12-13-2003

Automated Accident Detection In Intersections Via Digital Audio Signal Processing

Navaneethakrishnan Balraj

AUTOMATED ACCIDENT DETECTION IN INTERSECTIONS

VIA DIGITAL AUDIO SIGNAL PROCESSING

By

Navaneethakrishnan Balraj

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in Electrical Engineering
in the Department of Electrical and Computer Engineering

Mississippi State, Mississippi

December 2003

Name: Navaneethakrishnan Balraj

Date of Degree: December 13, 2003

Institution: Mississippi State University

Major Field: Electrical and Computer Engineering

Major Professor: Dr. Lori Mann Bruce

Title of Study: AUTOMATED ACCIDENT DETECTION IN INTERSECTIONS VIA DIGITAL AUDIO SIGNAL PROCESSING

Pages in Study: 62

Candidate for Degree of Master of Science

The aim of this thesis is to design a system for automated accident detection in intersections. The input to the system is a three-second audio signal. The system can be operated in two modes: two-class and multi-class. The output of the two-class system is a label of "crash" or "non-crash". In the multi-class system, the output is the label of "crash" or various non-crash incidents including "pile drive", "brake", and "normal-traffic" sounds. The system designed has three main steps in processing the input audio signal. They are: feature extraction, feature optimization and classification. Five different methods of feature extraction are investigated and compared; they are based on the discrete wavelet transform, fast Fourier transform, discrete cosine transform, real cepstrum transform and Mel frequency cepstral transform. Linear discriminant analysis (LDA) is used to optimize the features obtained in the feature extraction stage by linearly combining the features using different weights. Three types of statistical classifiers are investigated and compared: the nearest neighbor, nearest mean, and maximum likelihood

methods. Data collected from Jackson, MS and Starkville, MS and the crash signals obtained from Texas Transportation Institute crash test facility are used to train and test the designed system. The results showed that the wavelet based feature extraction method with LDA and maximum likelihood classifier is the optimum design. This wavelet-based system is computationally inexpensive compared to other methods. The system produced classification accuracies of 95% to 100% when the input signal has a signal-to-noise-ratio of at least 0 decibels. These results show that the system is capable of effectively classifying “crash” or “non-crash” on a given input audio signal.

DEDICATION

I would like to dedicate thesis work to my parents Mrs. Banumathy Balraj and Mr. Balraj, whose sincere support and prayer have made me achieve the desired, and also to my sister Nithya.

ACKNOWLEDGEMENTS

I would like to sincerely thank my advisor Dr. Lori Mann Bruce, for her support throughout my research work. Her dedicated support and suggestions have helped me complete this research. Each and every stage of the thesis made me learn something from her. I thank Dr. Roger L. King, for serving as my committee member.

I also thank Dr. Yunlong Zhang from Department of Civil Engineering for his support during this entire project. I thank Mr. Qingyong Yu, my research colleague in this research. I also thank my friends and Signal Processing Research Applications Laboratory (SPiRAL) group for their cooperation and help.

I am indebted to the Mississippi Department of Transportation (MDOT) for their financial support in doing this research as well as in my academic assistance. With all these people behind, my success in this research gave an amazing, ever achievable satisfaction.

TABLE OF CONTENTS

LIST OF TABLES

TABLE Page

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

The traffic system prevailing in the U.S national highway system was not adequately modeled to take care of congestion occurring due to abnormal traffic incidents. Traffic incidents are any non-recurring events that cause congestion on the traffic flow. This traffic congestion in turn causes traffic delay, more fuel consumption, air pollution and secondary accidents. Common types of incidents like accidents, breakdowns, construction and maintenance activities, bad weather, and structural failures occurring on the roads account for about 57 percent of the delays [1]. Approximately 50 to 60 percent of the delay on urban freeways is associated with incidents. Traffic congestion caused by these incidents is estimated by 2005 to cost the nation over $75 billion and 8.4 billion gallons of wasted fuel as lost productivity [1].

At intersections, vehicular flow is such that all approaches making left-turn, through, and right-turn movements leads to a majority of incidents at intersections when they try to get at the same time. Among these incidents, accidents occurring at intersections are the most serious ones and are estimated to be 2500-5000 vehicle-hours of delay per incident [2]. Thus, identifying such accidents in the intersections as early as possible and avoiding congestion at the intersections is bound to improve the chances of safety of victims. Therefore, timely and accurate accident detection at intersections is important in any traffic system.

Research on incident detection started in the early 60's [3]. Most of the previous work on incident detection [3,4] did not place an emphasis on the need to develop a system to detect accidents at intersections. Rather, they focused more on detecting accidents in highways, freeways, etc. To provide a necessary medical and emergency response, quick and accurate detection of accidents is necessary, thereby reducing congestion and delay. A system needs to be developed to automatically detect accidents at intersections, which will help in the whole traffic management system.

Detecting a real-time accident at an intersection is a very challenging task. In recent years, technological innovations have provided many advanced traffic sensors. Many detection systems are implemented in the Advanced Traffic Management System (ATMS). Instruments like magnetic, ultrasound, microwave, infrared light, and optical beam sensors are used in these detection systems. These sensors mainly provide direct measurement for counting, occupancy measurement, presence detection, queue detection, speed estimation, and vehicle classification [5]. In the traffic management system, the inductive loop detector is the most common sensor, but it has a high failure rate. One of the main defects of many of the commonly used sensors is high weather sensitivity. For example, systems that use video cameras suffer higher error rates when there are poor lighting conditions including darkness, precipitation, fog, or dust. One other important consideration in many of these types of sensors is the implementation can be costly in terms of equipment, installation, maintenance, and management.

Traffic accidents have characteristic sounds that can differentiate them from the normal traffic sounds such as vehicle passing, vehicle braking, vehicle sirens,

construction noise, thunder, etc. Audio sensors like microphones are, as compared to video sensors, cost effective, easy to install, and require less maintenance and management. Audio sensors are very adaptable to environmental conditions like variations in lighting, temperature, and humidity as compared to other sensors.

The goal of this thesis is to develop an automated system that can detect accidents at intersections using audio signals. Figure1 shows the proposed system block diagram. The input to the system is a three second segment signal recorded with a audio sensor, like a simple microphone, at an intersection. The system can perform in two modes: Two-class and multi-class. The output of the system is a label of “crash” or “non-crash” for a two-class system, and for a multi-class mode, the system identifies and labels “crash” and non-crash incidents like “pile drive”, “brake”, etc.

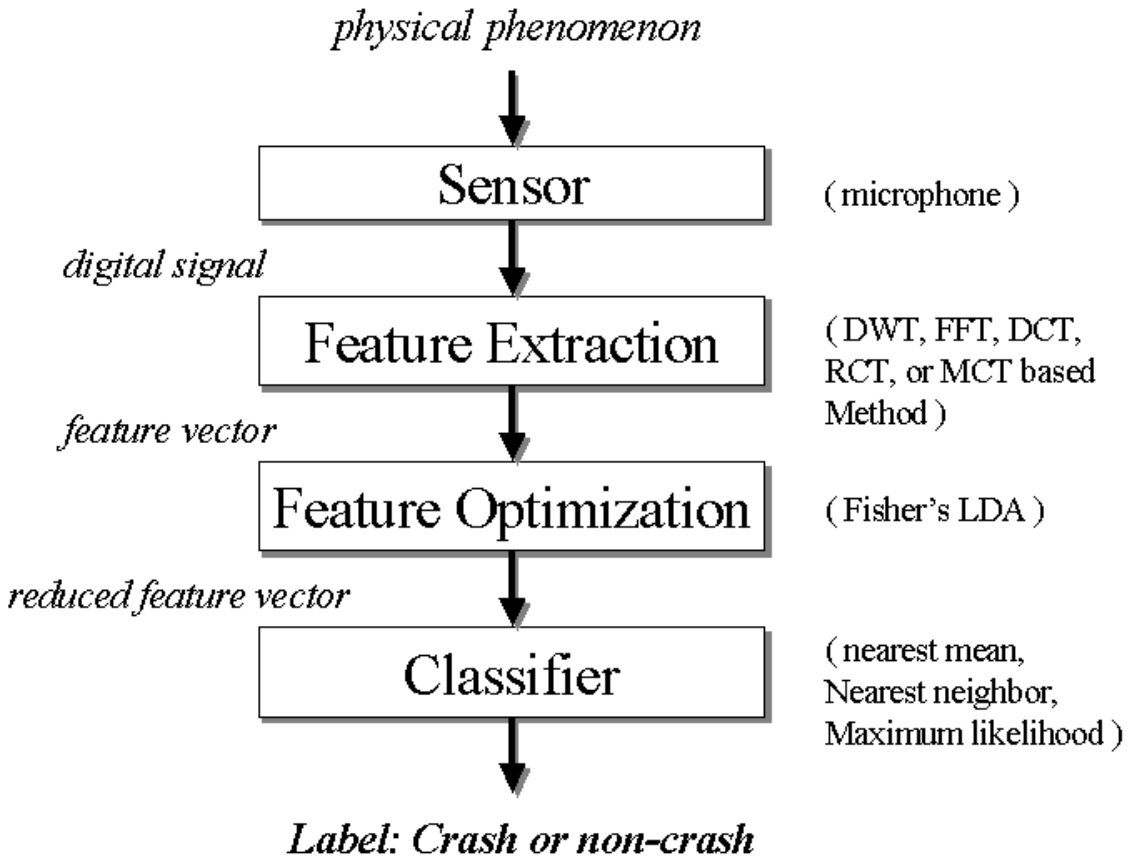


Figure 1 System block diagram

The three-second digital input signal is used to extract features using various transform methods like Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT), Mel frequency Cepstrum Transform (MCT), Real Cepstrum Transform (RCT), and Fast Fourier Transform (FFT). While using DWT as the feature extractor, various mother wavelets like Haar, symlet, or Coiflets are investigated. The feature vector obtained using the above said transform methods are optimized using Fischer's Linear Discriminant Analysis (LDA). The feature optimizer essentially reduces the size of the feature vector while maximizing class separation. Furthermore, the reduced features are input to the statistical classifier to finally assign a label of "crash" or "non-crash". Classifiers like nearest neighbor, maximum likelihood, and nearest mean methods are

investigated and compared. Performance of the overall system is measured using the leave-one-out testing method. A database of recorded audio signals of normal traffic and traffic accidents was used to train and test the system. The results obtained by using all the above methods showed that the combination of DWT with LDA and maximum likelihood classifier produced the best classification accuracy.

The accuracy rate showed that the system could classify “crash” and “non-crash” signals. Moreover, the system could also classify the type of non-crash signal. Thus the implementation and installation of the proposed system could greatly save lives and property by reducing traffic congestion and reducing secondary accidents.

CHAPTER II

CURRENT STATE OF KNOWLEDGE

2.1 Accident detection systems:

Prior research has been done for the development of accident detection systems at intersections. Sensors are the basis for these accident detection systems. Various types of sensors have been investigated including magnetic, ultrasound, infrared, video cameras and moving vehicle detectors that use microwave. These sensors are for surveillance. The prevailing systems mainly use video cameras as the sensors, which helps to visually identify the crash at a particular intersection or on a freeway [6]. Installation of video cameras at each and every intersection would be relatively costly. Researchers did make an attempt to use video sensors, when accident detection at freeways and at intersections first became important to avoid traffic congestions. Later, they tried using ultrasonic waves and microwaves [6,7] to detect the incidents at intersections and vehicle detections [8].

A commonly used system for incident detection is the inductive loop detector (ILD) [8]. ILDs are designed such that rectangular loops of cable are laid below the surface of the carriageway, with all the cables a set distance from the signal stop line. Though ILDs are the most commonly used detectors, the system overall has so many

disadvantages that researchers need a better alternative in accident detection systems. One of the main problems for the increase in the failure rate of detecting the accidents is

damaged loops and feeder cables. It was estimated that to repair the damaged loops and feeder cables by digging and resurfacing of the carriageway, would cost around 3 million dollars per annum in London, U.K[8]. The only alternative system that very well reached was microwave vehicle detectors (MVD).

