2011

Classification Of Civilian Vehicle Sounds Using A Large Database Of Vehicle Sounds

David Tettey
North Carolina Agricultural and Technical State University

CLASSIFICATION OF CIVILIAN VEHICLE SOUNDS USING A LARGE DATABASE OF VEHICLE SOUNDS

by

David Tettey

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE

Department: Electrical and Computer Engineering
Major: Electrical and Computer Engineering
Major Professor: Dr. M. Bikdash

North Carolina A&T State University
Greensboro, North Carolina
2011

School of Graduate Studies
North Carolina Agricultural and Technical State University

This is to certify that the Master's Thesis of

David Tettey

has met the thesis requirements of
North Carolina Agricultural and Technical State University

Greensboro, North Carolina
2011

Approved by:

Dr. M. Bikdash Major Professor	Dr. J. C. Kelly Committee Member
Dr. Robert Li Committee Member	Dr. J. C. Kelly Department Chairperson

Dr. Sanjiv Sarin
Dean of Graduate Studies

DEDICATION

This work is dedicated to my wife and children, my dad, late mum, mother-in-law and siblings for their prayers and unflinching support through my years of education.

BIOGRAPHICAL SKETCH

David Tettey was born in Accra, Ghana, on July 10, 1975. He did his undergraduate studies in Electrical Engineering at the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana, from 1997 to 2001. He joined the Department of Electrical and Computer Engineering at North Carolina Agricultural and Technical State University, Greensboro, North Carolina, in fall 2008 for the Master of Science degree in Electrical Engineering. He has been working in the Advanced Robotics Laboratory at North Carolina Agricultural and Technical State University under the supervision of Dr. M. Bikdash since 2009.

ACKNOWLEDGMENTS

Many people have contributed to this thesis in diverse ways. To all these people, I would like to express my genuine thanks and gratitude. First, my gratitude goes to God for seeing me through this work successfully. Secondly, I would like to express my deepest gratitude to my advisor, Dr. M. Bikdash for sharing his expertise to help complete this work. His insightful contributions and suggestions have helped me to better understand the problem statement for this work. He has been a major source of encouragement to me.

I also want to thank all the staff members of the Department of Electrical and Computer Engineering at North Carolina Agricultural and Technical State University, in Greensboro, North Carolina, for their help and encouragement in pursuit of my graduate education. My keen appreciation goes to my committee members Dr. J. C. Kelly and Dr. Robert Li for their invaluable contribution to the success of this research work.

I am very grateful to my wife and family for their prayers and encouragement. I also want to thank Marvin Aidoo and Daniel Opoku for helping in the data collection process. Finally, I would like to thank all my friends at North Carolina Agricultural and Technical State University for their help and support in making my graduate education very successful and memorable.

This work was supported by US Army RDECOM under contract W 911QX–07–C–0062.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES **PAGE**

LIST OF TABLES

TABLES **PAGE**

LIST OF SYMBOLS

STFFT	Short Time Fast Fourier Transform
DFT	Discrete Fourier Transform
PCA	Principal Component Analysis
PC	Principal Component
MLPN	Multi-layer Perceptron Network
USB	Universal Serial Bus
dB	Decibel
RPM	Revolution Per Minute
PSD	Power Spectral Density
SK	Spectral Kurtosis
S_x	Covariance matrix
ω	Frequency
μ_x	Spectral Centroid
γ_1	Nondimensional Skewness
γ_2	Nondimensional Kurtosis
T_1	First Tristimulus
T_2	Second Tristimulus
T_3	Third Tristimulus

ABSTRACT

Tettey, David. CLASSIFICATION OF CIVILIAN VEHICLE SOUNDS USING A LARGE DATABASE OF VEHICLE SOUNDS. **(Major Advisor: Dr. M. Bikdash)**, North Carolina Agricultural and Technical State University.

We have completed the building of an extensive database of civilian vehicle sounds. The database consists of correlated acoustic and seismic signatures of a large number (exceeding 850) of civilian vehicles. Each acoustic signature is obtained through two high-quality microphones separated by 25 feet, and whose signals are exactly synchronized. In this work, spectral and tristimulus features of civilian vehicle sounds are computed and then submitted to further processing using principal component analysis. The "super" features, derived after principal component analysis is performed, are then used for classification. In this research effort, the performance of the quadratic classifier with that of the neural network classifier is compared. Results presented here show that the neural network classifier out-performs the quadratic classifier in distinguishing different and same branded vehicle sounds. The classification usually has small (at times 0%) classification errors.

CHAPTER 1

INTRODUCTION

1.1 Literature Survey

Significant research work has been carried out recently on civilian vehicle classification using spectral features [1], neural networks and many other signal processing algorithms and pattern recognition tools. Mgaya et al. [2] for instance designed and implemented a data-fusion software for vehicle surveillance and applications using a distributed network of acoustic sensors. They used 3 civilian vehicles namely a Toyota Corolla, a Mitsubishi Mirage and a Toyota Sienna, 2 Sports Utility Vehicles (SUV), and 2 High Mobility Multipurpose Wheeled Vehicles (HMMWV) for their data analysis. The SUVs and the HMMWV data were provided by the military. Takechi et al., [3] also introduced a system of detecting a car by measuring the automobile sound of three vehicles. Thus far, research has suffered from limited availability of vehicle sound data for analysis. The goal of this study is to build a clean database of civilian vehicle sounds. We believe that for a good statistical analysis of vehicle sounds, it is prudent to use as much data as possible to validate findings.

One significant addition to this study is the use of an accelerometer to collect seismic data. Moving vehicles generate a succession of disturbances over the ground [4]. These ground disturbances propagate away from the source as seismic waves which are usually difficult to detect acoustically. Accelerometers become useful devices for taking such seismic measurements. In [5], Krishnan et al. used accelerometers for activity

detection. Nishkam et al. also used it in the detection of eight activities such as standing, walking, running, climbing up stairs, climbing down stairs, sit-ups, vacuuming and brushing of teeth [6]. Apart from activity detection, accelerometers can be used for context awareness. In [7], Seon et al. used it to determine a user's location and detect transitions between preselected locations. They used the measured and angular velocity data gathered through an accelerometer. In this experimental research work, the measurements from the accelerometer will provide understanding on the contribution of tire noise to the total vehicle sound.

A very important area that has gone through extensive research work is the study of higher order statistics [8] and [9]. These statistics are useful in problems where non-Gaussianity, nonminimum phase systems, colored noise, or nonlinearities are important and must be accounted for. In [9], Mendel described the application of higher-order statistics to the identification of nonminimum phase channels from noisy output measurements. Because many real world problems are non-Gaussian, higher-order statistics have found applicability in fields such as sonar, radar, seismic data processing etc. Work in higher-order statistics has led to several analysis tools complementary to classical second order methods. One of such useful tools is the fourth-order cumulant based kurtosis, which provides a measure of distance to gaussianity. It was used to detect randomly occurring signals in [10] and [11]. In the frequency domain, the Spectral Kurtosis (SK) of a signal is defined as the kurtosis of its frequency components. In [12] and [13], the SK approach is generalized by using the frequency component modulus. In this work, we focus heavily on spectral feature models such as the spectral kurtosis and

the spectral skewness to compute the seismic and acoustic spectral features of moving vehicles. We believe that vehicle attributes can be derived from these computed features. In addition, attention has been paid to other features such as the tristimulus features for the vehicle engine sound classification problem at hand. The primary objectives of this work are as follows:

1. To build a database of acoustic/seismic vehicle engine sounds.
2. To compute spectral shape and tristimulus features from the database of vehicle engine sounds.
3. To use statistical and data reduction techniques to extract super features for classification.
4. To use the super features as input to a quadratic and an artificial neural network classifier.
5. To compare the performance of the quadratic classifier and the artificial neural network classifier.

1.2 Synopsis

This thesis is organized as follows. Chapter 2 presents background material. Particular emphasis is given to techniques that were applied during our experimental work. It has three sections. Section 2.1 presents Principal Component Analysis (PCA). Section 2.2 discusses the Quadratic classifier. Section 2.3 presents the Artificial Neural Network classifiers. Chapter 3 describes the technical details of the equipment used, the experimental setup and the data collection methodology followed in the recording of acoustic/seismic vehicle sounds. The features used for classification are also presented. Chapter 4 describes seismic/acoustic spectral feature computation. Chapter 5 focuses on

efforts to extract super features from the acoustic/seismic spectral features. The mean of the spectral features are compared with the principal components of the spectral features. Chapter 6 discusses the tristimulus feature computation. Chapter 7 presents some results based on the Quadratic and Neural Network classifiers. Chapter 8 presents our conclusions and future work.

CHAPTER 2

BACKGROUND

2.1 Principal Component Analysis

In [14], Yang et al. demonstrated the Berkeley mote-based implementation of an acoustic vehicle system which was based on the Short Time Fast Fourier Transform (STFFT) and PCA. They converted the time-domain sound signal to a time-frequency domain signal using STFFT. The resulting high dimensions of the time-frequency domain features were reduced using PCA. Thus, they utilized the dimension reduction property of PCA to enable their application fit to a Berkeley-mote device. In [15], Junwen et al. presented a novel PCA classifier which calculated the principal components of each class of vehicle images to solve the problem of vehicle recognition and detection using static images. They performed this experiment on both artificial and real world data and obtained good classification. Furthermore, Meta et al. presented a novel vehicle-classification algorithm in [16]. In their work, they performed a Discrete Fourier Transform (DFT) on the signal generated by a single inductive loop detector and transformed the DFT features using PCA. They exploited PCA for decorrelation and dimensionality reduction. Clearly, research is replete with many such efforts where PCA has been extensively used in solving the problem of classification.

PCA is a technique for reducing the number of variables in a dataset while at the same time accounting for as much of the variation in the data set as possible. Its earliest descriptions were given by Pearson [17] and Hotelling [18]. PCA operates by

transforming a set of correlated variables into a new set of uncorrelated variables that are called Principal Components (PCs). In addition to being uncorrelated, the principal components are orthogonal and are ordered in terms of the variability they represent. That is, the first principal component represents the greatest amount of variability in the original data set. Each succeeding orthogonal component accounts for as much of the remaining variability as possible.

PCA is considered an eigenproblem. In [19], Nikon et al. summarized this eigenproblem in the following steps:

1. Obtain the feature matrix f_x from the data. Each column of the matrix defines a feature vector.

2. Compute the covariance matrix $S_x = f_x^T f_x$. ^{T} is the transpose. This matrix gives information about the linear independence between the features.

3. Obtain eigenvalues by solving the characteristic equation

$$\det|(\lambda_i I - S_x)| = 0,$$

where I is the identity matrix. The eigenvalues form the diagonal covariance matrix S_y. Since the matrix is diagonal, each element is the variance of the transformed data.

4. Obtain the eigenvectors by solving for b_i in

$$(\lambda_i I - S_x) b_i = 0$$

for each eigenvalue. Eigenvectors should be normalized and linearly independent.

5. The transformation matrix B is obtained by considering the eigenvectors as their columns.

6. Obtain the transform features by computing $f_y = f_x B^T$. The new features are linearly independent.

7. For classification applications, select the features with large values of λ_i. λ_i measures the variance and features that have large range of values will have large variance.

8. To reduce the dimensionality of the new feature vectors, the components of λ_i are set to zero. Features in the original data space can be obtained by $f_x^T = B^T f_y^T$.

2.2 Quadratic Classifier

One standard approach to a supervised classification problem is the Quadratic Discriminant Analysis (QDA). It models the likelihood of each class as a Gaussian distribution, then uses the posterior distributions to estimate the class for a given test point [20]. The QDC is based on Bayes's rule [21] and [8]. For *M* pattern classes, the multivariate normal density function is governed by

$$p(x/c_i) = \frac{1}{(2\pi)^{n/2}|S_i|^{1/2}} \exp[-\frac{1}{2}(x-\bar{x}_i)^T \pi S_i^{-1}(x-\bar{x}_i)],$$

for $i = 1, 2, ..., M,$ where each density is specified by the mean vector $\bar{x}_i$ and covariance matrix S_i of the class. They are defined as

$$\bar{x}_i = E_i\{x\},$$

and

$$S_i = E_i\{(x-\bar{x}_i)(x-\bar{x}_i)^T\}.$$

$E_i\{\cdot\}$ denotes the expectation operator over the patterns of class c_i. The covariance matrix S_i is symmetric and positive semi-definite. The diagonal element s_{kk} is the variance of the k^{th} element of the pattern vectors. The off-diagonal element s_{jk} is the

covariance of x_j and x_k. When x_j and x_k are statistically independent, $s_{jk} = 0$. The decision function for class c_i may be chosen as

$$\ln D_i(x) = \ln[\, p(x / c_i)\, p(c_i)].$$

Then

$$d_{\,i}(x) = \ln[\, p(x / c_i)\, p(c_i)] = \ln\, p(x / c_i) + \ln\, p(c_i).$$

Substituting the equations above, we have

$$d_i(x) = -\frac{n}{2}\ln 2\pi - 0.5\ln|S_i| - 0.5(x - \overline{x}_i)^T S_i^{-1}(x - \overline{x}_i) + \ln\, p(c_i).$$

Since the term $\frac{n}{2}\ln 2\pi$ does not depend on i, it can be eliminated from the expression to yield

$$d_i(x) = -0.5\ln|S_i| - 0.5(x - \overline{x}_i)^T S_i^{-1}(x - \overline{x}_i) + \ln\, p(c_i).$$

The above equations were used in computing the quadratic decision boundary for the quadratic classifier.

2.3 Neural Network Classifier

In [22], Wei et al. presented a novel method of vehicle classification using vertices and their topological structure as feature vectors and adopted the multi-layer perceptron network (MLPN) to recognize vehicles. Experimental results gave over 91% success rates for the neural network classifier used. In [23], Abdullah et al. used Forward Scattering Radar (FSR) data to automatically classify vehicles into one of three categories; namely, “small vehicle”, “medium vehicle” and “large vehicle”. They used the vehicle length as a feature for a back-propagation neural network and achieved good

classification. These and many other works emphasize the uniqueness and the ability of the neural network classifier to accurately classify vehicle signatures.

An artificial neural network is a network of artificial neurons that are connected through synapses or weights. It is primarily based on the emulation of our biological neural system. Warren McCulloch, a neuro-physiologist, and Walter Pittsit, a mathematician, began the pioneering work in artificial neural network in 1943 [24]. They composed a theme on the working of neurons and used electrical circuits to model neural networks. Work in neural network was motivated by the way the brain performs a particular function. Sumathi et al. have estimated that the human brain has about 10 billion neurons. Each of these neurons can connect up to about 200,000 other neurons.

The neuron is the basic processing element of a neural network. Figure 2.1 shows an artificial neuron with input and output elements, and processing functions. From Figure 2.1, the three basic elements we can identify are:

1. A set of input elements x_j multiplying connecting links characterized by a weight w_{kj}.

2. An Adder that performs a linear combination of the input elements, weighted by the respective synapses of the neuron.

3. An activation function that limits the amplitude of the output of the neuron to a permissive range of either [0 1] or [-1 1]. The most common activation function is the sigmoid function.

From Figure 2.1, initial weights w_{kj} are randomly chosen. These weights are updated until the network converges.

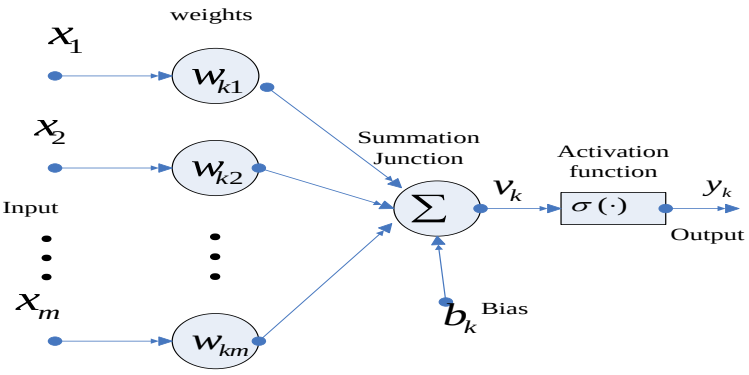

Figure 2.1. A nonlinear model of a neuron

In [25], Latha et al. used the back propagation neural network for face recognition. In our work, we implemented the multilayer feedforward network with the back propagation learning algorithm in Matlab. Two hidden layers were used. Figure 2.2 shows the multilayer feedforward network with the input layer denoted by x_i, the hidden layer h_i and the output layer o_i.

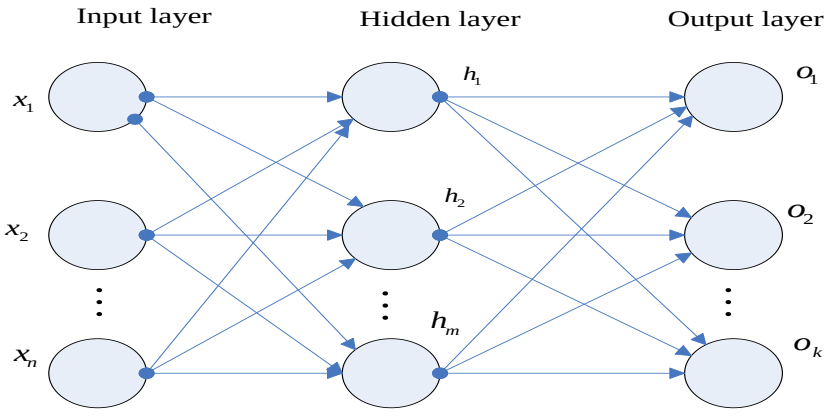

Figure 2.2. Multilayer feed forward network

CHAPTER 3

DATA-COLLECTION METHODOLOGY

3.1 Survey of Data collection sites

A survey was conducted by visiting a number of sites for the data collection process. The sites are:

1. Walmart shopping centers located at Wendover Avenue and at East Cone Boulevard in Greensboro, NC.
2. Two different highway exits on I40 East.
3. Best Buy located on South 40 Drive, Greensboro, NC.

