AUDIO BASED DETECTION OF REAR APPROACHING VEHICLES ON A BICYCLE

BY VANCHESWARAN KODUVAYUR ANANTHANARAYANAN

A thesis submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Computer Science

Written under the direction of

Liviu Iftode

and approved by

____

____

____

New Brunswick, New Jersey

January, 2012

ABSTRACT OF THE THESIS

Audio Based Detection Of Rear Approaching Vehicles On A Bicycle

by Vancheswaran Koduvayur Ananthanarayanan

Thesis Director: Liviu Iftode

Cycling is an efficient mode of travel widely used for transport, recreation and sport all over the world. In addition to being environmentally friendly, it also affects the health of the cyclist favorably. Safety is an important concern for a cyclist because, during an accident with a motor vehicle, the cyclist is exposed to higher risk of injury than the vehicle driver. Improving bicycle safety is an important factor in saving lives and promoting the use of this environmentally friendly mode of transport.

The Cyber-Physical Bicycle system was introduced as a concept bicycle that can alert the cyclist of dangerously approaching vehicles from the rear. The system aims to accurately detect and track vehicles approaching from the rear, differentiate dangerously approaching vehicles, and alert the cyclist early enough to take preventive measures. This is achieved through video based detection. The traditional bicycle is extended with computational capabilities and a rear facing video camera, which constantly monitors vehicular traffic behind the bicycle. Research indicates that the system is feasible, works with good accuracy and generates timely alerts, though it cannot operate at full efficiency while running in realtime.

In this thesis, we present an approach that augments a bicycle with audio based detection of rear approaching vehicles. The audio based Cyber-Physical Bike continuously listens to the environment behind the bicycle with a microphone, detects rear approaching vehicles and alerts the biker to their presence. We describe the design for an audio based Cyber-Physical Bike and demonstrate its feasibility through evaluation of our prototype. We found that distinguishing the directionality of vehicle approach is a significant problem in case of audio, due to the similarity in sounds. Subsequently, we identified several audio features that help us differentiate rear and front approaching vehicles accurately. We also used a rear facing microphone to improve detection. Results show that our approach works with comparable accuracy to the video based approach, performs real time detection at lower energy and hardware costs, and is more efficient. However, the system sacrifices on timeliness of alerts, and the alerts are generated much later when compared to video based detection.

Acknowledgements

Firstly, I would like to sincerely thank my advisor Professor Liviu Iftode, for being the constant source of guidance and encouragement without which it would have been impossible to bring this research to its rightful conclusion. I am deeply indebted to him for introducing me to systems research and he is a great source of inspiration to learn and work better, wherever I go. I also thank him for providing me with this opportunity to build something valuable during this research. In addition to this, many thanks for being a superb teacher of Distributed Systems, and for all the valuable insights during lectures and presentations.

Secondly, my sincere thanks to Stephen Smaldone, who as a mentor and colleague on this project helped me with suggestions and ideas whenever I was stuck, and for being an overall guiding light all through this project. My heartfelt thanks to him for his huge contributions in ideas and design of the project, and for letting me cite his work on the Cyber-Physical Bike system and for the use of the data that he collected with much difficulty. I am deeply indebted to him for letting me use his work as the basis for my research.

I would like to thank Professor Ahmed Elgammal and Chetan Tonde, without whose excellent work, this research would have been impossible. My sincere thanks to you in letting me cite your work. Many thanks to the all the interesting discussions and ideas we had, which were very enriching. Special thanks to Chetan Tonde for the all the fun and long hours spent working on the project, especially during the evaluations, thanks for sharing the load, thanks for all the understanding. Many thanks to both of you for letting me use your work as a basis for my research.

My sincere thanks to Prof. Badri Nath, for his kind guidance through all my research endeavors and for being a part of the committee. Thanks a lot Professor, for teaching me Computer Networks, the way only you could do, with plenty of insight, interesting facts and fun added to it. Thank you for your constant guidance during my other research projects, and for many interesting insights during presentations.

Many thanks to Lu Han, Matthew G. Muscari, Mohan Dhawan, Shakeel Butt, Pravin Shankar and others in the DisCo Lab for being so understanding and for your constant support and companionship.

I sincerely thank all my teachers here, including Prof. Thu Nguyen, Prof. Eric Allender, Prof. Muthu Muthukrishnan, Prof. Szemeredy, Prof. Sesh Venugopal and Prof. Alex Borgida. You have all been very inspiring and greatly supportive. You have constantly prompted me to think about things, and to see things in a different light.

Many thanks to my family, who has been supportive all through my studies. I am truly grateful for their love and support. Many thanks to my friends here, for your companionship and affection. Thank you for making my student life easier and enjoyable.

Dedication

To my parents, my brother.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Cycling is an important activity that contributes positively to the health of an individual. As an alternate mode of transportation, it affects the environment favorably compared to other forms of vehicular traffic. A significant number of users ride bikes for recreation and health reasons, while many others use it to commute short trips from home and run errands[30]. Bicycles provide a means of transport without burning fossil fuels, causing zero air and noise pollution, and are comparatively much more affordable than a motor vehicle. According to a survey conducted by National Household Travel Survey Administration, of all the trips made in the United States in 2009, about 1.0% were made on a bicycle[16]. Even though cycling is a much older tradition than driving, the advent of motor vehicles has marginalized the use of bicycles. Motor vehicles, though less energy efficient and polluting, are much more convenient, and are effective over longer trips, and since, have become the dominant mode of transport over roads. Consequently, cyclists have had to share the road with a growing number of motor vehicles over the years. This has put them at an increased risk of accidents.

Road safety is a key concern for a cyclist. A cyclist is exposed to higher risk of injury when compared to other vehicle drivers due to their vulnerable disposition on roadways. Bike helmets and other safety gear can provide some amount of protection against serious injuries, though they cannot avoid accidents. In the year 2009, 630 unfortunate pedalcyclists rode to their death[33] all over the United States. The same statistic also states that there were close to 51000 reported cases of cyclist injuries during these accidents. When a road accident involving a bicycle and a motor vehicle occurs, it presents an asymmetric risk situation to those who are involved. Clearly, the cyclist is more likely to be injured than the motor vehicle driver. Improving the safety situation of the cyclist is very crucial towards saving valuable lives and avoiding injuries, as well as promoting this environmentally friendly and healthy means of transportation.

There are several existing efforts that try to improve the safety situation of bicyclists. Exclusive bike lanes and laws mandating use of bike helmets are good examples, though they are mostly not very effective. Dedicated bike paths are a good preventive option, but they only provide partial coverage. A 2002 statistic shows that about 48% of cyclists in the United States use paved roads for travel[30], which they have to share with other mechanized forms of transport like cars and trucks. Bike helmet laws exist in many states, though they may not be applicable to all riders. Laws are however difficult to enforce and are not effective enough in preventing accidents. There is a need for a biker centric approach to safety that can prevent accidents from happening.

1.1 Bicycle Safety

This section describes the current state of bicycle safety and summarize various existing approaches to improving the situation. A brief summary of the fatality and injury statistics of bicycle accidents is presented, followed by the state of improvement in bicycle safety. Subsequently presented is a review of existing approaches to improving bicycle safety, including legal solutions, infrastructure solutions, biker education and biker protection.

As discussed earlier, in 2009, there were 630 cyclist fatalities and 51000 reported cases of injuries in the United States. Reports indicate that these fatalities account for about 2% of the total traffic fatalities of that year[33]. This is important considering trips on bicycles only made up 1% of all trips in the country that year[16]. According to some estimates, the total cost of cyclist injuries and death is over $4 billion per year[45]. This is in addition to the fact that bicycle crashes are severely underreported and only as low as 10% of all injury causing crashes are recorded by the police[45]. These facts highlight the importance of bicycle safety and the need to address it proactively. Over the years, States have initiated various reforms to improve the situation. Despite these efforts, the state of bicycle safety has consistently not improved[32, 31, 29] across the

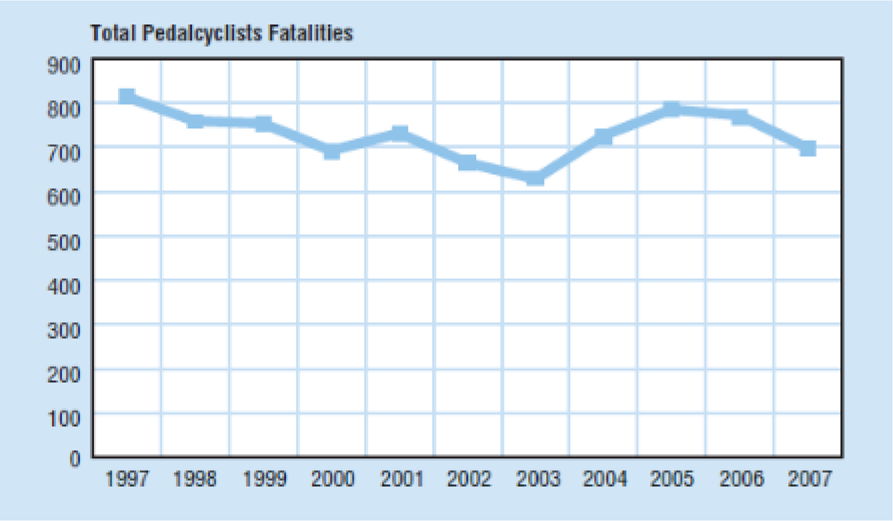

Figure 1.1: Pedalcyclists Fatalities 1997-2007
The graph plots the yearly number of pedalcyclist deaths over the period of 1997 to 2007. This is taken from Traffic Safety Facts 2007 published by the Department of Transportation, United States.[32]

decade as indicated by Figure 1.1.