Dickinson *et al* [8] evaluated how well MVDs perform at intersections. A narrow beam of microwave energy is transmitted, and the frequency of the reflected beam from the passing vehicles helps to detect incidents (using the Doppler effect). It was found that a vehicle that travels at a speed of 30 mph would have a difference frequency of 1000 Hz. Installation of these types of detectors is cheap and easy, because detectors are mounted on a single pole where as the loop detectors are laid below the surface of the carriageway. Also, the maintenance cost is cheaper when compared to loop detectors. MVDs have their own disadvantages, however. Dickinson *et al* evaluated MVDs by comparing the performance with respect to ILDs. They found that ILDs work better with 0% (missing vehicles), where as MVDs make a 3.5% miss. The main reason for this miss is found to be the speed with which the vehicles move. MVDs cannot detect vehicles in slow-moving or stationary traffic. Thus, the evaluation results show that while MVDs are less expensive, are operational for a longer duration in time, and are cheap to maintain, ILDs provide a more reliable system for detection.

Subramaniam *et al* [6] developed an algorithm to detect an object in motion using a wavelet transform. They used a video camera as the sensor and tried to detect objects in motion. An image flow field was estimated using the Gabor wavelet transform. This transform produced the complex coefficients. The phase difference between the wavelet coefficients in two successive frames helped to estimate the flow, and doing this at each pixel provided the estimated image flow. From this image flow field, motion hypotheses were extracted. A histogram with the flow field vectors was drawn, and a low pass filter was applied to the histogram. Finally, the maxima of a certain threshold level were considered. This resulted in a displacement vector, which is similar to feature optimization, wherein the features obtained using the image flow field were reduced. After obtaining the displacement vector, the Mallat wavelet transform was used to evaluate the edges on two successive frames. The image flow field obtained using the Gabor wavelet transform was used in this stage. These two algorithms, Gabor wavelet and Mallat wavelet transform, improved the accuracy of moving vehicle detection. The main disadvantage of the system was that when two vehicles move closely at the same speed, it was difficult for the system to detect both of the vehicles.

Many different methodologies are still being researched to construct a better system that will help the traffic management system detect intersection accidents and thereby reduce the traffic delays and congestions. In addition to video, ultrasound, and microwave sensors, audio sensors are being investigated.

2.2 Traffic analysis using acoustic sensors:

There are two types of traffic monitoring systems: those that detect traffic accidents and, more generally, those that detect overall traffic conditions. Accident

detection reduces the traffic congestions and traffic delays, but traffic management by detecting the traffic conditions would considerably reduce major traffic accidents as well as secondary accidents. In both scenarios, the existing sensors like video, ultrasound, and microwave, have their respective problems.

Researchers have investigated an alternative approach of using the acoustic signal [9,10]. Chen *et al* developed an algorithm to detect traffic conditions (but not necessarily accidents) with the help of sound signals recorded from moving vehicles [9]. The main advantage of the system is the flexible mounting of the microphones; they can be fixed on the road-side. The system can detect the speed and number of vehicles that have passed within a particular time. The system has two microphones set a small distance apart on the road-side; the sound wave from a vehicle is recorded with both the microphones. The cross correlation of these two signals is used to measure the time difference. The time difference and the speed of the sound give the air path difference (APD), and with this APD the vehicle is easily detected. The authors experimented with the system and obtained encouraging results. The main advantages of such a system are low installation and maintenance costs, flexibility in installing the sensor, and lower operational cost since the system is completely automated. However, the system is not designed to detect accidents in intersections.

Brockmann *et al* developed a technique to detect and count vehicles in motion based on the acoustic energy radiated from the axels, so that the information can be used to control traffic at intersections [10]. The acoustic energy radiated from the vehicle axels has unique characteristics, and these characteristics are matched with the signal model. The authors compared the results obtained using Fourier analysis, correlation methods,

and a developed signal model; the proposed technique was shown to be computationally efficient. However, the system is not designed to detect accidents in intersections.

2.3 Feature extraction from audio signals

Extracting important features from a crash or non-crash signal is one of the first stages in the system, as this would help to reduce the processing time of the system and increase the accuracy of the system. The basic idea of feature extraction is to represent the important and unique characteristics of each signal in the form of a series of numeric values, i.e., a feature vector. These feature vectors can be further used to classify the signal as crash or non-crash using a statistical classifier or a neural network. Researchers have tried using wavelet and cepstral transforms to extract features from audio signals such as speech signals [11,12]. These methods have achieved good results, which have stimulated investigations for a best transform method.

Kadambe *et al* developed a pitch detector using a wavelet transform [11]. One of the main properties of the dyadic wavelet transform is that it is linear and shift-variant. This property is useful when speech or audio signals are analyzed, as they are typically modeled as a linear combination of shifted and damped sinusoids. Another important property of the dyadic wavelet transform is that its coefficients have local maxima at a particular time when the signal has sharp changes or discontinuities. These two important properties of the dyadic wavelet transform help to extract the unique features of a particular speech or audio signal. Kadambe *et al* made a comparison of the results obtained from using dyadic wavelet transforms, autocorrelation, and cepstral transforms. The investigation showed that the dyadic wavelet transform pitch detector gave 100% accurate results. One reason for the difference in the results was that the other two methods assume stationarity within the signal and measure the average period, where as the dyadic wavelet transform takes into account the non-stationarities in the signal.

Hence, the dyadic wavelet transform method would be the best to extract feature when the signals are non-stationary.

Harlow *et al* developed an algorithm to detect traffic accidents at intersections [12]. The authors use an audio signal as the input to the system. The algorithm uses the Real Cepstral Transform (RCT) as a method to extract features. The audio signal is recorded using a digital audio tape (DAT) recorder with a microphone as the sensor. The signals recorded at intersections include brake, pile drive, construction and normal traffic sounds. These signals are segmented into three-second sections. Each of these three second segmented signals is analyzed using RCT. RCT is a method where the signal is windowed for every 100msec using a hamming window with an overlap of 50 msec. Thus, for a given three-second signal, there will be almost 60 segments of 100 msec duration each. RCT is applied to each of these segments, and the first 12 coefficients are used as the features. The features obtained using the RCT are then classified as "crash" or "non-crash" using a neural network.

The above analysis corroborates the fact that previous research studies have been done in the past to develop an efficient traffic incident detector system. From the analysis of traffic detection systems that are based on audio signals, it is clear that wavelet transform could potentially out perform other transformation methods, due to its ability to analyze non-stationary signals.

CHAPTER III

METHODOLOGIES

Traffic accidents have unique characteristics that help to differentiate them from the normal traffic sounds. Traffic accidents occur within a very short time duration. Hence in order for a system to detect accidents, the system should be capable of processing short time duration signals. Windowing the incoming audio signal is an obvious way for a system to perform analysis on a short time signal. Various transform methods are used to extract features from each of the windowed signals.

The main focus of this research was on selecting an optimum feature extraction method. Five different feature extraction methods including DWT, FFT, DCT, RCT, and MCT are investigated and compared. For the DWT approach, the feature vector is obtained by computing the root-mean-square energy of the wavelet coefficients at each scale. The number of DWT features is dependent on the number of scales in the DWT decomposition. The number of scales, and hence features, is dependent on the type of mother wavelet utilized. For the FFT method, the features are the magnitude of the FFT coefficients. The number of FFT features is equal to the order of the FFT. In order to have a fair comparison between methods, the FFT order is selected such that the number of FFT features is equal to the number of DWT features. Similar to the FFT, the DCT coefficients are used as the feature vectors. For the RCT and MCT approaches, the transform coefficients are used as features. These methods typically resulted in 12-14

features. A detailed analysis of each of these feature extraction methods is provided in this thesis.

3.1 Feature Extraction Methods

3.1.1 Real Cepstrum Transform:

Cepstrum is a term first used in speech analysis by Boger et al; it is a method of speech analysis based on the spectral representation of the signal [13]. One of the main properties of this method is that it is a homomorphic transform. A homomorphic transformation is one in which the convolution of two signals

$$x_1[n] * x_2[n] \tag{1}$$

becomes equivalent to a sum, through the use of logarithms,

$$\ln(x_1[n] * x_2[n]) = \hat{x}_1[n] + \hat{x}_2[n] \tag{2}$$

of the cepstra of the signals. The real cepstrum $c[n]$ of a digital signal $x[n]$ is defined as

$$c[n] = \frac{1}{2n} \int_{-\Pi}^{\Pi} \ln\left|X\left(e^{jw}\right)\right| \, e^{jwn} dw \tag{3}$$

where $X\left(e^{jw}\right)$ is the Fourier transform of $X(n)$.

Practical implementation of the real cepstrum (in Matlab) is the inverse Fourier transform of the real logarithm of the magnitude of the Fourier transform. If x is the signal, then the output of the real cepstrum will be

$$y = \mathrm{Re}\left(F^{-1}\left(\log\left(|F(x)| \right) \right) \right) \tag{4}$$

where F^{-1} is the inverse Fourier transform and F is the Fourier transform. Typically in speech processing analysis, the first 12-14 coefficients are considered as features [14].

3.1.1.1 Utilization of real cepstrum:

Cepstral analysis is used extensively in speech recognition. The given signals are windowed using a Hamming window with 100 msec intervals, where adjacent 100 msec intervals have an overlap of 50 msec, which helps the thorough analysis of the signal. Then, for each of these overlapping intervals, a real cepstrum transform method is applied, and the first few resulting coefficients are taken into account for analysis. Typically in speech processing, only the first 12-14 coefficients are taken into account. So, only the first 12 coefficients are utilized. This process is repeated for each 100 msec interval. Even for relatively short duration signals, like three-second signals, this method requires a very large amount of processing time. To make a comparison with the DWT method, the first $(M+1)$ coefficients are used as features, since the DWT approach results in $(M+1)$ features, where M is the maximum number of DWT decomposition levels. Thus, to analyze a three-second signal, sixty 100 msec segments are transformed using the real cepstrum transform, and all these 60 segments will produce a feature vector formed from the first 12 coefficients. The extracted RCT features are analyzed using the feature optimization method and then a statistical classifier like a maximum likelihood method. All these 60 segments are analyzed, and if a "crash" is detected for at least three of the segments, an overall classification of "crash" is assigned to the three-second signal.

3.1.2 Mel Frequency Transform:

Mel frequency cepstral transform is a depiction of RCT of a windowed short-time signal derived from the fast Fourier transform of the signal [14]. The basic difference between the RCT and the MCT is that a non-linear frequency scale (triangular filter) is

used in the MCT as opposed to (linear frequency scale) Hamming window in the RCT. Let the DFT of a given input signal $x[n]$ be

$$X_a[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi nk/N}, \qquad 0 \le k \le N \tag{5}$$

A filter bank with M' filters is defined, where filter (m= 1,2,…M') is a triangular filter, which is given by

$$H_m[k] = \begin{cases} 0 & k < f[m-1] \\ \dfrac{(k - f[m-1])}{f[m] - f[m-1]} & f[m-1] \le k \le f[m] \\ \dfrac{(f[m-1] - k)}{(f[m+1] - f[m])} & f[m] < k \le f[m+1] \\ 0 & k > f[m+1] \end{cases} \tag{6}$$

which satisfies $\sum_{m=1}^{M'} H_m[k] = 1$.

Let the lowest and the highest frequencies of the filter bank be f_l and f_h respectively, Fs be the sampling frequency in Hz, M' be the number of filters, and N be the order of the FFT. The boundary points $f[m]$ are uniformly spaced in the mel-scale

$$f[m] = \left(\frac{N}{F_S}\right) B^{-1}\left(B(f_l) + m\frac{B(f_h) - B(f_l)}{M+1}\right) \tag{7}$$

where the mel-scale B is given by $B(f) = 1125\ln(1 + f/700)$ and $B^{-1}(b) = 700\left(e^{(b/1125)} - 1\right)$.