The Walmart shopping centers had lots of vehicular activities. There were also lots of human activities such as the pushing of shopping carts, loitering and chatting by shoppers. These activities generated a lot of noise. The two Walmart shopping centers were therefore not suitable for data collection. Furthermore, high speeding vehicles on the I40 East highway also generated very loud noises. The exits on the I40 East highway were therefore not suitable for the data collection.

The back of Best Buy shopping center has a lot of vehicular activity. It serves as the entry and exit point to Best Buy. It is free from the noise generated due to the pushing of shopping carts because all the shopping carts are left at the parking lot which is in front of the Best Buy building. There is also very minimal loitering and chatting about as compared to the entrance of the building. It was therefore more reasonable to do the data collection at the back of the Best Buy building. Figure 3.1 shows the layout at the Best

Buy site for the data collection. The layout shows the parking lot in front of the entrance where all the carts are left. It also shows the position where the data collection process was undertaken.

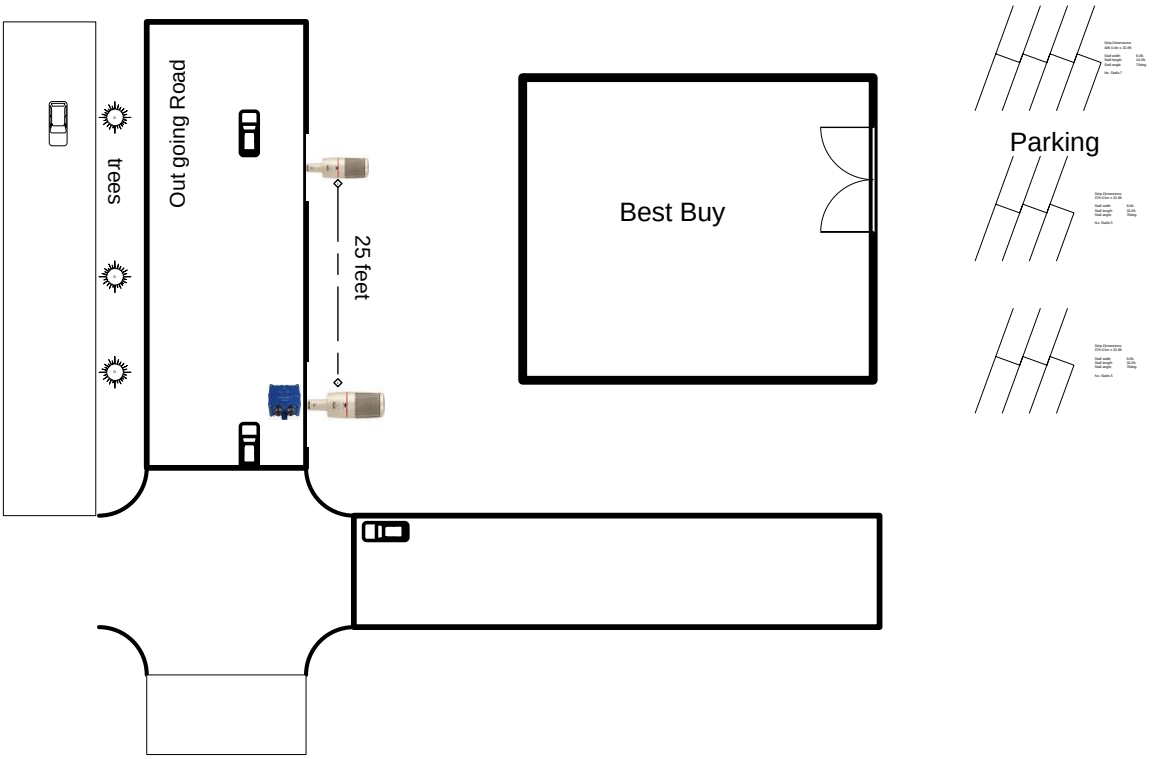

Figure 3.1. Site layout for data collection exercise

3.2 Instrument Setup

As shown in Figure 3.2, two microphones were mounted on two different tripods and connected to an AudioBox. The AudioBox was then connected to the USB port of a laptop. The accelerometer was bolted to a steel plate and connected to the second USB

port of the same laptop. The accelerometer was placed on the road from a distance of about 3 feet away from the path of moving vehicles. Acoustic gel was applied between the area of contact of the accelerometer and the road for good conductivity. One microphone was placed near the accelerometer and the other about 25 feet away.

A mark was drawn on the road 30 feet away from each microphone. These marks helped in determining when to start the recording and at what point to stop. The total distance between the marked portions on the road was 85 feet. Figure 3.3 shows the chalk markings on the street. Vehicle sounds and vibrations of different moving cars have been recorded for this study using the setups shown in Figure 3.2 and Figure 3.3.

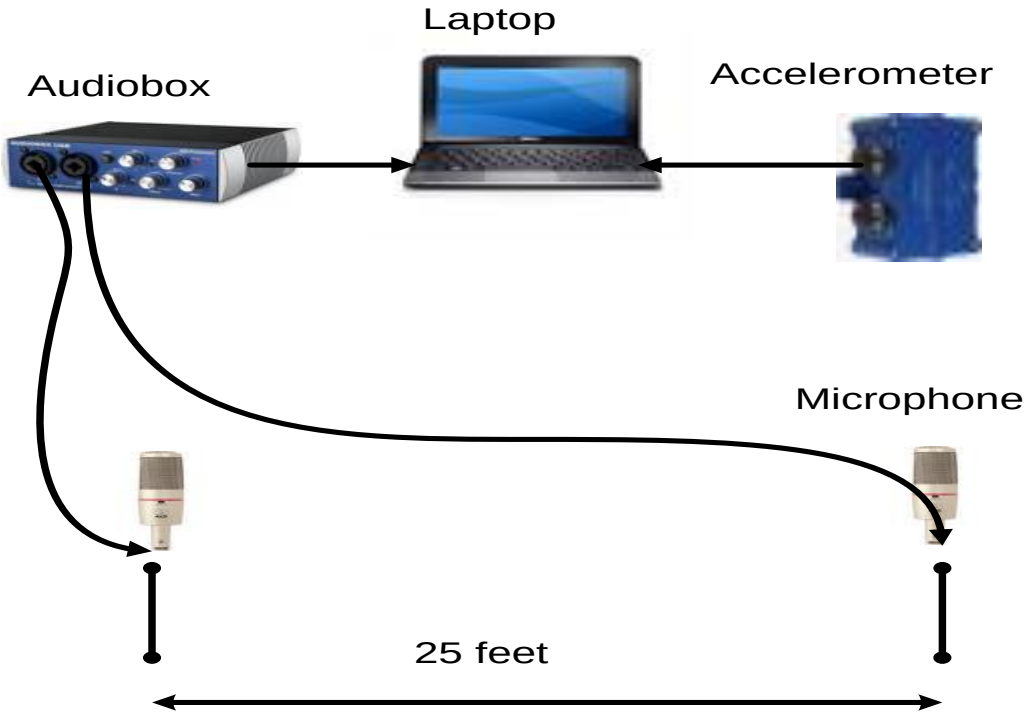

Figure 3.2. Instrument setup for data collection

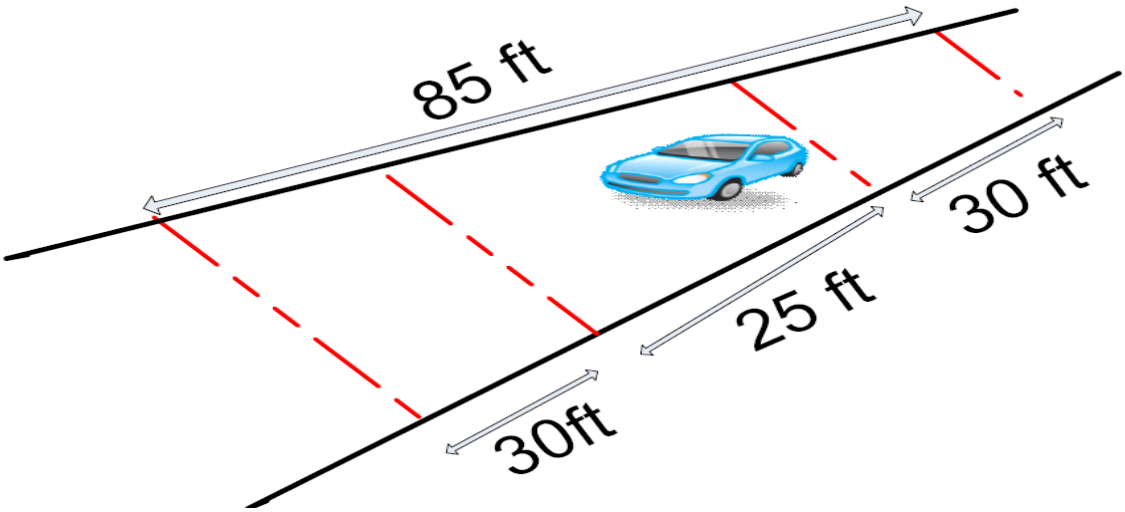

Figure 3.3. Chalk markings on the street

3.3 Description and some technical specifications of instruments used

Two microphones, a USB powered AudioBox and a triaxial accelerometer have been purchased and tested for the data collection. The first microphone is an AT2020 Cardioid [26] condenser microphone with a custom-engineered 16 mm low-mass diaphragm. It provides an extended frequency response ranging from 20 Hz to 20 kHz with superior transient response. The second microphone is a Shure SM81 [27] condenser microphone with the same frequency response as the first. It has low noise characteristics with a signal to noise ratio of 78 dB at 94 dB SPL. It also has low radio-frequency susceptibility.

The AudioBox is an audio recording interface with two microphone inputs. It has a 48 volts dc phantom power that is used in powering the microphones. It has microphone preamps with frequency supply ranging from 14 Hz to 70 kHz (-3 to 3 dB) and delivers

low noise with high gain to give pure and rich sound signals. It uses a professional-grade analog-to-digital converter with a dynamic range of 102 dB and 24-bit sample rate so that the digital conversion is big and quiet making it one of the best sounding USB bus-powered interfaces. The AudioBox has a sampling rate of 44100 Hz. It comes with Studio One software, which is used for recording purposes. It is worth noting that the microphones and the AudioBox device are able to operate below frequencies of 100 Hz and above 10 kHz. The significance of this is to ensure that signals useful for classification within these stated ranges are not omitted.

The CX1 [28] sensor device combines accelerometer, inclinometer and temperature sensors into one small sensing module. It measures acceleration in the x, y and z directions (triaxial). It has a range of -1.5 to 1.5 g. 1 g is a unit of acceleration equal to the earth's gravity at sea level, where 1g=9.8 m/s². The sampling rate for the accelerometer is 2000 Hz. The inclinometer sensors enable the CX1 to measure tilt positions. Its sensitivity to tilt which is also known as resolution is 10^{-5}g. The CX1 sensor device comes with *SensrView* software for capturing the acceleration, tilt and temperature values of objects. Accelerometers can be used to measure vibration on cars, machines and buildings. Seismic activity can also be measured using accelerometers. The CX1 accelerometer was used in measuring the vibration due to the movement of vehicles. Figure 3.4 shows the software interfaces for *Studio One* (left) and *SensrView* (right) softwares. The *SensrView* software has the capability of displaying the root mean square values of signals at different frequencies.

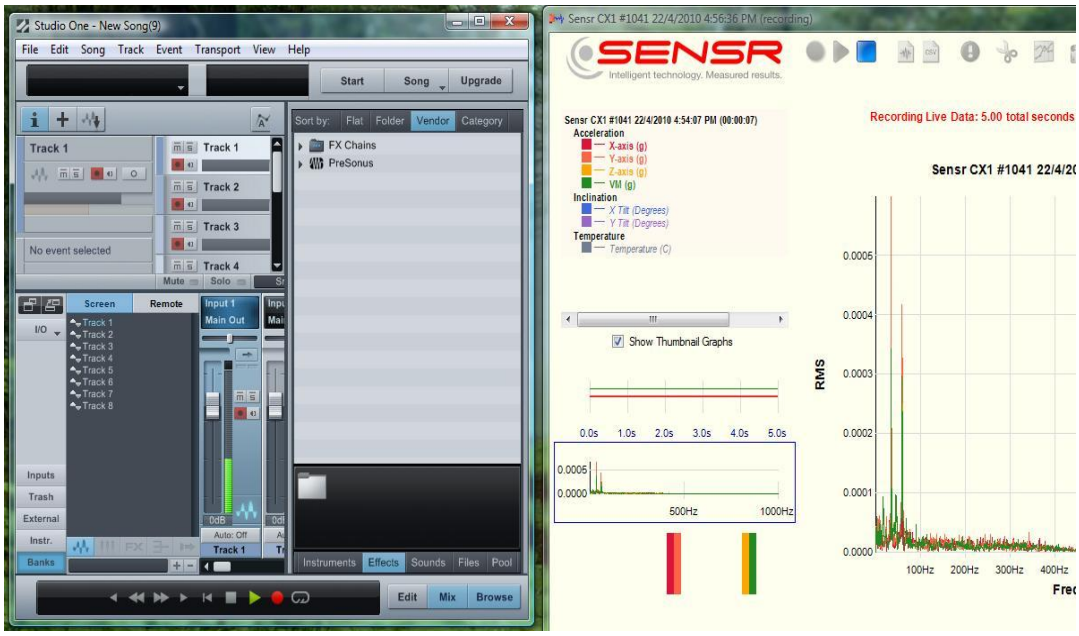

Figure 3.4. Studio One and SensrView software

3.4 Acoustic and Seismic Database

In the selection of an appropriate site for this data collection exercise we ensured that:

1. The site had a low level of noise and was significantly free from other forms of interference. Exits of highways were avoided because fast moving vehicles produced loud car noise that was a major source of interference.

2. The site had a low level of noise and was significantly free from other forms of interference. Exits of highways were avoided because fast moving vehicles produced loud car noise that was a major source of interference.

3. Vehicles whose signals were measured were sufficiently isolated thus avoiding the collection of sound generated from several vehicles at the same time

4. Vehicle speeds were in the range of 5-15 miles per hour in order to reduce Doppler effect between the two acoustic devices.

5. The wind effect was minimized, thus data collection was done on days that ambient wind speed was roughly 2 miles per hour.

Figure 3.5 shows the process for the creation of the database. Raw accelerometer data was recorded on the hard drive of the laptop shown in Figure 3.2 earlier. *SensrView* software was then used in converting the raw accelerometer data to a comma separated value (•csv) file format which is readable by Matlab. In the case of the acoustic signals, *Studio One* software was used in converting the raw acoustic signal to a •wav file. This format is also readable by Matlab.

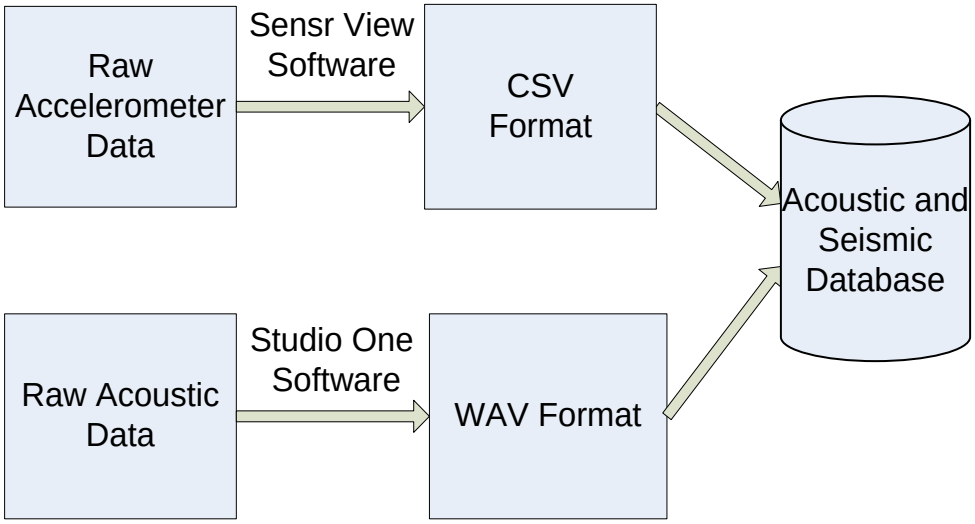

Figure 3.5. Process for database creation

About 860 vehicle sounds were collected for each microphone and the accelerometer. Toyota cars accounted for the largest percentage (14.30%) of vehicle sound data that was gathered. This percentage includes brands like Avalon, Corolla,

Camry and so forth. Ford cars were the second largest with a total of 12.81%. Figure 3.6 is a pie chart of the distribution of vehicles in the database.