Several laws that provide various rights and restrictions to cyclists and motorists who share the same road exist today. Some states restrict the slow moving bicycles to the extreme right lanes of the roads[41]. Similar laws also mandate the use of proper safety equipment like bike lamps, good brakes and reflectors on wheels[41]. In effect, many of these laws formally mandate several common sense practices for safe riding. Many states also stipuate the use of a bike helmet to prevent serious head injuries[21], though unfortunately, they are only applicable to riders under 18. This is despite the fact that majority of the bicycle accidents (70%) involve head injuries and only 20-25% of cyclists wear helmets[33]. There are also laws that regulate the motorists. Many states mandate a three feet rule, under which a motorist cannot approach within three feet of a cyclist[9]. Though many of these laws make sense, formulating them is not the same as enforcing them effectively. A study conducted in the city of New York indicates that close to 22% of cyclist accidents involve hit and run motorists, and nearly three quarter of faulty drivers go unpunished[22]. In essence, several promising legal solutions towards improving biker safety are deterred by their limited coverage and difficulty in enforcement. Additionally, they cannot proactively prevent accidents from happening, and only proscribe procedures to follow after the fact.

Another approach of improving bicycle safety is through infrastructure based solutions. This involves designing bicycle friendly towns and cities. This approach includes earmarking space for bicycles on paved roads, constructing dedicated bicycle lanes, providing clear warning signs to both cyclists and motorists at crossings, and actively maintaining these resources. One example for this approach is Portland, OR which has a large network of bicycle lanes[1]. Unfortunately, this is not the usual case and not all cities favor bicycle paths. The costs for adapting existing roads to add bicycle lanes can be extremely prohibitive. Building and maintaining such infrastructure requires consistent support and commitment to the cause of bicycle safety from the government, which tends to favor motorists over the minority of cyclists.

A different way to prevent accidents is by educating the cyclists formally about the risks and safe practices while sharing the road. In most states, acquiring a driver license involves rigorous testing of the applicant's knowledge of vehicle laws and safe practices. By contrast, learning to ride a bicycle is an informal activity and does not involve any formal knowledge testing. Various awareness programs that are organized by bicycling enthusiasts and other organizations do exist, but they are limited by their reach and impact.

Alternatively, to prevent injuries and fatalities, we can increase the protection of the cyclist. This involves wearing protective armor like helmets, elbow and knee guards while riding. They serve to protect the rider from fatal or serious injuries. Unfortunately, any amount of protection worn by the cyclist seems insufficient when it comes to an accident involving a powerful motor vehicle. The cyclist is always at a disadvantage compared to the motorist during an accident because of the inherent asymmetry in the safety situation. Though helmets have been effective in saving lives by reducing the chances of dangerous head injuries, a more effective approach would have been to prevent the accident itself.

Clearly, from the above discussion, none of the described approaches provide an adequate option to the cyclist to take preventive action against accidents.

1.2 The Problem

This section describes the specific safety issue addressed by this thesis: vehicles passing the cyclist dangerously from behind. It also describes the existing approaches used by cyclists in addressing this scenario.

One of the common risky scenarios for a cyclist is a motor vehicle approaching from rear dangerously close to the bicycle. Bikers tend to ride with the flow of the traffic and are usually forced to share the road with other motor vehicles. In such a case, there are usually a number of vehicles passing the biker from behind. A similar situation arises also when the cyclist drifts to the left to avoid obstacles in front. For example, a parked vehicle on the shoulder may cause the biker to swerve left in an attempt to pass it. This can cause a dangerous situation when another vehicle is already trying to pass the cyclist from behind. Since any error while passing can catch the cyclist unawares, the cyclist must constantly scan behind for approaching vehicles. In order to do this, the cyclist may have to turn her head to look behind consistently. Thus, the cognitive load on the cyclist riding amongst traffic becomes high, as she has to balance the vehicle, avoid obstacles in the front and also watch out for rear approaching vehicles. This is in addition to the physical load of actually pedaling the bicycle forward.

In addition to the constant distraction of looking behind, the physical action of turning one's head back affects the safety of the cycling negatively. When turning back, the action may cause the cyclist to drift into the traffic or into the shoulder. In addition to that, during this time, the ability to track obstacles in front of the bicycle is hampered and may cause a possible collision. Both are dangerous situations for the biker.

The common remedy for this is the use of a rear-view mirror, either attached to the helmet or the handlebar of the bicycle. The cyclist scans the mirror instead of turning her head back from time to time. Though, this simplifies the physical act, mirrors still distract the cyclist and only provide an inconsistent and incomplete picture of the rear, since they are limited by their field of view and position. Digital replacements like the Cerevellum [11] provide consistent view of the rear by placing a video camera behind the seat and providing a digital display in front. However, both these options fail to provide any alerts to the cyclist about rear approaching vehicles. The cyclist is still dependent on periodic scanning of the display for rear approaching danger. When there is no vehicle behind, these options simply become distractions to safe riding.

Another approach to address this issue is to wear reflective clothing while riding. One could also attach flashing lights or bike reflectors to the rear of the bike. They try to warn the motorist behind to the presence of the cyclist by making the cyclist standout on the road. The onus of avoiding the accident is thus pushed to the motorist. Obviously, this cannot prevent accidents caused by error on the motorist's part. Again, the cyclist has no option to avoid the accident, since this approach does not provide any warning to the cyclist of the danger.

1.3 Cyber-Physical Bike

Motivated by the safety issue of rear approaching vehicles discussed above, in 2010, the Cyber-Physical bike project[40] presented by Stephen Smaldone, et. al. proposed a bicycle concept that adds computational capabilities as well as audio, video and other sensors to a road bicycle. It aims to constantly monitor the environment behind the bicycle to detect dangerously approaching vehicles and alert the cyclist of any impending danger so that preventive action can be taken. Figure 1.2 illustrates the key components of the system.

The ordinary road bicycle is fitted with an array of sensors including a rear facing camera, accelerometer and microphone. It is then augmented with an embedded computer, which provides computational capabilities to the system. In this project, using the video camera fitted behind the seat, the system constantly captures video images from behind the bike and subsequently streams the video to the embedded computer. The embedded computer then processes these streams of sensory information and makes an inference about dangerously approaching vehicles from behind. Consequently, upon positive detection, an alert is generated, which warns the cyclist of impending danger. Specifically, the research introduced the application of computer vision techniques on

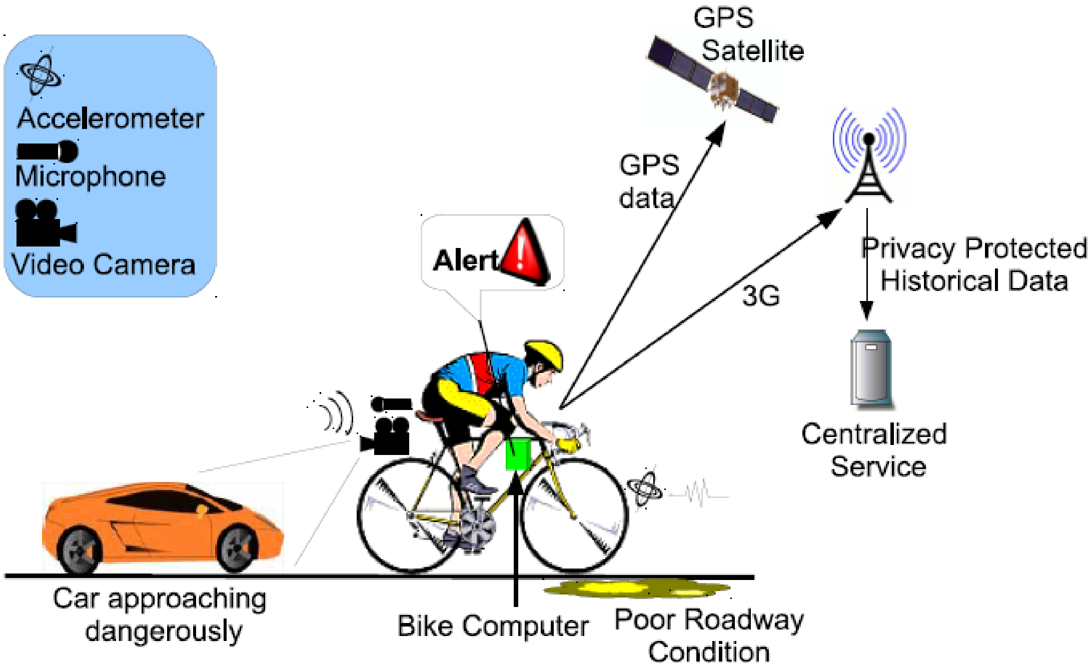

Figure 1.2: Cyber-Physical Bike
The concept bicycle presented by the Cyber-Physical Bike project with accelerometer, audio and video sensors attached. Also represented is a GPS sensor for reporting location and network connectivity to the cloud. Attached to the bike is a small embedded computer and an alert mechanism. Figure taken from publication by S.Smaldone, et.al.[40].

the captured stream from the rear facing video camera to make a determination of dangerously approaching vehicles. The rest of the sensors including the microphone were not used in this project. The design of a microphone based detection system is the main subject of this thesis.

The aims of the Cyber-Physical Bicycle are three fold. Firstly, it describes a feasible design for a Cyber-Physical Bike augmented with a rear facing video camera and an embedded computer to perform realtime detection of dangerously approaching vehicles from the rear. Secondly, by alerting the cyclist of dangerous vehicles from behind in a timely fashion, it aims to provide the cyclist with an opportunity to avoid an accident. Finally, it strives to reduce the cognitive load of the cyclist by assisting her in keeping track of the environment behind the bike. To achieve this, the system was designed with the following goals in mind. Most importantly, the system must detect and track vehicles approaching from rear accurately. Secondly, it must distinguish dangerous rear approaches from the benign front approaching vehicles to be effective as an alert system. Finally, it must realize this within the limited power and computational resources available on the bicycle and also account for platform instability due to bad road conditions.

The prototype system presented in the paper[40] is built using a light-weight HP Mini 311 netbook, which closely matches embedded hardware in terms of performance and cost. The video processing functionality makes use of both the CPU and GPU capabilities of the system and performs with an overall accuracy of 73.1%. The system produces timely alerts and has reasonable power requirements. However, the presented system cannot yet achieve full real time status without trading off on accuracy.