The log-energy at the output of each filter is

$$S[m] = \ln\left[\sum_{k=0}^{N-1} |X_a[k]|^2 H_m[k]\right], \qquad 0 < m \le M' \tag{8}$$

The mel-frequency cepstrum is then the discrete cosine transform of the M filter outputs:

$$c[n] = \sum_{m=0}^{M'-1} S[m]\cos(\pi n(m-1/2)/M') \ , \qquad 0 \le n < M' \qquad (9)$$

where M' varies for different implementations from 24 to 40. In speech analysis, typically only the first 13 cepstrum coefficients are used.

3.1.2.1 Utilization details of MCT:

The MCT uses a triangular window and a nonlinear frequency scale. The number of triangular filters varies from 24 to 40. In the system designed for this thesis, 40 triangular filters are used, *i.e.* M'=40 in the above equations. Similar to the RCT, the MCT method requires a windowing of the input audio signal. The three-second audio signal is partitioned into 40 segments. As compared to the RCT, the duration of the windowed signal varies as MCT uses a nonlinear frequency scale. For each segment, 13 of the MCT coefficients are used for classification. For each of the feature vectors formed with 13 cepstrum coefficients obtained from each segment, the feature optimization method and statistical classifiers are applied. The method is repeated 40 times, and an overall classification of "crash" is assigned to the signal if at least three of the segments are classified as "crash". Like the RCT, the MCT is much more computationally intensive than the DWT, FFT, or DCT methods.

3.1.3 Fast Fourier Transform:

Fast Fourier transform is a method that transforms the signal from the time domain to the frequency domain. Discrete Fourier transform is a Fourier representation of

a discrete time signal in a discrete sequence. The Fourier transformation for the signal $x(n)$ is given as

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j(2\pi / N)nk} \tag{10}$$

where $k = 0,1,2,\ldots,\ N$ -1.

The signal is transformed into a weighted sum of sinusoids. This transformation makes the implementation easy and thereby reduces the computational expense. The number of operations required for implementation using FFT is O($N \log_2 N$), where one operation is one real multiplication and one real addition. The number of operations required for implementation depends on N. Commonly used fast Fourier algorithms need N to be equal to 2^p where p is any positive integer.

3.1.3.1 Utilization of FFT:

The FFT is used to extract features from the signal. The FFT can be computed for different orders ($N = 2^p$) and the magnitude of FFT is used as the feature vector. To have a fair comparison with other transform methods like RCT, MCT, DWT and DCT, the order of the FFT was set such that the number of FFT features would be equal to the number of features obtained with the other methods. In this system, the transformation methods RCT, MCT, and DWT result in 12-14 features. So, the system used a 16 order FFT.

3.1.4 Discrete Cosine Transform (DCT):

The DCT is a method that decomposes a signal into a weighted sum of cosines. Let the given signal be $x[n]$, where n is an integer in the range 0 to N-1, then the forward DCT is:

$$C(k) = \sum_{n=0}^{N-1} x[n] \cos\left((\pi k / N)(2n+1)/2\right) \qquad (11)$$

where $k = 0,1,2,\ldots N\text{-}1$

The cosine transform can be calculated in $O(N \log_2 N)$, through an N-point FFT. The computational complexity of DCT and FFT are same, as the number of operations is the same. These transforms have high-energy compaction that helps in image compression techniques like JPEG. However, it is not clear whether or not this characteristic will be beneficial to feature extraction applications.

3.1.4.1 Utilization of DCT:

The DCT is used to extract features from the audio signal. The DCT can be computed for different orders. To have a fair comparison with other transform methods like RCT, MCT, DWT and DFT, the order of the DCT was set such that the number of DCT features would be equal to the number of features obtained with the other methods. In this system, the transformation methods RCT, MCT, and DWT result in 12-14 feature vectors. So, the system used a 16 order DCT.

3.1.5 Discrete Wavelet Transforms:

The DWT method decomposes the signal into a weighted sum of wavelet functions. The wavelet transform is the inner product of a set of wavelet basis functions

with an input signal. The mother wavelet is used to generate a set of wavelet functions. A set of wavelet basis functions, $\{\psi_{a,b}(t)\}$, can be generated by shifting and scaling the mother wavelet

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{12}$$

where $a > 0$ is the scaling factor and b is the translation variable, and both a and b are real numbers. Depending on the scaling factor a, the functions are dilated (if $a > 1$) or contracted (if $a < 1$). The coefficient, $\frac{1}{\sqrt{a}}$, is included to normalize the energy of the wavelets. The function must satisfy the admissibility condition to be called a mother wavelet. An important property of the wavelet system is the multiresolution analysis (MRA) property. This property makes the implementation easier and allows for development of a faster algorithm similar to fast Fourier transform.

Different types of mother wavelets and mother bases exist, like Haar, Coiflet, symlet, and Daubechies. Of these, the Haar wavelet is the simplest. For the dyadic DWT, the discrete wavelet basis functions are represented as

$$\psi_{j,k}(n) = 2^{\frac{-j}{2}}\psi\left(2^{-j}n - k\right), \tag{13}$$

and the wavelet coefficients are obtained by

$$W_{j,k} = \left\langle x[n], \psi_{j,k}(n)\right\rangle. \tag{14}$$

where *x[n]* is the signal in discrete time and $\psi_{j,k}(n)$ is the mother wavelet.

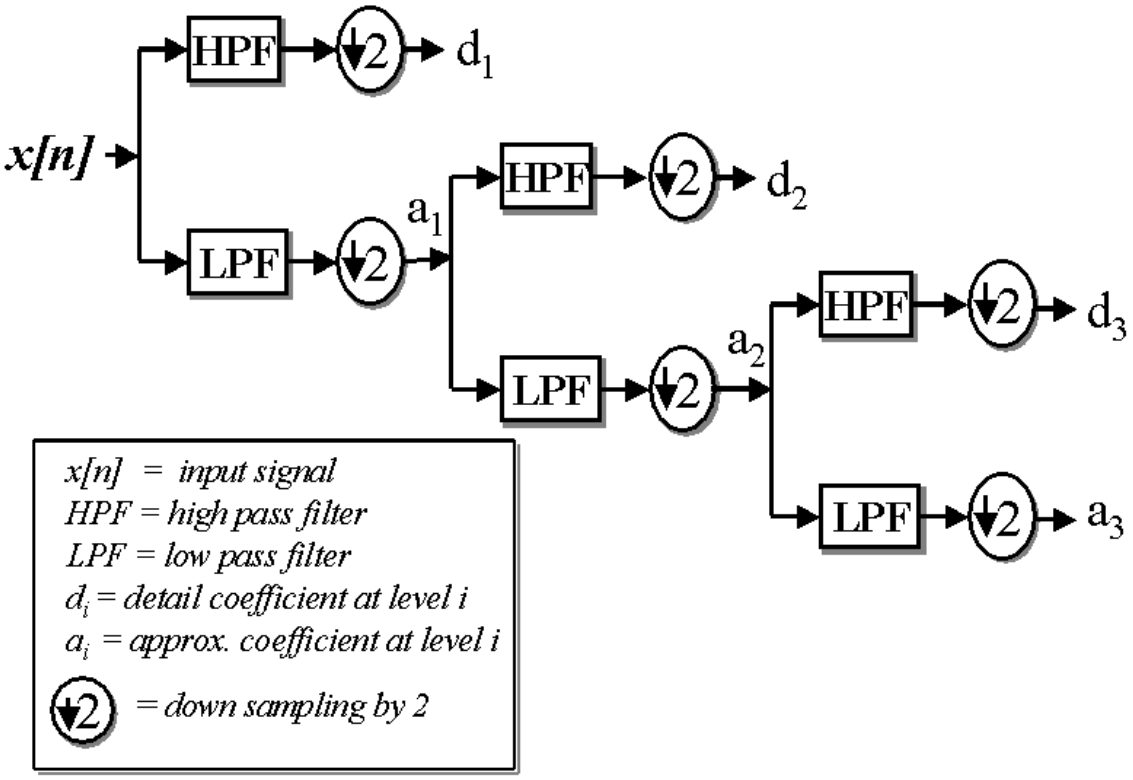


Figure 2 Dyadic filter tree implementation.

Many implementations of the DWT have been developed, and the type that is used most extensively is the dyadic filter tree. Depending on the mother wavelet, the high pass and the low pass filters are designed. At each stage of the filter tree, a set of approximation and detail coefficients are produced. Figure 6 shows the dyadic filter tree implementation. In this dyadic DWT, the scaling variables are powers of 2, and only dyadic shifts are used. In each shift, there will be an inner product of the wavelet and the input function. For example, at scale $a=2^{-1}$ there will be two shifts and hence, two inner products, and when the scale $a=2^{-3}$ there will be 8 inner products. The outputs of the inner product are called the wavelet coefficients. Thus, at scale $a=2^{-1}$ there will be two wavelet coefficients, and at scale $a=2^{-3}$ there will be 8 wavelet coefficients.

3.1.5.1 Utilization of DWT:

The output of the DWT is a set of wavelet coefficients, comprised of the detail coefficients, $D_j(i)$ at each scale and a set of approximation coefficients, $A_M(i)$. With these the root-mean-square (RMS) energy for the detail coefficients at each scale, ED_j, and the approximation coefficients at the final scale, EA_M, are calculated. A feature vector, $\vec{F}$, is formed

$$\vec{F} = [EA_M \quad ED_M \quad ED_{M-1} \quad \cdots \quad ED_1]^T \tag{15}$$

where the superscript T denotes a vector transpose,

$$EA_M = \frac{1}{P_M^A}\sqrt{\sum_{i=0}^{P_M^A-1}[A_M(i)]^2}\,, \tag{16}$$

and

$$ED_j = \frac{1}{P_j^D}\sqrt{\sum_{i=0}^{P_j^D-1}[D_j(i)]^2} \tag{17}$$

for $j = 1,2,3,\cdots,M$. Here M is the maximum wavelet decomposition level; P_j^D is the number of detail coefficients at level j; and P_M^A is the number of approximation coefficients at level M. The DWT feature extraction reduces the signal dimension to $M+1$. The value of M is dependent on the length of the input signal and the choice of mother wavelet. For the Haar mother wavelet, the value of M is calculated to be $M = \log_2(N)$, where N is the length of the original signal. These energy feature vectors are later used as an input to the statistical classification system.

3.1.6 Lifting Scheme:

The lifting scheme is a method used to construct wavelets. The main advantage of this method is that it does not use the Fourier transform *i.e.,* the wavelet construction is done in the spatial domain. It is not necessary to translate and dilate a mother wavelet and compute inner products with the input signal.

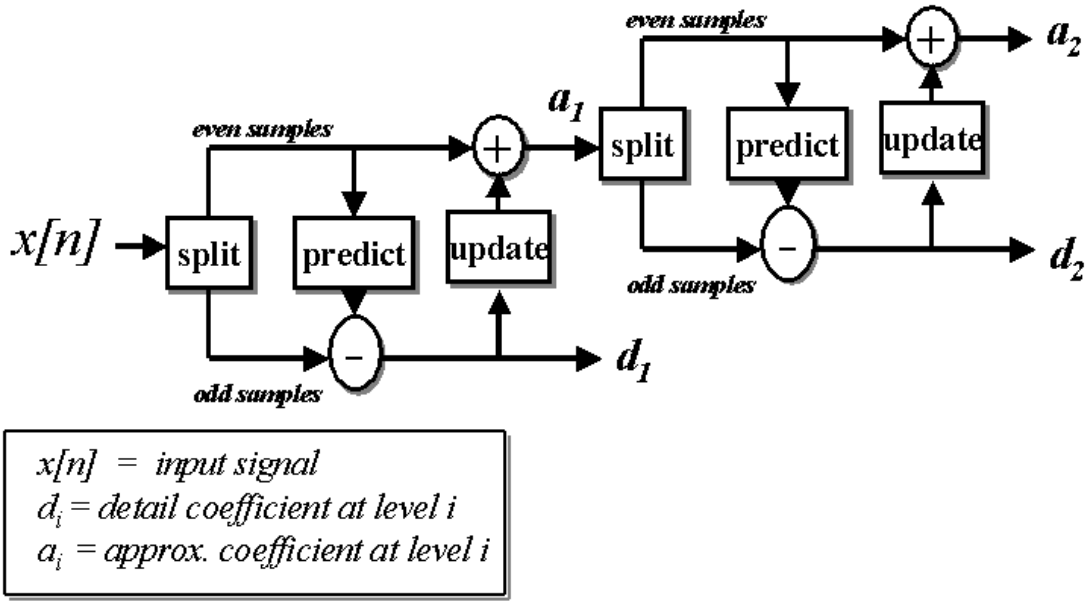


Figure 3 Lifting scheme implementation.

