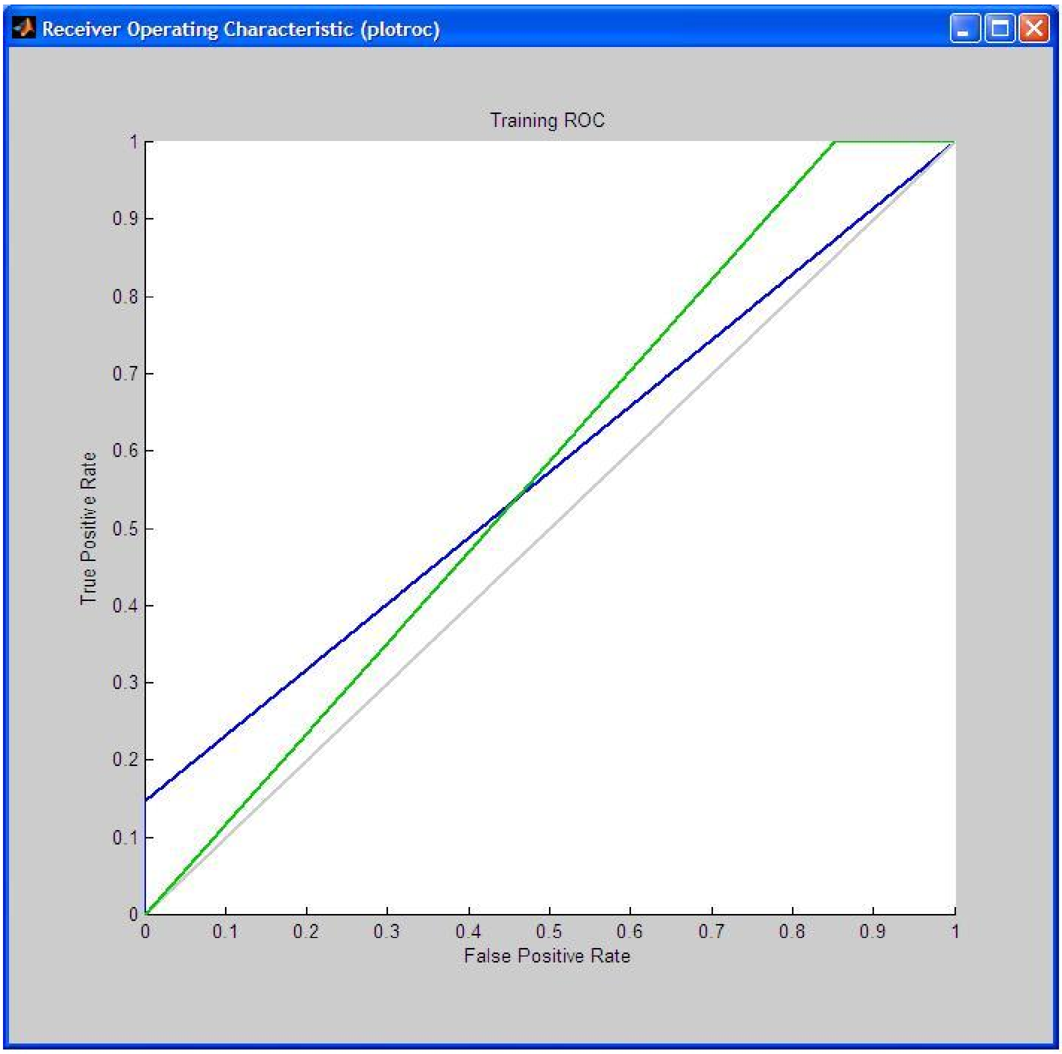


Figure 7.14: Example ROC of Co-Occurrence Matrix LVQ Classification (2-class) with Spectral Subtraction plus FIR Pre-Processing, Node=140, Learning Rate=0.01, Frame Size=32768.

| Frame Size | 32768 | 16384 | 8192 | 4096 | 2048 | 1024 |
|---|---|---|---|---|---|---|
| no-processing | 50.00% | 50.00% | 50.00% | 50.39% | 55.16% | 61.73% |
| normalised | 58.82% | 59.31% | 60.02% | 60.00% | 60.11% | 60.00% |
| PCA(3) | 50.00% | 50.00% | 50.00% | 50.00% | 47.67% | 49.98% |
| normalised & PCA(3) | 58.82% | 50.25% | 50.00% | 49.02% | 47.67% | 58.11% |

Table 7.30: Worst: ERB Filterbank Classification (2-class) Results with FIR Band-pass Filtering. (The numbers in brackets for PCA indicate the number of principle components kept.)

Table 7.31 is the results with LVQ classifier for ERB filterbank features.

| Rate | 10 nodes | 20 nodes | 30 nodes | 40 nodes | 50 nodes | 60 nodes | 70 nodes |
|---|---|---|---|---|---|---|---|
| 0.005 | 56.86% | 57.84% | 58.82% | 58.82% | 57.84% | 56.86% | 58.82% |
| 0.01 | 56.86% | 52.88% | 57.84% | 56.86% | 59.80% | 56.86% | 56.86% |
| 0.02 | 55.88% | 58.82% | 57.84% | 57.84% | 58.82% | 59.80% | 58.82% |

Table 7.31: ERB Filterbank LVQ Classification (2-class) Results with Wavelet Packet Pre-Processing, Frame Size=32768.

Again, Table 7.32 and Figure 7.15 are example confusion matrix and ROC for the best case achieved with LVQ.

| | Target Class: Truck | Target Class: Car |
|---|---|---|
| Output Class: Truck | 9 | 0 |
| Output Class: Car | 45 | 54 |

Table 7.32: Example Confusion Matrix of ERB Filterbank LVQ Classification (2-class) with Wavelet Packet Pre-Processing, Node=50, Learning Rate=0.01, Frame Size=32768.

### 7.3.7 Combination of Acoustic and Seismic

So far, different feature extraction methods have been examined individually so as to identify those with higher potential. In order to obtain better results, combinational effects were also examined. After investigating the three simple time domain feature extraction methods, the feasibility of the combination of them were investigated for both acoustic and seismic signals although the initial indication led to an expectation that the acoustic signal ones would not be good. Figures 7.16 and 7.17 show graphs of three methods plotted together. After all, the separation between truck and car and indeed noise by using only these features extracted from acoustic signals did not appear to be feasible although the advantage of inexpensive computation was attractive. However, that of seismic signals appeared to be much better, therefore further investigation was performed about these.

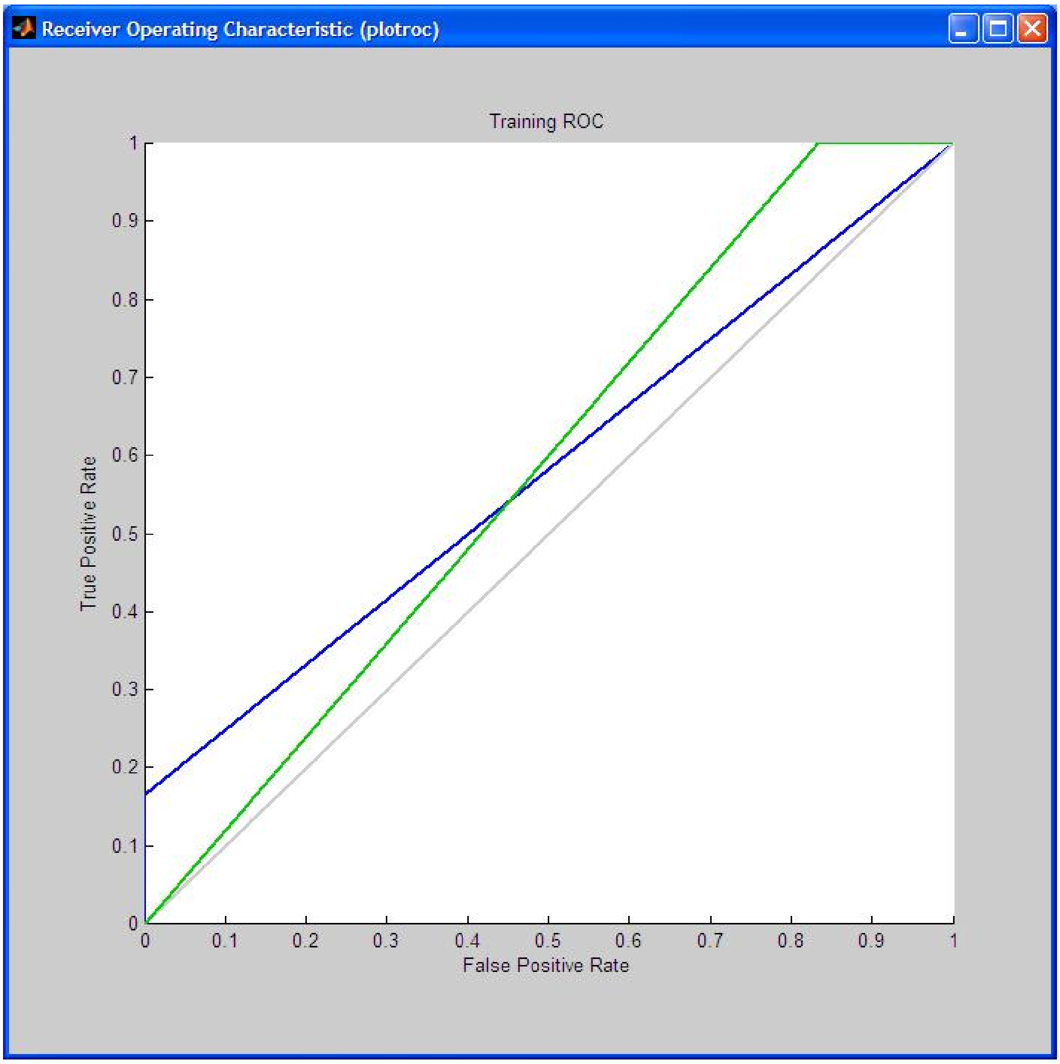


Figure 7.15: Example ROC Curve of ERB Filterbank LVQ Classification (2-class) with Wavelet Packet Pre-Processing, Node=50, Learning Rate=0.01, Frame Size=32768.

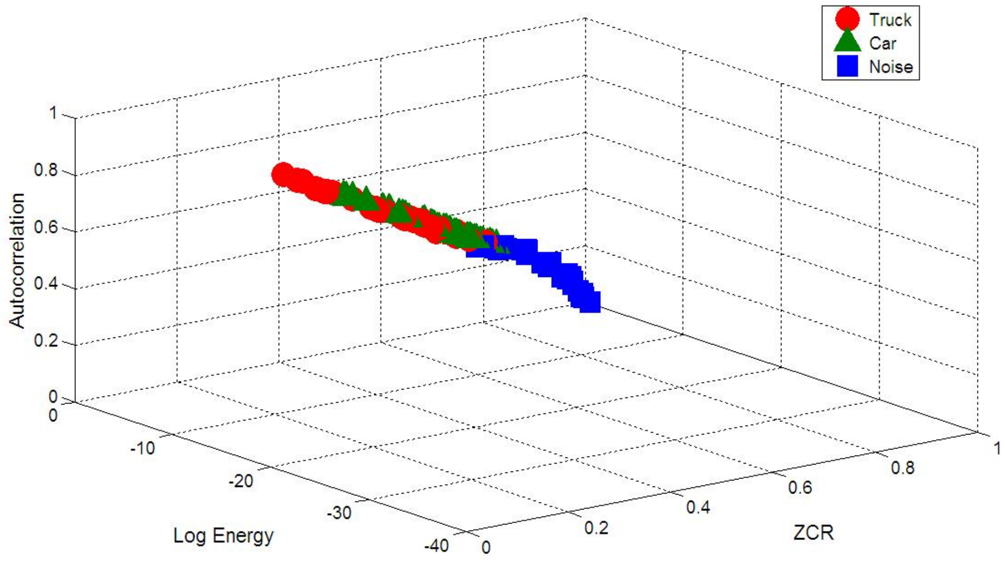


Figure 7.16: Combined Acoustic Time Domain Feature Vectors of Training Set A: ZCR, Log Energy, Autocorrelation of Truck (red), Car (green), and Noise (blue). Frame Size=32768, Fs=44.1kHz.

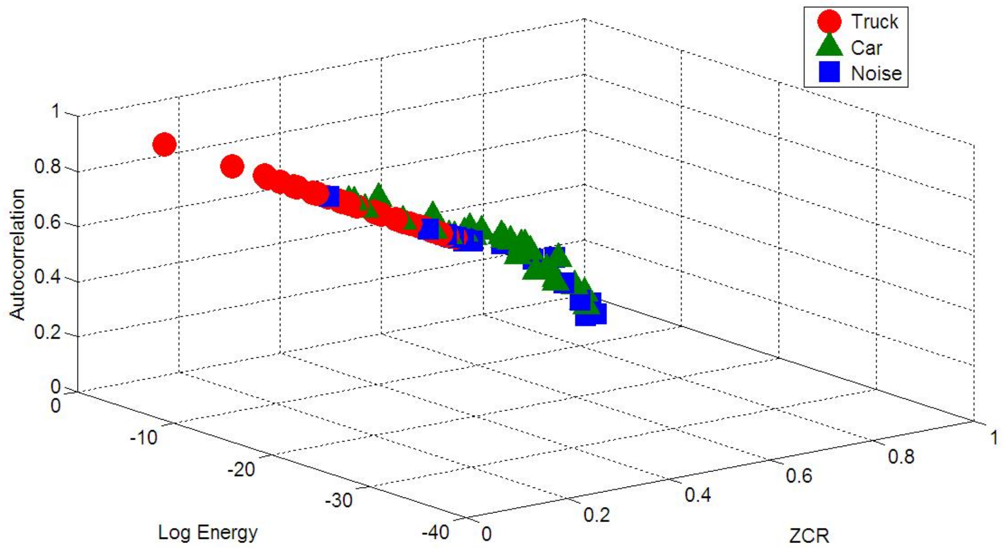


Figure 7.17: Combined Seismic Time Domain Feature Vectors of Training Set B: ZCR, Log Energy, Autocorrelation of Truck (red), Car (green), and Noise (blue). Frame Size=32768, Fs=44.1kHz.