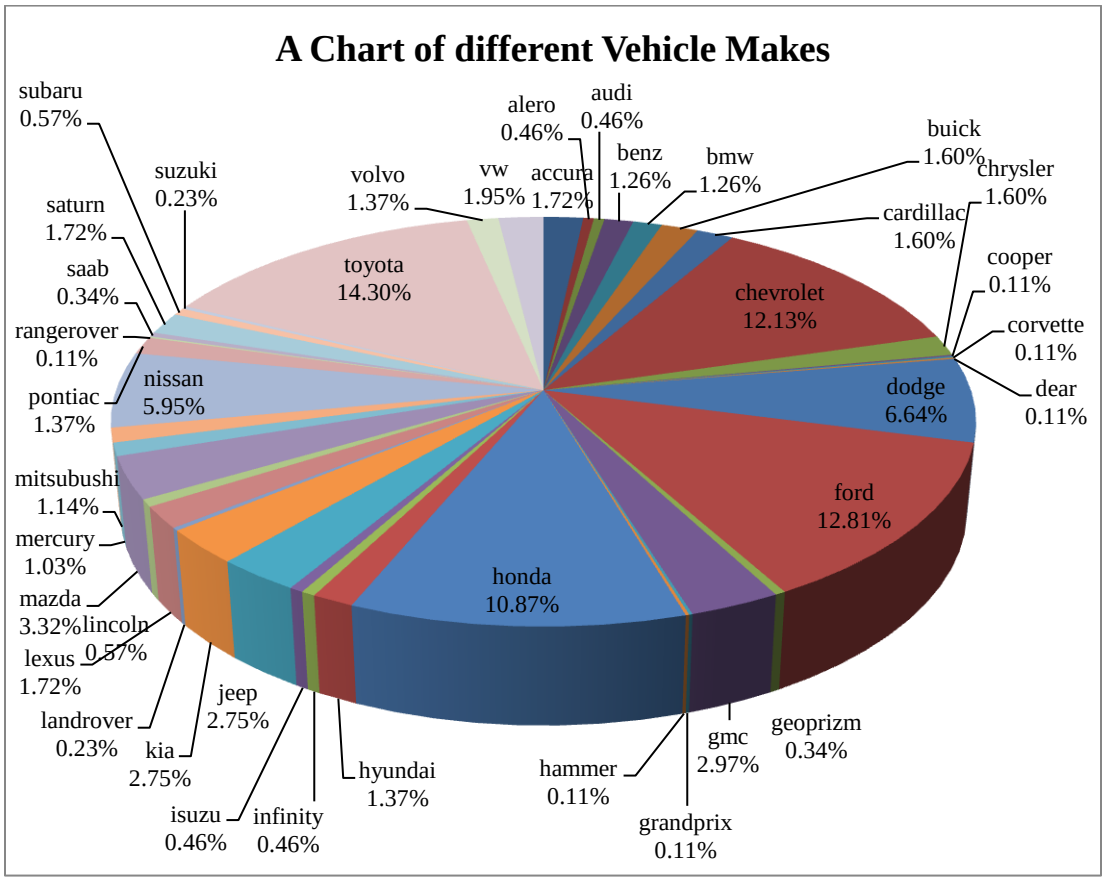

Figure 3.6. Pie chart of the distribution of different vehicle types

3.5 Acoustic and Seismic signal pre-processing for spectral features

The vehicles whose signals were collected were either moving toward or away from a fixed set of recording microphones. The vehicles could also be either accelerating or decelerating and therefore making their signals generally not stationary. Thus, the

acoustic signals of the vehicles are largely non-stationary [29] over a given time frame. It is therefore necessary to consider the effect of Doppler frequency shift.

Let Δv be the Doppler frequency shift, v the original frequency, ΔV the vehicle travelling speed, and c the speed of propagation of sound in air (343.4 m/s), then

$$\frac{\Delta v}{v} \simeq \frac{\Delta V}{c}.$$

Assuming the vehicles move at a speed of 15 mi/h (25 km/h), then the Doppler shift is given by

$$\frac{25000}{3600} \times \frac{1}{343.4} \times 100 = \pm 2.0223\%.$$

The resulting Doppler shift is ±2.0223%, which is less than 5%, and can be assumed negligible. Thus, the acoustic signals can be reasonably modeled under stationary conditions when the signals are segmented in short time durations with an overlap.

Figure 3.7 shows 3 seconds of a digitized acoustic signal sampled at 44100 Hz. The signal is segmented into short time durations with 50% overlap. The choice of the 50% overlap is arbitrary. It must however be reasonable enough to ensure that the signal is modeled under stationary conditions. The acoustic signal is normalized to zero-mean to remove any direct current (dc) bias. Figure 3.7 also shows that a 3 seconds signal segmented at 1 second interval with 50% overlap produces 5 short frames of the signal. Similarly, a seismic signal of 3 seconds duration, segmented at 1 second interval with 50% overlap and sampled at 2000 Hz produces 5 short frames. The seismic signal is also normalized to remove any dc bias. The spectral features were computed for each segment of signal and used as attribute in this work.

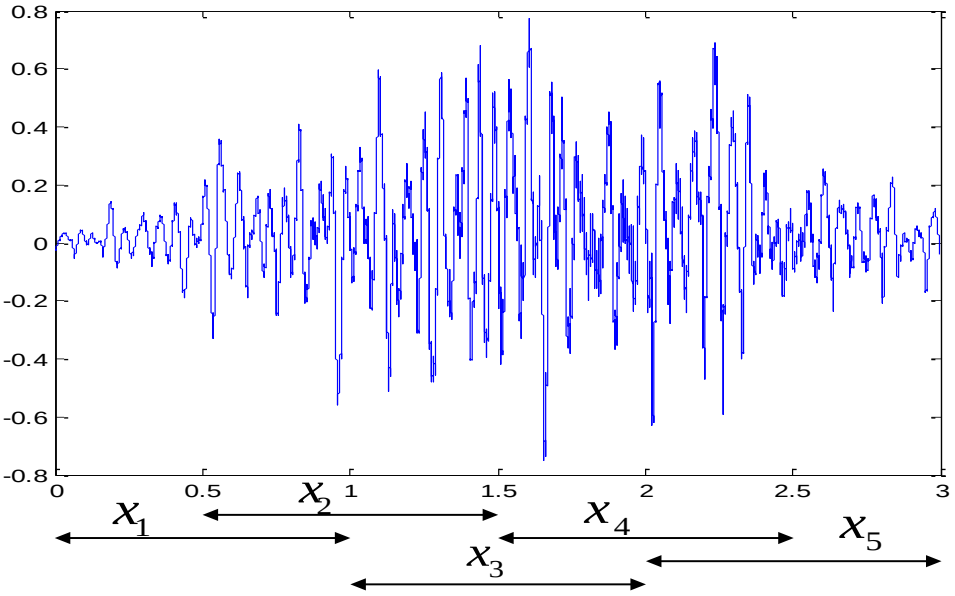

Figure 3.7. Signal segments with 50% overlap

.

3.6 Features used for classification

The process in which raw acoustic/seismic data is transformed into data that can be used as input to a classifier is called feature extraction. One commonly known feature extraction method is the wavelet transform which gained widespread application with the development of Daubechies [30]. Another frequency domain feature extraction method is the Fast Fourier Transform. In [1], new and efficient features for vehicle classification using various concepts were developed. Tristimulus features for example, were developed based on concepts from music theory. There was special emphasis on the isolation of tonal components related to the engine Revolution Per Minute (RPM) and the need for the robust estimation of the fundamental frequency that is free of integer ambiguity. There was also a discussion on the extraction of spectral features from

acoustic data. Investigations indicated that vehicle sounds were rich enough to distinguish two different cars using spectral and tristimulus features. In this work, the newly developed and efficient features are applied to a large database of vehicle sounds.

Table 3.1 shows the data structure of meaningful features for the problem of vehicle classification. These features can be grouped as spectral shape features, harmonic features and energy features. Some of the spectral features are spectrum centroid, spectrum spread, the spectrum skewness and spectrum kurtosis. Features dependent on fundamental frequencies are tristimulus features. The energy features indicate both the harmonic and non-harmonic energy content.

Table 3.1. Data structure of features in the database

Spectral Shape Features	**spec.cent**	**Spectral centroid**
	spec.spread	**Spectral spread**
	spec.skew	**Spectral skewness**
	spec.kurt	**Spectral kurtosis**
Harmonic Features	**harm.f0**	**Fundamental frequency**
	harm.tristimulus	**Tristimulus**
	harm.inharmonicity	**Inharmonicity**
Energy Features	**energy.tot**	**Total energy**
	energy.harm	**Harmonic energy**
	energy.inharm	**Non-harmonic energy**

CHAPTER 4

ACOUSTIC AND SEISMIC FEATURE COMPUTATION

4.1 Mathematical computation of spectral features

Figure 4.1 shows the general approach we adopted in computing different spectral shape features. Each 3 seconds seismic/acoustic signal was segmented and the Power Spectral Density (PSD) calculated. The PSD describes how the energy of a signal is distributed with frequency. There are many methods for the computation of the PSD of a signal. For instance we have the following fundamentally different approaches:

1. Parametric methods, i.e., Burg, Covariance and Yule-Walker [31];
2. Non-parametric method such as FFT and Welch [32] and [33].
3. Subspace methods, i.e., eigenvector and Multiple Signal Classification (MUSIC) [34].

Spectral features computed using each of the above power density computation methods may be different and the differences can be exploited in the classification. This richness will be pursued in future work. The Welch method was used in computing the power spectral density of the vehicle sounds. The Welch method for evaluating the power spectrum divides the data equally in several possibly overlapping segments and performs a Fast Fourier Transform (FFT) on each segment. Any remaining entry in the data that cannot be included in the segment of equal length is discarded. It then computes the average spectra of the segments. In this work, we used the Hamming window with a 50% overlap.

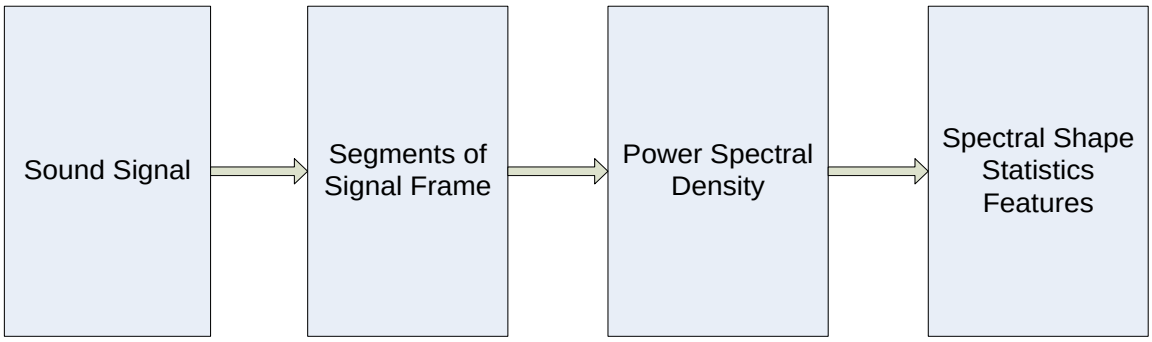

Figure 4.1. General approach to the computation of spectral features

The mathematical model used in the computation of the spectral features for both the seismic and acoustic signals is presented. The dependence of the spectral features on Doppler shift [1] is also discussed. For a given signal $x(t)$, the energy, E, is given by

$$E = \int_{\Omega} p_x(\omega) d\omega$$

where Ω is the frequency domain support on which $p_x(\omega) \le \varepsilon \ \max|p_x(\omega)|$. Limiting Ω is crucial to alleviating the effect of noise and to make the features more reliable. The spectral centroid is the center of mass of the PSD $p_x(\omega)$ of the signal $x(t)$:

$$\mu_x = \frac{\int_{\Omega} \omega p_x(\omega) d\omega}{E},$$

where ω is the frequency. A car moving with a radial velocity v with respect to the microphone, and emitting a tone of frequency ω, produces a tone whose frequency ω' undergoes a Doppler shift equal to $v\omega/c$ where c is the speed of sound. Hence the measured PSD is approximately $\hat{p}_x(\omega) = p_x(\omega') = p_x(\omega + \omega v/c)$.

We define $\beta = v/c$ and $\alpha = 1+\beta$. For a car moving at 80Km/hr, $\beta \cong \frac{80000/3600}{320} = 6.94\%$ and $\alpha = 1.0694$. The statistics of the measured spectrum are related to those of the emitted spectrum through the scaling factor α. For instance,

$$\begin{aligned}\int_0^\infty \hat{p}_x(\omega)d\omega &= \int_0^\infty p_x(\alpha\omega)d\omega = \int_0^\infty p_x(\alpha\omega)d(\alpha\omega)/\alpha \\ &= \frac{1}{\alpha}\int_0^\infty p_x(\omega')d\omega'.\end{aligned}$$

The first moment of the measured spectrum is

$$\begin{aligned}\int_0^\infty \omega\hat{p}_x(\omega)d\omega &= \int_0^\infty \omega p_x(\alpha\omega)d\omega \\ &= \frac{1}{\alpha^2}\int_0^\infty (\alpha\omega)p_x(\alpha\omega)d(\alpha\omega) \\ &= \frac{1}{\alpha^2}\int_0^\infty \omega' p_x(\omega')d\omega',\end{aligned}$$

Thus, the measured centroid is related to the emitted one like

$$\hat{\mu}_x = \frac{\int\omega\hat{p}_x(\omega)d\omega}{\int\hat{p}_x(\omega)d\omega} = \frac{\frac{1}{\alpha^2}\int\omega' p_x(\omega')d\omega'}{\frac{1}{\alpha}\int p_x(\omega')d\omega'} = \frac{1}{\alpha}\mu_x.$$

The spectral spread is the spread of the spectrum around its mean value. The variance of the spectrum is

$$\sigma^2 = \frac{\int_\Omega(\omega-\mu)^2 p_x(\omega)d\omega}{E}.$$

The effect of the Doppler shift is given by

$$\hat{\sigma}^2 = \frac{1}{E}\int(\omega-\hat{\mu}_x)^2.\hat{p}_x(\omega)d\omega$$

$$\hat{\sigma}^2 = \frac{1}{E}\int\left(\omega - \frac{\mu_x}{\alpha}\right)^2 .p_x(\alpha\omega)d\omega$$
$$= \frac{1}{E\alpha^2}\int(\omega' - \mu_x)^2 .p_x(\omega')d\omega' / \alpha = \frac{1}{\alpha^3}\sigma^2.$$

The standard deviation normalized by the centroid is then given by the nondimensional quantity

$$\hat{\beta} = \alpha^{-1/2}\frac{\sigma}{\mu}.$$

The spectral skewness is the asymmetry of a distribution around its mean. It can be computed by starting with the moment or cumulant,

$$m_3 = \frac{\int_\Omega(\omega - \mu)^3 p_x(\omega)d\omega}{E}.$$

The Doppler shift effect is

$$\hat{m}_3 = \frac{1}{E}\int(\omega - \hat{\mu})^3 .\hat{p}_x(\omega)d\omega = \frac{1}{E}\int\left(\omega - \frac{\mu_x}{\alpha}\right)^3 .p_x(\alpha\omega)d\omega$$
$$= \frac{1}{E\alpha^3}\int(\omega' - \mu_x)^3 .p_x(\omega')d\omega' / \alpha = \frac{1}{\alpha^4}m_3.$$

The nondimensional skewness is the ratio

$$\hat{\gamma}_1 = \frac{\hat{m}_3}{\hat{\sigma}^3} = \frac{\alpha^{-4}m_3}{(\alpha^{-3/2}\sigma)^3} = \sqrt{\alpha}\frac{m_3}{\sigma^3} = \sqrt{\alpha}\gamma_1.$$

Kurtosis is the ratio of the fourth-order central moment to the square of a second order central moment [11]. The spectral kurtosis of a signal is the kurtosis of its frequency components [35]. It measures the flatness of a distribution around its mean value and can be computed from the 4th order moment about the mean defined as

$$m_4 = \frac{\int_\Omega (\omega - \mu)^4 \, p_x(\omega) d\omega}{E},$$

and the Doppler shift effect is

$$\hat{m}_4 = \frac{1}{E}\int (\omega - \hat{\mu})^4 . \hat{p}_x(\omega) d\omega = \frac{1}{E}\int \left(\omega - \frac{\mu_x}{\alpha} \right)^4 . p_x(\alpha\omega) d\omega$$
$$= \frac{1}{E\alpha^4}\int (\omega' - \mu_x)^4 . p_x(\omega') d\omega' / \alpha = \frac{1}{\alpha^5} m_4 .$$

The perceived nondimensional kurtosis is given by:

$$\hat{\gamma}_2 = \frac{\hat{m}_4}{\hat{\sigma}^4} = \frac{\alpha^{-5} m_4}{\left(\alpha^{-3/2}\sigma\right)^4} = \alpha \frac{m_4}{\sigma^4} = \alpha\gamma_2 .$$

The spectral features were computed using the equations above. The Doppler shift effect was small in the experiments we conducted since the vehicles were moving at slow speeds. The Doppler shift will be important for fast moving vehicles (on highways). The sections that follow show the graphical representations of these features over the segments.

4.2 Seismic spectral feature representation

Figure 4.2 and Figure 4.3 show the spectral features of same branded cars for the accelerometer data. Figure 4.2 shows the plot of the spectral kurtosis versus the spectral skewness for all windowed segments of 2 Nissan Altimas. The mean of the skewness for the first car is 0.3960 and the mean of the kurtosis is 2.4389. For the second Nissan Altima, the mean of the skewness is 0.4931 and the mean of the kurtosis is 2.3727. The average of the skewness-kurtosis for the two cars is very close.