Clearly, the Cyber-Physical Bicycle system aims to be a biker centric and preventive approach to improving bike safety. Once the danger is detected, the cyclist gets an opportunity to take preventive action against a potential accident. Notably, this approach does not rely exclusively on the motorist to avoid accidents. Accurate and consistent alerts also tend to reduce the cognitive load on the cyclist while riding.

1.4 Video Based Detection

This section briefly describes the video based detection proposed by the Cyber-Physical Bike project[40]. A short description of the design and architecture of the system is provided, followed by a brief summary of results. This section is not to be considered a part of this thesis, and is presented here for the sake of clarity.

The video based detection module proposed by the Cyber-Physical Bike project aims to achieve the following:

- Detect rear approaching vehicles as early as possible, in order to provide the cyclist with enough time to avoid accidents.
- Once detected, track these vehicles in realtime and make an accurate prediction whether they will pass the biker dangerously. To differentiate between safe and unsafe approaches, the video detection subsystem applies a *virtual safety zone* around the biker of size three feet.

These aims are achieved through three separate modules.

Optical flow analysis As a bicycle moves forward, stationary objects and rear-approaching objects produce distinctive patterns in the rear field of view. This pattern of apparent motion caused by relative motion between observer and the objects is called *optical flow*. There are distinctive differences between the optical flow patterns produced by stationary background objects and objects approaching the bicycle from the rear. The optical flow analysis module uses this difference in patterns to detect rear approaching vehicles.

Roadway segmentation This module reduces the overall computational load of the system by selectively focusing on parts of the field of view where rear approaching vehicles tend to appear. In essence, the module identifies the roadway lane the biker is on and restricts further processing to the area behind the biker on that lane. For this purpose, it identifies the road edges and roadway lane markers using a Gaussian Mixer model in Hue Saturation Value colorspace, from the field of view behind the bike.

Vehicle Tracking Once a rear approaching vehicle is detected, this module tracks them in realtime to determine if they are about to enter the virtual safety zone. This is performed by adaptive algorithms using Principle Component Analysis based appearance models which can handle rapid changes in pose, scale and illumination.

The results presented by Stephen Smaldone, et. al[40] indicate that the system performs with high accuracy under a capture rate of 30 frames per second. The *timeliness* of the generated alerts, which indicate how much time the cyclist gets to take preventive action before the closest point of approach, is about 3.5 seconds. The results also indicate that the system has only reasonable power requirements, comparable to other multimedia applications like movie players. However, the system is not yet capable of running at the full frame rate on the prototype system. The system can only function at a severely reduced frame rate of close to 2 frames per second, trading off on accuracy as a result.

1.5 Thesis

An audio based detection system can be used by the Cyber-Physical Bicycle for detecting dangerously approaching vehicles from behind. It can provide improved energy efficiency and performance when compared to a video based detection system. Audio based detection can also be used in combination with video based detection, to improve overall accuracy of the existing system.

1.6 Audio Based Cyber-Physical Bike

The audio based detection subsystem uses a microphone attached to the rear of the bicycle to constantly listen to the environment behind the biker. The captured audio stream is forwarded to the embedded computer on the bike, which performs analysis on the stream to detect the approach of vehicles from behind. Once a rear approaching vehicle is detected, the subsystem generates an alert to the cyclist making her aware of the danger behind.

Rear approaching vehicles produce a recognizable audio pattern when they attempt to pass the cyclist. Cyclists regularly use audio cues to instinctively determine the approach of a vehicle from behind. A suitably wind shielded microphone attached to the rear of the bicycle is sufficient to capture these audio patterns that correspond to a dangerous vehicle approach from the rear.

There are some obvious advantages to using audio over video.

Performance Audio processing traditionally consumes much less computational power when compared to video processing. The amount of data while dealing with audio is much lesser compared to video, which also contributes to reduced memory costs. Our results reflect this very well. Such reduction in computational costs indicate that audio based detection systems may be feasible on smaller and wide-reaching platforms like smartphones and other hand-helds.

Cost Audio processing algorithms can usually be realized on much cheaper hardware when compared to video processing. For instance, the video processing part of the original Cyber-Physical Bike system uses a GPU for processing and requires a comparatively costlier video camera. Audio can usually be realized on just the general purpose processor and that too on reduced frequencies. Microphones are cheaper compared to video cameras. Also, costlier equipment attached to the bike tend to increase concerns about their theft.

Darkness or Failing light conditions Accuracy achieved by audio based detection remains unaffected by variations in light conditions on the road, which may not be the case with video. According to data from 2008[18], 21% of bicycle fatalities happened during 6-9PM in the evening. This is usually the time when the light is failing and video based detection accuracy may not be consistent during this crucial period. Audio based detection will work just fine during these times, as they do not depend on light conditions.

Low power requirements Audio processing traditionally consumes much less power compared to video processing. This is especially handy when it comes to designing an end product, where low power requirements become much more convenient for the cyclist. The product can be used for much longer trips as well as more frequently. The number of recharges required over long period of usages also decreases, which affects user experience favorably. Low power requirements may also drive down costs by allowing lower capacity batteries.

On the other hand, audio processing can also be used to augment the existing system. Alerts from both audio and video based detection systems could be combined to arrive at more accurate or consistent results. Audio can be used to improve the overall accuracy for cases where video may not be consistent. A good example is during twilight in the evening, where video may not perform consistently. The system accuracy could be reinforced with audio based detection during this time. Since the hardware costs, power requirements and computational costs for audio based detection are very low, such an addition will not be prohibitive. This thesis does not propose a combined design for audio and video based detection, but discusses some design approaches as future work in Section 5.3.

1.7 Related Work

Improvement of road safety is a well researched field, but unfortunately a majority of the mind-share has been spent on improving safety from the perspective of a motor vehicle. This section tries to describe the current state of road safety research in brief, with an added focus on improving bicycle safety. The opening section focuses on research innovations that assist the driver of a motor vehicle in order to operate more safely on the road. This is followed by a discussion of available research in the specific field of improving cycling experience and safety. Finally, we discuss other research based on audio detection techniques, which may not be related to vehicle safety, but nevertheless, are related to our own audio detection approaches.

Driver Assistance

A significant amount of research has been done towards improving road safety, while driving a motor vehicle. Most of these systems focus on assisting the driver towards a safer driving experience. For example, the SAFELANE system[6] tries to alert a driver upon accidentally straying from a road way lane. This is achieved by the use of video based detection algorithms, supplemented by other sensors like radar and laser scanners. To further assist the driver, various computer vision based techniques were applied in realtime detection of traffic signs[13, 7, 8]. Various other research initiatives focus on detection of obstacles on the road[19],while some specifically detecting pedestrians[15, 46] crossing the road, using computer vision techniques.

Most of these systems monitor the constantly changing external environment of a vehicle and try to alert the driver of any sudden unseen changes. Instead of looking out, some systems try to look inward in order to monitor the state of the driver. For example, various systems were introduced to detect driver drowsiness and inattention, using a variety of sensors. Some use an electroencephalogram(EEG) based approach[24] to detect drowsiness, while others use movement sensors attached to seat-belts, car-seats and steering wheels[47], or image-processing techniques based on pictures of the driver's face[42].

Autonomous driving has also received plenty of attention. A number of DARPA grand challenges where conducted in recent years, where many teams competed in building and racing autonomous vehicles[12, 28]. Google has recently developed technology for cars that can drive themselves, which use video cameras, radar sensors and laser range finders[34].

Most of the above systems use a range of sensors including video, radar and laser scanners to perform detection. They operate from the context of a motor vehicle where they can integrate with more powerful hardware, and afford higher power requirements. From the context of a bicycle, these are significant drawbacks. Any system built for a bicycle is limited by cost and power constraints. This also adds realtime performance constraints on detection systems, since cheaper hardware is less powerful. Additionally, the increased weight is a significant constraining factor for bicycle based systems, where as, this is less of a concern for motor vehicles.

Bicycle Safety

While plenty of research is oriented towards assisting motor vehicle drivers, little or no research focusing on improving bicycling experience and safety exists. In this section, we will explore the available advances presented in the field of bicycle safety.

The BikeNet[14] project describes a system to map the overall experience of the cyclist by applying sensor networking principles to the road bicycle. It presents the first working mobile networked sensing system for bicyclists, in order to map biking experiences. To achieve this, they include multiple sensors on their bikes, including accelerometers, photo-diodes, thermistors and microphones, in addition to camera and other radio sensors provided through the cyclist's mobile phones. The aim of the system is to extensively map the cyclist experience through these sensors, and provide measurements for user fitness and performance. The consolidated data is stored in the web, so that cyclists can visualize the experiences, and share routes between each other. The focus of the research is mostly on biker fitness and the social aspects of biking rather than improving safety. In a similar approach, the Copenhagen Wheel project[35] aims to add sensors to the bike in order to enable urban sensing applications. They also aim to provide motor assist to the biker while pedaling, and provide health and environmental monitoring facilities on the bike.

Another initiative similar in spirit to the Cyber-Bike Project, is the Biketastic project[36], which uses sensing through mobile phones to map bicycle routes. The system utilizes GPS to track the routes, and uses microphones and accelerometer data to record road roughness and noise levels. These routes can later be rated for experience and shared with other bikers. Through this information exchange, a safer and better cycling route can be selected by the user in future. This is similar in extent to various other route sharing services provided by web applications specific to bicycling, like Bikely[2], Veloroutes[5], MapMyRide[4] and Cyclopath[3].

The commercial product Hindsight 35[11], introduced by Cerevellum, is an attachment to the bike that acts as a digital rear view mirror. It consists of a video camera and a digital display unit attachable to the handlebar. In addition to this, the system can keep a video record of the last ten minutes, which can be useful during accidents.

Similar to the Cyber-Bike Project, many of these research initiatives augment the ordinary road bike with multiple sensors. However, it becomes clear from the discussion that there are very few directly addressing the problem of bicycle safety. Instead, most projects aim to improve the experience of bicycling as a whole. The Cyber-Physical Bike project[40] is possibly the first one to address the issue of biker safety directly and suggest solutions toward improving it.