Similar to the other cases explained so far, three-dimension seismic time domain feature vectors extracted by ZCR, Log Energy and normalised autocorrelation were classified by SVM with RBF kernel functions. Note, two more smaller frame sizes were also tested just in case although the results were worse than the others after all.

| Frame Size | 32768 | 16384 | 8192 | 4096 | 2048 |
|---|---|---|---|---|---|
| no-processing | 85.71% | 80.36% | 80.16% | 79.38% | 78.40% |
| normalised | 55.36% | 56.25% | 75.00% | 80.45% | 79.46% |
| PCA(3) | 83.92% | 82.14% | 80.56% | 79.02% | 78.70% |
| normalised & PCA(3) | 83.92% | 75.89% | 80.16% | 80.27% | 79.21% |

Table 7.33: Seismic Classification (2-class) Results (ZCR, Log Energy and Normalised Autocorrelation) with Spectral Subtraction. (The numbers in brackets for PCA indicate the number of principle components kept.)

Based on the findings discussed so far, acoustic LPC coefficients and these three time domain features extracted from the seismic signals appeared to be the strongest candidates for an optimum system. Hence, the effect of combining them was investigated next. This approach made good sense especially since the pre-processing found to be the best for these feature extraction methods were also the same, spectral subtraction of estimated noise. Because the spectral subtraction algorithm utilises FFT, frame length variations of powers of 2 have been adopted here.

Care must be taken when merging feature vectors extracted by various methods as their value ranges differ, more trials about normalising and dimensionality reduction were therefore examined with these cases. Table 7.34 shows the SVM classification results for data set B. It was observed the performance for the feature vectors normalised to -1 and 1 before classification reached 93.30% accuracy for the frame sie of 16384, making it the best case.

Following the results by SVM, the best case was then examined by LVQ also by

| Frame Size | 32768 | 16384 | 8192 | 4096 | 2048 |
|---|---|---|---|---|---|
| no-processing | 78.57% | 75.00% | 77.18% | 83.39% | 83.67% |
| normalised | 87.50% | **93.30**% | 91.07% | 89.20% | 89.33% |
| PCA(3) | 73.21% | 70.54% | 73.61% | 32.59% | 33.21% |
| normalised & PCA(3) | 78.57% | 81.70% | 53.57% | 87.68% | 88.56% |
| normalised & PCA(2) | 71.43% | 74.11% | 66.86% | 87.86% | 88.44% |
| normalised & PCA(3) on LPC only | 76.79% | 75.00% | 74.21% | 79.64% | 80.44% |

Table 7.34: SVM Classification (2-class) Results by Combination of Seismic ZCR, Log Energy and Normalised Autocorrelation as well as Acoustic LPC coefficients; all with Spectral Subtraction Pre-Processing. (The numbers in brackets for PCA indicate the number of principle components kept.)

using the data set B, and the results are in Table 7.35. The results are similar for various settings of both the node and the learning rate, reaching up to 80 %.

| Rate | 10 nodes | 20 nodes | 30 nodes | 40 nodes | 50 nodes | 60 nodes | 70 nodes |
|---|---|---|---|---|---|---|---|
| 0.005 | 80.80% | 80.86% | 77.23% | 77.68% | 76.79% | 78.56% | 79.46% |
| 0.01 | 80.80% | 78.57% | 78.57% | 78.57% | 78.57% | 75.00% | 76.34% |
| 0.02 | 79.91% | 77.68% | 78.57% | 79.02% | 80.36% | 79.02% | 77.23% |

Table 7.35: LVQ Classification (2-class) Results by Combination of Seismic ZCR, Seismic Log Energy and Seismic Normalised Autocorrelation as well as Acoustic LPC Coefficients with Spectral Subtraction Pre-Processing for All, Frame Size=16384.

Example confusion matrix and ROC can be found in Table 7.36 and Figure 7.18 for the the combination of acoustic and seismic signal processing.

| | Target Class: Truck | Target Class: Car |
|---|---|---|
| Output Class: Truck | 106 | 20 |
| Output Class: Car | 16 | 104 |

Table 7.36: Example Confusion Matrix of the Acoustic and Seismic Combination LVQ Classification (2-class) with Spectral Subtraction Pre-Processing, Node=20, Learning Rate=0.005, Frame Size=16384.

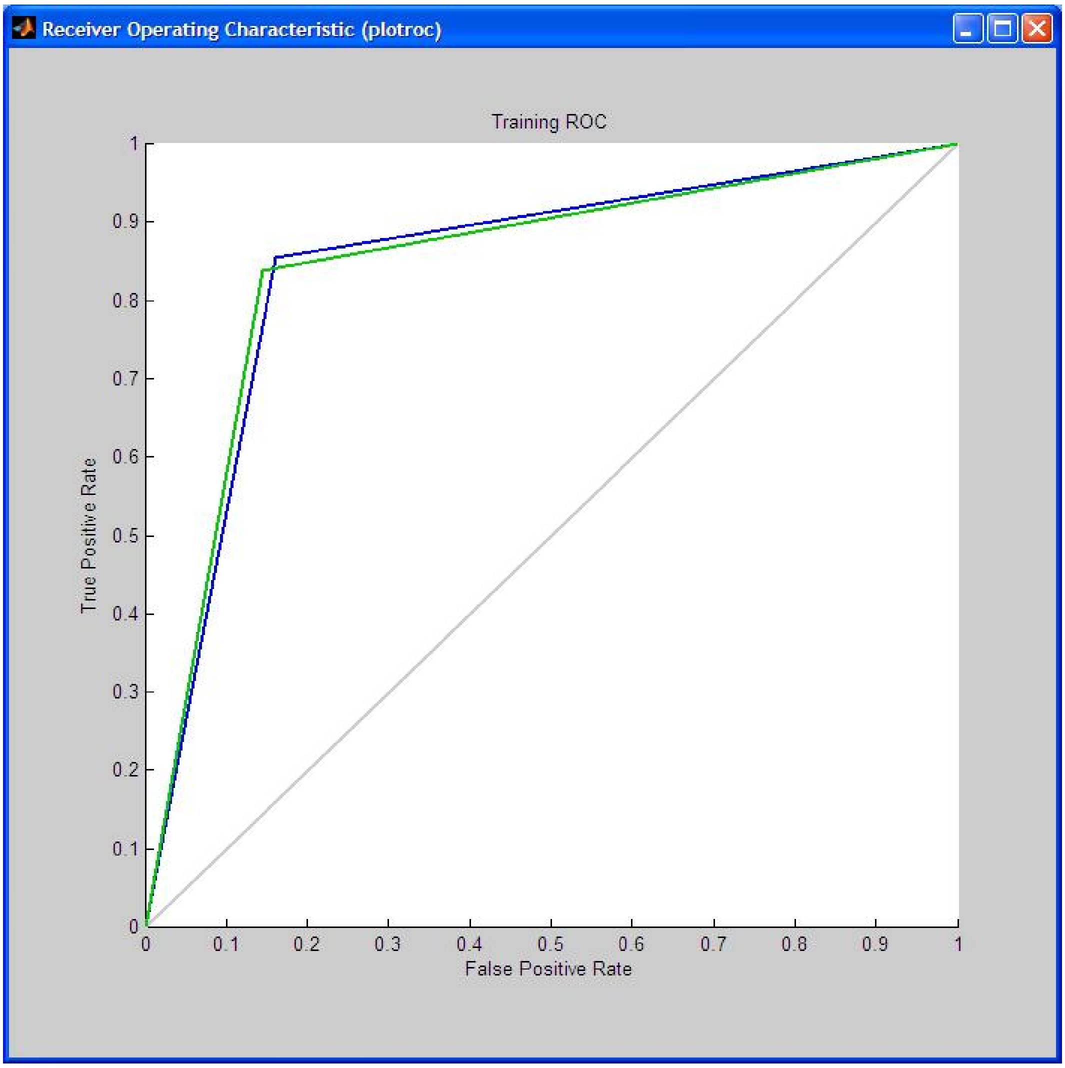


Figure 7.18: Example ROC Curve of the Acoustic and Seismic Combination LVQ Classification (2-class) with Spectral Subtraction Pre-Processing, Node=20, Learning Rate=0.005, Frame Size=16384.

## 7.4 Chapter Summary

Based on the literature review (Chapter 2) and all the preliminary findings about various methods, an investigation to develop a novel and optimum automated vehicle recognition system has been carried out. In order to improve the processing speed and efficiency, the use of a seismic detection algorithm has been suggested, using the combination of features extracted by Log Energy and TDSC methods. The outcome of the such method was highly accurate especially when it was tested with larger set of data, reaching up to 97.71 % for SVM and 99.02 % for LVQ. The experiment, however, revealed how influential the size of data was; in terms of both the total number of samples and the inequality. It was also noted that the effect of the sample size inequality seemed to be reduced when the total number of samples became larger.

For classification, the feasibility of feature extraction methods mainly in the time domain and a filterbank were examined. The method choices were made owing to combination of their class separability observed with training samples as well as how well each method generalised while processing samples in evaluation sets. After assessing each method, an algorithm that combines those indicated more accurate classification abilities were combined for further examination, which subsequently produced results reaching up to 93.30 % for two-class classification. The algorithm consists of three relatively simple time domain features extracted from seismic signals as well as LPC coefficients and the minimum error obtained from acoustic signals. What found to be the best pre-processing method for these were the same, thus, such combination should be easily implemented. Regarding the classifier, overall SVM outperformed LVQ. Consequently, Figure 7.19 suggests a structure of the system found to be the optimum in this research.

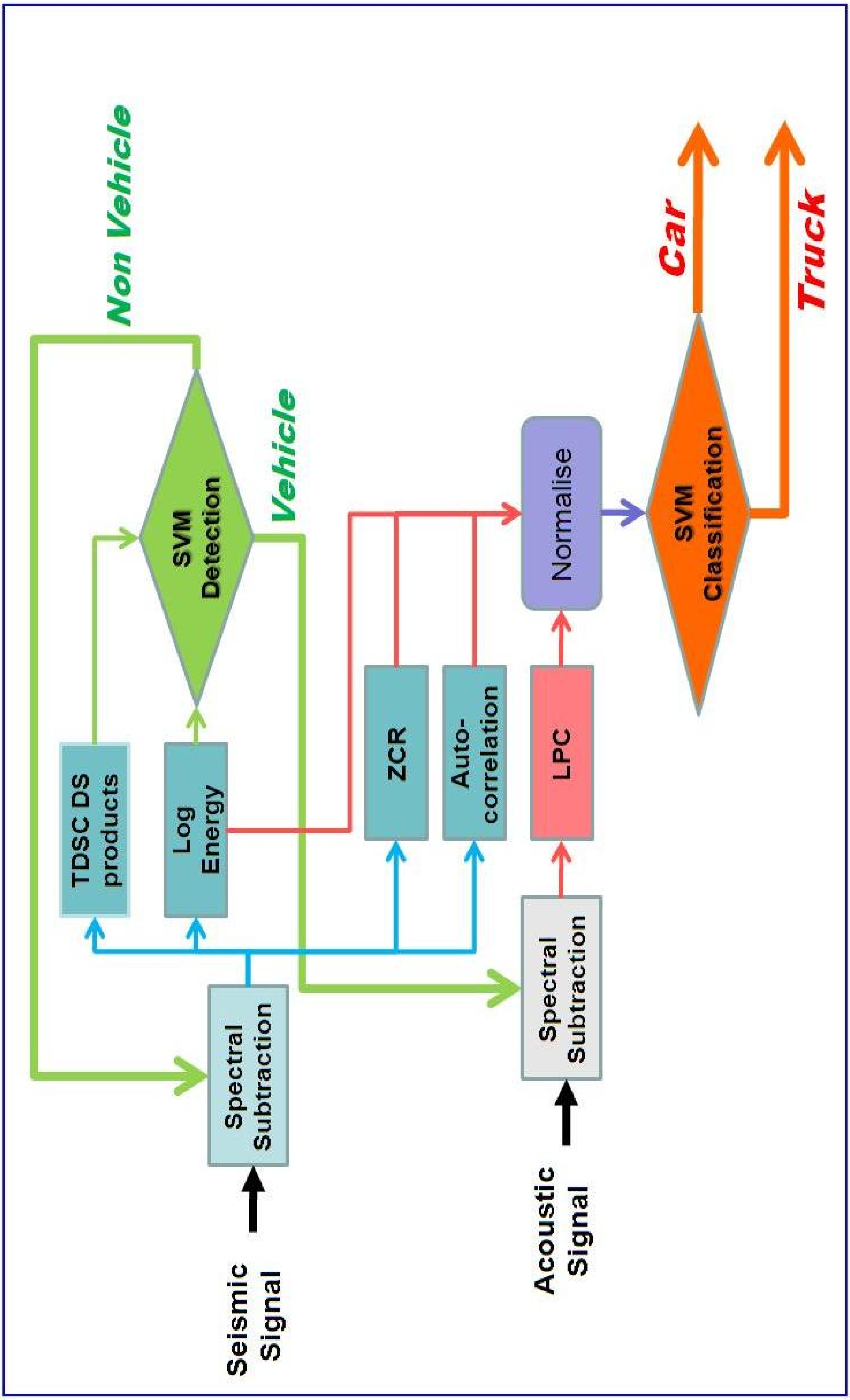


Figure 7.19: Structure of the Optimum System

# Chapter 8

# Conclusions and Future Study

## 8.1 Conclusion

In order to improve security of important infrastructure, the development of an automated vehicle recognition system is sought. Although the ultimate aim will be to include fusion of multiple types of sensors, acoustic and seismic sensors have been the choice of this study due to many of their advantages; these are passive sensors require relatively low installation, operation and maintenance costs as well as low power consumption.

A literature survey regarding acoustic and seismic vehicle recognition research has been conducted. Firstly it found few peer reviewed journal papers discussing more or less self contained recognition algorithms in the field. Secondly, most have focused on military vehicles that tend to generate louder, more distinctive signals than road vehicles. Thirdly, to the author's knowledge, no other study on road vehicle recognition algorithm that utilises the fusion of acoustic and seismic signal processing has been carried out so far. The potential of seismic signals in enhancing acoustic vehicle recognition, however, has already been proved to a certain extent regarding military vehicles. In Chapter 3, the mechanisms of target source generation as well as sound propagation have been discussed. It is realised there are so many factors that affect characteristics of signals received by a sensor. All of these findings have been exploited to determine the scope of this research.