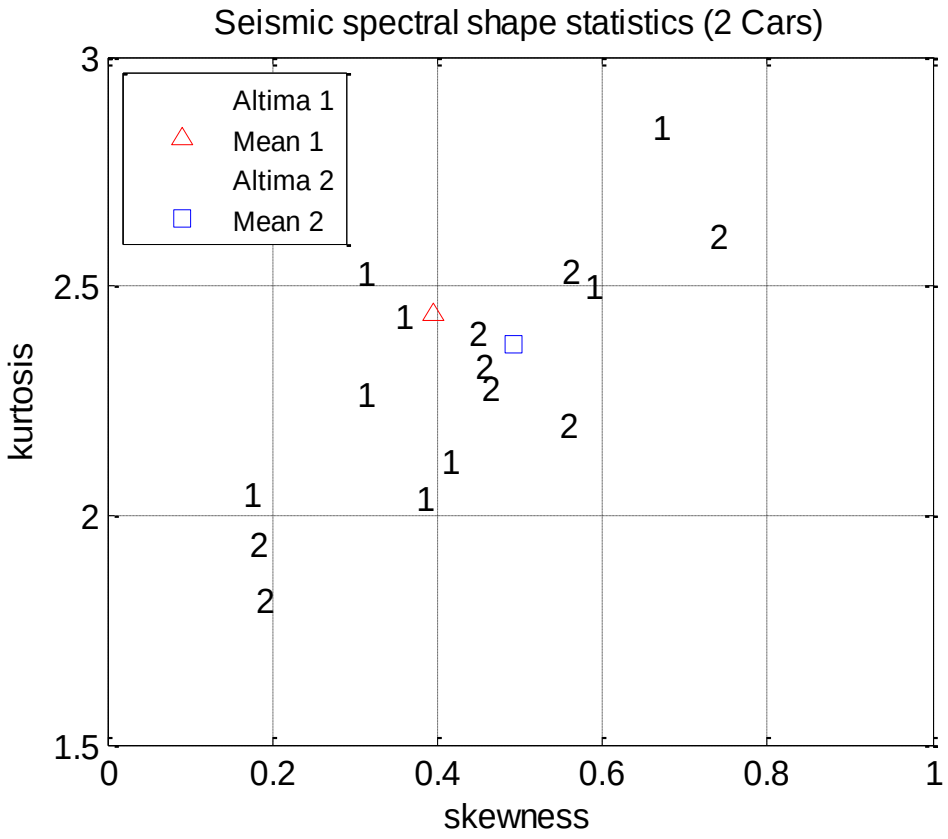

Figure 4.2. Seismic spectral shape statistics for two Altimas

Figure 4.3 shows the spectral features of all windowed segments of seismic signals for two Toyota Camrys. The spectral representation of different segments of Camry 3 is indicated by the number '3' and the number '4' show the distribution of the different segments of Camry 4. The mean of the skewness for Camry 3 is -1.2577 and the mean of the kurtosis is 4.6457. The mean of the skewness of Camry 4 is -0.9710 and the mean of its kurtosis is 3.4720. From Figure 4.2 and Figure 4.3, the distribution of the skewness-kurtosis for all segments is fairly clustered around the means.

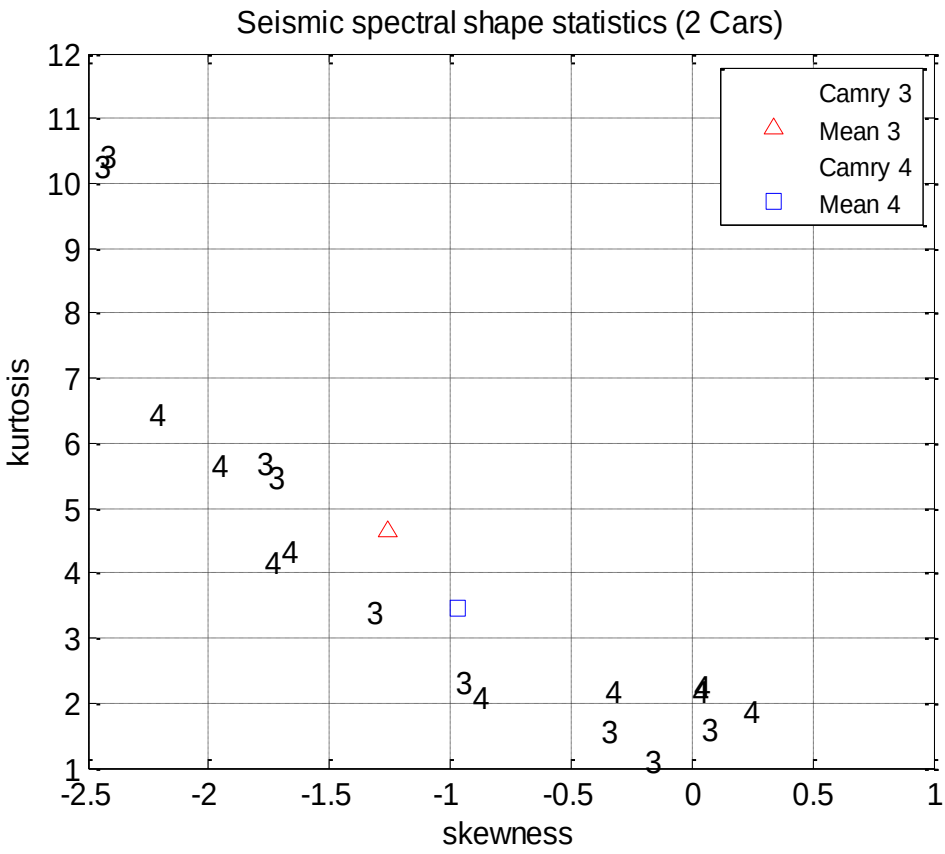

Figure 4.3. Seismic spectral features for two Camrys

4.3 Acoustic spectral feature representation

In Figure 4.4, Figure 4.5 and Figure 4.6, the spectral features for the acoustic signals are compared. Figure 4.4 shows the plot of the spectral features for 2 Nissan Altimas. The mean skewness of Altima 1 is 2.6474 and the mean of the kurtosis is 11.5118. The mean skewness of Altima 2 is 2.6653 and the mean of the kurtosis is 13.0882. Figure 4.5 shows a mean skewness of 1.1514 and a mean kurtosis of 4.6048 for Camry 3. Camry 4 has a mean skewness of 0.9729 and a mean kurtosis value of 4.4901.

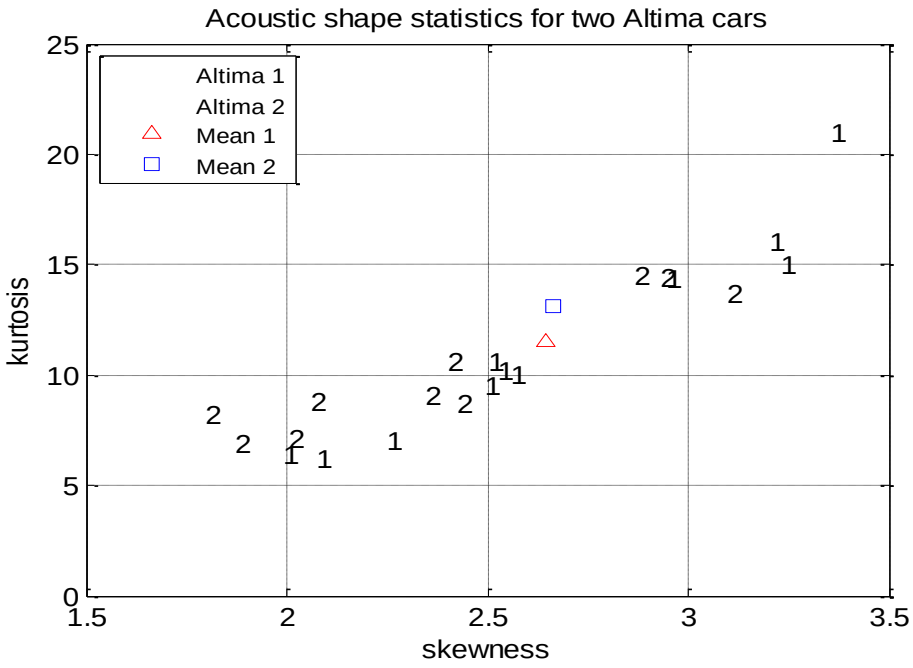

Figure 4.4. Acoustic shape statistics for two Altima cars

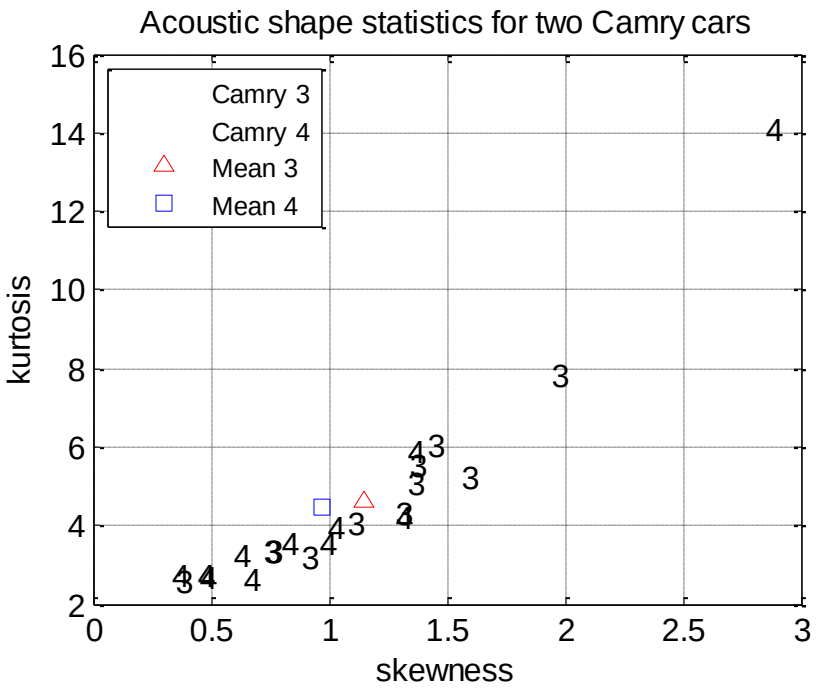

Figure 4.5. Acoustic shape statistics for two Camry cars

Figure 4.6 shows a cluster of the segments of the spectral values for 2 Accords. The segments of vehicle sounds from the first and second Accords are represented by '5' and '6' respectively. For Accord 5, the average skewness is 0.2356 and the average kurtosis is 2.6429. It also shows an average skewness of 0.3754 and average kurtosis of 3.4540 for Accord 6. Clearly, the mean of the spectral skewness and spectral kurtosis are close for same branded cars.

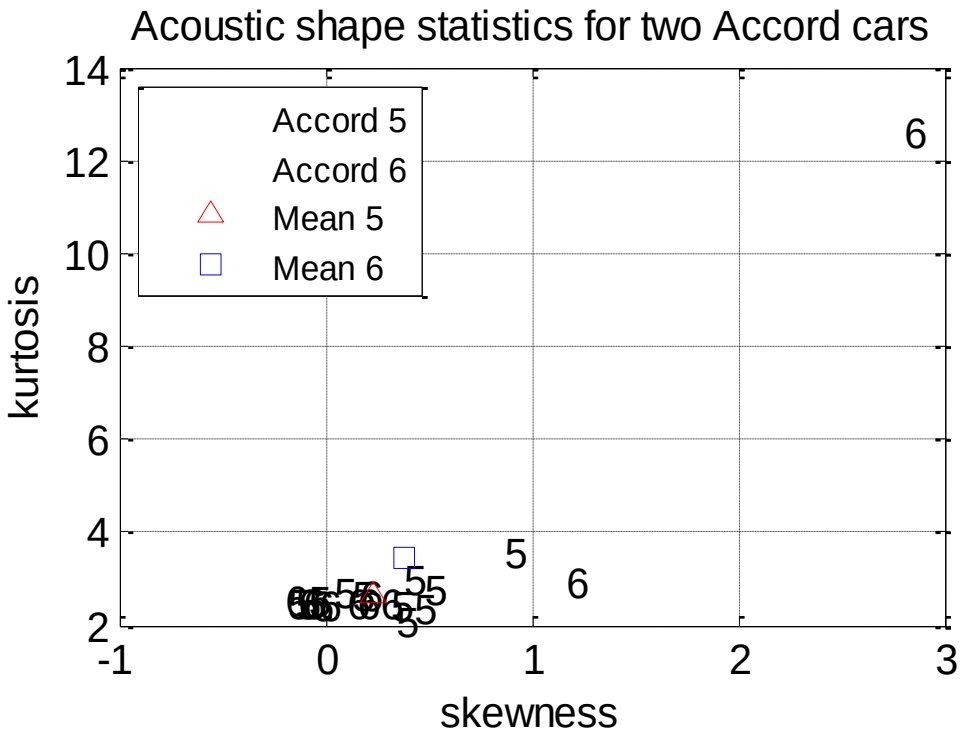

Figure 4.6. Acoustic shape statistics for two Accord cars

Figure 4.7 compares the spectral skewness-kurtosis of different branded cars for the microphone signals. Segments of Accord vehicle sound are represented as '1', Altima is represented by '2' and Camry is '3'. The spectral means have been computed and indicated in Figure 4.7 . The statistical means of the spectral skewness-kurtosis are quite different for the 3 different cars. The relationships that exist in the spectral features using the mean-skewness and mean-kurtosis for several vehicle sounds are investigated in the next chapter.

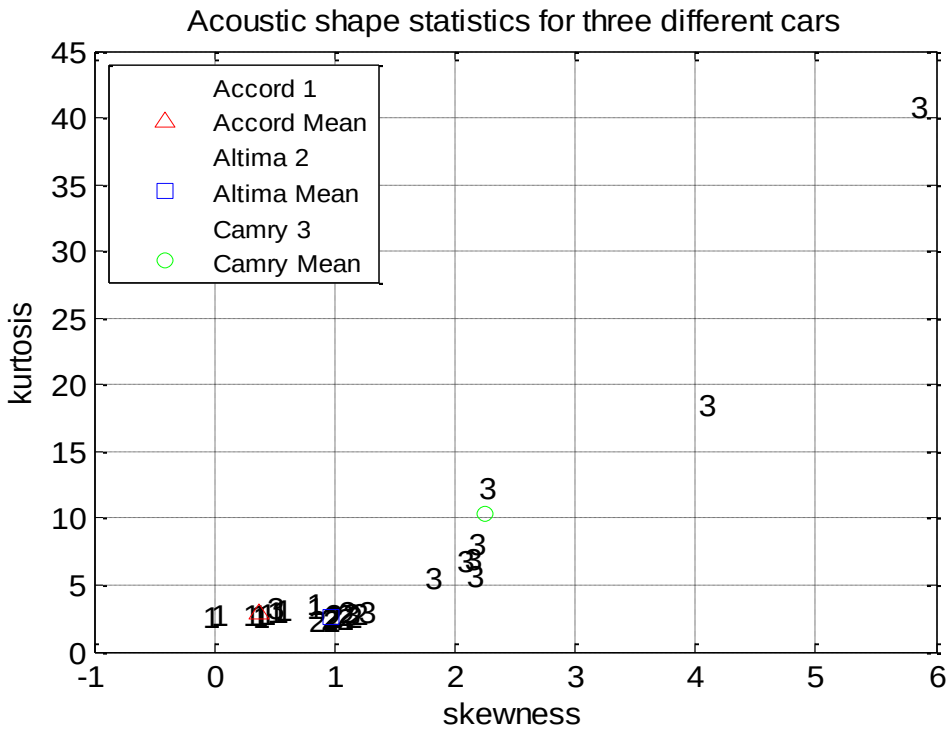

Figure 4.7. Acoustic shape statistics for three different cars

CHAPTER 5

ACOUSTIC SPECTRAL SUPER FEATURE EXTRACTION

After computing the features, further processing is usually required to extract more abstract features with a stronger classification power. Indeed, there are many algorithms for obtaining these super features. Entropy minimization and divergence maximization are just but a few of feature selection algorithms. In this work, two simple processes for extracting more powerful features are adopted. First, the use of descriptive statistics such as the average of the spectral features is considered. Secondly, a dimension reduction technique such as PCA is applied to the spectral features to find the relationships within the dataset.

5.1 Statistical analysis of spectral features from the database

The statistics of vehicle sounds are somewhat nonstationary and hence it is often beneficial to average the spectral features over the signal segments. Figure 5.1 shows the general approach adopted in exploring the mean spectral features as input to different classifiers. The mean skewness vs. the mean kurtosis for different brands of vehicles is represented graphically. Figure 5.2 shows the mean-skewness vs. mean-kurtosis for 24 Kias and 29 Mazdas. Figure 5.3 shows the mean-skewness vs. mean-kurtosis for 14 Corollas and 14 Maximas. Figure 5.4 shows the mean-skewness vs. mean-kurtosis for 21 Civics and 14 Maximas. Figure 5.5 shows the mean-skewness vs. mean-kurtosis for 39 Chevrolets, 33 Camrys and 34 Accords.