Audio Detection Techniques

When it comes to applying audio based algorithms to realtime vehicle approach detection, little or no work has been completed. However, our work utilizes a lot of previous research in the area of acoustic signal processing and classification. Our approach of looking at differentiating audio features in order to classify sounds, is derived from similar prior architectures.

A lot of work has been done in the field of sound classification. For instance, the problem of distinguishing music from speech is tackled in [38]. Their approach works by examining numerous audio features that differentiate properties of music from speech signals. Yet another approach focuses on using features that enable realtime discrimination[37] of speech and music on broadcast FM radio. This would be useful in differentiating talk-radio with music programming automatically. Both approaches utilize machine learning algorithms to make an inference based on a select group of audio features. In addition to this, several advances were made in classification of audio into more general categories like music, speech, multiple speakers, silence, etc. for content-based retrieval[23, 17]. In all these approaches, finding the right audio features was important.

There have been many initiatives to classify the acoustic environment in general[27, 26]. This can be useful in determining the context of a user, with respect of the user's surroundings, general location and mode of communication (speaking to a single person or to a group, etc). In a more recent work, the SoundSense[25] project applies audio based detection algorithms in order to detect a user's context using audio from their mobile phone. Their approach utilizes the microphone in the user's mobile phone to capture contextual audio and classify them into general sound types like music and voice, and also discover new sound based events. They utilize a combination of supervised and unsupervised learning techniques for classification of events happening in a user's everyday life. Their architecture is also based on feature extraction from audio and classification using the features.

Our research is heavily indebted to the large body of prior work done in sound classification. We directly benefit from the discovery and description of large number of audio features in these works. Our approach shares the general architecture of finding distinguishing audio features and using machine learning based techniques to classify different events, with similar approaches in the past. However,to the best of our knowledge, our work is unique in applying these techniques to predict vehicle approaches from rear, in the context of bicycle safety.

1.8 Contributions

This thesis makes the following contributions:

- It describes the design of a Cyber-Physical Bicycle system based on audio based detection of rear approaching vehicles. Using a microphone attached to the rear of the cyber-physical bike, the system continuously captures audio from behind, and uses audio processing techniques to alert the biker about dangerously approaching vehicles from the rear.

- Through an experimental prototype, it demonstrates the feasibility of an audio based Cyber-Physical Bicycle system. Experimental results suggest that the audio based system can work with comparable accuracy to the video based system, providing realtime results with higher efficiency. However, alerts are generated much later than video and hence, less time is provided to the cyclist to take preventive action. This can be alleviated by a combined approach which uses both audio and video based detection.

1.9 Contributors To The Thesis

The following is a list of people who co-authored the paper from which material was used in this thesis. Stephen Smaldone, Professor Liviu Iftode and Professor Ahmed Elgammal (Rutgers University) created the idea and provided the motivation for the Cyber-Physical Bike project based on both audio and video detection. Professor Elgammal and Chetan Tonde (Rutgers University) are responsible for the design and implementation of the video based detection subsystem, with Chetan Tonde being responsible for the evaluation and generation of results. The video related results included in this thesis are only for the purposes of comparison. Stephen Smaldone contributed to the design and implementation of the audio based detection subsystem. He also collected the experimental data for both audio and video by riding out numerous miles on his bike. The metric *timeliness*, which defines the amount of time available to the biker, to take evasive action after the alert is generated, was suggested by Stephen Smaldone.

Chapter 2

Audio Based Detection

This chapter describes the design of the audio based Cyber-Physical Bicycle. In the beginning section, it defines the design goals for an audio based detection subsystem, and subsequently, discusses the associated challenges and strategies. This is followed by a detailed description of the various components of the system.

2.1 Design Goals

For making an audio based detection of rear approaching vehicles, the Cyber-Physical Bike must have the capability to listen to the environment behind the bike. For this purpose, the bike is augmented with a microphone behind the seat, which can continuously capture the sounds of approaching vehicles. In order to become an effective alert system, the audio based detection system must achieve the following goals:

1. The audio based detection system must identify rear approaching vehicles accurately.

2. The system must generate alerts in realtime, and as early as possible, so that the cyclist gets enough time to take preventive action.

3. The system must differentiate vehicle approaches from the front, which are less dangerous, so that the alerts are consistent and not distracting.

2.2 Design Overview

The design of an audio based detection system hinges on the following observations:

- When a vehicle approaches a biker from behind, it generates a clear and recognizable audio signature. This is what traditionally enables an undistracted biker to turn her head and take action.

- This signature is noticeably different from what is caused by vehicles passing the bike from the forward direction. This is due to the directionality of the sound and due to the fact that vehicles approaching from the rear are more closer to the bike than those passing from the other lane. The directionality of the the microphone will also have an impact on their differences, since the microphone is turned towards the rear.

- Vehicles passing from the rear also tend to produce a slower build up in sound. This is due to the difference in relative velocities between the bike and the motor vehicle, during rear approaches and front approaches.

Our design intends to utilize these discernible audio patterns produced by rear approaching vehicles in identifying them. This specific audio signature can be envisioned as a set of feature values extracted from the audio, which vary over time with correlation, when a vehicle is about to pass the biker from the rear. A significant part of this research hinged on the search for these features and their identification. We analyze the incoming stream of audio, generate multiple feature values and create a vector of these values. A prediction module looks at this feature vector, and makes an inference whether a rear approaching vehicle is present or not.

We utilize machine learning techniques to predict rear approaching vehicles from a feature vector. The prediction module utilizes a classification model to detect rear approaching vehicles from the feature vector. We built this model statically using a training set of cycling traces that we collected. This data consists of audio captured from many instances of vehicles approaching the biker from the rear and passing the bicycle closely. We also included multiple occurrences of front approaching vehicles as well as several short intervals of audio where no vehicle was present behind. In order to better train, we annotated this data according to the type of the vehicle approach, or the lack there of. Subsequently, we created a stream of feature vectors corresponding to this data, and built a classification model statically based on them. Our prediction module utilizes this model in order to make predictions. This stream of predictions is used to generate alerts to the cyclist whenever a rear approaching vehicle is detected.

2.3 The Audio Detection Pipeline

Our approach to the audio detection system proposes a pipeline architecture for detection. The overall architecture of the audio pipeline is presented in Figure 2.1. Firstly, the audio is captured by a microphone and forwarded to the processing system. The captured data is in the form of a Pulse coded Modulated (PCM) audio stream, which is a digital representation of the analog audio signal. The audio signal is sampled at a fixed frequency by the microphone and quantized to a set of digital steps to generate a stream of PCM data. This stream of data is then split up into audio frames of fixed size, and fed into a feature generation component. The feature generation component acts on these frames and generate a vector of feature values per frame. Feature generation can be thought of as applying various mathematical functions to the sequence of samples in the frame. From a broader perspective, the feature generation module can be seen as consuming the audio stream in realtime, and generating a vector of features for each audio frame. The prediction component acts on this feature vector and makes a per-frame prediction. An alert generation system acts on this continuous flow of predictions and generates an alert to the rider after taking minor fluctuations into account.

2.3.1 Feature Selection

There are numerous challenges that an audio based detection system must conquer in order to succeed effectively. Firstly, the microphone is always exposed to wind and background noise. Secondly, the audio environment is constantly polluted by the noise caused by pedaling activity. Finally, there will always be differences in sound when the bicycle travels over different road surfaces. The identified features must accommodate for such factors in order to provide a good chance of detection. After scouring through

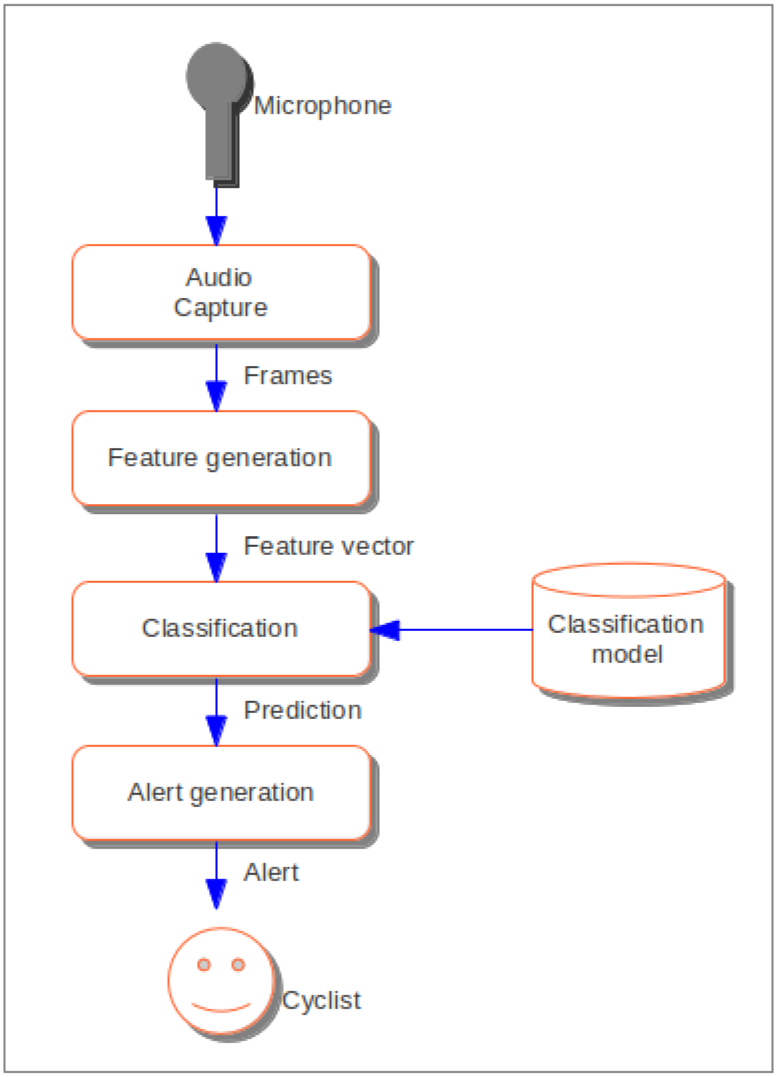

Figure 2.1: Architecture Diagram For The Audio Pipeline

large number of standard audio features described in [25, 23, 20, 38], a number of features were selected that clearly differentiate a rear approaching vehicle from other circumstances.