After establishing the data collection method for gathering real-life data; three signal pre-processing methods have been introduced, such as spectral subtraction of estimated noise, FIR filter and WP noise reduction algorithms. In Chapters 5 and 6, methods that could be employed to develop an optimum system for the study have been discussed in relation to their feasibility. Together with preliminary studies, the outcomes have led to identification of the techniques that should be investigated further with real-life data, for both feature extraction and decision making. These are:

- ZCR
- Log Energy
- Normalised Autocorrelation
- TDSC
- LPC
- Co-Occurrence Matrix
- ERB Filterbank

On selecting these methods, class separability has been the main aspect. In addition, a modification of TDSC has been introduced especially for vehicle detection algorithm using seismic signals.

Furthermore, evaluation of each of the listed feature extraction methods has been carried out in regards to its generalisation by using real-life data. The results have been presented about a range of combinations regarding pre-processing techniques, frame lengths and parameter settings combined with either SVM or LVQ classifier.

As a consequence, an optimum algorithm for an automated vehicle recognition system has been proposed. It consists of two-step algorithm, one for vehicle detection and another for classification to improve processing speed and efficiency to meet the demand of security applications.

Both detection and classification algorithms employ spectral subtraction to reduce noise from acoustic and seismic signals. In terms of feature extraction for vehicle detection, combination of seismic Log Energy and seismic TDSC DS products have been recommended as it produced highly accurate performance, reaching up to 97.71 % with SVM and also 99.02 % with LVQ.

For classification between trucks and cars, acoustic LPC coefficients have been combined with three relatively simple time domain methods using seismic signals: ZCR, Log Energy and normalised autocorrelation. On the subject of pre-classification signal processing, the effect has depended on the feature vectors to which the method applied. The best result obtained by using this feature vector combination was 93.30 %, classified by SVM when the vectors were normalised to -1 and 1 in advance. On the whole, SVM outperformed LVQ classifier.

## 8.2 Future Study

### 8.2.1 Sensor Fusion

As has been pointed out that fusion of multiple types of sensors as well as implementation of multiple sensor nodes should be the way forward, particularly for security purposes. It is likely to achieve more accurate detection and classification, as well as increasing not only the coverage area for surveillance but also the kinds of threat a system is able to respond to.

### 8.2.2 Multiple Target Recognition

As the research in acoustic and seismic road vehicle recognition is still at its emerging stages, there have not been enough grounds to be based on in order to challenge more complex issues. Hence, some assumptions have been made for the current study, and one of which is to limit the investigation to single target recognition only. Nonetheless, this limitation should be removed especially with sensor fusion.

### 8.2.3 Greater Recognition Range

Increasing the detection and recognition range for a system can be a desirable improvement for some practical applications. Nonetheless, the signals of interest are inevitably influenced by multiple factors while travelling between source and sensor as described in Section 3.2.2. Such factors include attenuation and/or modification due to geometry, meteorological, ground surface and terrain conditions. In addition, effects of diffraction and scattering should also be considered as it is likely to have some obstacles in and around the propagation path when the distacne to the sensor is widened. All of which could lead to a poorer performance without taking appropriate actions. Still, the advantage of having greater detection and recognition range is appealing, as long as the associated challenges are suitably addressed. As it has been seen in some of the military vehicle recognition studies, use of sound propagation models may be an attractive option for road vehicle.

### 8.2.4 Acoustic and Seismic Signal Coupling

Although the classification results obtained for the current study are good, consideration of acoustic and seismic coupling (Section 3.2.4) may further improve, especially if any of the above recommendations is put into practice as the system could consequently become more complex.

### 8.2.5 More Data or More Information

More data collection can be conducted to improve the study. Especially, if it is possible to obtain information such as vehicle models and engine types so that objective labelling of samples as well as classification between more classes can be achieved. For example, this could also be beneficial if the algorithm was to be utilised for traffic monitoring purpose. During the evaluation, it appeared that for both detection and classification experiments, the results were affected by both the total number of samples as well as the construction of the sample sets, i.e. the effect of inequality. Regarding the latter, it was noticed that the performance of classifiers becomes less susceptible to the inequality when larger set of data was utilised. Therefore, having more data may improve the current study. Nevertheless, this has been discussed often in machine learning whether the problem can be really solved by gaining more data or not. This leads to a debate about the problem known as “bias-variance tradeoff” or “bias-variance dilemma” [Marsland, 2009, p.177], and it should always be considered before performing more data collection [Geman et al., 1992].

# References

J. Agre and L. Clare. An integrated architecture for cooperative sensing networks. *Computer*, 33(5):106–108, 2000. ISSN 0018-9162.

D. G. Albert. The effect of snow on vehicle-generated seismic signatures. *Journal Of The Acoustical Society Of America*, 81(4):881–887, Apr. 1987.

D. G. Albert, S. N. Decato, and D. L. Carbee. Snow cover effects on acoustic sensors. *Cold Regions Science and Technology*, 52(2):132–145, Apr. 2008. ISSN 0165-232X.

J. Altmann. Acoustic and seismic signals of heavy military vehicles for co-operative verification. *Journal of Sound and Vibration*, 273(4-5):713–740, June 2004.

J. Altmann, S. Linev, and A. WeiB. Acoustic-seismic detection and classification of military vehicles–developing tools for disarmament and peace-keeping. *Applied Acoustics*, 63(10):1085–1107, 2002.

B. Anami and V. Pagi. An acoustic signature based neural network model for type recognition of two-wheelers. In *IMPACT '09: International Multimedia, Signal Processing and Communication Technologies*, pp. 28–31, 2009.

M. R. Anderberg. *Cluster Analysis for Applications*. Academic Press, 1973.

T. Anderson, M. Moran, S. Ketcham, and J. Lacombe. Technical validation of high-fidelity seismic signature simulations in support of fcs network ground sensors. In *User Group Conference, 2003. Proceedings*, pp. 62–67, 2003.

A. Arora, P. Dutta, S. Bapat, V. Kulathumani, H. Zhang, V. Naik, V. Mittal, H. Cao, M. Demirbas, M. Gouda, Y. Choi, T. Herman, S. Kulkarni, U. Arumugam, M. Nesterenko, A. Vora, and M. Miyashita. A line in the sand: a wireless sensor network for target detection, classification, and tracking. *Computer Networks*, 46 (5):605–634, Dec. 2004. ISSN 1389-1286.

B. S. Atal and L. R. Rabiner. A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24 (3):201 – 212, 1976.

K. Attenborough. Review of ground effects on outdoor sound propagation from continuous broadband sources. *Applied Acoustics*, 24 (4):289–319, 1988.

K. Attenborough. Sound propagation close to the ground. *Annual Review of Fluid Mechanics*, 34:51–82, 2002.

A. Averbuch, E. Hulata, V. Zheludev, and I. Kozlov. A wavelet packet algorithm for classification and detection of moving vehicles. *Multidimensional Systems And Signal Processing*, 12(1):9–31, Jan. 2001.

A. Averbuch, V. Zheludev, N. Rabin, and A. Schclar. Wavelet based acoustic detection of moving vehicles. Submitted, 2007.

A. Averbuch, V. A. Zheludev, N. Rabin, and A. Schclar. Wavelet-based acoustic detection of moving vehicles. *Multidimensional Systems and Signal Processing*, 20(1):55–80, 2009.

H. E. Bass, L. N. Bolen, D. Cress, J. Lundien, and M. Flohr. Coupling of airborne sound into the earth: Frequency dependence. *The Journal of the Acoustical Society of America*, 67(5):1502–1506, 1980.

A. Bateman and I. Paterson-Stephens. *The DSP handbook: algorithms, applications and design techniques*. Prentice Hall, 2002.

H. Bateman. The lift and drag functions for an elastic fluid in two dimensional irrotational flow. *Proceedings of the National Academy of Sciences of the United States of America*, 24(6):246–251, Jun. 15, 1938. ISSN 00278424.

G. Becker and A. Gudesen. Passive sensing with acoustics on the battlefield. *Applied Acoustics*, 59(2):149–178, Feb. 2000.

J. Benesty, M. M. Sondhi, and Y. Huang, editors. *Springer Handbook of Speech Processing*. Springer, 2008.

P. Bentley and J. McDonnell. Wavelet transforms: an introduction. *Electronics & Communication Engineering Journal*, 6(4):175–186, 1994. ISSN 0954-0695.

M. Berouti, R. Schwartz, and J. Makhoul. Enhancement of speech corrupted by acoustic noise. In *IEEE International Conference on Acoustics, Speech, and Signal Processing: ICASSP '79.*, volume 4, pp. 208–211, 1979.

M. Bertozzi and A. Broggi. Gold: a parallel real-time stereo vision system for generic obstacle and lane detection. *IEEE Transactions on Image Processing*, 7 (1):62–81, 1998. ISSN 1057-7149.

C. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

S. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 27(2):113–120, 1979. ISSN 0096-3518.

R. Braunling, R. M. Jensen, and M. A. Gallo. Acoustic target detection, tracking, classifiaction, and location in a multiple target environment. In *Proceedings of SPIE: Peace and Wartime Applications and Technical Issues for Unattended Ground Sensors*, volume 3801, pp. 57–66. SPIE, 1997.

E. O. Brigham. *The Fast Fourier Transform*. Prentice-Hall, Inc., 1974.

G. W. Brooks, G. M. Reel, and M. Huffnagle, Norman P. andMarietta. Acoustic/seismic vehicle locator. *The Journal of the Acoustical Society of America*, 62(S1):S58–S58, 1977.

C. J. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

M. J. Caruso and L. S. Withanawasam. Vehicle detection and compass applications. using amr magnetic sensors. Technical report, Honeywell International Inc., 1999.

V. Cevher, R. Chellappa, and J. McClellan. Joint acoustic-video fingerprinting of vehicles, part i. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 2, pp. II–745–II–748, 2007.

C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

R. Chellappa, G. Qian, and Q. Zheng. Vehicle detection and tracking using acoustic and video sensors. In *ICASSP '04: IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pp. 793–796, 2004.

J. Chen, K. Yao, and R. Hudson. Source localization and beamforming. *IEEE Signal Processing Magazine*, 19(2):30–39, 2002. ISSN 1053-5888.

H. Cheng, B. Shaw, J. Palen, B. Lin, B. Chen, and Z. Wang. Development and field test of a laser-based nonintrusive detection system for identification of vehicles on the highway. *IEEE Transactions on Intelligent Transportation Systems*, 6(2): 147–155, 2005. ISSN 1524-9050.

D. Chesmore. Robust in-field automated bioacoustic identification of species for rapid biodiversity studies. In *Proceedings of the Institute of Acoustics Fourth International Conference on Bio-Acoustics*, 2007.

E. D. Chesmore. Application of time domain signal coding and artificial neural networks to passive acoustical identification of animals. *Applied Acoustics*, 62: 1359–1374, 2001.

E. D. Chesmore. Automated bioacoustic identification of species. *Anais da Academia Brasileira de Cincias*, 76 (2):435–440, 2004.

H. C. Choe, R. E. Karlsen, G. R. Gerhart, and T. J. Meitzler. Wavelet-based ground vehicle recognition using acoustic signals. In *Proceedings of the SPIE–The International Society for Optical Engineering.on Sound Signal Processing*, volume 2762, pp. 434–445, 1996.

C. T. Christou and G. M. Jacyna. Simulation of vehicle acoustics in support of netted sensor research and development. In J. McConnel, editor, *SPIE Defense & Security Symposium*, 2005.

P. Ciosek, K. Brudzewski, and W. Wroblewski. Milk classification by means of an electronic tongue and support vector machine neural network. *Measurement Science & Technology*, 17 (6):1379–1384, 2006.

L. Cohen. *Time-Frequency Analysis*. Prentice-Hall, Inc. New Jersey, New Jersey, 1995.

J. Cooley and J. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

M. Cowling and R. Sitte. Comparison of techiniques for environmental sound recognition. *Pattern Recognition Letters*, 24:2895–2907, 2003.

A. Czyzewski. Automatic identification of sound source position employing neural networks and rough sets. *Pattern Recognition Letters*, 24(6):921–933, Mar. 2003.

L. Debnath, editor. *Wavelet Transforms and Time-Frequency Signal Analysis*. Birkhaser Boston, 2001.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977. ISSN 00359246.

T. G. Dietterich. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7):1895–1923, 1998.

C. Ding and T. Li. Adaptive dimension reduction using discriminant analysis and k-means clustering. In *Proceedings of the 24th international conference on Machine learning*, 2007.

J. Ding, S.-Y. Cheung, C.-W. Tan, and P. Varaiya. Signal processing of sensor node data for vehicle detection. In *Proceedings of The 7th International IEEE Conference on Intelligent Transportation System*, 2004.

S. Ding, J. Huang, D. Wei, and A. Cichocki. A near real-time approach for convolutive blind source separation. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 53(1):114–128, 2006. ISSN 1057-7122.

D. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, 1995. ISSN 0018-9448.

M. F. Duarte and Y. H. Hu. Vehicle classification in distributed sensor networks. *Journal of Parallel and Distributed Computing*, 64(7):826 – 838, Jul. 2004.

R. O. Duda and P. E. Hart. *Pattern Classifiaction and Scene Analysis*. John Wiley & Sons Inc., 1973.

R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classifiaction*. John Wiley & Sons, Inc., 2nd ed. edition, 2001.

R. P. W. Duin. A note on comparing classifiers. *Pattern Recognition Letters*, 17(5): 529–536, May 1996. ISSN 0167-8655.

K. Dyrdal, Idarand Lie, M. Huseby, and R. Rahimi. Modeling of acoustic ground vehicle signatures and outdoor sound propagation. In E. M. Carapezza, editor, *Proceedings of SPIE: Unattended ground sensor technologies and applications*, number 5090 in 1, pp. 111–121. Society of Photo-optical Instrumentation Engineers., SPIE, 2003.

K. B. Eom. Analysis of acoustic signatures from moving vehicles using time-varying autoregressive models. *Multidimensional Systems And Signal Processing*, 10(4):357–378, Oct. 1999.

N. Evans and D. Chesmore. Automated acoustic identification of vehicles. In *Proceedings of Institue of Acoustics*, volume 30, pp. 238–245, 2008.