Construction of a first-order wavelet using the lifting scheme consists of three steps. First, the signal is split into even and odd samples. Second, the detail coefficient is found by subtracting even samples from odd ones (predict). Third, the detail coefficients obtained in the first step are added to the even samples and averaged giving the approximation coefficients (update). This process is repeated, through all the three steps for the approximation coefficients that were obtained in the third step, until there is one sample in the signal. Figure 7 shows the lifting scheme implementation

Implementation of this lifting scheme is faster than the dyadic filter tree for the DWT. The number of floating point operations is reduced by a factor of 2 [15]. The main advantage in implementation is that it does not require auxiliary memory.

3.1.6.1 Utilization of lifting scheme

In lifting scheme, the number of floating point operations is reduced by a factor of 2 [15]. The main idea of utilizing lifting scheme is to increase the computational speed, as the number of floating point operations is reduced by a factor of 2 as compared to discrete wavelet transform.

3.2 Feature reduction method

Feature reduction is the process of projecting *N*-dimensional feature vectors onto an *n*-dimensional space, where *n* is less than *N*. These *n*-dimensional feature vectors should maximize the class separation. The feature reduction process may be linear or non-linear. The reduced features can be obtained using supervised or unsupervised methods. A supervised method uses a data set where the classification of the signals is known to train the method. Fisher's linear discriminant analysis (LDA) is an example of a supervised method. LDA is a transformation method that yields the transformation matrix *W* that can maximize the between-class separation and minimize the within-class variability [16].

A within-class scatter matrix is defined as

$$S_W = \sum_{i=1}^{c} \sum_{x \in C_i} (\vec{F} - m_i)(\vec{F} - m_i)^t \tag{18}$$

where c is the number of classes, C_i is the data set that belongs to the i^{th} class, and m_i is the mean of the i^{th} class. The above equation of within-class scatter matrix is the summation of the covariance matrices of all the classes.

The between-class variance can be defined as

$$S_B = \sum_{i=1}^{c} n_i (m_i - m)(m_i - m)^t \tag{19}$$

where n_i is the input feature vector of the i^{th} classes, m_i is the mean of the i^{th} class and m is the total mean vector.

To maximize the between-class separation and minimize the within-class variability, transformation matrix, *W,* should maximize the following criterion,

$$J(W) = \frac{\left| W^t S_B W \right|}{\left| W^t S_W W \right|} \tag{20}$$

The matrix W that maximizes the above criterion must also satisfy

$$S_B W = \lambda S_W W \tag{21}$$

for some constant λ. This is a generalized eigenvalue problem. The eigen decomposition of the below equation results in the transformation matrix W

$$S_W^{-1} S_B W = \lambda W \tag{22}$$

For the case, when the within-class scatter matrix is singular, the transformation method fails. In these cases, other transformation methods like the Karhunen-Loeve transform can be adopted [16].

3.2.1 Utilization of LDA:

The input to the LDA is a set of feature vectors, $\vec{F}$. The output is an optimum linear combination weight matrix W, so as to maximize the between-class separation and minimize the within-class variability. With the weight matrix W, the reduced feature vector, $\vec{F}_r$, can be computed as $\vec{F}_r = W^T \cdot \vec{F}$. When the within-class scatter matrix S_w is singular, Karhunen-Loeve transformation method is used as an alternate to LDA in this system.

3.3 Classification methods

In this system three types of statistical classifiers are investigated: nearest mean, nearest neighbor, and maximum likelihood. These three are supervised classifiers, as they are trained using data whose classifications are known. For this thesis, all three classifiers are implemented, and the best out of the three is selected based on the overall accuracy.

3.3.1 Nearest mean classifier:

The nearest mean classifier is a parametric classifier, as it requires the first order statistics of the training data *i.e.,* individual means of all the different classes in the database like crash or non-crash. The mean of each class is computed, and the test data is compared with the computed class means. Comparison is based on the Euclidian distance between the test data and each of the class means. The test data is determined to be in the class where the distance between the class mean and the test data is minimum.

3.3.2 Nearest neighbor classifier:

The nearest neighbor classifier is a non-parametric classifier, as it does not require statistics of the training data's distribution in the feature space. The Euclidian distance is computed between the test data and each signal in the training data set. The test data is assigned to the class of the nearest signal. The main advantage of this classifier is that when there are a number of outliers in the feature space, the nearest neighbor method classifies the test data more accurately. Unfortunately, this method requires more memory as it needs to store all the training data to compare with the test data.

3.3.3 Maximum likelihood classifier:

The maximum likelihood classifier is a parametric classifier. It requires second order statistics of the training data *i.e.,* the individual class means and the variances. Based on the class statistics of the training data, boundaries for all the classes are formed. Then, the test data is compared with all the class boundaries and classified based on the boundary within which it falls.

3.4 Testing methods:

The leave-one-out testing method is used to evaluate each of the systems investigated. When the amount of training and testing data is very limited, the leave-one-out testing method is the best approach to evaluate a supervised system. This testing method leaves one signal under investigation as the testing data and the rest of the database as the training data, thereby maximizing the training data set. The testing data, a single signal, is classified using the rest of the samples in the database whose classifications are known. This testing is done for each signal in the whole database, thus giving unbiased classification results, as both the training and the testing data are mutually exclusive. The disadvantage of this method is the computational complexity.

In the automated detection system, features are extracted from the signals in the database whose classifications are known using one of the transform methods and leaving one signal to be investigated. Using the extracted features of the training data set, LDA is used to determine the optimal transformation matrix W that will be used to reduce the dimensionality of the test feature vector. A statistical classifier like maximum likelihood is used to compare the test signal with the training data set. The classification accuracy is determined by referring to the truth of the test signal. The whole process is repeated for each signal in the whole database. The classification accuracy is obtained by determining the percentage of signals correctly classified.

3.5 Data Collection /Processing

3.5.1 Data collection:

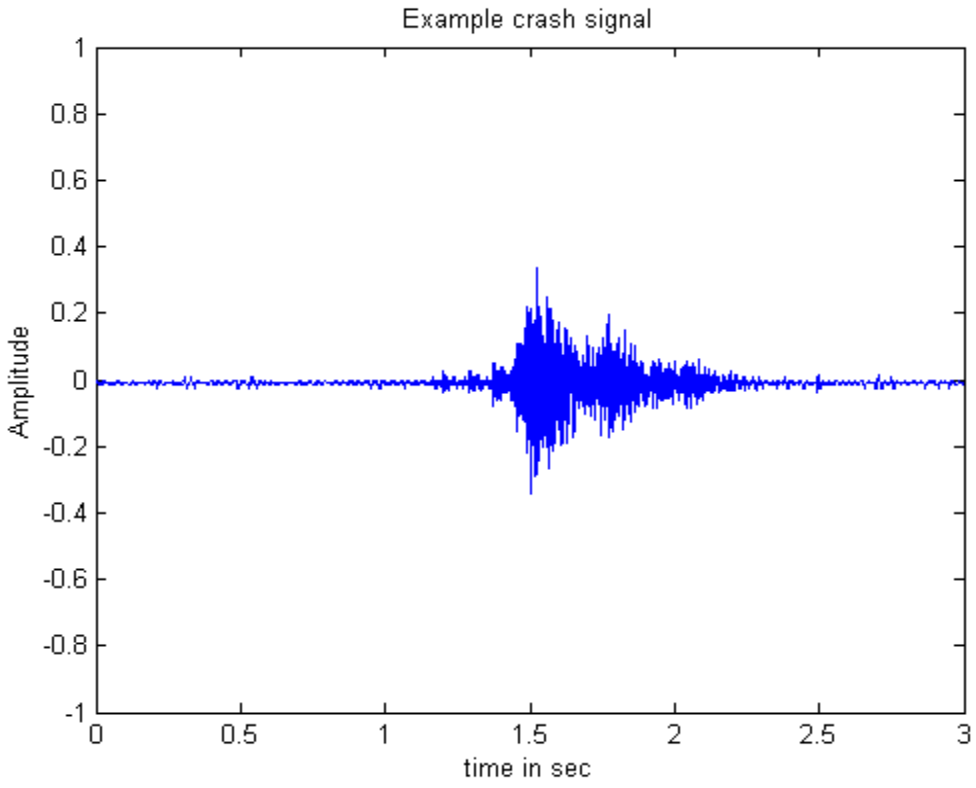
To test the feature extraction methods and the overall accident detection systems, a database containing a variety of traffic sound signals needed to be collected. These testing data needed to have accident sound signals as well as normal traffic sound signals. A device to record the signal at an intersection needed to be chosen considering the installation, maintenance, and the advantages and disadvantages. A Sony TCD-D8 DAT Walkman, which was a digital audio tape (DAT) recorder, was used to record signals. Using the DAT recorder, sound signals were recorded with a sampling frequency of 44.1 KHz. The DAT recorder output was a digitized signal in ".wav" format.

Video and audio recording of various types of traffic incidents were obtained from the Kentucky Transportation Cabinet. Figures 2, 3, 4, and 5 show a few samples of these images and audio plots.

The Mississippi Department of Transportation (MDOT) recommended intersections with high traffic flow and a high probability of an accident occurring. Traffic signals were collected from Jackson, MS and Starkville, MS under various traffic conditions, so that the data collected would have sounds from trucks, cars, motorcycles and buses as well as brake and horn sounds. In addition, Dr. Charles Harlow from Louisiana State University provided recorded signals of traffic accidents obtained from the Texas Transportation Institute (TTI) crash test facility. He also gave some different traffic sounds that are unique and difficult to collect like pile drive sounds, severe brake sounds, and other construction sounds. Since the crash sounds obtained from the crash

test facility were not recorded in actual traffic scenarios, the signals were used to synthesize realistic traffic accident sounds recorded in an intersection.