Among the features selected, some are time domain features and others are in the frequency domain. For calculation of features, the audio stream is segmented into frames of fixed size, composed of PCM sample values. We represent frame size usually by n and each individual PCM sample is represented as x_i where i is the index of the sample within the frame. To compute frequency domain features, we must first compute the Fast Fourier Transform (FFT) for the frame, which provides us the frequency response of the frame. FFT is an algorithm for efficiently computing the Discrete Fourier Transform of an analog signal. It provides us with a measure of phase as well as amplitude of sound as a function of frequency. We apply a hanning window over the frame before computing FFT in order to reduce the adverse effect of frame boundaries[20]. We disregard the zeroth component, which indicates the average volume of the sound (the DC component), before performing any more analysis. We represent $p(i)$ as the magnitude of the ith frequency bin in the spectrum.

Strongly Correlating Features

In this subsection, we discuss the main features that show excellent correlation when a motor vehicle passes the cyclist from rear. These features are primarily responsible for distinguishing dangerous rear approaches from other benign situations.

Root Mean Square Amplitude (RMS) [25] is a time domain feature which stands as a good approximation for loudness of captured sound. It is calculated as

$$RMS_f = \sqrt{\frac{1}{n}\sum_1^n x_i^2}$$

where RMS_f stands for the RMS value per frame, n stands for the number of samples per frame, and x_i stands for the i^{th} PCM sample in the frame. Clearly, when a rear approaching vehicle passes a bicycle, it makes a significantly loud sound, which rises through the event and then falls as the vehicle speeds away.

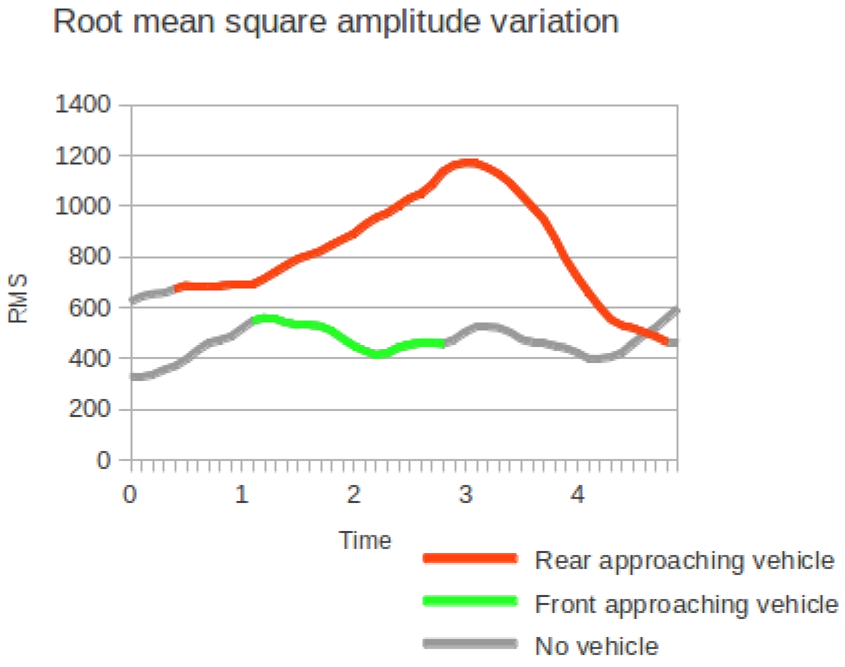

Figure 2.2: Root Mean Square Amplitude Variation When Vehicle Passes The Biker. Rear approach event causes the RMS value to rise. Front approach event does not cause a loud enough change.

The RMS feature utilizes this change in loudness as a measure to identify rear approaches. Figure 2.2 depicts a graph that shows both a vehicle approaching from rear as well as front to illustrate the difference in RMS levels. The front approach even hardly registers while a rear approach causes visible rise in the value. The rear approach is indicated in red and the front approach is indicated in green. It should be noted that RMS is susceptible to events that cause a similar rise in loudness level and duration.

Spectral Centroid (SPC) [23] is a frequency domain feature that describes the central point in the power distribution within the spectrum. When a vehicle approaches a cyclist from the rear, the balance of the spectrum slowly shifts to the high frequency ranges. It can be computed as:

$$SPC_{index} = \frac{\sum_1^n i.p(i)^2}{\sum_1^n p(i)^2}$$

where, i represents the frequency index, and $p(i)$ represents the magnitude of the

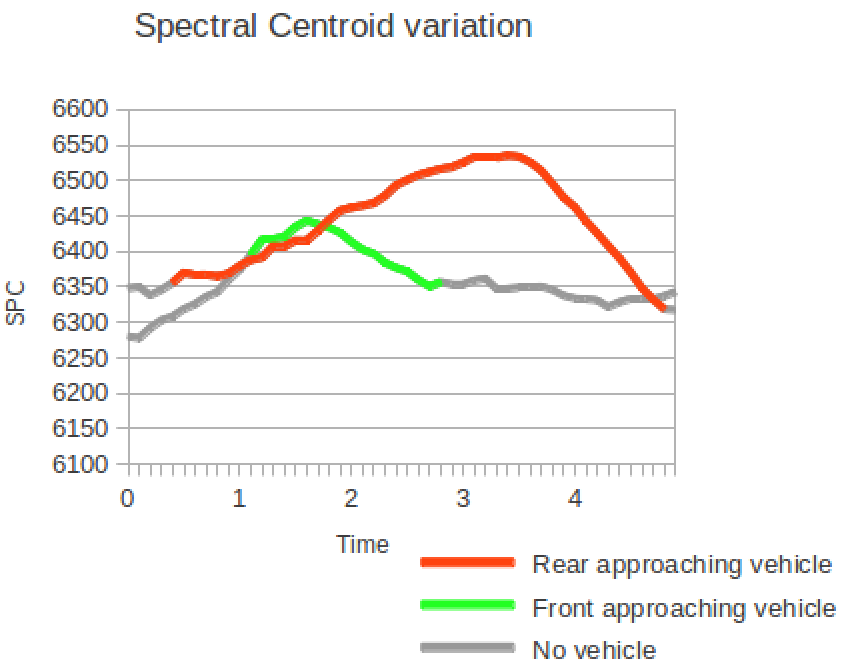

Figure 2.3: Spectral Centroid Variation When Vehicle Passes The Biker. The rear approach event causes the SPC value to rise gradually and more higher. The front approach event shows a shorter and lower rise in SPC.

frequency bin i. n indicates the number of frequency bins. SPC_{index} defines the index of the centroid frequency within the array of frequency bins.

As illustrated in figure 2.3, the centroid frequency rises whenever there is a rear approaching vehicle. A rear approaching event and a front approaching event are overlaid on top of each other and the figure clearly illustrates the differences between the events. The rear approach causes a slower and higher rise, when compared to the front approach.

Spectral Entropy (SPE) [25] is the average measure of entropy within the sound spectrum. Whenever there is a pattern in the sound, SPE tends to drop. When a vehicle approaches from the rear, it creates a definite pattern in sound, which drives the entropy down. It can be computed as:

$$SPE_f = -\sum_1^n p(i).log(\frac{p(i)}{\sum_1^n p(i)})$$

where,SPE_f represents the SPE for frame f, i represents the frequency index,

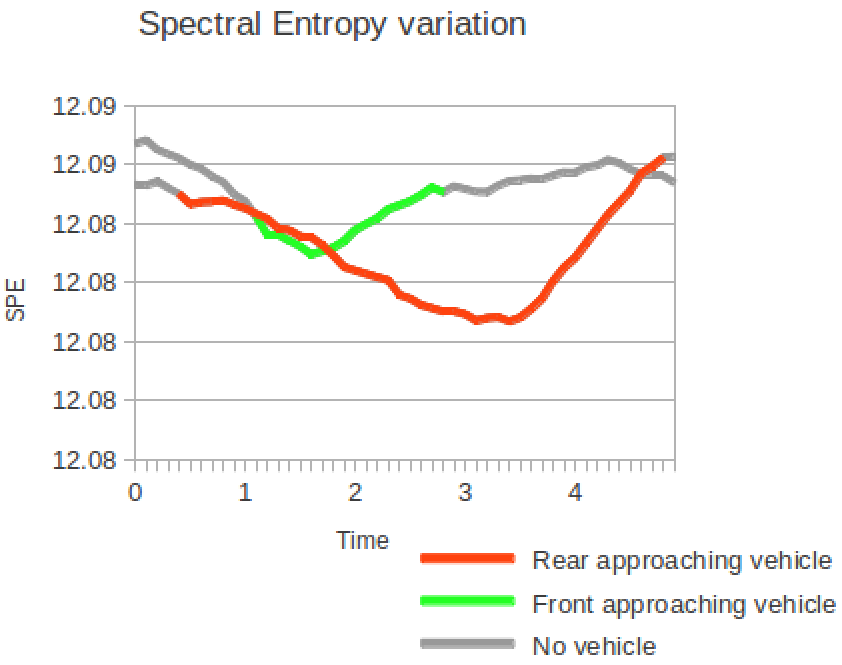

Figure 2.4: Spectral Entropy Variation When Vehicle Passes The Biker. SPE falls more gradually and deeply for a rear approaching vehicle, when compared to the short and shallow fall for a front approaching vehicle.

and $p(i)$ represents the magnitude of the frequency bin i. n indicates the number of frequency bins.

As the figure 2.4 illustrates, spectral entropy falls when there is a vehicle approaching the bicycle from rear. It is quite different from a front approach event, in which case the fall is shorter in duration, steeper and shallower in magnitude.