K. Faber and P. W. Maxwell. Geophone spurious frequency: What is it and how does it affect seismic data quality? *Canadian Journal of Exploration Geophysiscs*, 33 (Nos. 1 & 2):46 – 54, 1997.

I. J. Farr. *Automated Bioacoustic Identificaiton of Statutory Quarantined Insect Pests*. PhD thesis, Univeristy of York, 2007.

I. J. Farr and D. Chesmore. Automated bioacoustic detection and identification of wood-boring insects for quarantine screening and insect ecology. In *Proceedings of the Institute of Acoustics Fourth International Conference on Bio-Acoustics*, 2007.

H. Fastl and E. Zwicker. *Psychoacoustics: Facts and Models*. Springer Berlin Heidelberg, 3rd ed. edition, 2007.

B. Favre. Noise emission of road vehicles: Evaluation of some simple models. *Journal of Sound and Vibration*, 91(4):571–582, Dec. 1983. ISSN 0022-460X.

B. M. Favre and B. T. Gras. Noise emission of road vehicles: Reconstitution of the acoustic signature. *Journal of Sound and Vibration*, 93(2):273–288, 1984.

T. Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8): 861–874, June 2006. ISSN 0167-8655.

H. Fletcher. The mechanism of hearing as revealed through experiment on the masking effect of thermal noise. *Proceedings of the National Academy of Sciences of the United States of America*, 24(7):265–274, June 1938a. ISSN 00278424.

H. Fletcher. Loudness, masking and their relation to the hearing process and the problem of noise measurement. *The Journal of the Acoustical Society of America*, 9(4):275–293, 1938b.

H. Fletcher. Auditory patterns. *reviews of Modern Physics*, 12(1):47–65, Jan. 1940.

K. Fukunaga. *Intorduction to Statistical Pattern Recognition*. Academic Press, 2nd edition, 1990.

C. Gargour, M. Gabrea, V. Ramachandran, and J.-M. Lina. A short introduction to wavelets and their applications. *IEEE Circuits and Systems Magazine*, 9(2): 57–68, 2009. ISSN 1531-636X.

S. Geman, E. Bienenstock, and R. Doursat. Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1):1 – 58, 1992.

Y. Ghanbari and M. R. Karami-Mollaei. A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets. *Speech Communication*, 48(8):927–940, Aug. 2006. ISSN 0167-6393.

A. Ghosh, M. Biehl, and B. Hammer. Performance analysis of lvq algorithms: A statistical physics approach. *Neural Networks*, 19(6-7):817–829, 2006.

D. D. Greenwood. Auditory masking and the critical band. *The Journal of the Acoustical Society of America*, 33(4):484–502, 1961a.

D. D. Greenwood. Critical bandwidth and the frequency coordinates of the basilar membrane. *The Journal of the Acoustical Society of America*, 33(10):1344–1356, 1961b.

A. Grossmann and J. Morlet. Decomposition of hardy functions into square integratable wavelets of constant shape. *SIAM J. MATH. ANAL.*, 15(4):723 – 736, 1984.

S. R. Gunn. Support vector machines for classification and regression. Technical report, Image Speech and Intelligent Systems Research Group, University of Southampton, 1997.

K. Gurley and A. Kareem. Applications of wavelet transforms in earthquake, wind and ocean engineering. *Engineering Structures*, 21(2):149–167, 1999.

A. L. Hagedoom, E. Kruithof, and P. Maxwell. A practical set of guidelines for geophone element testing and evaluation. *First Break*, 6(10):325, 1988.

D. J. Hand. *Construction and Assessment of Classifiaction Rules*. John Wiley & Sons Ltd., 1997.

R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6):610–621, 1973. ISSN 0018-9472.

F. Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–83, 1978. ISSN 0018-9219.

J. A. Hartigan. *Clustering Algorithm*. John Wiley & Sons, Inc., 1975.

S. Haykin. *Neural Networks: A comprehensive Foundation*. Prentice-Hall, Inc., 2nd ed. edition, 1999.

S. Hinz and U. Stilla. Car detection in aerial thermal images by local and global evidence accumulation. *Pattern Recognition Letters*, 27(4):308–315, Mar. 2006.

J. Holmes and W. Holmes. *Speech Synthesis and Recognition, 2nd ed.* Taylor & Francis, 2 edition, 2001.

J. A. M. Howard, David M.; Angus. *Acoustics and Psychoacoustics (3rd ed).* Elsevier Science & Technology, 2006.

J.-W. Hsieh, S.-H. Yu, Y.-S. Chen, and W.-F. Hu. Automatic traffic surveillance system for vehicle tracking and classification. *IEEE Transactions on Intelligent Transportation Systems*, 7(2):175–187, 2006. ISSN 1524-9050.

W. W. Hsieh. Nonlinear principal component analysis by neural networks. *Tellus*, 53A:599–615, 2001.

E. C. Ifeachor and B. W. Jervis. *Digital Signal Processing: A Practical Apprach*. Pearson Education Limited, 2nd ed. edition, 2002.

J. Jackowski and R. Wantoch-Rekowski. Classification of wheeled military vehicles using neural networks. In *18th International Conference on Systems Engineering, ICSEng 2005.*, pp. 212 – 217, 2005.

G. M. Jacyna, C. T. Christou, B. George, and B. F. Necioglu. Netted sensors-based vehicle acoustic classification at tier 1 nodes. In *Proceedings- SPIE The International Society For Optical Engineering*, volume 5796, 2005.

S. Jaffard, Y. Meyer, and R. D. Ryan. *Wavelets : Tools for Science & Technology*. Society for Industrial and Applied Mathematics, 2001.

A. Jain and D. Zongker. Feature selection: evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):153–158, 1997. ISSN 0162-8828.

J. Jan. *Digital Signal Filtering, Analysis and Restoration.* Institution of Electrical Engineers, 2000.

J. Jarnicki, J. Mazurkiewicz, and H. Maciejewski. Mobile object recognition based on acoustic information. In *Industrial Electronics Society, 1998. IECON '98. Proceedings of the 24th Annual Conference of the IEEE*, volume 3, pp. 1564–1569 vol.3, 1998.

X.-Y. Jing, D. Zhang, and Y.-Y. Tang. An improved lda approach. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(5):1942–1951, 2004. ISSN 1083-4419.

I. T. Jolliffe. *Principle Component Analysis*. Springer-Verlag New York, Inc., 2002.

H. Jonasson. Source modelling of road vehicles. Technical report, Harmonoise, 2003.

H. Jonasson. Acoustical source modelling of road vehicles. *Acta Acustica united with Acustica*, 93(2):173–184, 2007.

R. E. Karlsen, G. R. Gerhart, T. J. Meitzler, R. Goetz, and H. C. Choe. Wavelet analysis of ground vehicle acoustic signatures. In *Proceedings of the SPIE–The International Society for Optical Engineering.*, 1995.

P. Kearey and M. Brooks. *An Introduction to Geophysical Exploration*. Blackwell Scientific Publications, 1984.

S. Ketcham, M. Moran, J. Lacombe, R. Greenfield, and T. Anderson. Seismic source model for moving vehicles. *IEEE Transactions on Geoscience and Remote Sensing*, 43(2):248–256, 2005. ISSN 0196-2892.

R. A. King and W. Gosling. Time-encoded speech. *Electronics Letters*, 14 (15): 456–457, 1978.

R. A. King and M. T. C. Phipps. Shannon, tespar and approximation strategies. *Computers & Security*, 18(5):445–453, 1999.

W. Koch, J. Koller, and M. Ulmke. Ground target tracking and road map extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(3-4):197–208, Dec. 2006.

T. Kohonen. Improved versions of learning vector quantization. In *IJCNN International Joint Conference on Neural Networks*, 1990.

T. Kohonen. *Self-Organizing Maps*. Springer, 2001.

Y. Koren and L. Carmel. Robust linear dimensionality reduction. *IEEE Transactions on Visualization and Computer Graphics*, 10(4):459–470, 2004. ISSN 1077-2626.

J. Kotecha, V. Ramachandran, and A. Sayeed. Distributed multitarget classification in wireless sensor networks. *IEEE Journal on Selected Areas in Communications*, 23(4):703–713, 2005. ISSN 0733-8716.

M. A. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37 (2):233–243, 1991.

W. Kropp, F.-X. Bcot, and S. Barrelet. On the sound radiation from tyres. *Acustica*, 86:769–779, 2000.

A. Kujipers and G. van Blokland. Tyre/road noise models in the last two decades: a critical evaluation. In *Internoise 2001*, pp. 2593 – 2598, 2001.

B. V. K. V. Kumar, A. Mahalanobis, and R. D. Juday. *Correlation Pattern Recognition*. Cambridge University Press, CambridgeKumar2005, 2005.

C. F. Lam and D. Kamins. Signature recognition through spectral analysis. *Pattern Recognition*, 22(1):39–44, 1989. ISSN 0031-3203.

J. Lan, S. Nahavandi, T. Lan, and Y. Yin. Recognition of moving ground targets by measuring and processing seismic signal. *Measurement*, 37(2):189–199, Mar. 2005.

B. Lerner, H. Guterman, M. Aladjem, and I. Dinstein. A comparative study of neural network based feature extraction paradigms. *Pattern Recognition Letters*, 20(1):7–14, Jan. 1999.

D. Li, K. Wong, Y. H. Hu, and A. Sayeed. Detection, classification, and tracking of targets. *IEEE Signal Processing Magazine*, 19(2):17–29, 2002. ISSN 1053-5888.

H.-Y. Lin, K.-J. Li, and C.-H. Chang. Vehicle speed detection from a single motion blurred image. *Image and Vision Computing*, 26(10):1327–1337, Oct. 2008. ISSN 0262-8856.

L. Liu. Ground vehicle acoustic signal processing based on biologicl hearing models. Master's thesis, University of Maryland, 1999.

C. G. Looney. *Pattern Recognition Using Neural Networks: Theory and Algorithms for Engineers and Scientists*. Oxford University Press, Inc. New York, 1997.

J. E. Lopez, J. C. Lo, and J. Saulnier. Vehicle signal enhancement using packet wavelet transform and nonlinear noise processing techniques. In *Meeting of MSS Specialty Group on Battlefield Acoustic and Seismic Sensing, Magnetic and Electric Field Sensors*, 1999.

B. Lu, A. Dibazar, and T. Berger. Nonlinear hebbian learning for noise-independent vehicle sound recognition. In *IEEE International Joint Conference on Neural Networks, IJCNN 2008 (IEEE World Congress on Computational Intelligence).*, pp. 1336–1343, 2008a.

B. Lu, A. Dibazar, and T. Berger. Perimeter security on detecting acoustic signature of approaching vehicle using nonlinear neural computation. In *IEEE Conference on Technologies for Homeland Security*, pp. 51–56, 2008b.

W. Lucking. *The Application of Time Encoded Signals to Automated Machine Condition Classification using Neural Networks*. PhD thesis, University of Hull, 1997.

W. K. Lui and K. M. Li. A theoretical study for the propagation of rolling noise over a porous road pavement. *The Journal of the Acoustical Society of America*, 116(1):313–322, 2004.

R. Lyons. *Understanding Digital Signal Processing*. Addison-Wesley Longman, Inc., 1997.

H. Maciejewski, J. Mazurkiewicz, K. Skowron, and T. Walkowiak. Neural networks for vehicle recognition. In U. Ramacher, H. Klar, and A. Koenig, editors, *Proceedings of the 6th International Conference on Microelectronics for Neural Networks, Evolutionary and Fuzzy Systems*, pp. 292–296, 1997.

J. Makhoul. Linear prediction in automatic speech recognition. In *Speech Recognition: Invited Papers Presented at the 1974 IEEE Symposium*. Academic Press, 1975a.

J. Makhoul. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4): 561–580, 1975b. ISSN 0018-9219.

D. J. Mapps. Magnetoresistive sensors. *Sensors and Actuators A: Physical*, 59(1-3): 9–19, Apr. 1997.

J. D. Markel and A. H. Gray. *Linear Prediction of Speech*. Springer-Verlag Berlin Heidelberg, 1976.

S. Marsland. *Machine Learning: An Algorithmic Perspective*. Taylor & Francis Group, LLC, 2009.

P. W. Maxwell and K. Faber. Modern geophones, do they meet the demands of shallow seismic measurements? Technical Report NS2.6, Sensor Netherland bv., 1994. Paper presented at the 64th meeting of the SEG (The Society of Exploration Geophysics), Los Angeles, October 1994.

A. M. Mayer. Researches in acoustics. *Philosophical Magazine*, 31:500 –507, 1876.

G. Mazarakis and J. Avaritsiotis. Lightweight time encoded signal processing for vehicle recognition in sensor networks. In *Research in Microelectronics and Electronics 2006, Ph. D.*, pp. 497–500, 2006.

G. P. Mazarakis and J. N. Avaritsiotis. Vehicle classification in sensor networks using time-domain signal processing and neural networks. *Microprocessors and Microsystems*, 31(6):381–392, Sept. 2007.

J. Merelo, A. Prieto, F. Moran, R. Marabini, and J. Carazo. A ga-optimized neural network for classification of biological particles from electron-microscopy images. *Biological and Artificial Computation: From Neuroscience to Technology*, 1240:1174–1182, 1997.

F. Meyer, S. Hinz, A. Laika, D. Weihing, and R. Bamler. Performance analysis of the terrasar-x traffic monitoring concept. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(3-4):225–242, Dec. 2006.

G. F. Miller and H. Pursey. On the partition of energy between elastic waves in a semi-infinite solid. In *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 1955.

B. C. J. Moore. *An Introduction to the Psychology of Hearing*. Academic Press, London, 5th ed. edition, 2003.

B. C. J. Moore and B. R. Glasberg. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America*, 74(3):750–753, 1983.

M. Moran and R. Greenfield. Estimation of the acoustic-to-seismic coupling ratio using a moving vehicle source. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7):2038–2043, 2008. ISSN 0196-2892.

B. Mulgrew, P. Grant, and J. Thompson. *Digital Signal Processing: Concepts and Applcations*. Palgrave Macmillan, 2003.

M. E. Munich. Bayesian subspace methods for acoustic signature recognition of vehicles. In *Proceedings of the 12th European Signal Processing Conference (EUSIPCO 2004)*, Sep 2004.