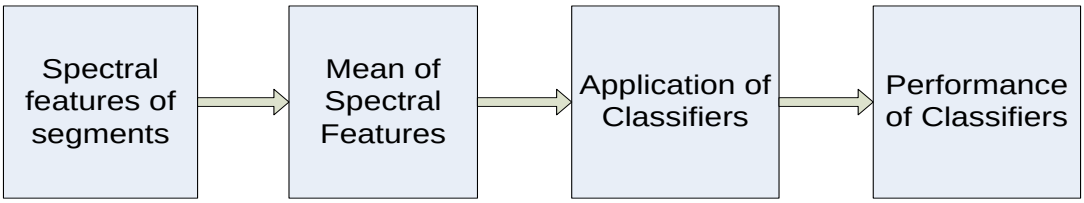

Figure 5.1. Use of spectral means as input to a classifier

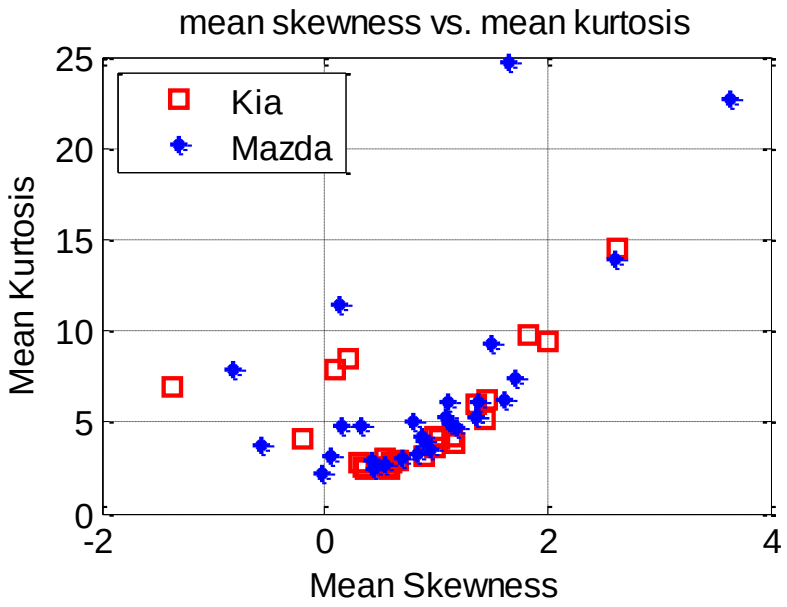

Figure 5.2. Mean skewness vs. mean kurtosis for Kia and Mazda

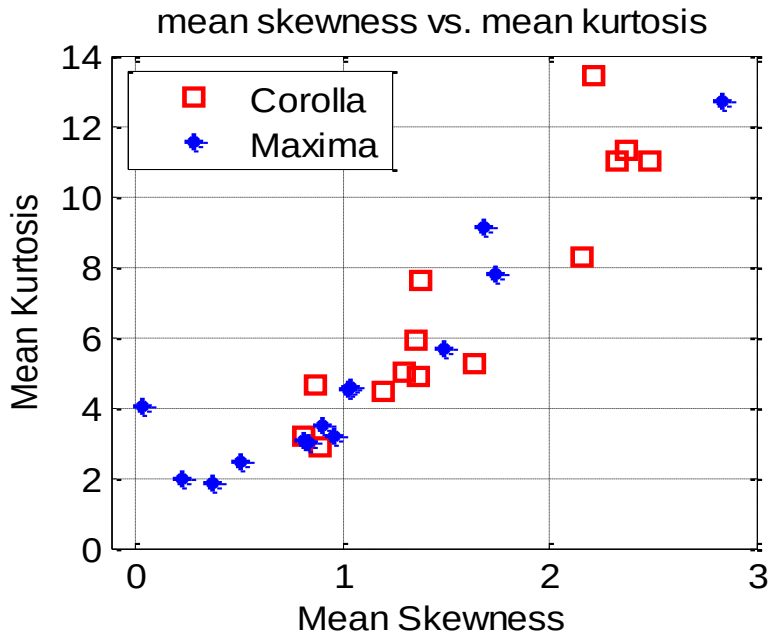

Figure 5.3. Mean skewness vs. mean kurtosis for Corolla and Maxima

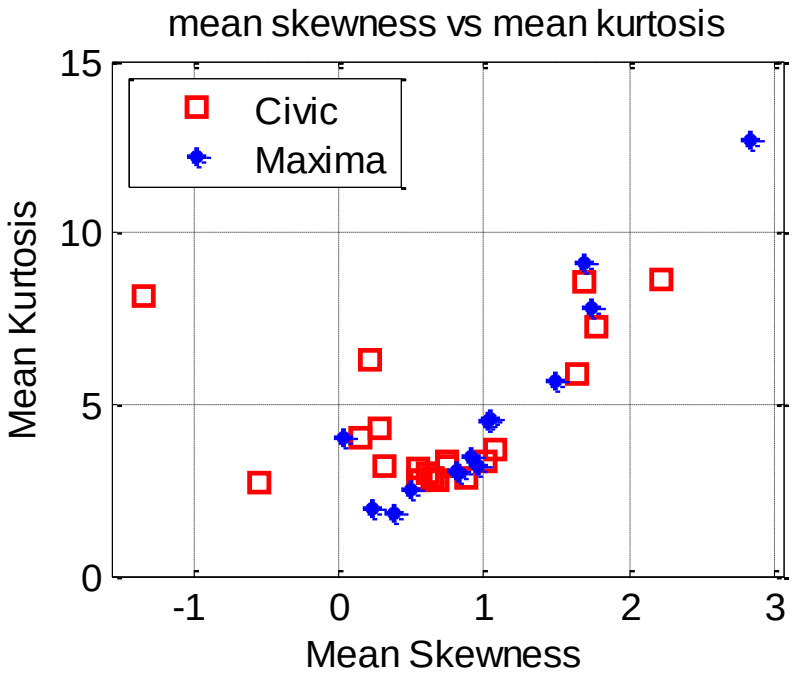

Figure 5.4. Mean skewness vs. mean kurtosis for Civic and Maxima

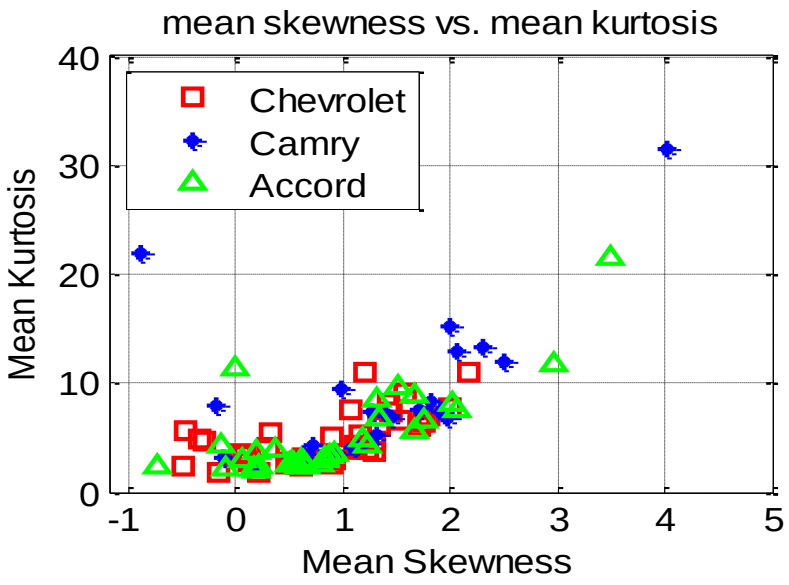

Figure 5.5. Mean skewness vs. mean kurtosis for Chevrolet, Camry and Accord

We also plot the mean-skewness vs. mean-kurtosis for vehicle sounds from the same manufacturer. Figure 5.6 shows the mean spectral representation for 21 Civics and 34 Accords. The Civics are denoted by 'red squares' and the Accords denoted by 'blue asterisks'. Figure 5.7 shows the mean spectral representation for 14 Corollas and 33 Camrys. The Corollas are denoted by 'red squares' and the Camrys are denoted by 'blue asterisks'. From Figure 5.6 and Figure 5.7, there is an overlap of the mean spectral features for vehicle sounds from the same manufacturer. Figure 5.8 shows the plot for a 5-class scenario for vehicle sounds from different manufacturers. Generally, these spectral features overlap. Therefore, additional processing is needed to obtain better "super" features. We use PCA to obtain these "super" features.

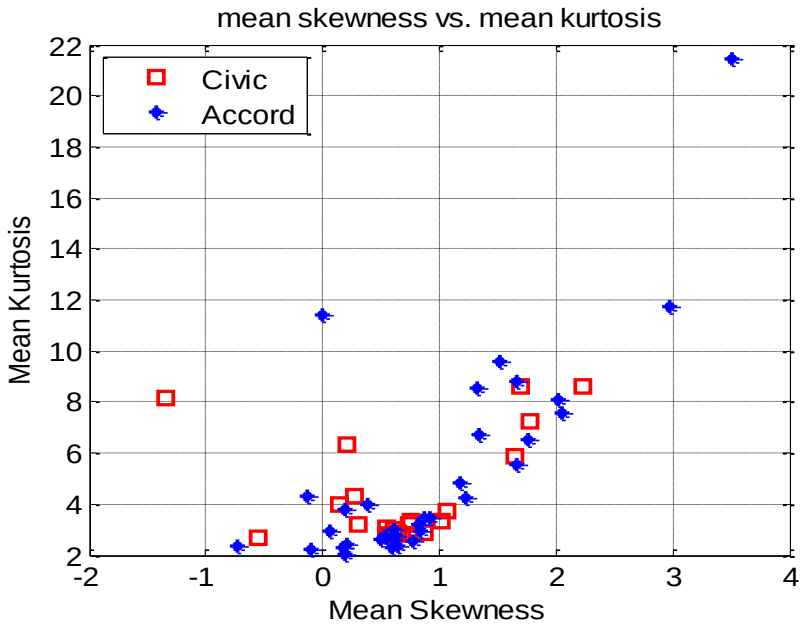

Figure 5.6. Mean spectral features for Civic and Accord

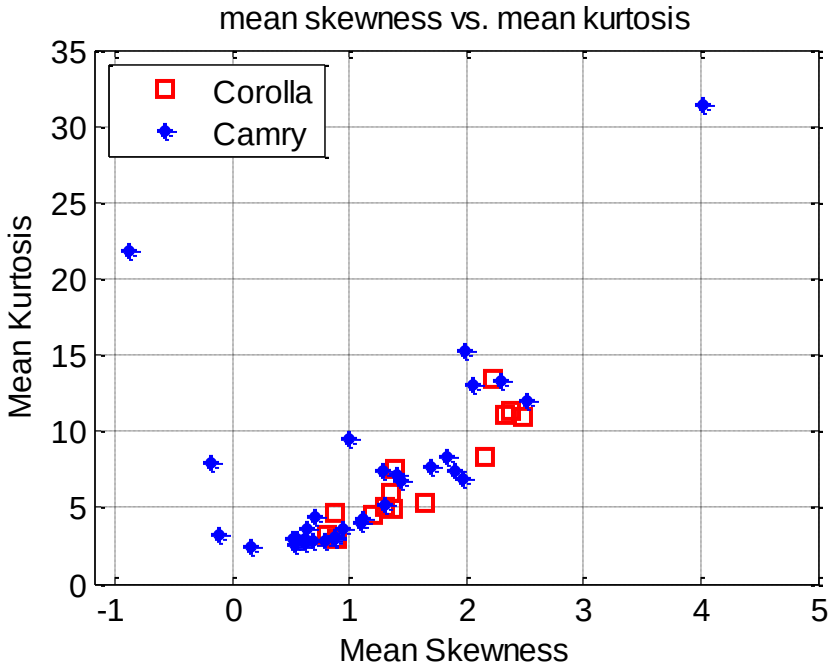

Figure 5.7. Mean spectral features for Corolla and Camry

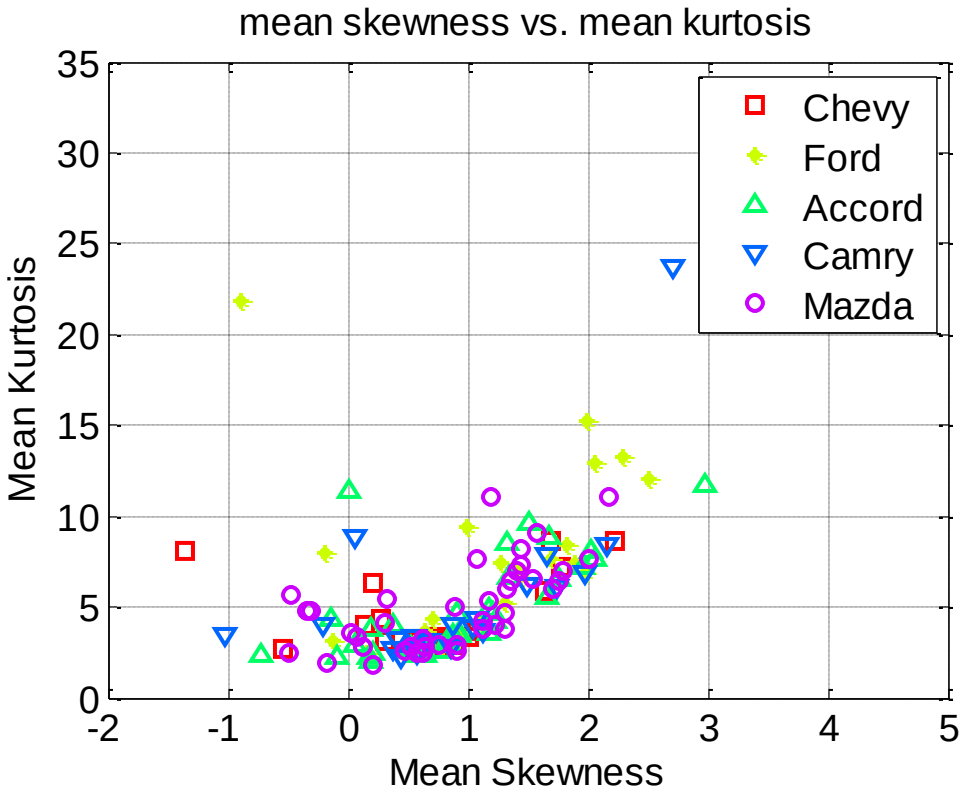

Figure 5.8. Mean spectral features for a 5-class scenario

5.2 PCA analysis of spectral features from the database

Mathematically, PCA is considered an eigenproblem [19] applied to the covariance matrix $S_x = f_x^T f_x$, computed from the collection of feature vectors for the different instances. The transformed "super" features are computed as $y = Bx$, where B is the transformed matrix derived from the eigenvalue problem. Only the few dominant super features are retained. The first 4 dominant super features are retained in this work. The dominant super features are called principal components.

Figure 5.9 shows the procedure we used in applying PCA to the spectral features from the database. The spectral skewness and spectral kurtosis were computed over several segments for each signal. Hence for a single vehicle, the feature vector has 20 elements, 10 skewness values for 10 segments and 10 values for kurtosis. Fifty elements are used when the tristimulus is included. PCA was then applied to the combined segments of the skewness-kurtosis and the dimensionality reduced such that the principal components retained over 90% of the information contained in the spectral features.

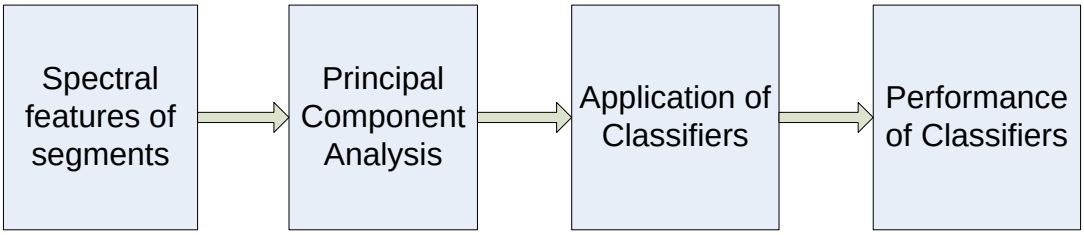

Figure 5.9. Use of PCA for feature dimension reduction

Table 5.1 shows the principal components of different vehicle sounds in terms of the percentage variance (Var. %) and the percentage cumulative variance (Cum. Var. %). The percentage cumulative variance for the first four principal components for Civic is 94.15% and the '**OTHERS**' account for 5.86%. Similarly, Maxima records 98.91% for the first four principal components of the percentage cumulative variance and 1.12% for the '**OTHERS**'. From Table 5.1, the cumulative variance of the first four dominant principal components account for more than 90% of the total variance in all the cases

considered. This means that the first 4 principal components explain more than 90% of the spectrum and hence are retained. Hence for a single vehicle sound, the feature vector has 20 elements. The elements are made up of ten skewness values for 10 segments of the vehicle sound and 10 values for kurtosis.

Table 5.1. Principal Components of different vehicle sounds

	1ST PC	2ND PC	3RD PC	4TH PC	OTHERS
Civic					
Var. (%)	46.04	26.78	12.41	8.92	5.86
Cum. Var (%)	46.04	72.82	85.23	94.15	100
Maxima					
Var. (%)	43.33	34.81	17.84	2.93	1.12
Cum. Var (%)	43.33	78.14	95.98	98.91	100
Accord					
Var. (%)	61.26	21.51	8.33	4.49	4.43
Cum. Var (%)	61.26	82.77	91.1	95.59	100
Camry					
Var. (%)	38.64	35.59	14.86	6.07	4.85
Cum. Var (%)	38.64	74.23	89.09	95.16	100
Kia					
Var. (%)	55.34	23.18	9.91	5.49	6.08
Cum. Var (%)	55.34	78.52	88.43	93.92	100
Mazda					
Var. (%)	57.27	24.48	12.04	4.6	1.61
Cum. Var (%)	57.27	81.75	93.79	98.39	100
Corolla					
Var. (%)	51.4	25.37	10.11	7.74	5.38
Cum. Var (%)	51.4	76.77	86.88	94.62	100

The scatter plots of the first two principal components for different brands of vehicles are shown in Figure 5.10, Figure 5.11, Figure 5.12, Figure 5.13 and Figure 5.14 where different car models exhibit significantly different patterns. For instance, Figure 5.10 shows the 1st and 2nd principal components of the spectral features for Civic and Maxima. Similarly, Figure 5.11 shows the 1st and 2nd principal components of the spectral features for Kia and Mazda. Comparing these figures to their counterpart mean spectral features, it is clear that the mean spectral features have significantly more overlap and hence little discerning powers. These identifiable patterns are explored in classification using the quadratic classifier and also the artificial neural network classifier in chapter 7.