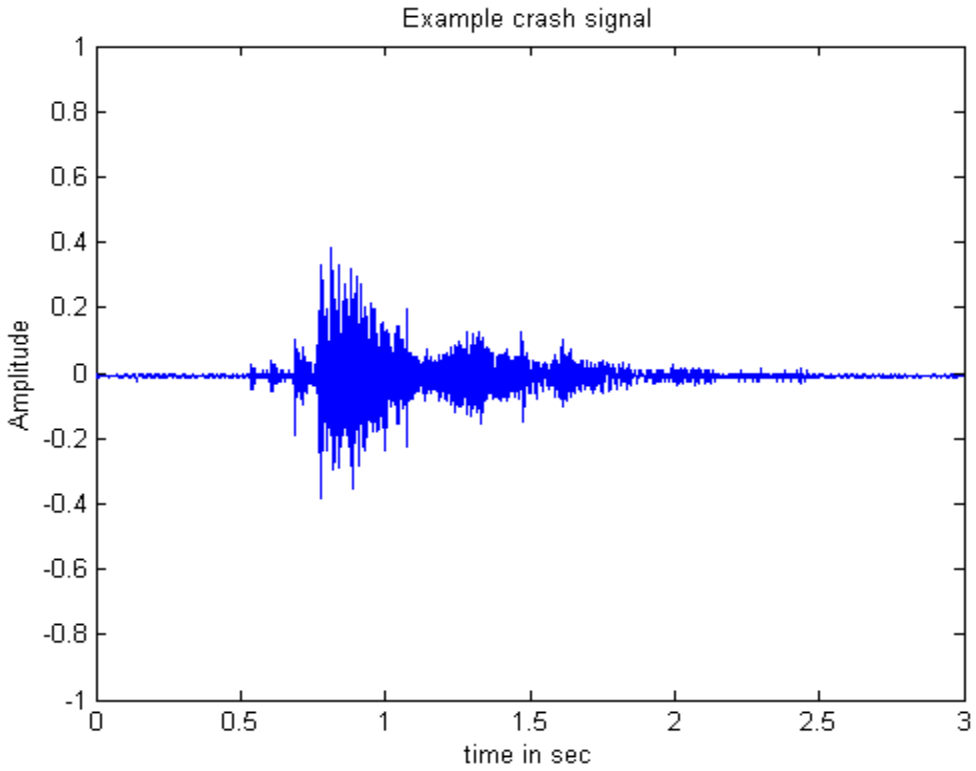
(a)



(b)

Figure 4 (a) Digital Image of crash incident at an intersection in Louisville, Kentucky, (b) Digital audio signal plot of the crash sound

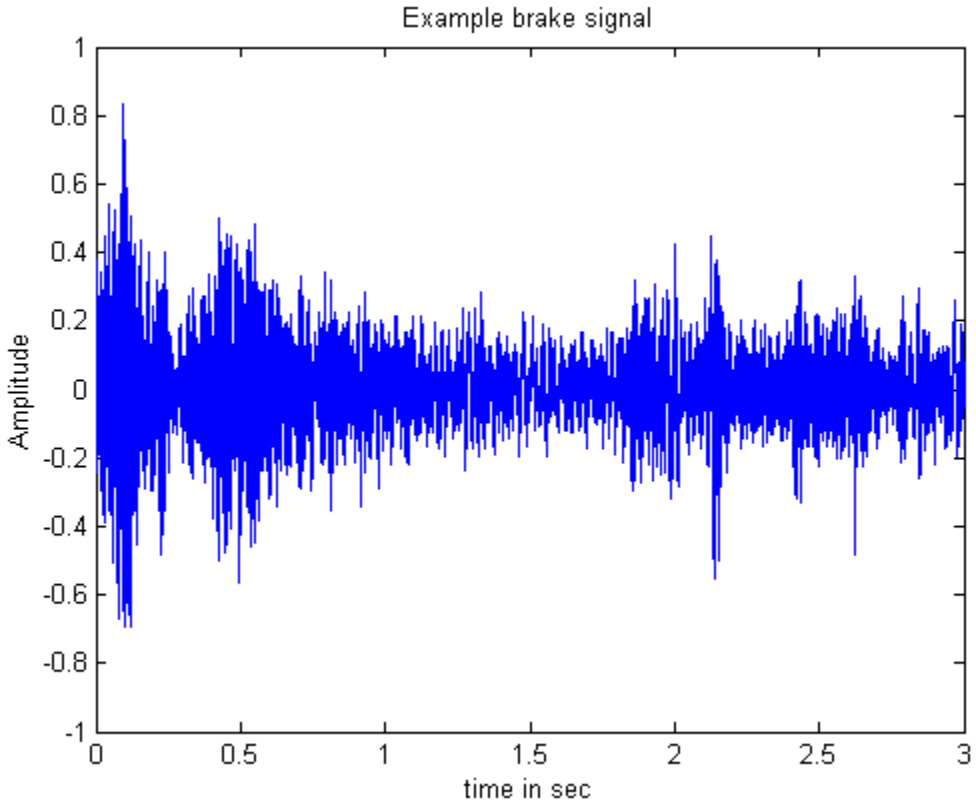
(a)



(b).

Figure 5 (a) Digital Image of Crash incident at an intersection in Louisville, Kentucky, (b) Digital Audio signal plot of the crash sound

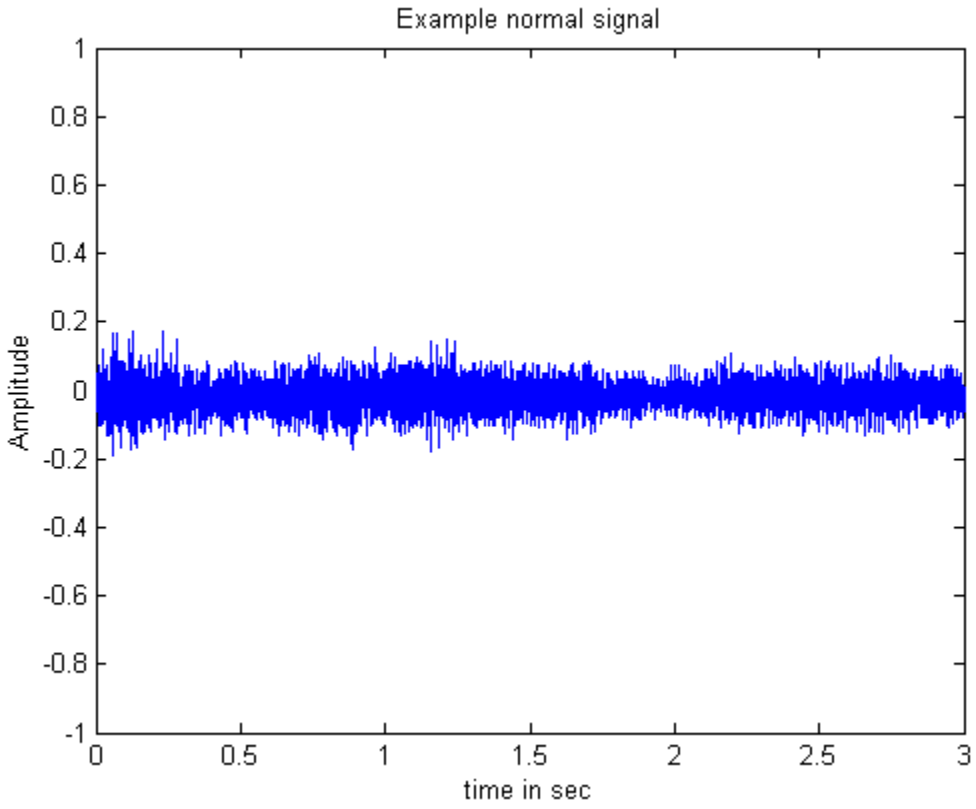
(a)



(b)

Figure 6 (a) Digital Image of brake incident at an intersection in Louisville, Kentucky, (b) Digital Audio signal plot of the brake sound.

(a)



(b)

Figure 7 (a) Digital Image of normal traffic incident at an intersection in Louisville, Kentucky, (b) Digital Audio signal plot of normal sound.

3.5.2 Data Preprocessing:

The sound signals were collected using the DAT recorder with a 44.1 KHz sampling frequency, 16-bit resolution, and stereo-channel mode. These were down sampled to 22.05 KHz with 8-bit resolution and mono-channel so as to make all the signals have the same format. Each signal in the database was windowed to have a three-second duration, which made every signal have 66176 samples. The volume level in terms of amplitude of all the signals would vary widely. In order for the volume level to be consistent in all the signals, the signals were normalized to have a maximum amplitude of one. The output of the DAT recorder was in ".wav" format, which could be stored in the computer using the sound card. Furthermore, Matlab, the software used in developing and testing the algorithm could read signals in ".wav" format. A database containing pile drive, brake, crash and normal traffic sounds were created. A total of 99 signals were used to test the system, the reason being that Matlab can allow a maximum of only 99 signals with the above-said format because of the file size. Since the data collected was more than required, various databases were created and tested.

3.5.3 Synthesized crash data:

The crash sound obtained from the TTI crash test facility varies from a crash sound recorded from an intersection due to the background traffic sounds at the intersections. In order to create a sound similar to a normal crash sound at an intersection, the pure crash sound obtained from the crash test facility is mixed with the normal background traffic sound. Varying weights are given to both the crash and background sound signals to create the synthetic intersection accident signal. Assume $f_{nc}(n)$ is the

recorded non-crash signal, $f_c(n)$ is the recorded crash signal, and α is the weighting variable. The new synthesized crash signal, $\tilde{f}_c(n)$, was computed as

$$\tilde{f}_C(n) = f_{nC}\ (n) + \alpha \cdot f_C(n) \tag{23}$$

If the variable α is high, then the crash sound will have a higher volume meaning it occurred very near the microphone at the intersection. Similarly, when the weighting variable α is low, then the crash has occurred somewhere far from the intersection. By varying α, the signal-to-noise-ratio (SNR) ranged from –50 decibels (dB) to +50dB. When computing the SNR of $\tilde{f}_c(n)$, the "noise" component is $f_{nc}(n)$ and the "signal" component is $f_c(n)$. When the SNR was –50dB, the crash signal was very low in amplitude as compared to the non-crash audio signals (crash occurs far away from the intersection). When the SNR was 50dB, the crash audio signal was very high in amplitude as compared to the non-crash audio signals (crash occurred at the intersection). When the SNR was 0dB, the crash signals and non-crash signals had the same volume.

3.6 Performance analysis

Performance of the system is analyzed based on the accuracy assessment and computational expense assessment. The accuracy assessment is based on the overall percentage classification accuracy. Overall percentage accuracy is calculated by dividing the total number of signals that are correctly classified by the total number of signals tested.

Computational expense assessment is based on the computational time required by the system to classify a single signal. Computational expense assessment is also based on the number of multiplies and adds required for a signal during the feature extraction method; the other processes like data preprocessing, feature optimization, and classification methods remain the same for all the systems. In the system analysis, the computational expense assessment is made for the DWT and lifting scheme based on the order of complexity. For the Discrete wavelet transform method the order of complexity is given as O(N. log N). The lifting scheme method is analyzed based on the additions and subtractions for a signal.

CHAPTER IV

RESULTS

In this thesis, testing results are analyzed based on accuracy and computational expense. Accuracy assessment is based on the results obtained using different feature extraction methods, LDA for feature reduction, and statistical classifiers with leave-one-out testing. Computational assessment is based on the computation time required to perform the feature extraction and classification.

4.1 Accuracy assessment:

Accuracy comparison will help to decide on the optimum feature extraction method and classifier; five different types of feature extraction methods are investigated and tested. Feature extraction methods investigated are DWT, DCT, RCT, MCT, and FFT. While using DWT as the feature extraction method various mother wavelets like Haar, Daubechies4, Daubechies15, Coiflet2, Coiflets5, Symlets2, Symlets8 are investigated. Statistical classifiers like maximum likelihood, nearest mean, and nearest neighbor are also investigated. Two types of classification modes are studied: two-class and multi-class. A two-class system labels each input signal as either crash or non-crash, and multi-class systems labels them as either crash or as several non-crash events like brake, pile drive, and other construction sounds.

4.1.1 Comparison of mother wavelets:

The feature extraction method with wavelet was investigated with different classifier methods and leave-one-out testing method. Mother wavelets like Haar, Daubechies, symlets, and coiflets are used for analysis. Depending on the accuracies the best mother wavelet is selected.

Tables 1 and 2 show the maximum likelihood classifier results for a two-class system (crash or non-crash) and a multi-class system (crash sounds, normal traffic sounds, and abnormal traffic sounds includes pile drive, brake and horn sounds.), respectively. The signals are normalized such that maximum amplitude of the signal is one.

In choosing the best mother wavelet, overall accuracy as well as the sensitivity and specificity are considered. Sensitivity is the proportion of crash signals correctly classified as crash. Specificity is the proportion of non-crash signals correctly classified as non-crash. The system should have high sensitivity, as the misses of the crash signals would make the system a failure. The specificity of the system needs to be tolerable, as the misclassification of many non-crash signals as crash would also make the system a failure.