Reinforcing Features

This section describes the reinforcing features that show lesser correlation individually, though in combination with the rest of the features, boost the accuracy of prediction.

Zero Crossing Rate (ZCR) [23] is a time domain feature that measures the number of zero-crossings within the frame. In other words, it is the number of times the sign of the value changes within the frame. This can be calculated as :

$$ZCR_f = \sum_1^{n-1} \frac{sign(x_{i+1}) - sign(x_i)}{2}$$

where x_i stands for the sample value at ith index within the time domain frame and n represents the size of the frame. The number of zero crossings tend to show a higher variance when there are no vehicles passing the biker.

Spectral Flux (SPF) [38] describes the change in shape of the spectrum. It is defined as the L2-norm of the difference in spectral amplitude vector for two adjacent frames. It is calculated as:

$$SPF_f = \sqrt{\sum_1^n (p_f(i) - p_{f-1}(i))^2}$$

where SPF_f stands for the spectral flux for frame number f, $p_f(i)$ represents the magnitude of frequency bin i for frame f, and n represents the number of frequency bins. When there is a vehicle approaching the bicycle from rear, there is a slow and gradual change in the shape of the spectrum. SPF shows a small bump at the onset of a vehicle pass. However, SPF is susceptible to sudden changes in the spectrum due to the pedaling noise.

Spectral Roll-off Frequency (SPROFF) [23] represents the frequency bin under which majority of the spectrum is concentrated. We calculate it as the frequency below which 92% of the spectrum distribution is located. It can be defined as:

$$SPROFF_{index} = min(j | \sum_0^j p(i) > 0.92 \sum_0^n p(i))$$

where j stands for frequency index, n stands for number of frequency bins, $p(i)$ stands for the amplitude represented by ith frequency index. $SPROFF_{index}$ represents the index of the Spectral roll-off frequency within the array of frequency bins. During a rear vehicle approach, the spectrum tends to shift to the right, and hence the SPROFF values tend to show a gradual rise.

In addition to these features, we also use the first order time differentials of the above discussed features in order to reinforce accuracy. They indicate the rate at which each of these features change over time. This is helpful, as rear approach events tend to be much longer in duration and show a different rate of change for many features when compared to front approach events. In order to account for short bursts of noise produced by pedaling as well as bumpy road conditions, these feature values are smoothed using an averaging window, so that sudden fluctuations are leveled off.

2.3.2 Model Generation

For the prediction module to work accurately, we built a classification model based on a subset of data, kept aside only for this purpose. The first step towards creating a model, involved the selection of this training set. All of our data consists of cycling traces captured using a rear facing video camera attached below the seat of an ordinary road bicycle, while riding through some cycling routes over central Jersey. From these traces, we selected a set of clips, which consisted of isolated vehicles passing the bicycle from rear, and some similar clips with front approaching vehicles. We also added some clips where no vehicle is present behind the biker or can be heard. Since it is important to differentiate between front and rear approaching vehicles in addition to performing accurate detection, all three event types were represented in equal duration.

As a next step, we annotated all available data with event based information. For this, we watched all the clips and marked the event-times for vehicle approaches from both front and rear. The following details were noted:

- The time at which the rear approaching vehicle appears behind the biker
- The time at which the rear approaching vehicle can be heard for the first time
- The time at which the rear approaching vehicle passes closest to the biker
- The time at which the front approaching vehicle is heard for the first time
- The time at which the front approaching vehicle can no longer be heard

Once we have this information, we generate per-frame feature vectors for all audio in the training set. Using the timing information available through annotations, we associated the feature vectors with three classes:

1. Vehicle approaching from rear
2. Vehicle approaching from front

3. No vehicle approaching.

We used the decision tree classifier [44] in order to build the classification model. A decision tree model is represented as a tree of its nodes, where each node describes a constraint check to be performed on the feature values. The leaf nodes of the decision tree model represent predictions or output classes. We chose the decision tree model for our classifier because it proved to be simple, efficient and indicated higher accuracy over other classifier models we tried, including the NaiveBayes[44] and BayesNet[44]. Our classifier model utilizes two output classes while making a prediction:

Alert : This output class indicates that there is a vehicle approaching dangerously from behind the biker.

No Alert : This class indicates a situation of no danger to the biker. For example, it could mean that no vehicles are currently passing the biker from either direction. Alternately, it could indicate benign vehicle approaches, like those from the front of the biker.

We utilized the ten-fold cross-validation [44] technique to estimate accuracy rates of the model. This involves splitting the training set randomly into ten parts, where each class is represented approximately in equal proportions. Subsequently, the model is built by training on nine parts and testing it on the 10th. This is repeated ten times in total, where each part gets a turn to be the 10th part. The error rates and accuracy rates are then averaged over the ten iterations. Once an accurate static model was built, it is utilized by the prediction module to detect rear approaching vehicles.

2.3.3 Prediction

The prediction module utilizes the classification model in order to make predictions. It accepts a vector of feature values provided by the feature generator, and arrives at a prediction of whether a motor vehicle is approaching the cyclist from rear. Our prediction module is an implementation of the decision tree classifier. It takes a per-frame feature vector as input, and its output is one of the output classes: *Alert* or *No*

Alert. The classification model consists of a tree of nodes, where each node represents a constraint check on feature values. The classifier implementation loads this tree in memory, reads the input feature vector and traverses it while performing constraint checking on each node. After walking through the tree, it reaches down to a leaf-node which indicates a prediction.

We chose to use decision tree classifier because it is simple and displayed high accuracy. Another advantage of using the decision tree is that it can be optimized and hence can be very efficient and fast. The tree can be flattened into a long ladder of *if-else* structures in code after optimization, which reduces memory usage as well as execution time. This is especially useful when trying to implement on cheap and low power hardware.

2.3.4 Alert generation

The alert generation module accepts the stream of predictions provided by the prediction module. The prediction module generates predictions at the frame level of the audio stream. The frames represent very small intervals of time, typically ranging in 10s of milliseconds. In order to reduce the effect of incorrectly classified results of the individual frames, this layer removes small fluctuations in the prediction stream by performing a window based averaging over the stream. This smoothens the prediction stream and removes incorrect predictions due to minor fluctuations in input data or classifier results due to noise. Once an alert determination is made by this layer, the alert is propagated to the user. The formula used by this layer to generate the alert is described below:

$$alert_f = if(\sum_{f-4}^{f}(\frac{prediction(i)}{4}) \geq 0.75)$$

where f represents the frame number, $prediction(i)$ is the prediction for frame number i.

Chapter 3

Implementation

This chapter describes the implementation details of the audio based detection subsystem of the Cyber-Physical Bicycle. In the opening section, the hardware used to implement the system is described. This is followed by implementation details of the audio pipeline, which was built using Python 2.6. It then continues to describe other submodules used in capturing the audio and performing the feature generation. Subsequently, it discusses the way the classifier model is built and is used by the audio pipeline for making predictions.

Hardware

The audio based subsystem was implemented over the exact same hardware used by the Cyber-Physical Bike project. This was done so that an actual comparison can be made between audio and video based subsystems. Another reason is that the resulting implementation can be easily integrated into the existing Cyber-Physical Bike solution. The audio pipeline is built on a HP Mini 311 netbook, whose specifications include an Intel Atom N280 1.67GHz processor, 3 GB of RAM, an NVIDIA ION GPU, 8 GB Solid state disk for secondary storage, and an internal 6-cell Li-ion battery. This netbook is equipped with a microphone, which is used to capture the required audio. According to [40], this hardware was chosen for the following reasons:

- The netbook hardware approximates embedded hardware closely in performance.
- The hardware costs are relatively cheap.
- The netbook weighs about 3.26lbs, which is a reasonably light burden while riding. This weight includes unwanted components of the netbook like the display and

keypad, so the final weight can be further optimized.

All these reasons are also applicable to the implementation of the audio based detection system. In the case where audio and video based detection systems are to be combined, implementation over this hardware is useful in studing the feasibility of the combined system. However, it should be noted that individually, the audio subsystem will require much less CPU requirements, as well as smaller power requirements. It also requires no GPU for functioning.

The Audio Pipeline

As discussed before, the audio pipeline is implemented in Python 2.6. It is responsible for the overall operation of the audio detection subsystem. In order to function correctly, it utilizes four different modules: (i) the capture module, (ii) the feature generation module, (iii) the classification module and (iv) the alert generation module. The pipeline code handles the coordination and transfer of data between each of these modules. The pipeline is currently implemented as a single thread calling each of these four modules sequentially. A single frame of raw audio data is read from the capture module buffers, and passed to the feature generation module, which gives out a feature vector for that frame. This feature vector is then passed to the classification module which generates a frame specific prediction. This prediction is then sent to the alert generation module, which produces alerts. Despite being a single threaded implementation, our evaluations show that the system works in a realtime fashion and there is no buffering delay across the pipeline. Each of these submodules are described in detail below.

The Capture Module

The capture module is implemented in Python 2.6 using the PyAudio module. It opens the audio capture device on the system and continuously reads the captured audio as a single channel (mono) stream at a sampling rate of 22050 Hz. The data is read as signed 16 bit PCM samples at a frame size of 4410 bytes. This represents 100 milliseconds worth of audio data.

For generating time domain features, the raw frame of size 4410 bytes is given out. For the calculation of frequency domain features, this frame is padded equally on both sides with data from earlier and subsequent audio frames, so that the effective size of the frame is 8820 bytes. Then a hanning window is applied on this frame and sent out. As discussed before, this is to reduce the adverse effects caused by frame boundaries while performing FFT calculation. Since the frames are padded and slided over 4410 bytes at a time each, each frame still represents 100 milliseconds worth of data.

Feature Generation

The feature generation module is mostly implemented using the NumPy (Numeric Python) module in Python. Both time domain features as well as frequency domain features are generated using various functions in this module. FFT values are however generated by the optimized C library ”fftw”. This C library is loaded by the python code using the ”ctypes” module as a shared library. All necessary features are generated for each frame and composed together to form a feature vector. This vector is then returned to the pipeline.