G. Nagy. Classification algorithms in pattern recognition. *IEEE Transactions on Audio and Electroacoustics*, 16(2):203–212, 1968. ISSN 0018-9278.

B. F. Necioglu, C. T. Christou, E. B. George, and G. M. Jacyna. Vehicle acoustic classification in netted sensor systems using gaussian mixture models. In I. Kadar, editor, *Proceedings of the SPIE, The International Society For Optical Engineering: Signal Processing, Sensor Fusion, and Target Recognition XIV*, volume 5809, pp. 409–419, 2005.

P. M. Nelson, editor. *Transportation Noise Reference Book*. Butterworth & Co. (Publishers) Ltd., 1987.

A. Y. Nooralahiyan, M. Dougherty, D. McKeown, and H. R. Kirby. A field trial of acoustic signature analysis for vehicle classification. *Transportation Research Part C: Emerging Technology*, 5(3-4):165–177, Aug. 1997.

A. Y. Nooralahiyan, H. R. Kirby, and D. McKeown. Vehicle classification by acoustic signature. *Mathematical and Computer Modelling*, 27(9-11):205–214, 1998.

R. Nota, R. Barelds, and D. van Maercke. Engineering method for road traffic and railway noise after validation and fine-tuning. Technical Report Technical Report HAR32TR-040922-DGMR20, Harmonoise, 2005.

C. Ofria. A short course on automobile engines. online, last accessed in 2009. URL `http://www.familycar.com/engine.htm`.

A. V. Oppenheim, R. W. Schafer, and J. R. Buck. *Discrete-Time Signal Processing*. Prentice-Hall, Inc., 2nd edition, 1999.

F. J. Owens. *Signal Processing of Speech*. The Macmillan Press Ltd, 1993.

P. W. Parfomak. Pipeline security: An overview of federal activities and current policy issues. Technical report, Congressional Research Service, 2004.

R. D. Patterson and B. C. J. Moore. *Auditory filters and excitation patterns as representations of frequency resolution*, chapter 3, pp. 123 – 177. Academic Press Inc., 1986.

R. D. Patterson, M. H. Allerhand, and C. Giguere. Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *The Journal of the Acoustical Society of America*, 98(4):1890–1894, 1995.

J. Peng, P. Zhang, and N. Riedel. Discriminant learning analysis. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 38(6):1614–1625, 2008. ISSN 1083-4419.

J. E. Piercy, T. F. W. Embleton, and L. C. Sutherland. Review of noise propagation in the atmosphere. *The Journal of the Acoustical Society of America*, 61 (6): 1403–1418, 1977.

M. Ploetner, J. Trivedi. A multimodal framework for vehicle and traffic flow analysis. In *IEEE Intelligent Transportation Systems Conference,*, 2006.

F. Press and M. Ewing. Theory of air-coupled flexural waves. *Journal of Applied Physics*, 22(7):892–899, 1951.

J. G. Proakis and D. G. Manolakis. *Digital Signal Processing: Principles, Algorithms, and Applications, 3rd ed,*. Prentice-Hall International, Inc., 1996.

H. Qi, X. Tao, and L. H. Tao. Multiple target recognition based on blind source separation and missing feature theory. In *1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, pp. 205–208, 2005.

S. Qian. *Introduction to Time-Frequency and Wavelet Transforms*. Prentice-Hall, Inc., 2002.

J. Qu, C. Ma, K. Li, and J. Ma. The technique of the target tracking and classification in multiple target environment. In *5th international symposium on test and measurement*, volume 1-6, pp. 3029–3032, 2003.

L. Rabiner and B.-H. Juang. *Fundamentals of Speech Recognition*. PTR Prentice Hall, Inc., 1993.

L. Rabiner and R. Schafer. Recursive and nonrecursive realizations of digital filters designed by frequency sampling techniques. *IEEE Transactions on Audio and Electroacoustics*, 19(3):200–207, 1971. ISSN 0018-9278.

L. R. Rabiner and R. W. Schafer. *Digital Processing of Speech Signals*. Prentice-Hall, Inc., 1978.

K. Ramchandran and M. Vetterli. Best wavelet packet bases in a rate-distortion sense. *IEEE Transactions on Image Processing*, 2(2):160–175, 1993. ISSN 1057-7149.

S. Ray and R. Turi. Determination of number of clusters in k-means clustering and application in colour image segmentation. In *Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques*, 1999.

M. Raymer, W. Punch, E. Goodman, L. Kuhn, and A. Jain. Dimensionality reduction using genetic algorithms. *IEEE Transactions on Evolutionary Computation*, 4(2):164–171, 2000. ISSN 1089-778X.

D. A. Reynolds. Speaker identification and verification using gaussian mixture speaker models. *Speech Communication*, 17(1-2):91–108, Aug. 1995.

E. S. Robinson and C. Coruh. *Basic Exploration Geophysics*. John Wiley & Sons, Inc, 1988.

W. Roper. Geospatial informatics applications for assessment of pipeline safety and security. *International Journal of Technology Transfer and Commercialisation (IJTTC)*, 4 (3):401–420, 2005.

T. D. Rossing, F. R. Moore, and P. A. Wheeler. *The Science of Sound*. Addison Wesley, 3rd ed. edition, 2002.

E. Rustighi, S. Elliott, S. Finnveden, K. Gulys, T. Mcsai, and M. Danti. Linear stochastic evaluation of tyre vibration due to tyre/road excitation. *Journal of Sound and Vibration*, 310(4-5):1112–1127, Mar. 2008. ISSN 0022-460X.

J. Sabatier and N. Xiang. An investigation of acoustic-to-seismic coupling to detect buried antitank landmines. *IEEE Transactions on Geoscience and Remote Sensing*, 39(6):1146–1154, 2001. ISSN 0196-2892.

J. M. Sabatier, H. E. Bass, L. N. Bolen, and K. Attenborough. Acoustically induced seismic-waves. *The Journal of the Acoustical Society of America*, 80(2):646–649, Aug. 1986a.

J. M. Sabatier, H. E. Bass, and G. R. Elliott. On the location of frequencies of maximum acoustic-to-seismic coupling. *The Journal of the Acoustical Society of America*, 80(4):1200–1202, 1986b.

S. Sampan. *Neural Fuzzy Techniques in Vehicle Acoustic Signal Classifcation*. PhD thesis, Virginia Polytechnic Institute and State University, 1997.

U. Sandberg. Tyre/road noise-myths and realities. In *Proceedings of the 2001 International Congress and Exhibition on Noise Control Engineering*, 2001.

U. Sandberg and J. A. Ejsmont. *Tyre/Road Noise Reference Book*. Informex, Kisa, Sweden, 1st edition, 2002.

R. Schalkoff. *Pattern Recognition: Statistical, Structual and Neural Approaches*. John Wiley & Sons Inc., 1992.

B. Scharf. *Critical Bands*, volume 1, chapter 5, pp. 159–202. Academic Press, Inc., 1970.

S. J. Searle. Efficient matched processing for localisation of a moving acoustic source. *Signal Processing*, 85:17871804, 2005.

E. G. Shower and R. Biddulph. Differential pitch sensitivity of the ear. *The Journal of the Acoustical Society of America*, 3(1A):7–7, 1931.

A. Singhal and D. E. Seborg. Clustering multivariate time-series data. *Journal of Chemometrics*, 19(8):427–438, 2005.

M. A. Sobreira-Seoane, A. R. Molares, and J. L. A. Castro. Automatic classification of traffic noise. *The Journal of the Acoustical Society of America*, 123(5):3823–3823, 2008.

N. Srour. Back propagation of acoustic signature for robust target identification. In *Proceedings of SPIE, The International Society for Optical Engineering*, volume 4232, pp. 399–408, 2001.

N. Srour and J. Robertson. Remote netted acoustic detection system: Final report. Technical Report ARL-TR-706, Army Research Laboratory, May 1995.

J. Stafsudd, S. Asgari, R. Hudson, K. Yao, and E. Taciroglu. Localization of short-range acoustic and seismic wideband sources: Algorithms and experiments. *Journal of Sound and Vibration*, 312(1-2):74–93, Apr. 2008. ISSN 0022-460X.

G. Succi, T. K. Pedersen, R. Gampert, and G. Prado. Acoustic target tracking and target identification - recent results. In *Unattended Ground Sensor Technologies and Applications, Proceedings of SPIE*, volume 3713, pp. 10 – 21, 1999.

G. P. Succi, G. Prado, R. Gampert, T. K. Pedersen, and H. Dhaliwal. Problems in seismic detection and tracking. In *Unattended Ground Sensor Technologies and Applications II, Proceedings of SPIE*, volume 4040, pp. 165 – 173, 2000.

S. Suchandt, M. Eineder, H. Breit, and H. Runge. Analysis of ground moving objects using srtm/x-sar data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(3-4):209–224, Dec. 2006.

Z. Sun, G. Bebis, and R. Miller. Monocular precrash vehicle detection: features and classifiers. *IEEE Transactions on Image Processing*, 15(7):2019–2034, 2006a. ISSN 1057-7149.

Z. Sun, G. Bebis, and R. Miller. On-road vehicle detection: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):694–711, 2006b. ISSN 0162-8828.

M. D. Swarbrick. *Acoustic Diagnosis of Heart Defects using Time Domain Signal Processing and Artificial Neural Networks*. PhD thesis, University of Hull, 2001.

A. Temko and C. Nadeu. Classification of acoustic events using svm-based clustering schemes. *Pattern Recognition*, 39(4):682–694, Apr. 2006.

D. Terzopoulos. Co-occurrence analysis of speech waveforms. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-33(1):5–30, 1985.

D. W. Thomas and B. R. Wilkins. Determination of engine firing rate from acoustic waveform. *Electronics Letters*, 6(7):193–194, 1970.

D. W. Thomas and B. R. Wilkins. Analysis of vehicle sounds for recognition. *Pattern Recognition*, 4(4):379–389, 1972.

Y. Tian, H. Qi, and X. Wang. Target detection and classification using seismic signal processing in unattended ground sensor systems. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) '02*, volume 4, pp. IV–4172, 2002.

M. B. Tmer and M. C. Demir. A genetic approach to data dimensionality reduction using a special initial population. In *IWINAC 2005: Artificial Intelligence and Knowledge Engineering Applications:A Bioinspired Approach: First International Work-Conference on the Interplay Between Natural and Artificial Computation*, volume 3562 of *Lecture Notes in Computer Science*, pp. 310–316. Springer Berlin / Heidelberg, June 15-18 2005.

C. Toth, A. Barsi, and T. Lovas. Vehicle recognition from lidar data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 23(23 (Part 3/W13)):163166, 2003.

L.-W. Tsai, J.-W. Hsieh, and K.-C. Fan. Vehicle detection using normalized color and edge map. *IEEE Transactions on Image Processing*, 16(3):850–864, 2007. ISSN 1057-7149.

M. Ture, I. Kurt, and Z. Akturk. Comparison of dimension reduction methods using patient satisfaction data. *Expert Systems with Applications*, 32(2):422–426, Feb. 2007.

M. F. Umer and M. S. H. Khiyal. Classification of textual documents using learning vecotr quantization. *Information Technology Journal*, 6 (1):154–159, 2007.

Unknown. Hundred killed in nigerian pipeline explosion. Guardian Unlimited, May 2006a. Accessed on 7th August, 2007.

Unknown. Nigeria mourns pipeline victims. BBC NEWS, 27th December, Decemebr 2006b. URL `http://news.bbc.co.uk/1/hi/world/africa/6211363.stm`. Last accessed on 7th August, 2007.

V. N. Vapnik. *The Nature of Statistical Learning Theory*. Statistics for Engineering and Information Science. Springer Science+BusinessMedia, LLC, 2nd edition, 2000.

T. Viangteeravat and A. Shirkhodaie. Multiple target vehicles detection and classification with low-rank matrix decomposition. In *System of Systems Engineering, 2007. SoSE '07. IEEE International Conference on*, pp. 1–8, 2007.

T. Viangteeravat, A. Shirkhodaie, and H. Rababaah. Acoustic signature analysis and data fusion of vehicles based on acoustic sensor arrays. In *Signal Processing, Sensor Fusion, and Target Recognition XVI, Proceedings of SPIE*, number 656705 in 6567, pp. 1 – 9, 2007a.

T. Viangteeravat, A. Shirkhodaie, and H. Rababaah. Multiple target vehicles detection and classification based on low-rank decomposition. In F. A. Sadjadi, editor, *Proceedings of SPIE: Automatic Target Recognition XVII*, number 6566 in 1, p. 65660R. SPIE, 2007b.

B. Walczak and D. L. Massart. Noise suppression and signal compression using the wavelet packet transform. *Chemometrics and Intelligent Laboratory Systems*, 36 (2):81–94, Apr. 1997. ISSN 0169-7439.

X. Wang and H. Qi. Acoustic target classification using distributed sensor arrays. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) '02*, volume 4, pp. IV–4186, 2002.

A. Webb. *Statistical Pattern Recognition*. Arnold, a member of the Hodder Headline Group, 1999.

R. L. Wegel and C. E. Lane. The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Phys. Rev.*, 23(2):266–, Feb. 1924.

M. V. Wickerhauser. *Adapted Wavelet Analysis from Theory to Software*. A.K. Peters, Ltd., 1994.

J. S. Wilson, editor. *Sensor Technology Handbook*. Elsevier Inc., 2005.

H. Wu and J. M. Mendel. Classifier designs for binary classifications of ground vehicles. In E. M. Carapezza, editor, *Proceedings of SPIE: Unattended Ground Sensor Technologies and Applications V*, number 5090 in 1, pp. 122–133. SPIE, 2003.

H. Wu and J. M. Mendel. Multi-category classification of ground vehicles based on their acoustic emissions. In E. M. Carapezza, editor, *Proceedings of SPIE: Unattended/Unmanned Ground, Ocean, and Air Sensor Technologies and Applications VI*, number 5417 in 1, pp. 31 – 42. SPIE, 2004.

H. Wu and J. M. Mendel. Classification of battlefield ground vehicles using acoustic features and fuzzy logic rule-based classifiers. *IEEE Transactions on Fuzzy Systems*, 15(1):56–72, 2007. ISSN 1063-6706.