Table 1 shows that a system with Haar, Daubechies4, or coiflets5 as the mother wavelet performs the best with high sensitivity. In the case of Table 2, in a multi-class system Haar, symlet2 and symlet8 seem to outperform the rest of the other mother

wavelets, the system has high specificity and the probability of crash “misses” is low when compared to other mother wavelets. As Haar seems to be the common mother wavelet that works better for two-classes and multi-class systems, Haar is the better wavelet that can be chosen to extract features using the DWT method.

TABLE 1 MAXIMUM LIKELIHOOD CLASSIFICATION ACCURACIES FOR TWO-CLASS SYSTEM USING DWT-BASED FEATURES

Wavelet	Crash	Non-crash	Overall Accuracy	Confidence Interval(95%)
Haar	1.0000	1.0000	1.0000	0.0000
Daubechies4	1.0000	1.0000	1.0000	0.0000
Daubechies15	1.0000	0.9722	0.9798	0.0232
Coif lets2	0.9259	1.0000	0.9798	0.0232
Coif lets5	1.0000	1.0000	1.0000	0.0000
Symlets2	0.964	0.9861	0.9798	0.0232
Symlets8	0.9259	1.0000	0.9794	0.0237

TABLE 2 MAXIMUM LIKELIHOOD CLASSIFICATION ACCURACIES FOR MULTI-CLASS SYSTEM USING DWT-BASED FEATURES

Wavelet	Crash	Normal-traffic	Abnormal-traffic	Overall Accuracy	Confidence Interval(95%)
Haar	0.963	1	0.5	0.9394	0.0393
Daubechies4	0.9259	1	0.8	0.9697	0.0283
Daubechies15	0.9259	0.9839	0.8	0.9495	0.0361
Coif lets2	0.9259	1	0.9	0.9697	0.0283
Coif lets5	0.9259	1	0.9	0.9697	0.0283
Symlets2	0.963	1	0.8	0.9697	0.0283
Symlets8	0.963	0.9839	0.875	0.9691	0.0288

Tables 3 and 4 show the nearest neighbor classifier results for a two-class system (crash or non-crash) and a multi-class system (crash sounds, normal traffic sounds, and abnormal traffic sounds includes pile drive, brake and horn sounds.), respectively. The signals are normalized such that maximum amplitude of the signal is one.

Table 3 shows that a system with all the mother wavelets except symlets2 performs the best with high sensitivity, but on the basis of overall accuracy Haar, Daubechies4, Coiflets2, and symlets8 are the best. In the case of Table 4, in a multi-class

system Daubechies4, Coiflets5, and symlet8 seem to outperform the rest of the other mother wavelets, the system has high sensitivity *i.e.* relative to the other mother wavelets.

TABLE 3 NEAREST NEIGHBOR CLASSIFICATION ACCURACIES FOR TWO-CLASS SYSTEM USING DWT-BASED FEATURES

Wavelet	Crash	Non-crash	Overall Accuracy	Confidence Interval(95%)
Haar	1	1	1	0
Daubechies4	1.0000	1	1	0
Daubechies15	1.0000	0.9722	0.9798	0.0232
Coif lets2	1.0000	1	1	0
Coif lets5	1	0.9861	0.9899	0.0165
Symlets2	0.9630	0.9861	0.9798	0.0232
Symlets8	1	1	1	0

TABLE 4 NEAREST NEIGHBOR CLASSIFICATION ACCURACIES FOR MULTI-CLASS SYSTEM USING DWT-BASED FEATURES

Wavelet	Crash	Normal-traffic	Abnormal-traffic	Overall Accuracy	Confidence Interval(95%)
Haar	0.9630	0.9839	0.9000	0.9697	0.0283
Daubechies4	1.0000	0.9839	0.9000	0.9798	0.0232
Daubechies15	0.9259	0.9839	0.8000	0.9495	0.0361
Coif lets2	0.9630	0.9839	0.9000	0.9697	0.0283
Coif lets5	1.0000	0.9839	0.9000	0.9798	0.0232
Symlets2	0.9259	1	0.9000	0.9697	0.0283
Symlets8	1.0000	0.9839	1	0.9899	0.0165

Tables 5 and 6 show the nearest mean classifier results for a two-class system (crash or non-crash) and a multi-class system (crash sounds, normal traffic sounds, and abnormal traffic sounds includes pile drive, brake and horn sounds.), respectively. The signals are normalized such that maximum amplitude of the signal is one.

Table 5 shows that a system with all the mother wavelets performs the best with high sensitivity, but on the basis of overall accuracy coiflets2, coiflets5, and symlets8 are the best. In the case of Table 6, in a multi-class system Haar, Daubechies4, Coiflets5, and symlet8 seem to outperform the rest of the other mother wavelets, the system has high sensitivity *i.e.,* probability of crash "misses" is low when compared to other mother wavelets.

TABLE 5 NEAREST MEAN CLASSIFICATION ACCURACIES FOR TWO-CLASS SYSTEM USING DWT-BASED FEATURES

Wavelet	Crash	Non-crash	Overall Accuracy	Confidence Interval(95%)
Haar	1.0000	0.9583	0.9697	0.0283
Daubechies4	1.0000	1	1	0
Daubechies15	1.0000	0.9167	0.9394	0.0393
Coif lets2	1.0000	0.9861	0.9899	0.0165
Coif lets5	1	0.9861	0.9899	0.0165
Symlets2	1	0.9583	0.9697	0.0283
Symlets8	1	0.9861	0.9899	0.0165

TABLE 6 NEAREST MEAN CLASSIFICATION ACCURACIES FOR MULTI-CLASS SYSTEM USING DWT-BASED FEATURES

Wavelet	Crash	Normal-traffic	Abnormal-traffic	Overall Accuracy	Confidence Interval(95%)
Haar	1	0.9355	0.8000	0.9394	0.0393
Daubechies4	1.0000	0.9032	1.0000	0.9394	0.0393
Daubechies15	0.963	0.9355	1.0000	0.9495	0.0361
Coif lets2	0.963	0.9194	1.0000	0.9394	0.0393
Coif lets5	1	0.9194	0.9	0.9394	0.393
Symlets2	0.963	0.9355	0.800	0.9293	0.0423
Symlets8	1	0.9032	0.9	0.9293	0.0423

In summary, three different classifier methods were analyzed with different mother wavelets. Out of these different methods and wavelets, the optimum classifier

method and a mother wavelet is selected based on the sensitivity and the implementation cost. Haar wavelet along with the maximum likelihood classifier is chosen as the best.

4.1.2 Comparison of transform-based feature extraction methods:

In order to investigate which transform method will give a higher accuracy, all the five different methods (DWT (Haar), FFT, DCT, RCT, and MCT) are investigated by using leave-one-out testing method with LDA as the feature optimization method. Table 7 shows the overall classification accuracies for both the two-class and multi-class systems. Here, the data is normalized so that the signal has a maximum amplitude of one. It shows that out of these methods, the DWT, RCT and MCT perform the best, giving accuracies greater than 98%. DWT and RCT give as much as 100% accurate results for a two-class system. However, when comparing and considering the time taken to classify the signals under analysis, the DWT with Haar as the mother wavelet is superior. Even in the case of a multi-class system the DWT, RCT and MCT perform the best, providing classification accuracies >= 94%. However, similar to the two-class system when comparing the processing times, the DWT with Haar as the mother wavelet is the best.

Figure 2 shows the classification results for the methods for both the two-class and the multi-class systems. It shows that the two-class system outperforms the multi-class system in all the five different methods. This is due to the fact that all the non-crash events are combined into a single class, thereby reducing the possibility of misclassifications.

TABLE 7 MAXIMUM LIKELIHOOD CLASSIFICATION ACCURACIES FOR TWO-CLASS AND MULTI-CLASS SYSTEMS

Feature Extractor	Two-class	Multi-class
DWT	1	0.9394
RCT	1	0.9897
MCT	0.9899	0.9596
FFT	0.9091	0.8788
DCT	0.9394	0.8485

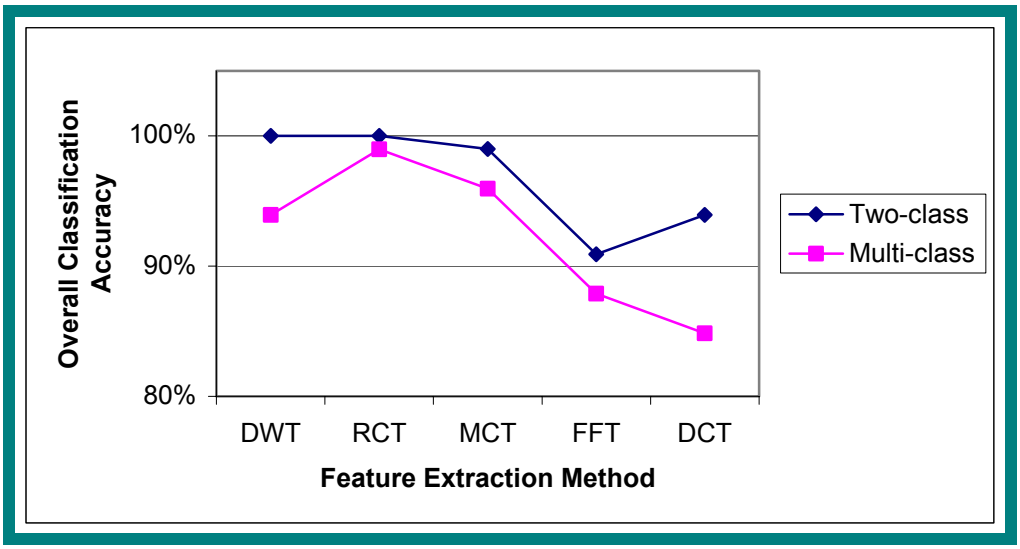


Figure 8 Maximum likelihood classification accuracies for two-class and multi-class systems

4.1.3 Comparison of statistical classifiers and system sensitivities:

A comparison based on the overall accuracies is done to choose a better classifier method. In order for that DWT (Haar) feature extraction method along with the leave-one-out testing method and LDA is used. For each of the three types of statistical classifiers, the two-class system and the multi-class system results are shown in Tables 8 and 9 and Figure 9 and 10. The audio signals used for analysis are manipulated to model various ambient noise conditions. The SNR's (signal-to-noise ratio) were varied between

the ranges of –50dB to 50dB. The –50dB case simulates a scenario where the crash signal is very low in amplitude as compared to the non-crash signals. For example, the crash occurred far from the sensor, so the local normal traffic sounds dominate. The 50dB case simulates a scenario where the crash signal is very high in amplitude as compared to the non-crash signals. For example, the crash occurred very near the sensor, so the crash sounds dominate over the local normal traffic sounds. The 0dB case simulates a scenario where the crash signals and non-crash signals have the same amplitude. The results show that the classification accuracies decrease with decreasing SNR. This is due to the fact that with decreasing SNR, the crash audio signal is becoming more and more like a non-crash audio signal. Note that of the three classifiers, irrespective of SNR value, the nearest neighbor and the maximum likelihood classifier performs best. However, when comparing the computational cost, and the system sensitivity maximum likelihood classifier is considered to be superior.