Classification Module

The classification module was built in two steps. During the first step, a classifier model was built using the Weka machine learning toolkit, version 3.6.2. We chose the J48 implementation of the decision tree classifier to build a static model. This model can be converted to a C source implementation using the Weka toolkit. The generated C source is then compiled to create a shared library, which provides methods for making a classification using a feature vector as parameter. This C library is subsequently loaded into classification module through the ctypes Python module. The classification module can thus take a feature vector as input and produce a prediction as output.

Alert Generation

This module is written in Python and takes a stream of predictions as input. A fixed window of older predictions is remembered and constantly updated by this module. An alert is generated based on this window and the incoming prediction.

Chapter 4

Evaluation

This chapter discusses the results of evaluation performed on the audio based Cyber-Physical Bike prototype. In the beginning section, it lays down the goals of the evaluation process and explains what we intend to achieve through this evaluation. Then, it briefly describes the details of the test data and the methodology which was used to perform the evaluation. Results of the evaluation are then laid out, described separately based on accuracy, timeliness, performance and energy efficiency. Subsequently, we compare the audio results with available results of the video based detection system. In the final section, we explore a simple combination of both audio and video based detection systems and explore the resulting accuracy changes.

4.1 Goals

For an audio based Cyber-Physical Bike to be effective, it must be highly accurate in detecting rear approaching vehicles. Not only must it be accurate, it must be able to detect danger as early as possible so that the cyclist can be warned in a timely fashion. We define a metric called *timeliness* to address this particular concern. The detection system must not get confused by other similar events on the road and must generate alerts consistently. In other words, false alerts can affect the effectiveness of system adversely. Finally, the system must perform efficiently on the given netbook hardware and must have reasonable power consumption requirements, without which the detection system may not be realizable as a useful real world product. Our evaluation goals check each for these factors in order to establish the feasibility and effectiveness of the audio based Cyber-Physical Bicycle.

The main goals of our evaluation can be listed as follow:

- How accurately does the audio based detection system perform in detecting rear approaching vehicles?
- How early can the cyclist be alerted of danger?
- Can the system accurately perform in realtime?
- What are the power requirements of the system?
- How does the audio based detection system compare with video for accuracy, timeliness, efficiency and energy requirements?

4.2 Test Data And Methodology

In order to perform an actual comparison with video based detection, we use the exact same test data utilized by the original Cyber-Physical Bike project. As described in [40], the collected data accumulates to over three hours of real roadway cycling traces captured during day time, along typical cycling routes in central New Jersey. This amounts to roughly 10GB worth of data. A rear facing Sony Handicam DCR SX40 digital video recorder (which includes a microphone) was mounted on an ordinary road bike (Trex FX7.5) in order to collect this data.

After annotating the data as described by Section 2.3.2, we split the data into two separate sets for training and testing, where each set contained roughly equal number of rear approaching vehicle events. The training set was used to build the classifier model. Evaluation was performed on the test set. The test set contained 20 rear approach events. Vehicles approaching from the rear were heard for an average of 2.83 seconds per event and where seen for an average of 3.8 seconds.

Using the annotations as ground truth, we measured the accuracy of the system as well as its timeliness by comparing the alerts produced with the ground truth. In order to measure performance, various components of the audio detection pipeline where timed by running them through 1000 frames of audio. This was done by adding timestamp generation code inside the audio pipeline. The measured time was then converted to frames per second for each component. Battery related results were measured by using the tool Battery Bar[43] for Windows 7. This tool can reveal both the discharge rate of the battery as well as the current battery charge levels. Measurements were made by charging the battery to the same level (75%) before each round, so that any variation in discharge rate at different levels does not interfere with our evaluation.

4.2.1 Results

This section describes the evaluation results in detail. The results are separated into detection accuracy, timeliness of alert generation, end-to-end realtime performance and energy efficiency.

Detection Accuracy

In this section, we evaluate the accuracy level of the audio based detection system. Table 4.1 describes the confusion matrix for audio at the event level. True positives (TP) describe the scenario where an alert was generated correctly for a rear approaching vehicle. True negatives (TN) represent cases where the absence of a rear approaching vehicle was correctly detected. False positives (FP) describe spurious alerts generated without any rear approaching vehicle behind the biker. False negatives (FN) describe the case where a vehicle approached the biker from behind and no alerts were generated. We are unable to provide values for true negatives here, since absence of vehicles cannot be quantified based on number of events.

	Positives	Negatives
True	18	N/A
False	2	2

Table 4.1: Detection Accuracy For Audio

We define accuracy using the formula provided in [40] as

$$Accuracy = \frac{TP}{TP+FP+FN}$$

Using this formula, we calculate the accuracy for audio based detection as **81.8%**.

In order to understand the weaknesses of the system, a specific examination of the various error cases was conducted. After closer inspection, we found that the first false negative is caused by a very slowly approaching car, which slows down after a while and turns away from the bike before ever reaching dangerously close to the biker. This was an instance where a rear approaching vehicle was annotated overly conservatively in the trace data, because the car never actually gets close enough to the biker to be a danger. In another interesting instance, there is a car approaching very slowly from behind and a number of vehicles pass the bike from the opposite direction in rapid succession. The sound from the rapid car passes drowns out the slow passing sounds of the behind car, confusing the classifier. This leads to a false positive where the classifier detects a front approaching car as danger and generates the alert. The classifier also produces a false negative here because it actually misses the rear approaching car altogether .The final false positive can be attributed to sudden noise generated by the bike passing over a really bad patch of road. From this analysis, it becomes clear that the audio based detection system shows high accuracy, but is susceptible to the noise created by front approaching vehicles. It can also be overwhelmed by loud noise caused by bad road conditions. The two definite areas of improvement for the system should be (i) better differentiation of front and rear approaches, and (ii) overcoming the adverse affects of suddenly changing road conditions. Some possible approaches to deal with these limitations are discussed in 5.1 and 5.2.

Timeliness Of Detection

We define *timeliness* of the detection system as the amount of time available to the cyclist after the alert is generated, before the approaching vehicle passes the biker at its closest. In a similar vein, we can also define *potential timeliness* as a percentage of the maximum possible *timeliness* achievable by a given detection system. For an audio based detection system, maximum *timeliness* is the entire duration for which the vehicle was heard before it passed the biker. The essence of measuring *timeliness* is that it represents the amount of time available to the cyclist after an alert, to avoid a possible accident.

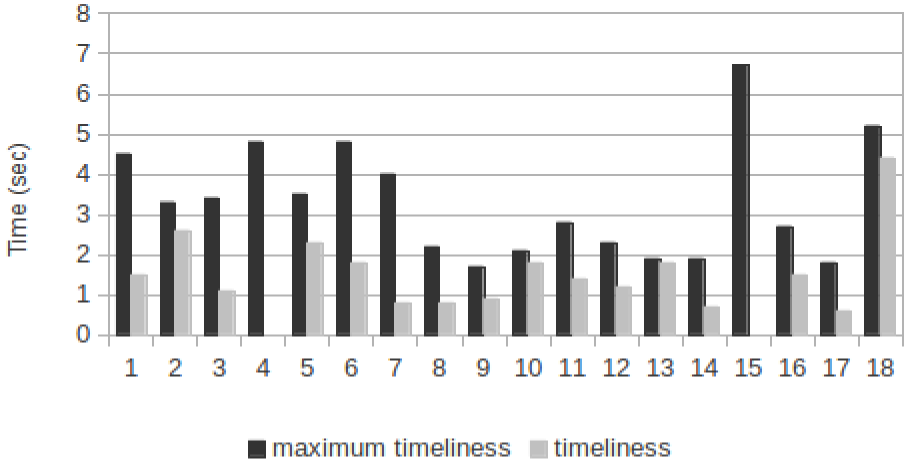

Figure 4.1: Timeliness Of Audio Based Detection
Graph plotting timeliness and maximum timeliness for audio based detection. Maximum timeliness indicates the maximum achievable timeliness for audio, which is the duration for which each vehicle approach was heard.

The figure 4.1 represents the timeliness of the audio based detection system for selected events. A particular set of events in our data correspond to a close succession of cars passing the biker from behind, one after the other. Their approach sounds overlapped and the system raised continuous alerts spanning the entire duration. Due to this, it was not possible to determine alert times specific to each car, except for the first one. It was also difficult to determine the exact time at which the cars behind became audible due to the overlapping sounds. Hence, to be conservative, we excluded such events in our evaluation of *timeliness* and *potential timeliness*.

The average *timeliness* for audio was 1.4 seconds, at an average *potential timeliness* of 47%. The median *timeliness* is 1.3 and median *potential timeliness* is 43.8%. This indicates that audio based detection system waits about half the available time before generating an alert. While it is debatable whether 1.4 seconds is sufficient time to take evasive action, future work must focus on improving the *potential timeliness* of the system to as close to 100% as achievable. The lowest observed *timeliness* was 0.6 seconds with a 33% *potential timeliness*. In addition to improving the average *timeliness*, it is also important to improve the minimum *timeliness* observed. This is because if this falls too low, the alert is effectually useless to the cyclist.

One of the reasons for *potential timeliness* being so low for audio is the fact that the classifier component tries to differentiate between front approaches and rear approaches using features that tend to depend on the length of the event. For example, both spectral centroid and spectral entropy show a longer variation for a rear approach event, while showing a similar, though shorter, variation for front approaches. This dependence on the length of the event causes the delay in prediction, which lowers *potential timeliness*. In order to improve *timeliness*, we may need to use other features that differentiate front and rear approaches without depending directly on the duration of the event. Microphone arrays[10] have previously been successfully used to determine the source location of sound. Parabolic microphones, which use a reflector to focus sounds, can be utilized to increase capture sensitivity. Even though this means vehicles can be heard earlier, it could mean reduced accuracy due to increased noise levels. Some other approaches for this are discussed in 5.1.

Realtime Performance

This section describes the end-to-end performance costs of the audio based detection system, as it runs on the netbook hardware. It also investigates whether the system is able to run in a realtime manner, without any buffering delay.