H. Wu, M. Siegel, and P. Khosla. Vehicle sound signature recognition by frequency vector principal component analysis. In *Proceedings of IEEE Conference on IMTC (Instrumentation and Measurement Technology Conference)*, volume 1, pp. 429–434 vol.1, 1998.

H. Wu, M. Siegel, and P. Khosla. Vehicle sound signature recognition by frequency vector principal component analysis. *IEEE Transactions on Instrumentation and Measurement*, 48(5):1005–1009, 1999. ISSN 0018-9456.

H. Xiao, C. Cai, Q. Yuan, X. Liu, and Y. Weni. A comparative study of feature extraction and classification methods for military vehicle type recognition using acoustic and seismic signals. In *Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues*, volume

4681 of *LECTURE NOTES IN COMPUTER SCIENCE*, pp. 810–819. Springer Berlin / Heidelberg, 2007.

H. G. Xiao, C. Z. Cai, and Y. Z. Chen. Military vehicle classification via acoustic and seismic signals using statistical learning methods. *International Journal Of Modern Physics C*, 17 (2):197–212, Feb. 2006.

S. Yella, N. K. Gupta, and M. S. Dougherty. Comparison of pattern recognition techniques for the classification of impact acoustic emissions. *Transportation Research Part C: Emerging Technologies*, 15(6):345–360, Dec. 2007.

G. Yen and K.-C. Lin. Wavelet packet feature extraction for vibration monitoring. *IEEE Transactions on Industrial Electronics*, 47(3):650–667, 2000. ISSN 0278-0046.

W. A. Yost. *Fundamentals of Hearing: An Introduction*. Academic Press, 4th ed. edition, 2000.

J. Yu, Q. Tian, T. Rui, and T. Huang. Integrating discriminant and descriptive information for dimension reduction and classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(3):372–377, 2007. ISSN 1051-8215.

M. Zhu and A. Martinez. Subclass discriminant analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1274–1286, 2006. ISSN 0162-8828.

# Appendix A

# Data Collection Additional Information

Table A.1 lists meteorological data for each recording session as mu as available.

| Recording | Date | Location | Temperature | Maximum Wind | Humidity |
|---|---|---|---|---|---|
| I | Jan. 2008 | A(York, U.K.) | n/a | n/a | n/a |
| II | Jul. 2008 | B(York, U.K.) | 21 °C | 7.8 km/h | n/a |
| III | Sep. 2008 | A(York, U.K.) | 21 °C | 12.6 km/h | n/a |
| IV | Oct. 2008 | C(York, U.K.) | n/a | 23.0 km/h | n/a |
| V | Oct. 2009 | D(York, U.K.) | n/a | n/a | n/a |
| VI | Jul. 2010 | E(Japan) | 31 °C | n/a | 65-71 % |

Table A.1: Meteorological Data for Recording Sessions

The equipment used during each of recording session was:

- Microphone: Rode NT5 condenser cardioid microphone for sessions I to V, Beyer M201TG dynamic hypercardioid microphone for session IV.
- Geophone: Sensor SM-4 vertical basic geophone unit for all.
- Recorder: Edirol portable 4-channel recorder and Marantz portable 2-channel recorder for session I. Edirol portable 4-channel recorder for sessions II to V. Marantz portable 2-channel recorder for session IV.

Figure A.1 is a picture of one of the recording locations with the equipment.

Figure A.1: Picture of a Recording Location (in outskirts of York)

Also, some of characteristics of the two Sensor SM-4 vertical basic geophones utilised for the study have been tested.

- Resistance: 388 Ω and 387 Ω (measured at 24.9 °C).
- Polarity: Both were correct.
- Natural Frequency: 10.2 Hz and 10.1 Hz.

# Appendix B

# Feature Extraction Additional Information

Other distribution graphs for various feature extraction algorithms are shown here.

## B.1 ZCR

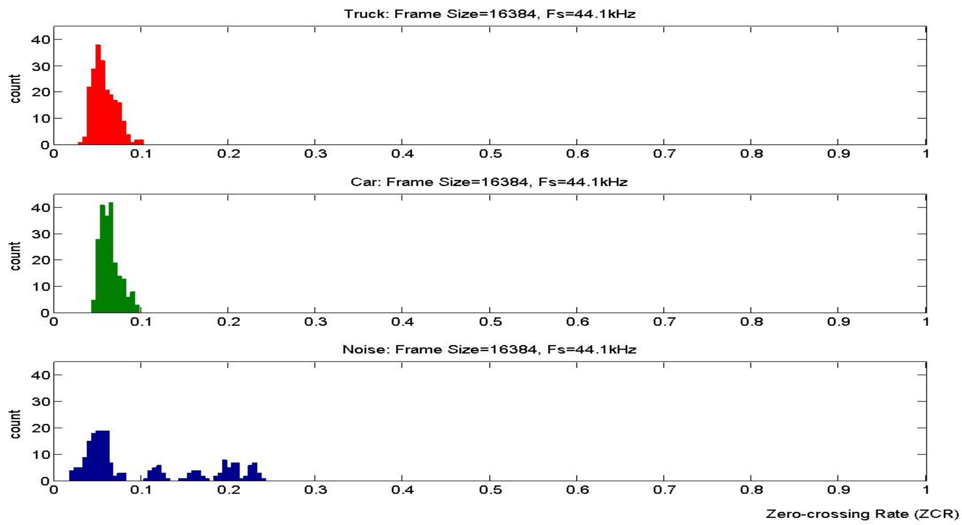


Figure B.1: Distribution of Acoustic ZCR, Frame Size=16384, Fs=44.1 kHz.

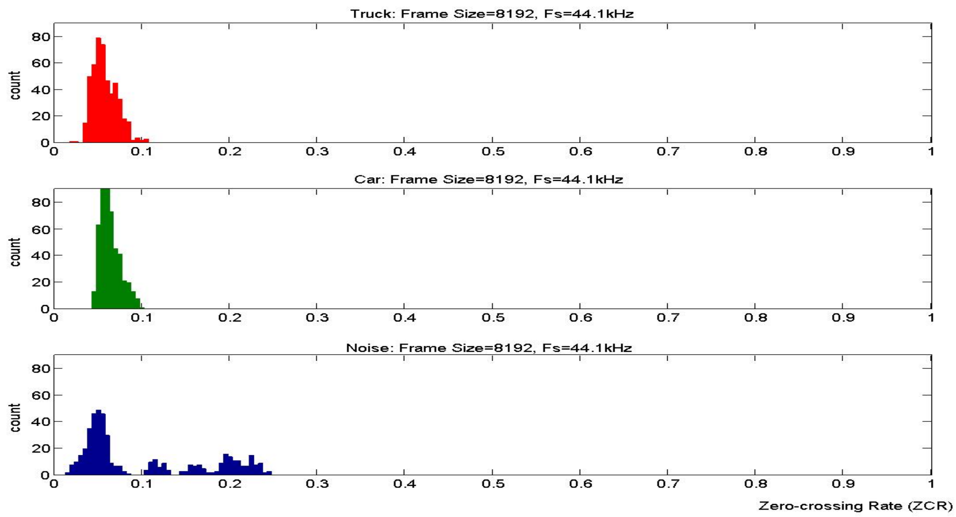


Figure B.2: Distribution of Acoustic ZCR, Frame Size=8192, Fs=44.1 kHz.

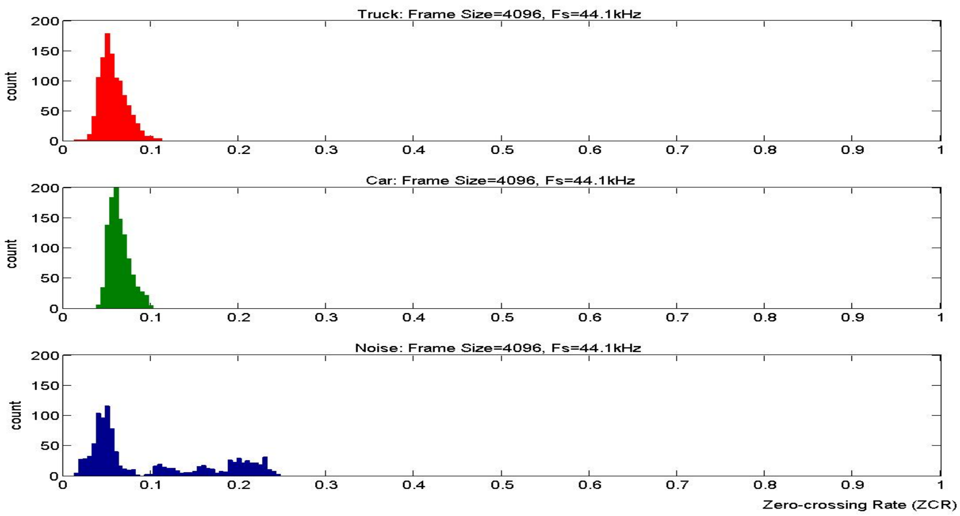


Figure B.3: Distribution of Acoustic ZCR, Frame Size=4096, Fs=44.1 kHz.

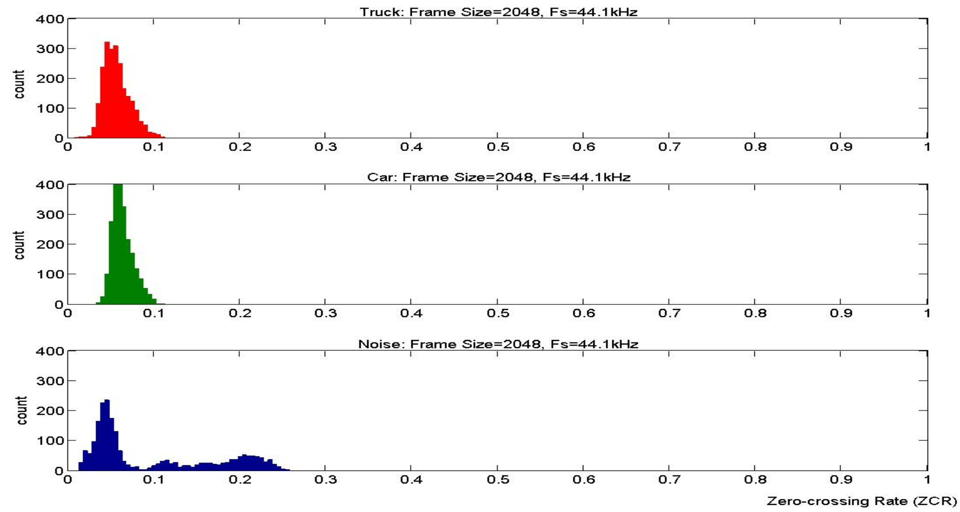


Figure B.4: Distribution of Acoustic ZCR, Frame Size=2048, Fs=44.1 kHz.

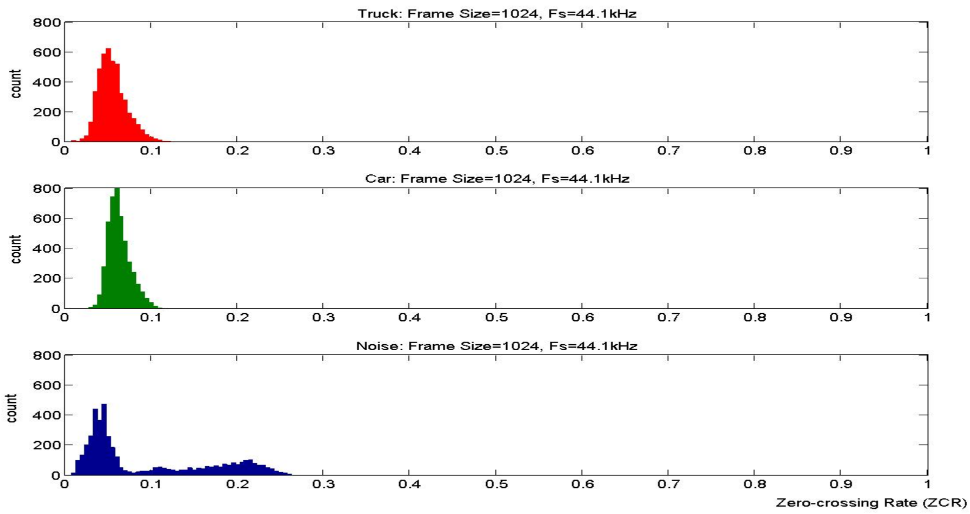


Figure B.5: Distribution of Acoustic ZCR, Frame Size=1024, Fs=44.1 kHz.

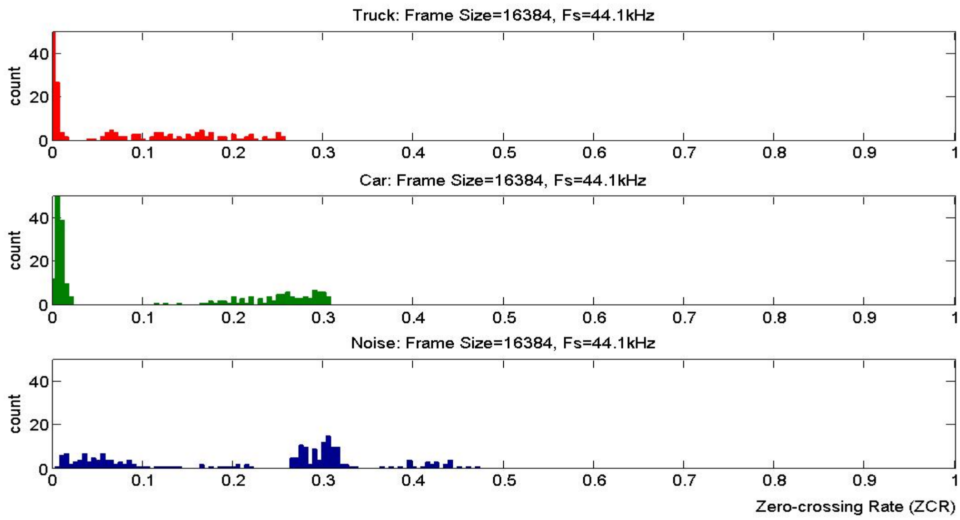


Figure B.6: Distribution of Seismic ZCR, Frame Size=16384, Fs=44.1 kHz.