TABLE 8OVERALL ACCURACY OF CLASSIFICATION WITH DWT
FOR THE TWO-CLASS SYSTEMS

SNR	Nearest Mean	Maximum Likelihood	Nearest Neighbor
50db	0.9697	1	1
20 db	0.9697	0.9899	1
10 db	0.9697	0.9899	1
0 db	0.9596	0.9899	0.9899
-10 db	0.9596	0.9495	0.9596
-20 db	0.7071	0.7273	0.7273
-50 db	0.5455	0.5556	0.4545

TABLE 9 OVERALL ACCURACY OF CLASSIFICATION WITH DWT
FOR THE MULTI-CLASS SYSTEMS

SNR	Nearest Mean	Maximum Likelihood	Nearest Neighbor
50db	0.9394	0.9394	0.9697
20 db	0.9293	0.9394	0.9697
10 db	0.9293	0.9293	0.9697
0 db	0.9293	0.9293	0.9697
-10 db	0.899	0.9091	0.9596
-20 db	0.697	0.7475	0.7576
-50 db	0.5354	0.5455	0.4545

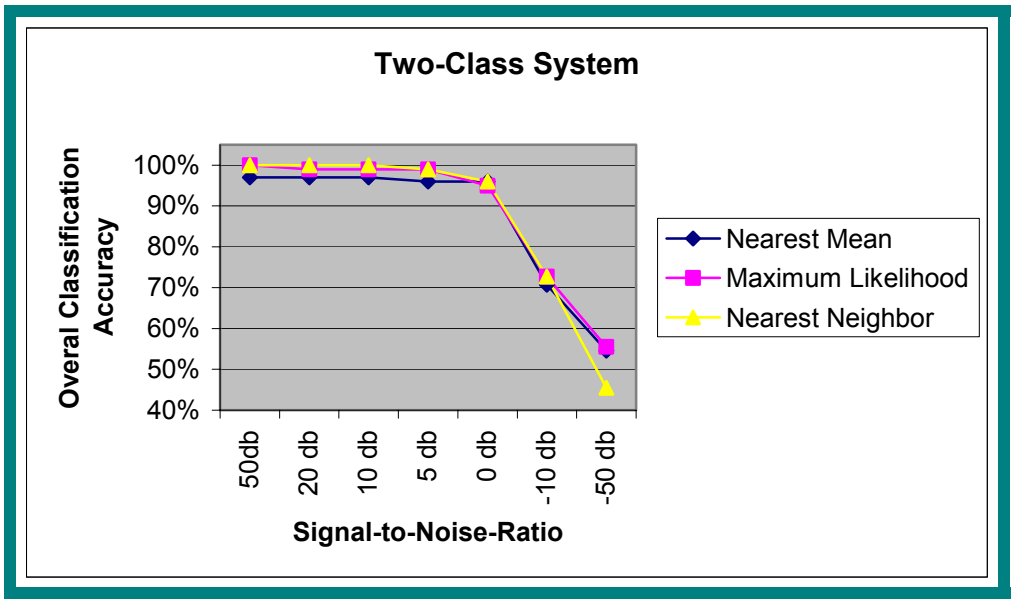


Figure 9 DWT-based feature extraction using Haar mother wavelet for two-class system.

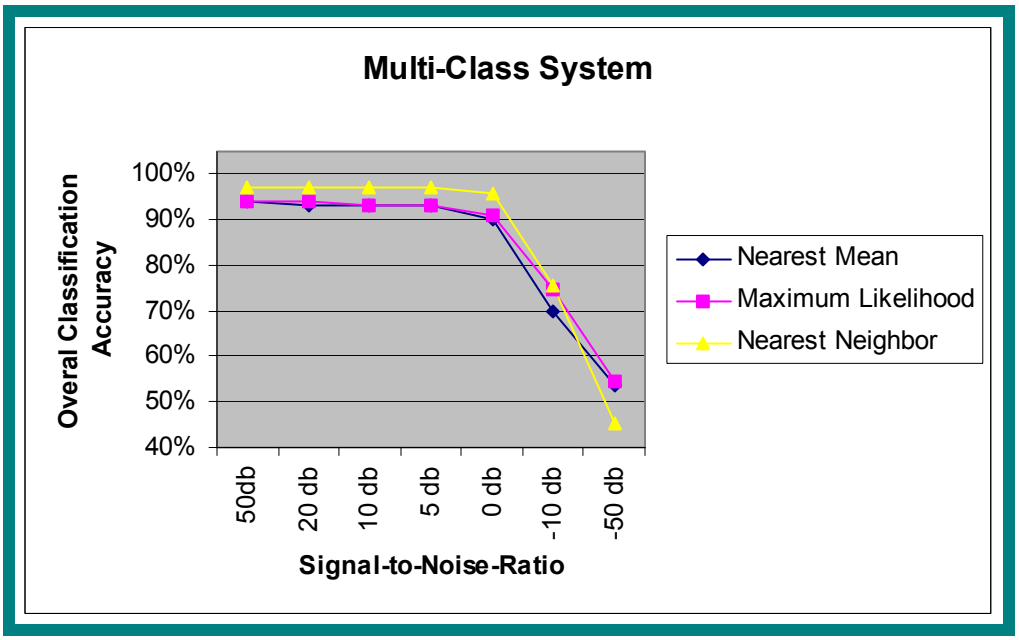


Figure 10 DWT-based feature extraction using Haar mother wavelet for multi-class system.

4.1.4 Comparison of feature extraction methods and system sensitivities:

Table 10 shows maximum likelihood classification accuracies for various feature extraction methods for a two-class system. The signals are manipulated to model various ambient noise conditions such that the signal-to-noise-ratio varies from –50 dB to +50 dB. At 0 dB, the accuracy ranges from 94% to 98% for the case of DWT (Haar), and RCT. MCT, FFT and DCT did not perform well when compared to the DWT and RCT. When the SNR is 5 dB and above, DWT (Haar), and RCT give an overall accuracy of 99% and above.

Table 11 shows the results for a multi-class system. When the SNR is –50 dB to +50 dB similar to the two-class system. For the case of the RCT method, when the SNR is 0 dB and above it gives an overall accuracy of 99%. The MCT also almost performs similar to the RCT method. DWT (Haar) has an overall accuracy of 93% when the SNR is 5 dB and above.

Figures 11 and 12 show that RCT performs the best out of all the five different methods. In both two-class and multi-class systems, when the SNR is 0 dB and above RCT gives an overall accuracy of 99%. After RCT, MCT performs the best in the two-class system as well as in the multi-class system. However, the DWT (Haar) seems to be almost equal in performance when compared to the RCT and the MCT methods. When the SNR is above 5 dB, DWT (Haar) gives an overall accuracy of 99%. Considering the computational expense, DWT (Haar) is selected as the optimum method.

TABLE 10 MAXIMUM LIKELIHOOD CLASSIFICATION ACCURACIES FOR VARIOUS FEATURE EXTRACTION METHODS FOR THE TWO-CLASS SYSTEM

SNR	DWT (Haar)	FFT	DCT	RCT	MCT
50db	1	0.9293	0.9394	1	1
10 db	0.9899	0.899	0.9091	1	0.9899
5 db	0.9899	0.8687	0.899	1	0.9495
0 db	0.9495	0.8384	0.8182	0.9899	0.8384
-10 db	0.7273	0.7273	0.7071	0.8384	0.6364
-50 db	0.5556	0.6566	0.697	0.4949	0.4242

TABLE 11 MAXIMUM LIKELIHOOD CLASSIFICATION ACCURACIES FOR VARIOUS FEATURE EXTRACTION METHODS FOR THE MULTI-CLASS SYSTEM

SNR	DWT (Haar)	FFT	DCT	RCT	MCT
50db	0.9394	0.8788	0.798	0.9899	0.9798
10 db	0.9394	0.8687	0.7475	0.9899	0.9576
5 db	0.9293	0.8687	0.7273	0.9899	0.899
0 db	0.9091	0.7576	0.6263	0.9899	0.8081
-10 db	0.7475	0.6465	0.5859	0.8485	0.5758
-50 db	0.5455	0.5859	0.404	0.4949	0.4343

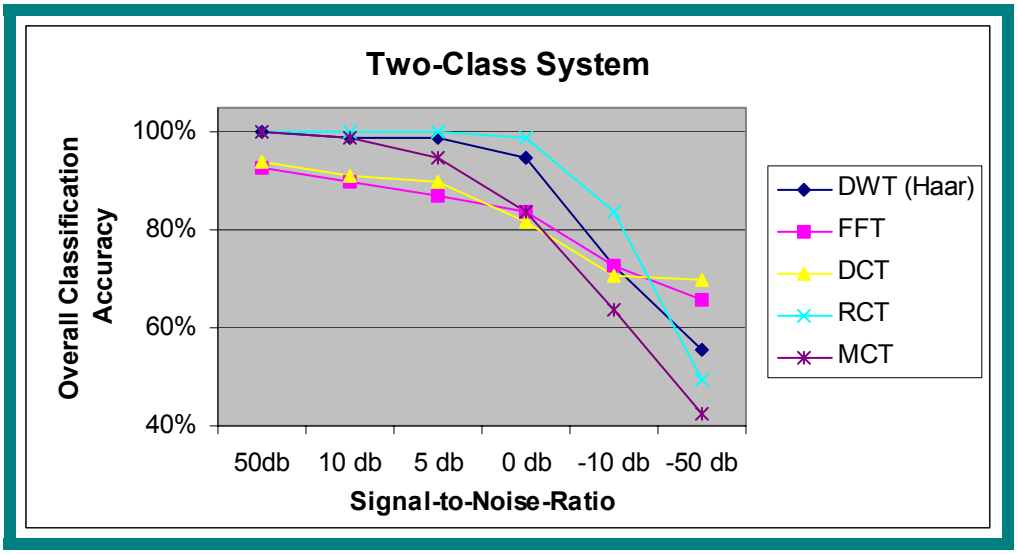


Figure 11. Maximum likelihood classification accuracies for various feature extraction methods for the two-class system.

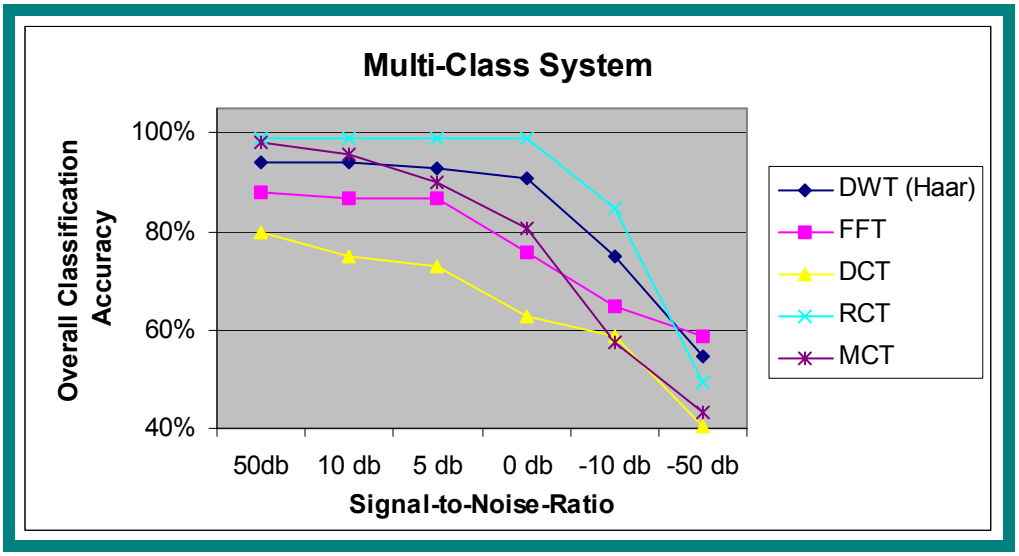


Figure 12 Maximum likelihood classification accuracies for various feature extraction methods for the multi-class system.

4.1.5 Results based on Jackson Data:

Based on the above results, DWT method using the Haar mother wavelet and the maximum likelihood classifier was selected for the system design. This optimal combination is tested using data collected from various intersections in Jackson, MS. Table 12 and Table 13 show the two-class and multi-class classification accuracies. From Tables 12 and 13, we can see that the classification accuracies are about 95% when SNR is greater than 0dB.

TABLE 12 CLASSIFICATION RESULTS WITH DWT AND MAXIMUM LIKELIHOOD CLASSIFICATION FOR THE TWO-CLASS SYSTEMS.