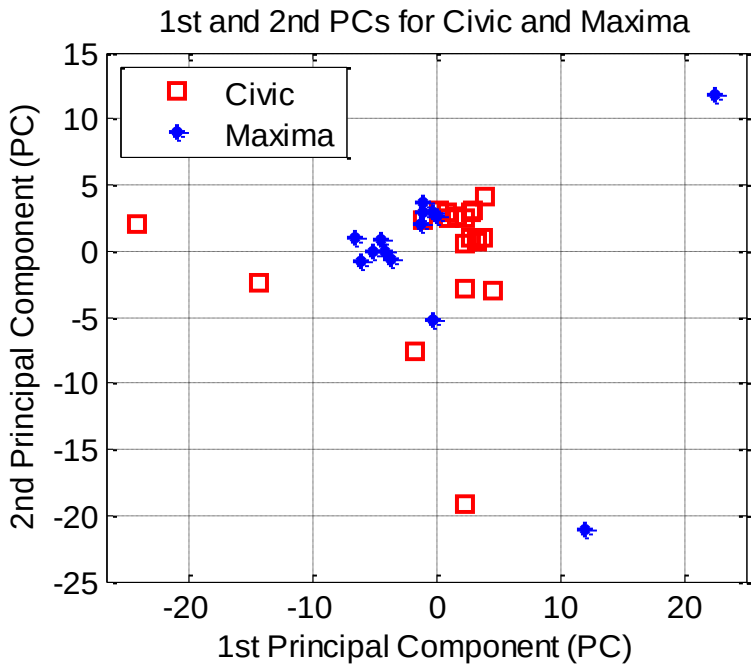

Figure 5.10. 1st and 2nd PCs of spectral features for Civic and Maxima

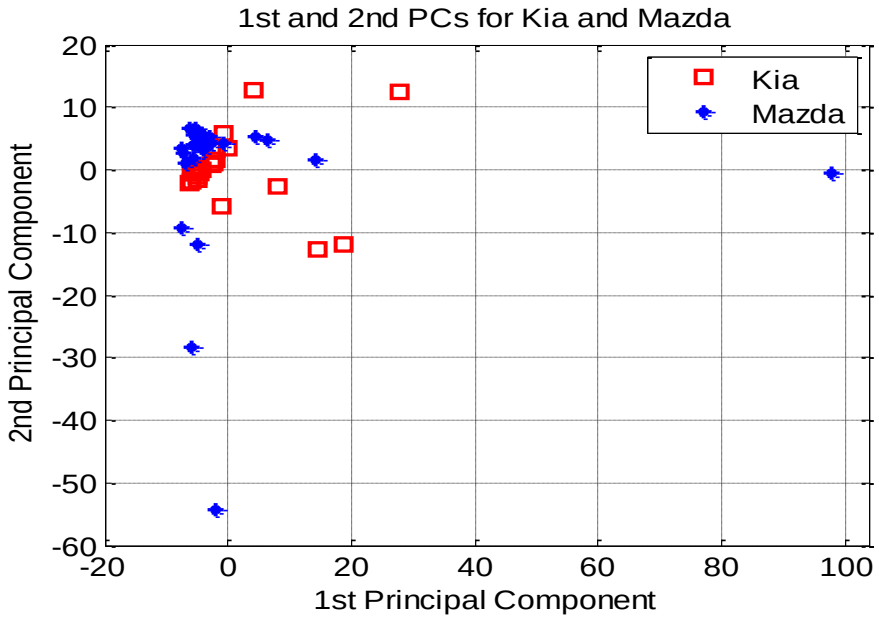

Figure 5.11. 1st and 2nd PCs of spectral features for Kia and Mazda

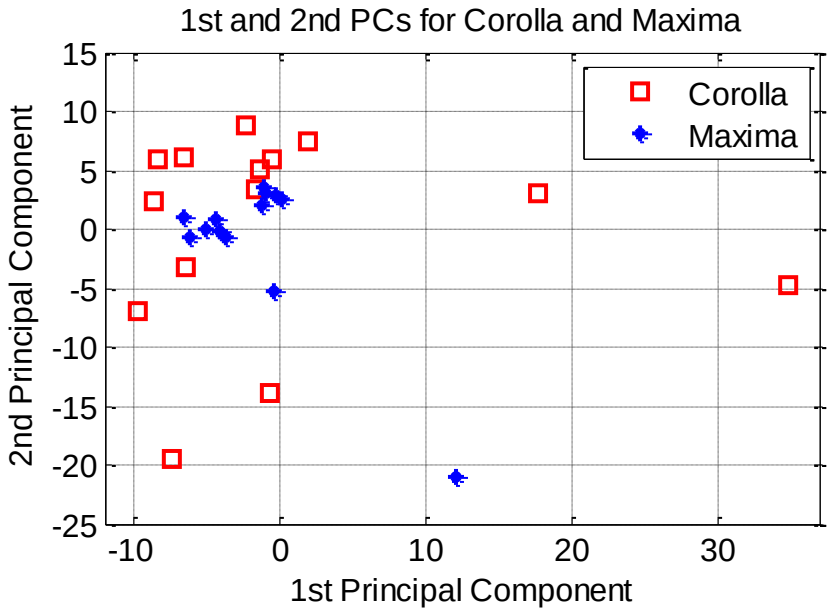

Figure 5.12. 1st and 2nd PCs of spectral features for Corolla and Maxima

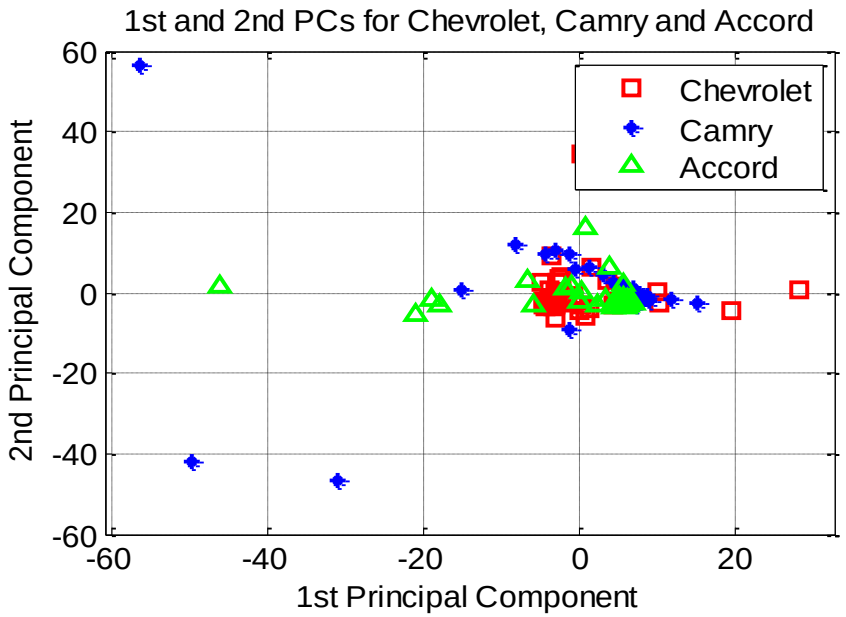

Figure 5.13. 1st and 2nd PCs of spectral features for Chevrolet, Camry and Accord

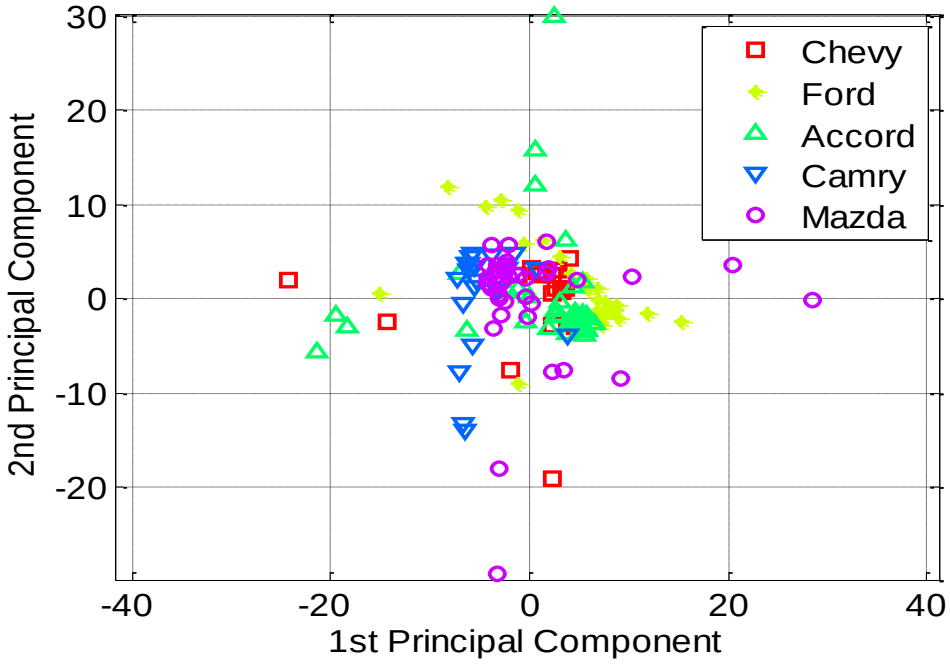

Figure 5.14. 1st and 2nd PCs of spectral features for a 5-class case

CHAPTER 6

TRISTIMULUS FEATURES

In [1], several physical phenomena relevant to the problem of acoustic vehicle classification were identified. The sound due to explosion pulses for instance produces periodic spectral components. The belt and gear mechanisms may also produce several quasi periodic frequencies in the spectrum of the engine sound. These and many other reasons clearly emphasize the need for a robust estimation for the fundamental frequency of vehicle engine sound. In this chapter, the tristimulus features for two different brands of vehicles sounds (e.g., Civic and Altima) are computed based on their fundamental frequency.

6.1 Fundamental frequency computation

In estimating the fundamental frequency of vehicle sounds, three different algorithms were used. The fundamental frequency estimation algorithms used are the Yin, the auto-correlation and angles of the poles of the Burg filter. The vehicle sound was segmented into 0.2 second frames with 0.05 second overlap. For the first segment, the first peak was found from the PSD Burg. Assume this peak is at $f_{0.}$ Then for each segment:

1. Window a signal segment using a Hamming window.
2. Using the Auto-correlation, Yin and poles of the Burg filter methods, search for an estimate of f_0 within ±20% of the previous estimate. All these methods can have parameters that depend on the most recent estimate of $f_{0.}$

3. Then the intermediate average of the f_0 estimates will be

$$f_{0i} = 0.3f_{0Auto} + 0.4f_{0Burg} + 0.3f_{0Yin}.$$

4. Smooth the time-variations of f_0 according to

$$f_{0new} = 0.2f_{0i} + 0.8f_{0previous}.$$

5. Based on the f_0 update the parameters for auto-correlation, Yin and Burg methods for the next segment.

Figure 6.1 shows the power spectral density for Altima. An initial peak of 60.39 Hz was observed. This initial peak is assumed to be the fundamental frequency, f_0. Using the auto-correlation, Yin and Burg methods, the f_0 for each segment was computed to arrive at a smoothed version of f_0. Figure 6.2 shows the f_0 for different segments using the three algorithms. It also shows the smoothed f_0 for the same Altima vehicle sound used in Figure 6.1. From Figure 6.1 and Figure 6.2, the initial fundamental frequency observed from the PSD method is quite consistent with the smoothed f_0 from the auto-correlation, Yin and Burg methods.

Figure 6.3 shows the power spectral density for Corolla. The initial f_0 is observed to be 103 Hz. Using the same approach as described above, the f_0 for different segments was found using the auto-correlation, Yin and Burg methods. Figure 6.4 shows the f_0s using the three different methods and the smoothed f_0 for the same Corolla vehicle sound used in Figure 6.3. The results shown in Figure 6.3 and Figure 6.4 are consistent.

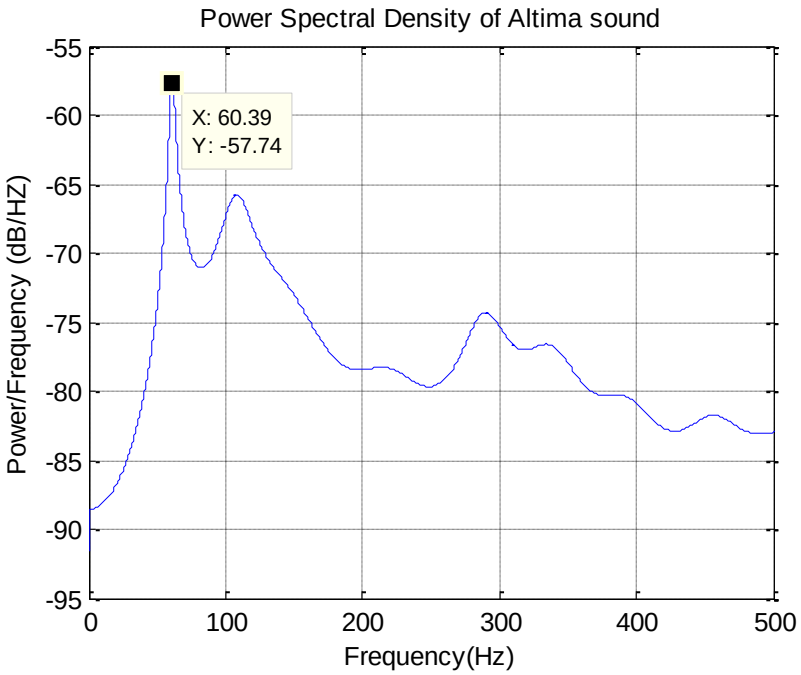

Figure 6.1. Power spectral density for Altima sound

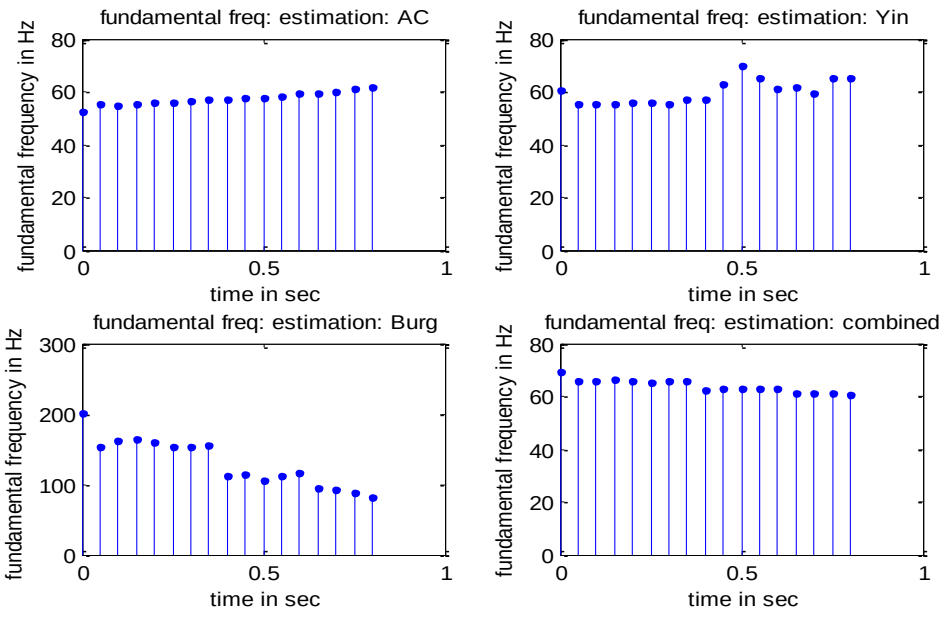

Figure 6.2. Fundamental frequencies using 3 methods

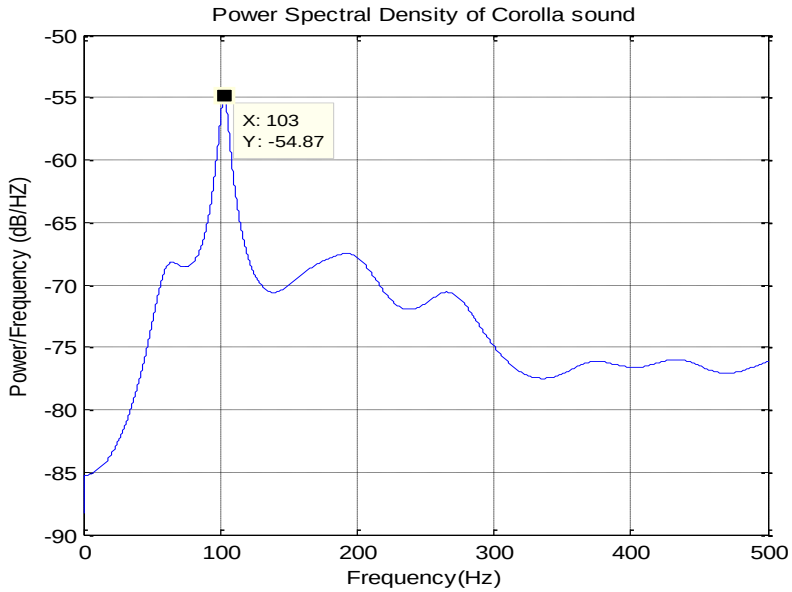

Figure 6.3. Power spectral density of Corolla sound

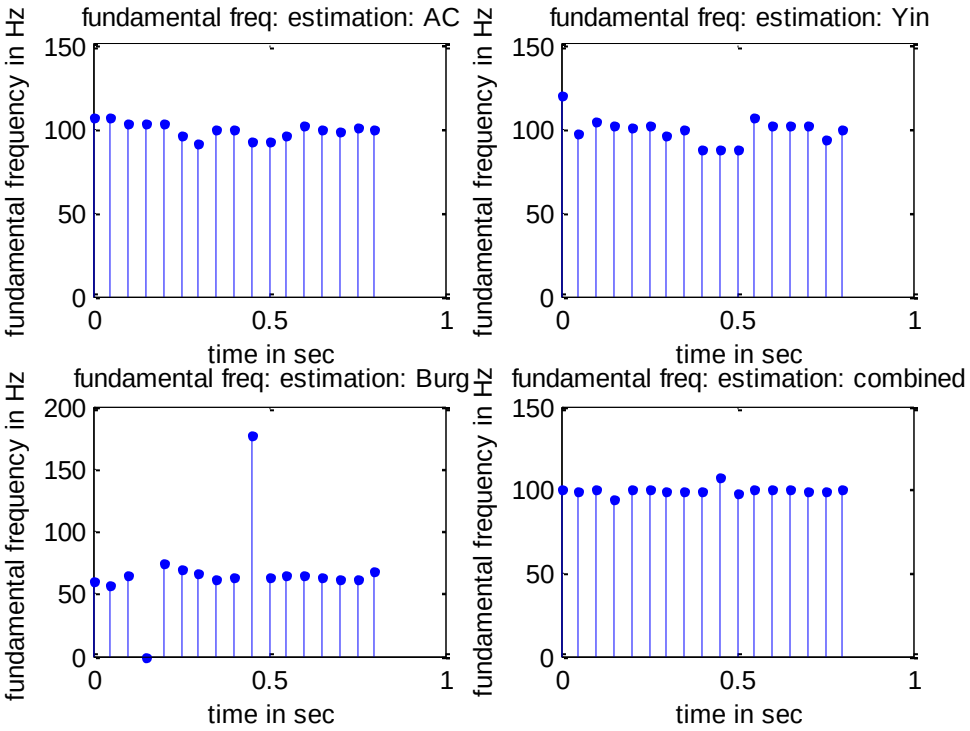

Figure 6.4. Fundamental frequencies for Corolla

6.2 Computation of Tristimulus features from the database

Tristimulus features give an indication of the mixture of harmonics in the sound by grouping the harmonics into three categories called stimuli. The first tristimulus calculates the relative weight of the first harmonic; the second tristimulus calculates the relative weight of the 2nd, 3rd, and 4th harmonics combined; and the third tristimulus calculates the relative weight of all the remaining harmonics. We computed the tristimulus based on the distribution of energy among the different harmonics using the equations:

$$T_1 = \frac{1}{E} RMS\{F_0\}$$

$$T_2 = \frac{1}{E} RMS\{2F_0, 3F_0, 4F_0\}$$

$$T_3 = \frac{1}{E} RMS\{5F_0, 6F_0, 7F_0, ...\}$$

where $E = RMS\{F_i, F_j, ...\}$ denotes the RMS (square root of the energy) of the signal made of the sum of the tones at the pitches $F_i, F_j, ...$

Figure 6.5 shows the tristimulus diagram for 2 Civic vehicle sounds over several segments. The distribution of the tristimulus for the 1st Civic vehicle sound is denoted by '1' and '2' for the 2nd Civic vehicle sound. Similarly, Figure 6.6 shows the tristimulus diagram for 2 Altima vehicle sounds. The distribution of the tristimulus for the 2 Altimas is denoted by '3' and '4'. The average of the tristimulus for each vehicle sound is different. The results from the tristimulus diagrams illustrate the usefulness of these features and therefore the tristimulus features can be used for classification.