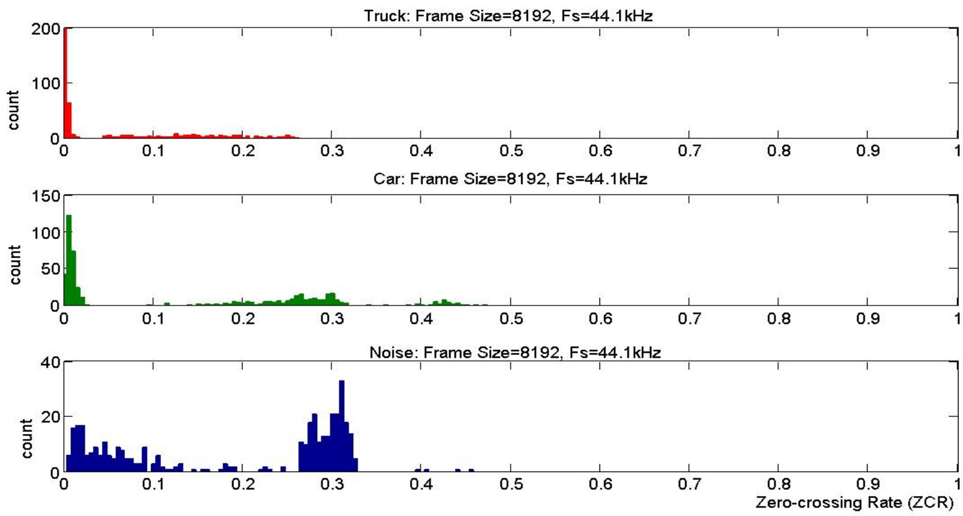


Figure B.7: Distribution of Seismic ZCR, Frame Size=8192, Fs=44.1 kHz.

# B.2 Energy / Log Energy

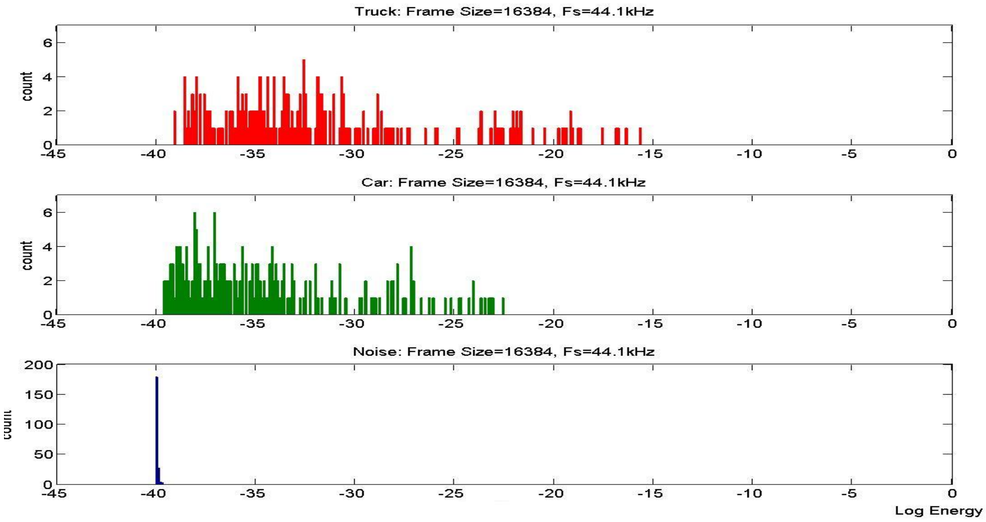


Figure B.8: Distribution of Acoustic Log Energy, Frame Size=16384, Fs=44.1 kHz.

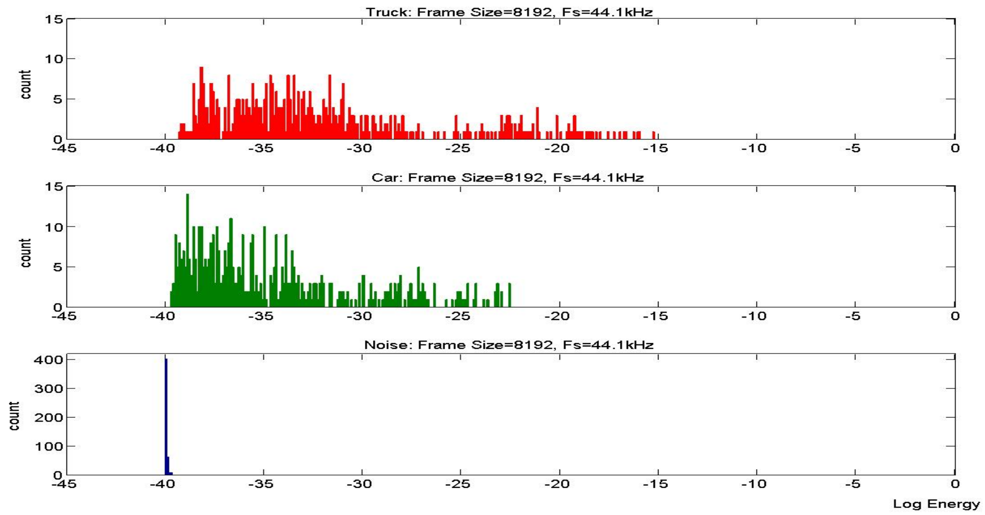


Figure B.9: Distribution of Acoustic Log Energy, Frame Size=8192, Fs=44.1 kHz.

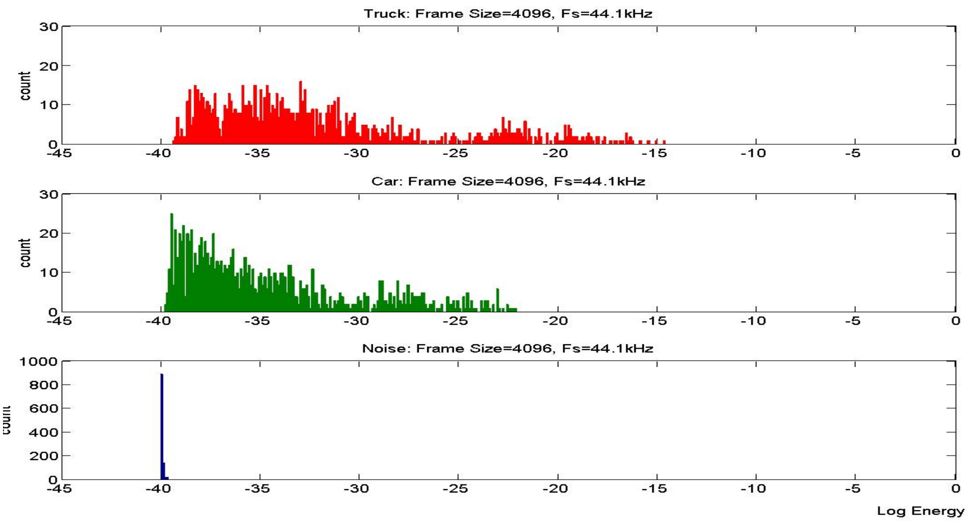


Figure B.10: Distribution of Acoustic Log Energy, Frame Size=4096, Fs=44.1 kHz.

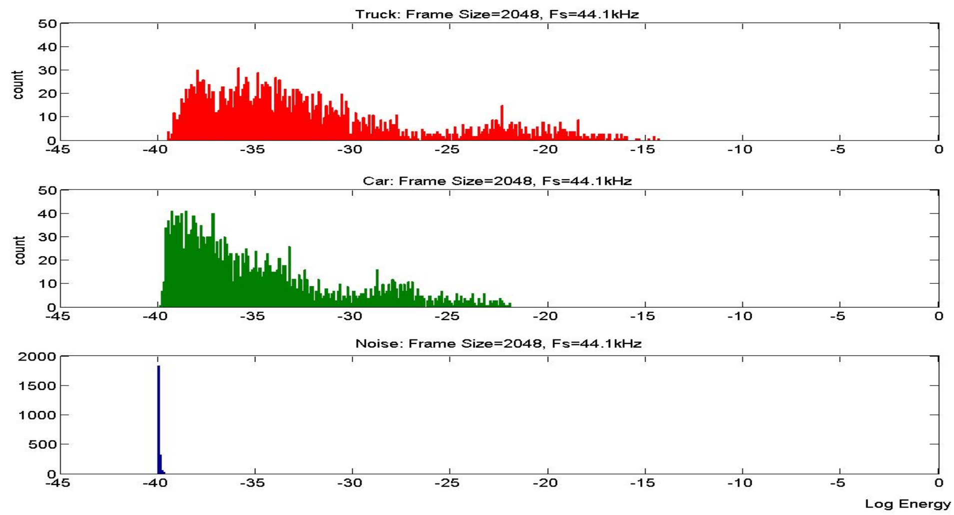


Figure B.11: Distribution of Acoustic Log Energy, Frame Size=2048, Fs=44.1 kHz.

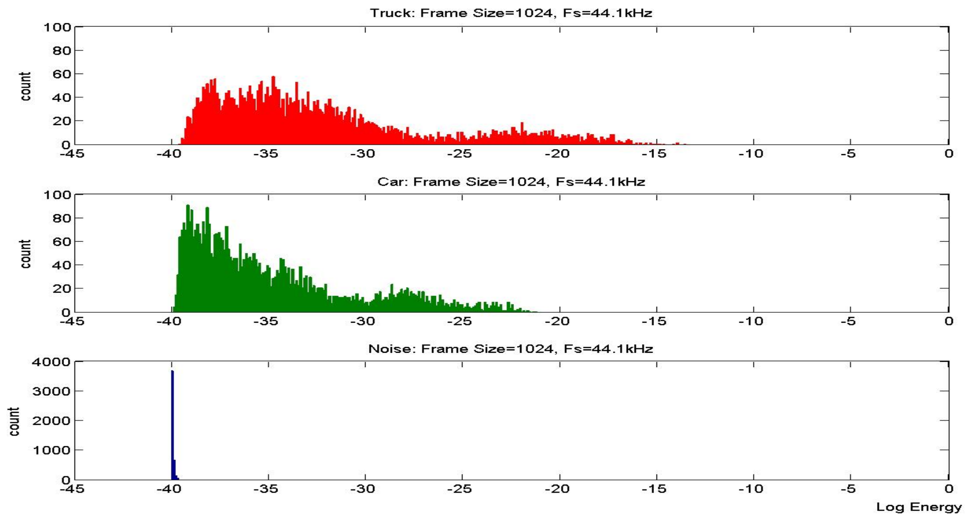


Figure B.12: Distribution of Acoustic Log Energy, Frame Size=1024, Fs=44.1 kHz.

## B.3 Time Domain Correlation

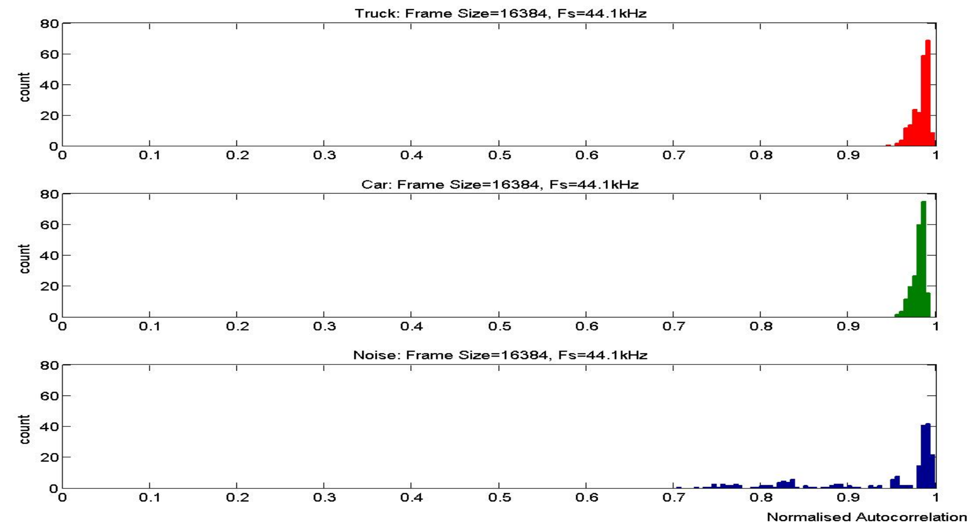


Figure B.13: Distribution of Acoustic Autocorrelation, Frame Size=16384, Fs=44.1 kHz.

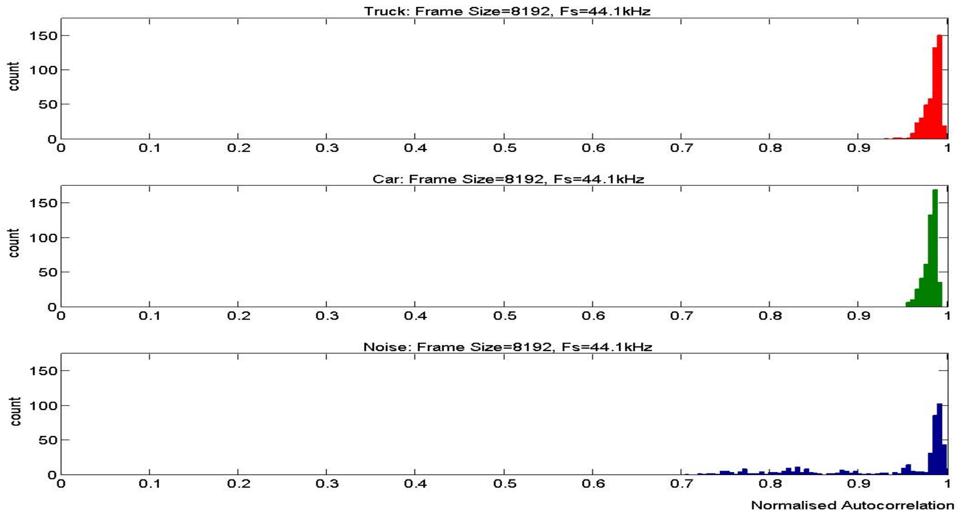


Figure B.14: Distribution of Acoustic Autocorrelation, Frame Size=8192, Fs=44.1 kHz.