SNR	Crash	Non-crash	Overall Accuracy	Confidence Interval (95%)
50db	0.9630	0.9444	0.9495	0.0361
20 db	0.9630	0.9444	0.9495	0.0361
10 db	0.9630	0.9444	0.9495	0.0361
0 db	0.9630	0.9167	0.9293	0.0423
-10 db	0.7770	0.8750	0.8485	0.0591
-20 db	0.3304	0.7500	0.6468	0.0788
-50 db	0.2963	0.6528	0.5586	0.0819

TABLE 13 CLASSIFICATION RESULTS WITH DWT AND MAXIMUM LIKELIHOOD CLASSIFICATION FOR THE MULTI-CLASS SYSTEMS

SNR	Crash	Normal-traffic	Abnormal-traffic	Overall Accuracy	Confidence Interval(95%)
50db	0.0963	1.0000	0.6000	0.9495	0.0361
20 db	0.0963	1.0000	0.6000	0.9495	0.0361
10 db	0.0963	1.0000	0.6000	0.9495	0.0361
0 db	0.8519	1.0000	0.7000	0.8788	0.0538
-10 db	0.6296	0.9576	0.8000	0.8485	0.0591
-20 db	0.2222	0.8387	0.6000	0.6465	0.0788
-50 db	0.0370	0.7581	0.6000	0.5455	0.0821

4.1.6 Results of lifting scheme feature extraction:

Another approach where in the processing time is a prime concern was investigated. The feature extraction method used in this method is the first-order lifting scheme. Table 14 and Table 15 show the two-class and multi-class classification accuracies, respectively. LDA is used for feature reduction and the maximum likelihood method is used for the classifier. Table 14 shows that for the two-class system, the classification accuracies are about 100% when SNR is >= 5dB, and Table 15 shows that above 0 dB the accuracy is above 93% for a multi-class system.

TABLE 14 CLASSIFICATION RESULTS WITH LIFTING SCHEME AND MAXIMUM LIKELIHOOD CLASSIFICATION FOR THE TWO-CLASS SYSTEMS

SNR	Crash	Non-crash	Overall Accuracy	Confidence Interval(95%)
50db	1	1	1	0.0000
20 db	1	1	1	0.0000
10 db	1	1	1	0.0000
0 db	0.9259	0.9444	0.9394	0.0393
-10 db	0.4815	0.8056	0.7172	0.0742
-20 db	0.0370	0.7083	0.5253	0.0823
-50 db	0.0741	0.7083	0.5354	0.0822

TABLE 15 CLASSIFICATION RESULTS WITH LIFTING SCHEME AND MAXIMUM LIKELIHOOD CLASSIFICATION FOR THE MULTI-CLASS SYSTEMS

SNR	Crash	Non-crash	Pile-drive	Overall Accuracy	Confidence Interval(95%)
50db	0.9259	1.0000	0.8000	0.9596	0.0325
20 db	0.9259	1.0000	0.8000	0.9596	0.0325
10 db	0.9259	1.0000	0.7000	0.9495	0.0361
0 db	0.9259	1.0000	0.6000	0.9394	0.0393
-10 db	0.4815	0.8871	0.5000	0.7374	0.0725
-20 db	0.0370	0.8548	0.5000	0.5960	0.0809
-50 db	0.0000	0.8226	0.5000	0.5657	0.0817

4.2 Computational assessment:

The computational efficiency of the system needs to be considered while implementing the algorithms on digital signal processing (DSP) chips. This can be obtained by calculating the number of multiplications and additions for every operation performed in the algorithm. Comparison of various DSP chip's speed is based on the number of multiplications and additions completed during the computation of the features for every incoming signal. The number of samples in every incoming signal is 66176. So the total number of additions and multiplications performed are $2.9779X10^{11}$ each (for a system with DWT as feature extraction method). This calculation is based on the order of complexity of the dyadic filter tree decomposition.

CHAPTER V

CONCLUSIONS

In this thesis, an automated system was designed and tested for accident detection at intersections using audio signals. Five different types of feature extraction methods were investigated: DWT, FFT, RCT, MCT, and DCT. Also the lifting scheme was analyzed as an alternative for the DWT in the real time implementation of the system. An investigation in the selection of an optimum mother wavelet was done, where seven different mother wavelets - Haar, Daubechies4, Daubechies15, Coiflets2, Coiflets5, Symlets2 and Symlets8 were analyzed. Three different statistical classifiers were also investigated: nearest mean, nearest neighbor, and maximum likelihood classifiers. The system used Fisher's LDA for feature optimization and the leave-one-out testing method. The system was designed to operate in two modes: two-class and multi-class. The two-class system was designed to identify crash or non-crash, and the multi-class system was designed to identify crash and several other non-crash events.

To test the system, a database was created containing recorded traffic audio signals like brake, pile drive, other construction sounds, normal traffic, and traffic accident (crash) audio signals. All signals were normalized, so that the maximum amplitude of each of the signal was one. The non-crash audio signals were collected from Jackson, MS, Starkville, MS, and Louisiana. Traffic signals were recorded using a DAT

recorder and microphone. The traffic accident audio signals were obtained from a crash testing facility in Texas.

The results obtained using various feature extraction methods, mother wavelets, and statistical classifiers are compared for both the accuracy and the computational expense. Also, the sensitivity of the algorithm was analyzed by varying the SNR, where the "signal" was the crash audio data and the "noise" was all other traffic audio data. The optimum methods were selected based on the accuracy and computational assessments.

5.1 Conclusions drawn from the results

The experimental results of the system showed that among the three different statistical classifiers investigated, maximum likelihood and nearest neighbor performed best. However due to the computational costs of the nearest neighbor, the maximum likelihood method was selected for the final system design. After choosing the classifier, different mother wavelets were analyzed. Haar, Daubechies4, and Coiflets5 provided the best classification accuracies for a two-class system. Haar was chosen because it is the simplest mother wavelet in terms of implementation. Among the five different feature extraction methods analyzed on the basis of the overall accuracy, RCT performed best. Also, when the SNR was greater than 0 dB for a two-class system, RCT, MCT and DWT gave an overall accuracy greater than ≈99%. Considering the computational time, DWT with Haar mother wavelet and maximum likelihood classifiers would be preferred. The second-generation wavelet method, the lifting scheme, was also investigated. It proved

computationally efficient when compared to DWT; it gave an overall accuracy of 100% for a two-class system when the SNR was greater than 5 dB.

Thus, the optimum design for an automated system would be a wavelet-based feature extractor with a maximum likelihood classifier. The DWT method gave a consistent classification accuracy of ≈95% to 100% when the SNR was at least 0 dB. The system operating in a two-class-system mode was superior to a multi-class system, when preference was based on the accuracy. The lifting scheme method would be the best choice when it comes to real time implementation of the algorithms. Compared to the DWT method, the lifting scheme performed equally well with an overall classification accuracy of ≈94% at SNR = 0 dB and 100% for SNR above 5 dB. Thus the choice of DWT or the lifting scheme would be preferred for a real-time system.

5.2 Suggestions for future work

The algorithms developed for the system were tested using pre-recorded signals. Though the overall classification accuracy obtained using the system was appreciably high, it was not tested in real-time. That is, the recorded signals were analyzed and classified in the lab. The signals from various intersections of Jackson, MS, Starkville, MS, and Louisiana were recorded without consideration for environmental conditions like weather, construction, and other traffic conditions. The testing was restricted to a certain variety of sound signals obtained from the intersections. Traffic signals with environmental conditions like inclement weather and a large variety of construction need

to be recorded and tested with the system. The next phase would be the testing and implementation of the developed algorithms, in a real-time system. An accident detection system could be modeled by recording a day-to-day traffic signal continuously, then processing a short duration of the recorded signal by extracting features from that windowed signal using a transform method, and finally classifying the signal as crash or non-crash. The information would be transmitted to the traffic management center (TMC) if the output of the system were labeled as a crash.

Two system architectures, centralized and de-centralized, should be analyzed for the implementation of the system. The centralized system architecture would be an approach where a centralized server would process the traffic audio signal at the TMC. The traffic audio signals would be transmitted from the intersection through a communication channel. At the intersection, signals could be recorded using a sensor, or microphone. The central server containing the accident detection algorithm would process the incoming traffic signal and identify whether it is a crash or a non-crash.

The de-centralized system architecture would be an approach where every intersection had a sensor like a microphone, along with a digital signal processor that contained the detection algorithm to process the incoming signal and identify the signal as crash or non-crash. A simple communication channel could be used to transmit an alarm indicating the accident to the TMC at a particular intersection.

The main advantage of the de-centralized system would be that much less information would need to be transmitted along the channel to the TMC. Using a

decentralized system would reduce the transmission cost drastically as opposed to the centralized system, where the whole signal needs to be transmitted. For the centralized system, the transmission of the whole signal to the TMS would require a communication channel with a large bandwidth. At the TMC, the signal would be processed, and any accidents detected. However, the processing time of a traffic signal would be high in a centralized system as compared to the de-centralized system. Therefore, transmission of the signal would be very important in both the system architecture and the communication channel decision would depend on the cost, performance, and reliability.

REFERENCES

[1] TTI 1999 Urban Mobility Report, Texas Transportation Institute. 1999

[2] INCIDENT MANAGEMENT PROGRAM BACKGROUND, Spring 2002. http://kdot1.ksdot.org/public/kdot/kcmetro/pdf/Ch1.pdf. Accessed Dec. 20, 2002

[3] H.J. Payne, E.D. Helfenbein, and H.C. Knobel, "*Development* and Testing of Incident Detection Algorithms: Volume 2"-*Research Methodology and Detailed Results, Report No. FHWA-RD-76-20, Washington, D.C.,* Federal Highway Administration, 1976.

[4] J. M. McDermott, "Incident Surveillance and Control on Chicago-Area Freeways," Special Report 153: *Better Use of Existing Transportation Facilities. TRB,* National Research Council, Washington, D.C., 1975,pp. 123–140.

[5] David A. Whitney and Joseph J. Pisano, TASC, Inc., Reading, Massachusetts. "AutoAlert: Automated Acoustic Detection of Incidents" December 26, 1995.

[6] K. Subramaniam, S.S. Daly, F.C. Rind, "Wavelet transforms for use in motion detection and tracking application," *Proc. Seventh Int. Conf. on Image Processing and Its Applications*, vol. 2, pp. 711-715, 1999.

[7] I. Ohe, H. Kawashima, M. Kojima, Y. Kaneko, "A Method for Automatic Detection of Traffic Incidents Using Neural Networks," *Proc. Vehicle Navigation and Information Systems Conf,* pp. 231 –235, 1995.

[8] K.W. Dickinson and C.L. Wan, "An evaluation of microwave vehicle detection at traffic signal controlled intersection," *Proc. Third Int. Conf. On Road Traffic Control*, pp. 153-157, 1990.

[9] S. Chen; Z.P. Sun; B. Bridge, "Automatic traffic monitoring by intelligent sound detection," *Proc. IEEE Conf. on Intelligent Transportation Systems*. pp. 171-6, 1997.

[10] E.M. Brockmann, B.W. Kwan, L.J. Tung, "Audio detection of moving vehicles," *Proc. IEEE Intt. Conf. on Systems, Man, and Cybernetics - Computational Cybernetics and Simulation,* vol.4. p. 3817-21, 1997.

[11] S. Kadambe, G.F. Boudreaux-Bartels, "Application of the wavelet transform for pitch detection of speech signals," *IEEE Trans. on Information Theory* , vol. 38, no. 2, part 2, pp. 917-924, 1992.

[12] C. Harlow and Y. Wang, "Automated Accident Detection," *Proc. Transportation Research Board 80th Annual Meeting*, pp 90-93, 2001.

[13] B. Bogert, M.Healy, and J.Tukey, "The Quefrency Analysis of Time Series for Echoes," *Proc. Symp. On Time Series Analysis*, 1963, New York, J.Wiley, pp.209-243.

[14] X. Huang, A. Acero, H. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*, pp. 306-318, Prentice-Hall, 2001.

[15] S. Gnavi, B. Penna, M. Grangetto, E.Magli, G. Olmo, "DSP performance comparison between lifting and filter banks for image coding," *Acoustics, Speech, and Signal Processing, 2002 IEEE International Conference on ,* Vol 3 pp. III-3144 -III-3147.

[16] R.Duda, P.Hart, D.Stork, *Pattern Classification*, Wiley-Interscience, 2001.