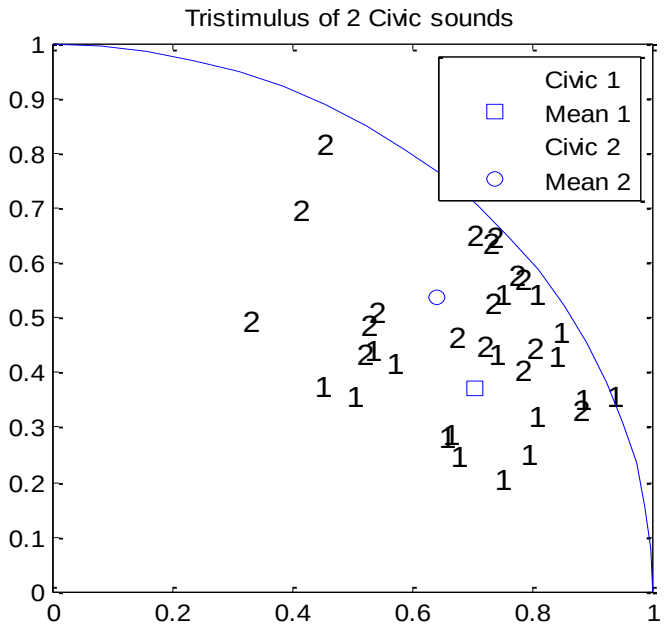

Figure 6.5. Tristimulus diagram for 2 Civic sounds

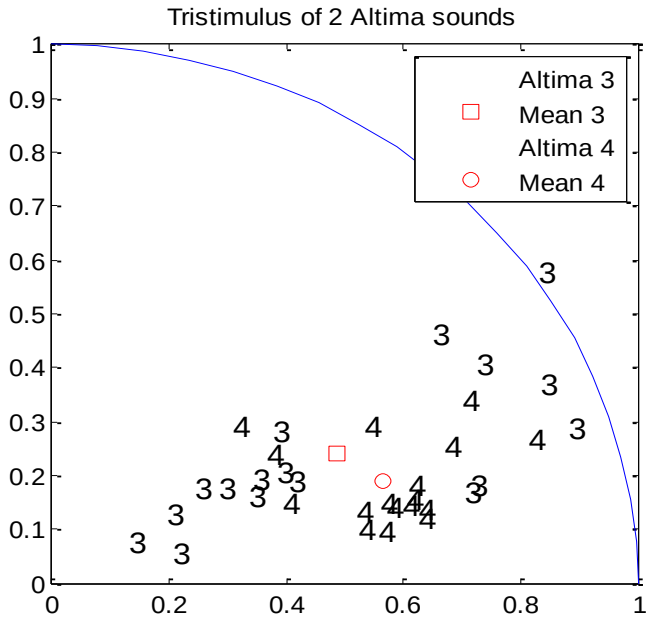

Figure 6.6. Tristimulus diagram for 2 Altima sounds

CHAPTER 7

RESULTS FROM CLASSIFICATION

7.1 Results from the quadratic classifier

In classification problems, we are usually interested in the test error, which is the expected prediction error on an independent set. The stratified 3-fold cross-validation was used in our work. The stratified 3-fold cross-validation randomly divides the data into 3 disjoint subsets of roughly equal size. Moreover, same-class proportions (e.g. ratio of Toyotas to Nissans) are roughly preserved in each of the sets as in the data to be classified. During the 3-fold cross-validation, one subset is removed and the rest two are used for training (including the minimization of classification error). The trained model is then used to classify the previously removed data set. This process is repeated for the next two. For instance, for 3 data sets A, B, and C of equal sizes, the optimization of the classifiers is performed in 3 stages. In the first, the data in A and B are used for training while C is used for post validation. Then the datasets A and C are used for training while B is used for testing and so forth. Then the discriminating functions are combined into one classifier whose performance is the unbiased average of the performance of the individual classifiers.

Data from our database was used to perform some two-class classifications using the stratified 3-fold cross-validation quadratic classifier. The performance of the quadratic classifier using the mean spectral features and the principal components of the spectral features were evaluated and compared. The confusion matrices for different

cases of two-class classifications are shown in the tables that follow. Table 7.1 shows the confusion matrix for Civic and Maxima. There were 21 Civics and 14 Maximas. Nine (9) Civics were misclassified as Maximas and 3 Maximas were misclassified as Civic. Twelve (12) Civics were correctly classified and 11 Maximas were correctly classified.

Table 7.2 shows the confusion matrix for Corolla and Maxima. There were 14 Corollas and 14 Maximas. Two (2) Corollas were misclassified as Maxima and 4 Maximas were misclassified as Corolla. Twelve (12) out of 14 Corollas were correctly classified and 10 out of 14 Maximas were correctly classified. Table 7.3 shows the confusion matrix for Accord and Camry. We used 34 Accords and 33 Camrys. Seven (7) Accords were misclassified and 19 Camrys were misclassified. Twenty-seven (27) Accords were correctly classified and 14 Camrys were correctly classified.

Table 7.4 shows the confusion matrix for Kia and Mazda. Out of 24 Kias, 19 were correctly classified and 5 misclassified as Mazda. Nine (9) Mazdas were misclassified as Kia. Twenty (20) out of 29 Mazdas were correctly classified. A summary of results using the mean spectral features and the principal components of the spectral features as input to the 3-fold stratified quadratic classifier can be seen in Table 7.5. For instance, from Table 7.5, the success rate using the mean spectral features and the PCA spectral features is 65.78% and 77.15% respectively for Civic-Maxima. The classification of Kia and Mazda achieved a very low correct classification rate of 54.72% using the mean spectral features. Performing PCA on the spectral features improved the correct classification rate to 73.58%. From Table 7.5, the quadratic classifier performs better with the principal components of the spectral features.

Table 7.1. Confusion matrix for Civic and Maxima

	Civic	Maxima	Total
Civic (Mean spectral features)	12	9	21
Maxima (Mean spectral features)	3	11	14

Table 7.2. Confusion matrix for Corolla and Maxima

	Corolla	Maxima	Total
Corolla (Mean spectral features)	12	2	14
Maxima (Mean spectral features)	4	10	14

Table 7.3. Confusion matrix for Accord and Camry

	Accord	Camry	Total
Accord (Mean spectral features)	27	7	34
Camry (Mean spectral features)	19	14	33

Table 7.4. Confusion matrix for Kia and Mazda

	Kia	Mazda	Total
Kia (Mean spectral features)	19	5	24
Mazda (Mean spectral features)	9	20	29

Table 7.5. Summary of results from Quadratic Classifier

2-Class Combination	SPECTRAL FEATURES	
	Mean %	PCs %
Civic-Maxima	65.78	77.15
Corolla-Maxima	78.58	82.15
Accord-Camry	61.7	65.67
Kia-Mazda	54.72	73.58

7.2 Results from the neural network classifier

Neural network classifiers usually exhibit superior performance over other classifiers. Neural networks are:

1. Nonlinear models. The nonlinear functionality of neural networks makes them more flexible in classifying complex real work data.
2. Data driven self-adaptive methods which enable them to adjust to the data.
3. Universal function approximators and thus able to approximate any function.

One common problem that occurs during neural network training is data overfitting, where the network tends to memorize the training examples without learning how to generalize to new situations. The default method for improving generalization is called early stopping and consists in dividing the available data set into three subsets. The sets are:

1. Training set, which is used for computing the gradient and updating the network weights and biases;
2. Test set, whose error is monitored during the training process. The error tends to increase when data is overfitted. The test set ensures that there is no overfitting;
3. Post training evaluation set, whose error can be used to assess the quality of the overall models. It serves as a completely independent test to the network generalization process.

The neural network model used in matlab, assigns 60% of the data set to the training set, 20% to the test set and 20% to the post training evaluation set. The confusion matrices shown in all the figures in this section are those for the 20% post training evaluation data set.

In this research effort, the first experimental work using the neural network examined the classification of vehicle sounds from the same manufacturer. From Table

7.6, 4 Civics and 7 Accords were used for the post training evaluation. From the confusion matrix in Figure 7.1, one (1) Accord was misclassified. We obtained 90.9% success rate. Similarly, from Table 7.7, 3 Corollas and 6 Camrys were used for the post training evaluation. The confusion matrix from Figure 7.2 shows 100% success rates.

Table 7.6. Vehicle sounds for testing and training for Civic and Accord

Vehicle	Training	Testing	Post training evaluation	Total
Civic	13	4	4	21
Accord	20	7	7	34

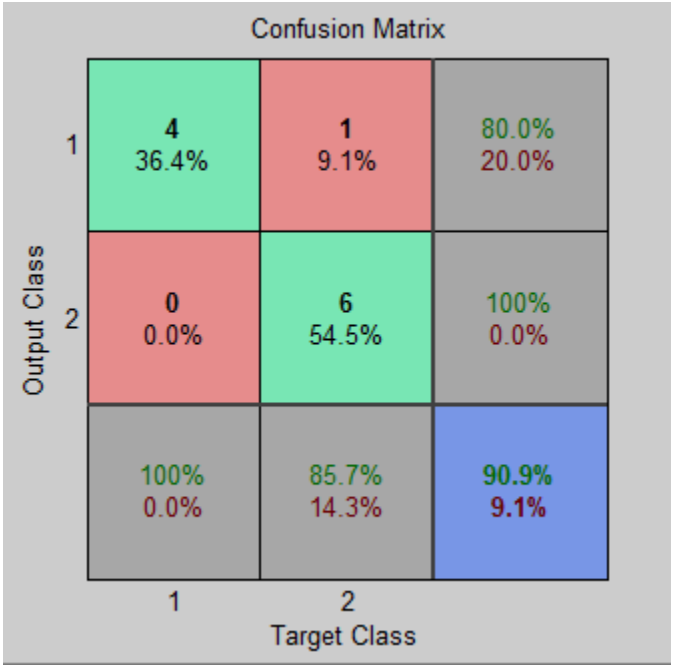

Figure 7.1. Confusion matrix for Civic and Accord

Table 7.7. Vehicle sounds for training and testing for Corolla and Camry

Vehicle	Training	Testing	Post training evaluation	Total
Corolla	8	3	3	14
Camry	20	7	6	33

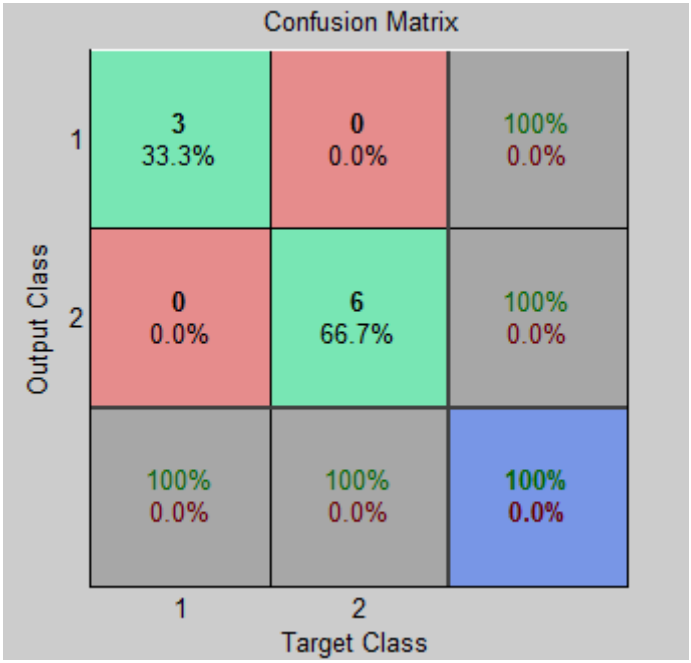

Figure 7.2. Confusion matrix for Corolla and Camry

The second experimental work examined the classification of vehicle sounds from different manufacturers. From Table 7.8, 4 Civics and 3 Maximas were used for the post training evaluation. From Figure 7.3, we recorded a 100% success rate for the classification of Civic and Maxima. Similarly, the number of vehicle sounds used for post

training evaluation is shown in Table 7.9, Table 7.10 and Table 7.11 for vehicles from different manufacturers. In Table 7.9, we used 3 vehicle sounds each for Corolla and Maxima for post training evaluation. Three Corollas and 3 Maximas were used for testing. The confusion matrices are shown in Figure 7.4, Figure 7.5 and Figure 7.6. A 100% success rate was achieved for all the cases considered as shown in the confusion matrices in Figure 7.4, Figure 7.5 and Figure 7.6.

Table 7.8. Vehicle sounds for training and testing for Civic and Maxima

Vehicle	Training	Testing	Post training evaluation	Total
Civic	13	4	4	21
Maxima	8	3	3	14

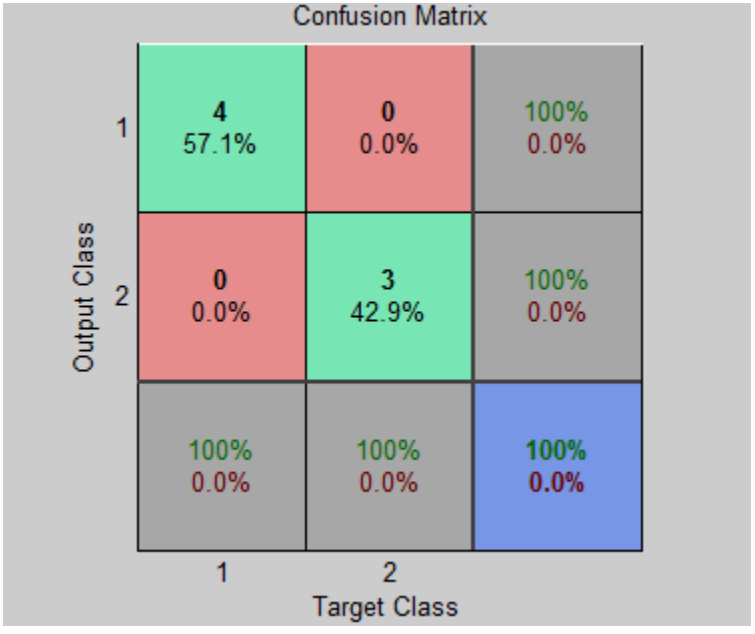

Figure 7.3. Confusion matrix for Civic and Maxima

Table 7.9. Vehicle sounds for training and testing for Corolla and Maxima

Vehicle	Training	Testing	Post training evaluation	Total
Corolla	8	3	3	14
Maxima	8	3	3	14

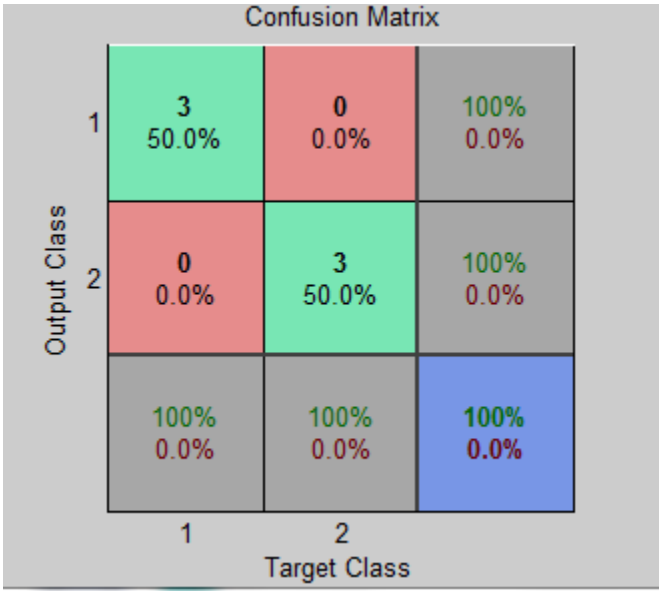

Figure 7.4. Confusion matrix for Corolla and Maxima

Table 7.10. Vehicle sounds for training for Accord and Camry

Vehicle	Training	Testing	Post training evaluation	Total
Accord	20	7	7	34
Camry	19	7	7	33