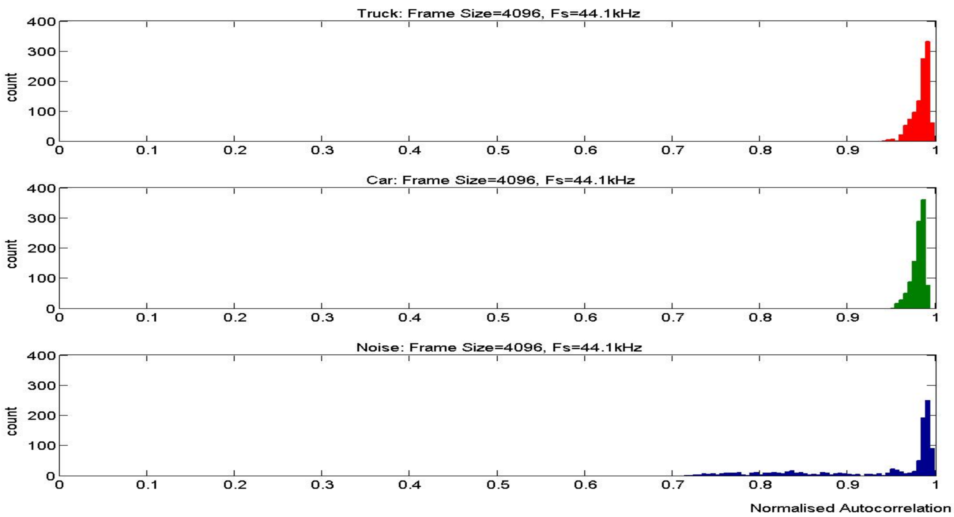


Figure B.15: Distribution of Acoustic Autocorrelation, Frame Size=4096, Fs=44.1 kHz.

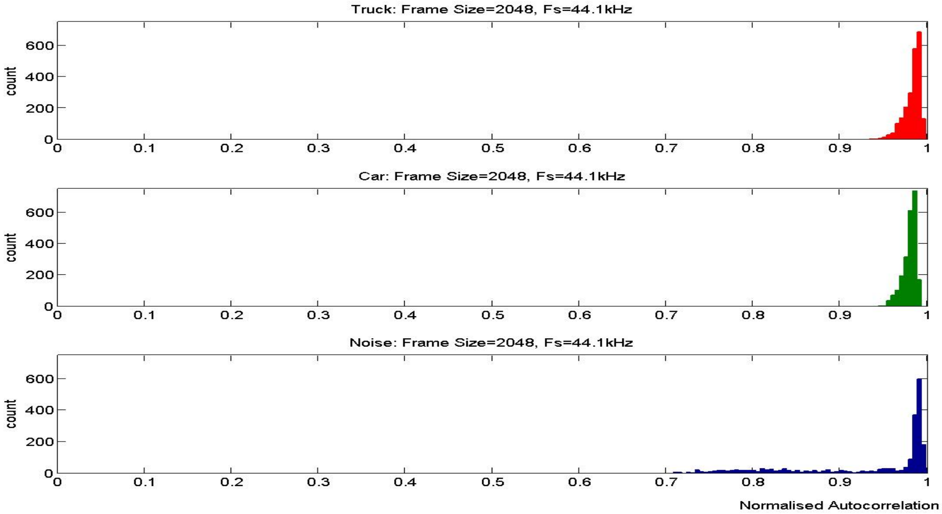


Figure B.16: Distribution of Acoustic Autocorrelation, Frame Size=2048, Fs=44.1 kHz.

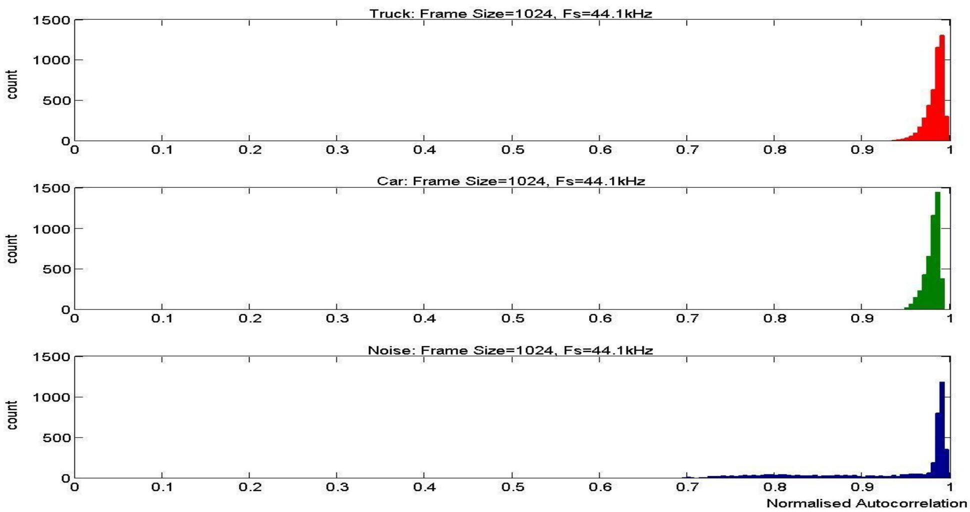


Figure B.17: Distribution of Acoustic Autocorrelation, Frame Size=1024, Fs=44.1 kHz.

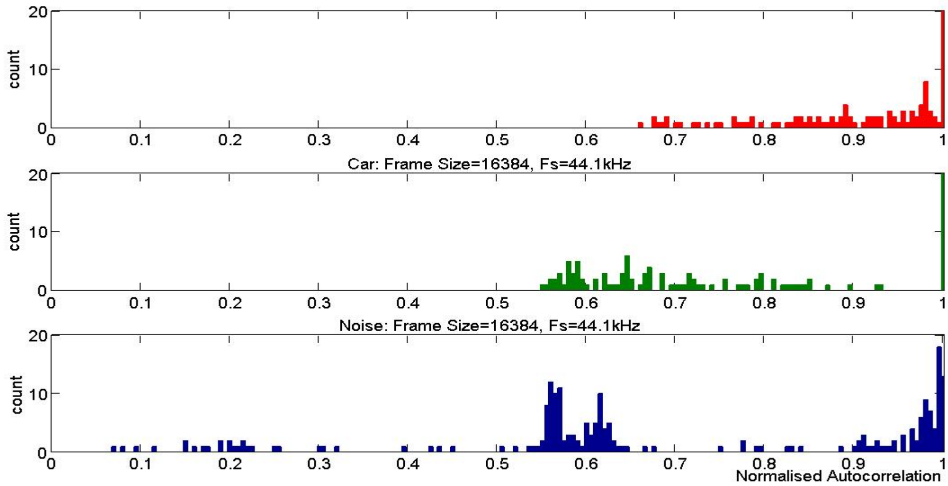


Figure B.18: Distribution of Seismic Autocorrelation, Frame Size=16384, Fs=44.1 kHz.

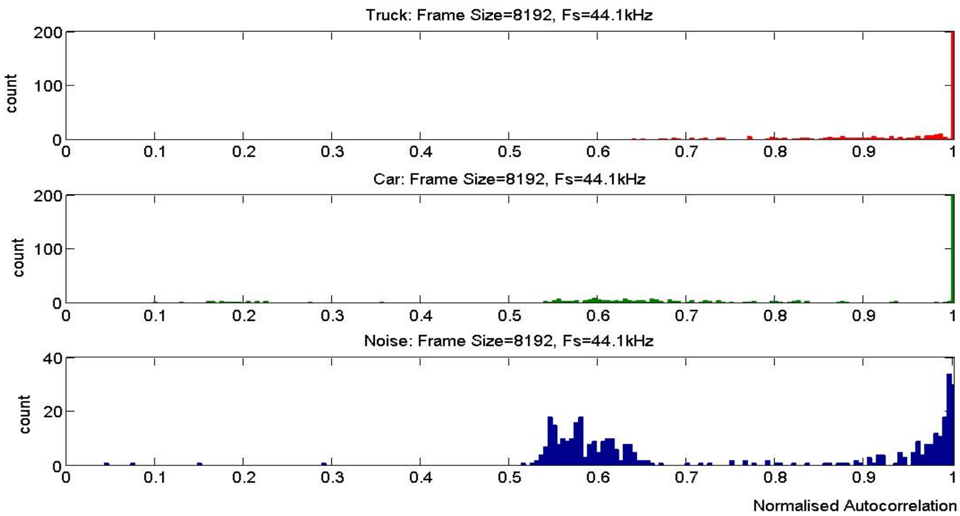


Figure B.19: Distribution of Seismic Autocorrelation, Frame Size=8192, Fs=44.1 kHz.

# Appendix C

# LPC Function

During the software evaluation for the LPC algorithm, by comparisons between the sets of coefficients collected with the author's code and with MATLAB showed that the polarity of each coefficient was of the opposite sign while the magnitude was the same. The minimum error values of two techniques differed but by a consistent ratio. Nonetheless, these discrepancies were insignificant for distinction between various vehicle classes since the condition is equal for any input signals. Therefore the author's code (below) has been used for the experiment presented in the main text.

Table C.1 shows results as in Table 7.34 but implemented with the LPC function of MATLAB. Most results are either the same or very similar, a part from those with PCA (without normalising) pre-classification signal processing for three longer frame lengths, producing less accurate results. Similarly, Table C.2 and C.3 are classification of two techniques but without using the minimum error values. Understandably, the two sets of results are identical. Moreover, all the three sets of results still indicated that the best pre-classification signal processing method and the best frame size are the same as the case explained in Section 7.3.7, with marginally better performance for the last two.

| Frame Size | 32768 | 16384 | 8192 | 4096 | 2048 |
|---|---|---|---|---|---|
| no-processing | 78.57% | 75.00% | 76.79% | 83.39% | 83.67% |
| normalised | 87.50% | **93.30**% | 91.07% | 89.20% | 89.33% |
| PCA(3) | 26.79% | 35.71% | 34.92% | 32.59% | 33.21% |
| normalised & PCA(3) | 78.57% | 81.70% | 53.57% | 87.68% | 88.56% |
| normalised & PCA(2) | 71.43% | 74.11% | 66.86% | 87.86% | 88.44% |
| normalised & PCA(3) on LPC only | 76.79% | 63.84% | 74.21% | 79.64% | 80.44% |

Table C.1: SVM Classification (2-class) Results by Combination of Seismic ZCR, Log Energy and Normalised Autocorrelation as well as Acoustic LPC coefficients as in Table 7.34 but calculated with MATLAB's code.

| Frame Size | 32768 | 16384 | 8192 | 4096 | 2048 |
|---|---|---|---|---|---|
| no-processing | 78.57% | 75.00% | 76.59% | 83.48% | 83.80% |
| normalised | 87.50% | 93.75% | 90.28% | 89.20% | 89.37% |
| PCA(3) | 33.93% | 61.61% | 62.70% | 63.13% | 81.25% |
| normalised & PCA(3) | 75.00% | 58.04% | 76.19% | 87.68% | 66.07% |
| normalised & PCA(2) | 76.79% | 74.11% | 66.87% | 87.86% | 67.26% |
| normalised & PCA(3) on LPC only | 71.43% | 75.89% | 73.81% | 79.82% | 80.44% |

Table C.2: SVM Classification (2-class) Results by Combination of Seismic ZCR, Log Energy and Normalised Autocorrelation as well as Acoustic LPC coefficients calculated with the author's code (using coefficients only).

| Frame Size | 32768 | 16384 | 8192 | 4096 | 2048 |
|---|---|---|---|---|---|
| no-processing | 78.57% | 75.00% | 76.59% | 83.48% | 83.80% |
| normalised | 87.50% | 93.75% | 90.28% | 89.20% | 89.37% |
| PCA(3) | 33.93% | 61.61% | 62.70% | 63.13% | 81.25% |
| normalised & PCA(3) | 75.00% | 58.04% | 76.19% | 87.68% | 66.07% |
| normalised & PCA(2) | 76.79% | 74.11% | 66.87% | 87.86% | 67.26% |
| normalised & PCA(3) on LPC only | 71.43% | 75.89% | 73.81% | 79.82% | 80.44% |

Table C.3: SVM Classification (2-class) Results by Combination of Seismic ZCR, Log Energy and Normalised Autocorrelation as well as Acoustic LPC coefficients calculated with MATLAB's code (using coefficients only).

```
function [LPCcoeff,MinimumError] = LPC_auto_Dec2007(Input,p)
%N Evans 04/12/2007-
%Function to calculate LPC coefficients (autocorrelation method)
%Based on [Rabiner and Schafer, 1978, p.402]
%this function only takes a column vector input
%*****************************************************************
%input error check
if (nargin == 0),
    error('At least, one parameter is required');
end
if p <1
    error('At least order p needs to be one');
end
[FrameSize,col]=size(Input); %find the size of input
if (col>=2)
    error('input needs to be a column vector');
end
%*****************************************************************
%process input signal
%prepare a matrix to store autocorrelation function matrix
Autocorrelation=zeros(p+1,1);
matrix_pxp=zeros(p,p);
matrix_px1=zeros(p+1,1);
%calculate autocorrelation per delay
%R(delay)=sum(S(m)*S(m+delay)) (for m=1:size(S)-delay)
p_Matrix=repmat(0:1:p-1, [p, 1]);
Dealy_Matrix=abs(p_Matrix-p_Matrix');

for Delay=0:p
   DelayedInput=circshift(Input, [Delay, 0]);
   DelayedInput(1:Delay)=false;
   Autocorrelation(Delay+1,1)=sum(Input.*DelayedInput);
  PositionIndex=Dealy_Matrix==Delay;
  matrix_pxp(PositionIndex)=Autocorrelation(Delay+1,1);
  matrix_px1(Delay+1,1)=Autocorrelation(Delay+1,1);
end
LPCcoeff=matrix_pxp\matrix_px1(2:p+1,1);
%*****************************************************************
%minimum mean squared prediction error
%E=R(0)-sum(LPCcoeff(k)*R(k)), where k=1:p
MinimumError=matrix_pxp(1,1)-sum(LPCcoeff'*matrix_px1(2:p+1,1));
```

LPC Code Written by the Author

# Appendix D

# Publication List

N. Evans and D. Chesmore. *Automated Identification of Vehicles using Acoustic Signal Processing*. Acoustics'08 (A peer reviewed conference by IEEE and Acoustic Society of America), Paris, 30th June - 4th July, 2008.

N. Evans and D. Chesmore. *Automated Acoustic Identification of Vehicles*. In Proceedings of the Institute of Acoustics, volume 30, pp. 238-245, 2008.