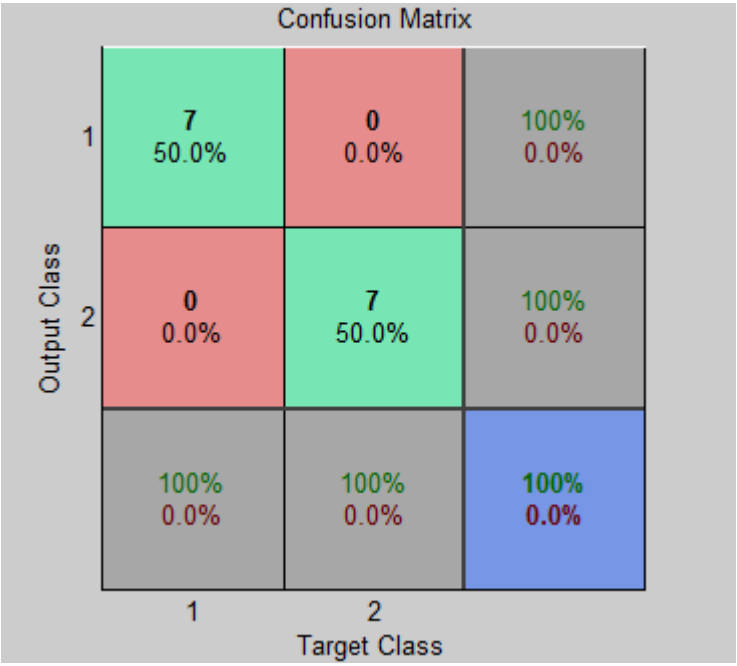

Figure 7.5. Confusion matrix for Accord and Camry

Table 7.11. Vehicle sounds for training and testing for Chevrolet, Camry and Accord

Vehicle	Training	Testing	Post training evaluation	Total
Chevrolet	18	6	6	30
Camry	19	7	7	33
Accord	18	8	8	34

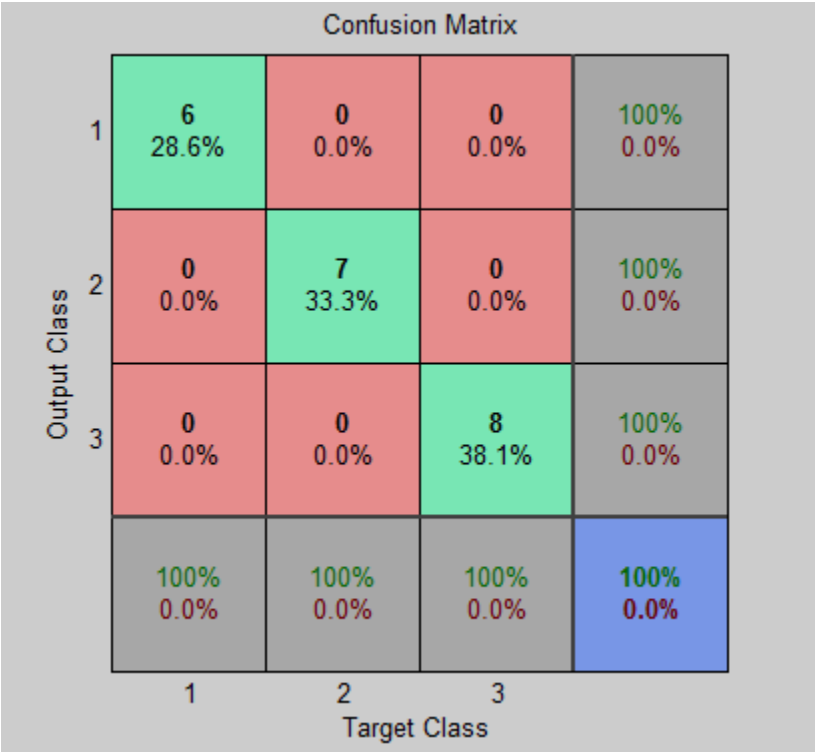

Figure 7.6. Confusion matrix for Chevrolet, Camry and Accord

A five-class scenario was also considered. Table 7.12 shows the investigation of a 5-class problem using the artificial neural network classifier. We used 41 Chevrolets, 21 Fords, 39 Accords, 33 Camrys and 21 Mazdas. Twenty-five Chevrolets, 11 Fords, 24 Accords, 19 Camrys and 12 Mazdas were used as training data. Eight Chevrolets, 5 Fords, 8 Accords 7 Camrys and 5 Mazdas were used for testing data.

Figure 7.7 shows the confusion matrix for the post training data set. There was zero error classification for all the 8 Chevrolets used in post training evaluation. Out of 5 Fords, 2 were misclassified as Chevrolet. Out of 7 Camrys, 1 was misclassified as an Accord. There was 40% and 14.3% error classification rate for Ford and Camry respectively. We achieved an overall success rate of 90.3% with the use of the principal components of the spectral features.

Table 7.12. Vehicle sounds for training and testing for a 5-class problem

Vehicle	Training	Testing	Post training evaluation	Total
Chevrolet	25	8	8	41
Ford	11	5	5	21
Accord	24	8	7	39
Camry	19	7	7	33
Mazda	12	5	5	21

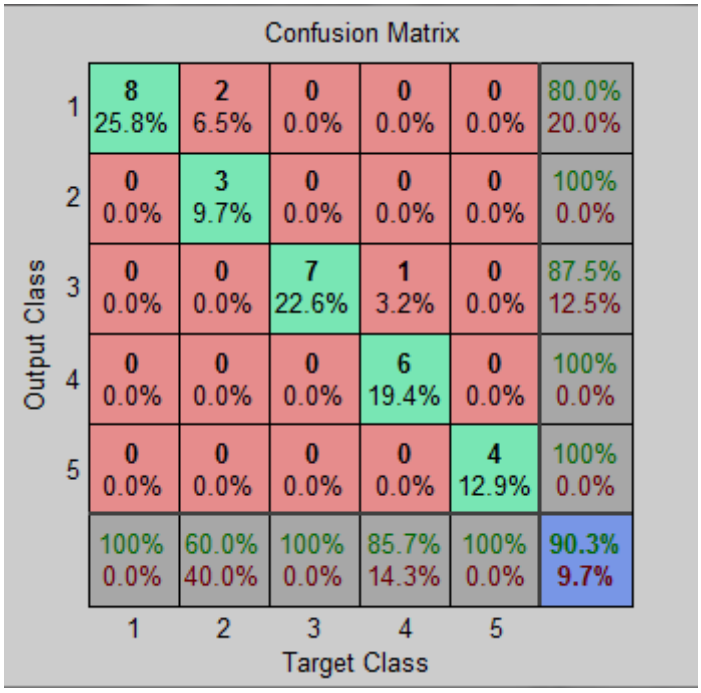

Figure 7.7. Confusion matrix for 5-class problem

The tristimulus features were also used for classification. Figure 7.8 shows a 100% classification when the tristimulus features are used alone as input to a neural network classifier for a two class problem. The result significantly degrades when there are more than two classes. We therefore examined the classification of the combined spectral and tristimulus features with the application of PCA on a five-class problem. Figure 7.9 shows the confusion matrix for the combined spectral and tristimulus features with the application of PCA. From Figure 7.9 we achieved a success rate of 93.8% which is a moderate increase over our previously reported 5-class problem using the principal components of the spectral features.

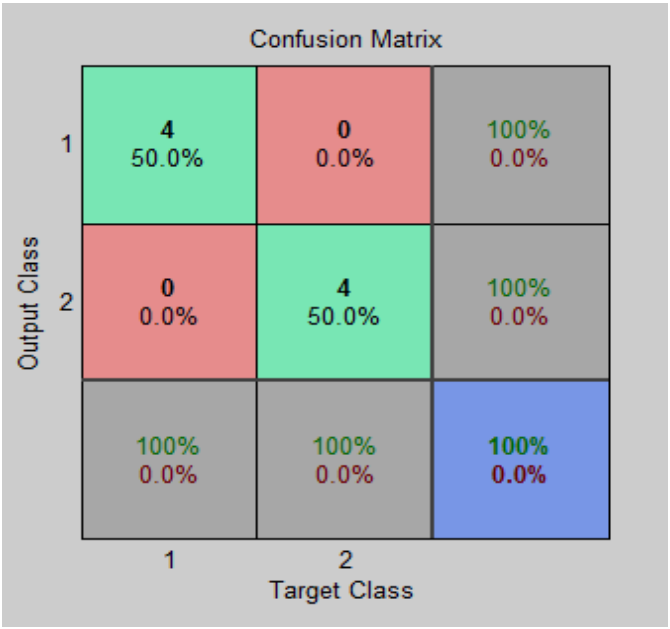

Figure 7.8. Confusion matrix for tristimulus features for Civic and Altima

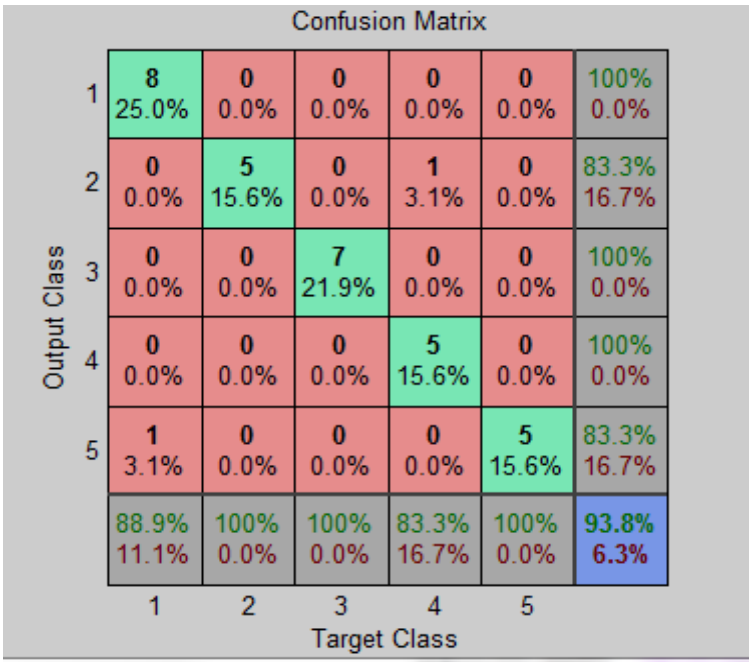

Figure 7.9. Confusion matrix for combined spectral and tristimulus features

CHAPTER 8

CONCLUSIONS AND FUTURE WORK

8.1 Conclusions

We have built a database of acoustic and seismic vehicle sounds. The mean and the principal components of the spectral features were used as inputs to the quadratic and the neural network classifiers. The quadratic and the neural network classifiers performed better with the use of the principal components of the spectral features than the mean of the spectral features. This is due to the fact that the principal components of the spectral features do not have as much overlap as the mean spectral features.

The neural network classifier out-performs the quadratic classifier with the use of the principal components of the spectral features. However, the quadratic classifier uses K-fold cross validation and hence our confidence in the performance is based on a larger sample and is unbiased and hence better.

The tristimulus features alone were sufficient to achieve low classification errors. We observed that the tristimulus features perform well only under 2-class scenarios for the neural network classifier.

We combined the spectral features and the tristimulus features and applied PCA on the combined features for a 5-class problem. The correct classification ratio from the neural network classifier was 93.8% which was slightly higher than 90.3% obtained by using the principal components of the spectral features.

These conclusions were validated using the NCAT database of vehicle acoustic and seismic signatures. We conclude that one can classify civilian vehicles into many makes when they are moving at about 15mph and when there is little interference from other vehicle sounds.

8.2 Future Work

Future work will test the interference from other overlapping vehicle sounds as well as classifying vehicles moving at higher speeds. Issues like Doppler shift, removing tire noise (seismic data) will be tested. We shall also consider the use of the seismic data (issues of synchronization) for classification.

REFERENCES

[1] Kozhisseri S., and Bikdash M., "Features for the Classification of Civilian Vehicles from Acoustic Data", *IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems*, April 2009, pp. 2-4.

[2] Mgaya R., Zein-Sabatto S., Shirkhodaie A., and Chen W., "Vehicle Identifications using Acoustic Sensing", *IEEE SoutheastCon, April 2007*, pp. 555-560.

[3] Takechi T., Sugimoto K., Mandono T., and Sawada H., "Automobile identification based on the measurement of car sound", *The 30th Annual Conference of the IEEE Industrial Electronics Society, 2*, May 2005, pp. 174-178.

[4] Succi George, Daniel Clapp, Robert Gambert and Gervasio Prado, "Footstep Detection and Tracking", *Proc. SPIE - The International Society for Optical Engineering, Acoustic and Seismic Unattended Ground Sensor (UGS) Systems*, April 2001, Vol. 4393.

[5] Krishnan N. C., and Panchanathan S., "Analysis of low resolution accelerometer data for continuous human activity recognition", *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008,* May 2008, pp. 3337-3340.

[6] Nishkam R., Nikhil D., Preetham M. and M. Littman, "Activity Recognition from Accelerometer Data", *American Association for Artificial Intelligence (www.aaai.org)*, pp. 1-6.

[7] Seon-Woo Lee and Mase K., "Activity and location recognition using wearable sensors", *Journal Pervasive Computing, IEEE*, Vol. 1, Issue 3, Dec 2002, pp. 24-32.

[8] Brillinger D.R, *Time Series, Data Analysis and Theory*, Expanded Edition, Holden-Day Inc., San Francisco, 1981.

[9] Mendel J. M., "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications", *Proc. of the IEEE*, Vol. 79, No. 3, March 1991, pp. 277-305.

[10] Dwyer R.F., "Use of the Kurtosis Statistic in the Frequency Domain as an Aid in Detecting Random Signals", *IEEE Journal of Oceanic Engineering*, Vol. OE-9, No. 2, April 1984, pp. 85-92.

[11] Dwyer R.F., "Detection Performance of discrete power and higher-order estimates", *IEEE Journal of Oceanic Engineering*, Vol. OE-10, No. 3, July 1985, pp. 303-315.

[12] Otonnello C. and S. Pagnan, "Modified Frequency Domain Kurtosis for Signal Processing", *Electronics Letters*, Vol. 30, No. 14, July 1994, pp. 1117-1118.

[13] Pagnan S., C. Otonnello and G. Tacconi, "Filtering of Randomly Occuring Signals by Kurtosis in the Frequency Domain", *Proc. of the 12th International Conference on Pattern Recognition*, Oct. 1994, Vol. 3, pp. 131-133.

[14] Yang Sun, and John N. Daigle, "A PCA-based Vehicle Classification System in Wireless Sensor Networks", *IEEE Wireless Communications and Networking Conference*, April 2006, pp. 2193.

[15] Junwen Wu, and Xuegong Zhang, "A PCA classifier and its application in Vehicle detection", *Proceedings IEEE International Joint Conference on Neural Networks*, July 2001, pp. 1-3.

[16] Meta S., Cinsdikici, M. G, "Vehicle-Classification Algorithm Based on Component Analysis for Single-Loop Inductive Detector", *IEEE Transactions on Vehicular Technology*, July 2010, Vol. 59, Issue 6, pp. 2795.

[17] Pearson K, *On Lines and Planes of Closest Fit to Systems of Points in Space*, Philosophical Magazine 2 (6), 1901, pp.559-572.

[18] Hotelling H., "Analysis of a complex of statistical variables into principal components", *Journal of Educational Psychology*, Vol. 24, pp. 417-441, 498-520.

[19] Nixon Mark S., and Alberto S. Aguado, *Feature Extraction and Image Processing*, Elsevier, 2nd Ed, 2008, pp. 332.

[20] Hastie T., Robert T., and Friedman J., *The Elements of Statistical Learning*, New York, Springer-Verlag, 2001.

[21] Redmond Stephen J. and C. Heneghan, "Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea", *IEEE Trans. Biomed. Eng.*, Vol. 53, No. 3, 2006, pp. 485-496.

[22] Wei Wu, Zhang QiSen, and Wang Mingjun, "A Method of Vehicle Classification Using Models and Neural Networks", *IEEE Vehicular and Technology Conference 53rd*, Vol. 4, 2001, pp. 3022-3026.

[23] Raja Abdullah, M. Saripan, M.Cherniakov, "Neural Network Based for Automatic Vehicle Classification in Forward Scattering Radar", *IET International Conference on Radar Systems*, Feb 2009, pp. 1-5.

[24] Sumathi S and Surekha P., *Computational Intelligence Paradigms Theory and Applications using Matlab*, Taylor & Francis Group, 2010, pp. 29-34.

[25] Latha P., Ganesan L., and Annadural S., "Face Recognition using Neural Networks", *Signal Processing An International Journal*, 2009, Vol. 3, Issue (5), pp. 153-160.

[26] Audio Technica, U.S., Inc. (Audio-Technica) Retrieved from www.audio-technica.com:http://www.audiotechnica.com/cms/resource_library/literature/49f63e6efc082082/at2020_english.pdf, 2009.

[27] Shure, Retrieved from www.shure.com/idc/groups/public/documents/webcont_ent/us_pro_sm81_specsheet.pdf, 2009.

[28] Sensr, SENSR Solutions. Retrieved from www.sensr.com/products/civil-and-structural/cx1.php, 2010.

[29] Kie B. Eom, "Analysis of acoustic signatures from moving vehicles using time-varying autoregressive models", *Multidimensional Systems and Signal Processing*, Vol. 10, No. 4, October 1999.

[30] Daubechies I. "The Wavelet Transform, Time-Frequency Localization and Signal Analysis", *IEEE Transactions on Information Theory*, 1990, Vol. 36, Issue 5, pp. 961-1005.

[31] Kay S. M., *Modern Spectral Estimation*: *Theory and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[32] Marple S. L. Jr., *Digital Spectral Analysis with Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[33] Stoica P. and R. Moses, *Introduction to Spectral Analysis*, Prentice-Hall, Upper Saddle River, NJ, 1997.

[34] Helstrom C. W., *Statistical Theory of Signal Detection*, 2nd ed., New York: Pergamon Press, 1968, Ch. IV, 4, pp. 124-130.

[35] Valeriu D. Vrabie, Pierre G and Christine S. "Spectral Kurtosis: From Definition to Application", *6th IEEE International Workshop on Nonlinear Signal and Image Processing*, 2003, pp. 1-5.