The audio based detection system was able to perform at the required 10 frames per second over three trials of 5 minutes each. Figure 4.2 indicates how fast each of the various components in the audio pipeline can be calculated. The slowest component is the calculation of FFT, which can only be processed at a speed of about 100 frames per second. We only need to process 10 frames per second to become realtime.

The real world implications of this result are that we can afford to build an audio based system using much slower hardware. Slower hardware usually uses less power, and is comparatively more affordable. This drives down the cost as well as saves much valued battery life. Lower costs and longer battery life translate to improved overall user experience when building a real world product.

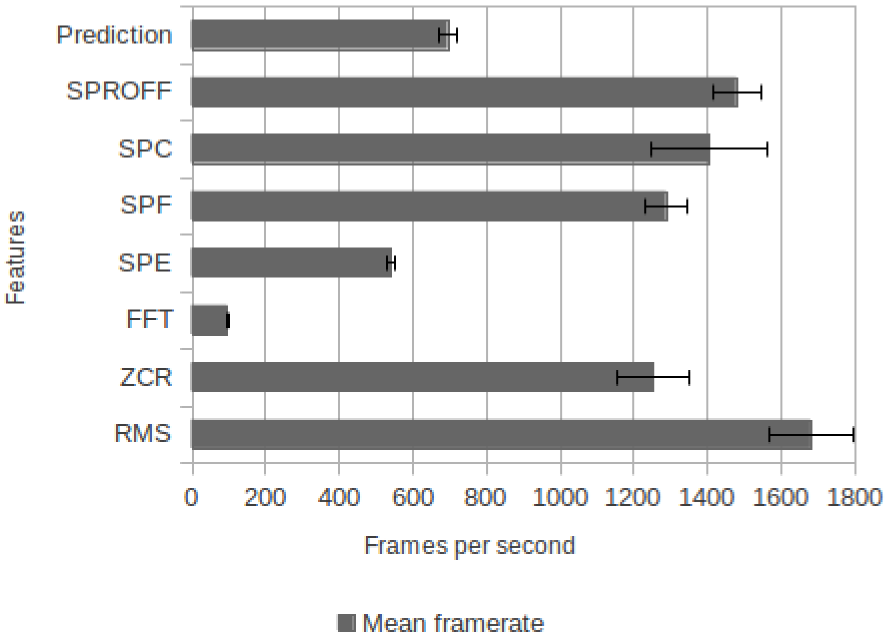

Figure 4.2: Audio Performance
Graph plotting maximum frames per second possible for various audio processing components. The top most bar represents prediction by the classifier model, and the rest of the bars represent various other audio features.

Energy Efficiency

This section describes the evaluation results related to energy requirements for audio based detection. Table 4.2 presents the power consumption rate for audio based detection in comparison with CPU idling, a movie player playing a 480p video as well as video based detection system (based on data from [40]). Clearly, audio based detection utilizes very low power in order to function. The power consumption is close to that of an idling machine, which represents a large advantage in terms of battery life.

	Power consumption in Watts
Idle	6.4
Audio	8.5
Movie player	12.4
Video	14.2

Table 4.2: Power Consumption For Audio Detection

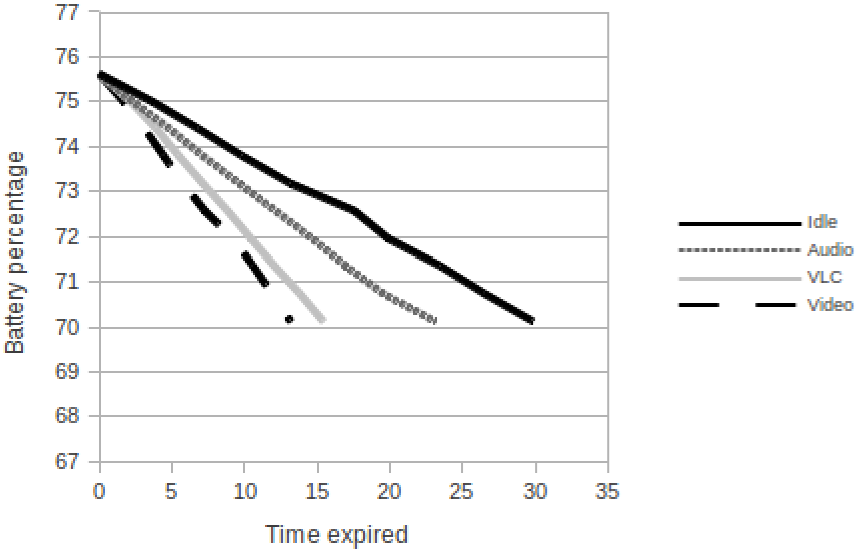

Figure 4.3: Battery Percentage Drain For Audio Based Detection
Graph which plots how fast battery is drained for audio based detection. For comparison, battery drain during idling, a movie player playing a 480p video and video based detection system are presented together.

Figure 4.3 represents the rate at which the battery discharges when audio based detection is running in the system. To avoid any adverse effects of variation in battery discharge rates, all tests were conducted by recharging the battery to the same initial charge levels. In comparison to other usecases like watching a movie, the battery discharge rate is much lower. This indicates that the system can keep on running without needing a recharge for much longer. Another advantage is that when building a real world product based on audio, the system can afford to use a much lower capacity battery. This reduces both the size and weight of the system as well as the cost of the system. In other words, higher energy efficiency and performance imply that an actual product built based on audio can be comparably lighter, smaller and cost-effective than a video based system.

4.2.2 Comparison With Video

In this section, we will investigate how the existing video based detection system compares with the audio based detection system. We will make the comparison based on four different aspects of the system: (i) accuracy, (ii) timeliness, (iii) realtime end-to-end performance, and (iv) energy requirements.

From [40], we know that the accuracy rate for video based detection is 73.1%. Table 4.3 compares the accuracy results of video obtained from [40] with that of audio. As we can see, the results are very close to each other. Video tends to perform marginally better at detecting rear approaches compared to audio, while audio tends to perform more consistently with less false alerts generated.

	True Positives	False Positives	False Negatives
Audio	18	2	2
Video	19	6	1

Table 4.3: Detection Accuracy: Audio Vs Video

When it comes to timeliness of alerts, video clearly outperforms audio. From [40], we know that the average *timeliness* for video alerts is 3.5 seconds, which is well above the 1.4 seconds observed for audio. Any additional time available to the cyclist can make crucial differences when it comes to avoiding accidents. One of the reasons video outperforms audio is the fact that vehicles tend to become visible much more earlier than when they become audible. They can be seen by the camera from much larger distances before the sound from them can make a noticeable impact on the microphone. This can potentially be rectified by using more sensitive microphones to some level, though we may have to deal with an increase in the associated noise levels. Video also outperforms audio when it comes to *potential timeliness*. Video based detection produced alerts as early as 92% of maximum achievable time, where as audio only managed 47%. In order to become as effective as video, audio based detection must improve on both these metrics. It is possible that audio may never achieve the high *timeliness* levels achievable by video, since vehicles can be seen at much longer distances in case of video. However, it should be enough for audio to achieve a consistently high level of *timeliness*, where the cyclist has a reasonable amount of time to take preventive action. This can be achieved by improving *potential timeliness* to close to 100%, as the average time that audio was heard for a rear approaching vehicle was 2.83 seconds.

Another significant difference between audio and video is the inability of audio to determine the level of danger of a rear approaching vehicle. According to [40], the video based system tracks the rear approaching vehicle and produces an alert only when the vehicle enters the *virtual safety zone*, which is defined by a 3-feet zone around the biker. However, in our evaluations, we observe that audio alerts are generated much later than the corresponding video alerts, and hence the vehicles were much more closer to the biker compared to video. This indicates that any approach to generate alerts based on level of danger must first improve overall *potential timeliness*. To further differentiate rear approaches based on danger, a possible approach is to annotate the data specifically based on the vehicle entering the *virtual safety zone* rather than when they are heard. We can further create a prediction model based on these annotations, which can differentiate a benign rear approaching vehicle from a dangerous one. However, it is quite possible that the vehicle begins to be heard only after it has entered well into the virtual safety zone. In such cases, we propose the use of more sensitive microphones to capture audio. This could lead to a trade-off in accuracy, because the increased sensitivity also increases the amount of noise captured by the system.

An area where audio outperforms video is realtime end-to-end performance. Audio runs at full realtime frame rate, while the prototype implementation of video based detection is running at a much lower rate of 3.6 FPS [40]. At full speed, video must operate at 30 frames per second. This indicates that video based must detection must tradeoff accuracy in order to run at full rate on the given hardware[40]. Compared to this, audio already runs at full speed on this hardware, and result indicates that it can be realized on much slower hardware. The main reason for this disparity is that audio processing algorithms tend to handle much less data when compared to video and tend to be computationally much cheaper compared to video algorithms.

Audio also outperforms video when it comes to power requirements, as is clear from table 4.2. This can be attributed to the large difference in computational requirements, which are very low for audio. Audio also tends to use much less Random Access Memory (RAM), as the amount of data involved in the process is much low in comparison. Another key difference is the usage of GPU by the video based detection model, which is used by the tracking component[40].

From this discussion, it becomes clear that audio based detection is much more hardware efficient and power saving than video. It displays comparable accuracy to video, but needs improvement in the aspect of generating alerts early.

4.2.3 Augmenting Video

In this section, we will discuss the advantages and disadvantages of combining audio and video detection systems in a simple manner. For the purpose of this evaluation, we imagine that both systems were installed on the bike separately and were generating alerts independently of each other. The purpose of this experiment is to investigate the motivations for a combined multi-modal approach as proposed in [40]. In order to evaluate, we investigate each event case-by-case for both audio and video and record the results of their simple combination. Obviously, it is sufficient for one of the systems to generate a correct alert for successful functioning. On the other hand, both systems need to be consistent in order to achieve overall consistency. As both systems are attached to the bike as two separate systems, their energy and efficiency constraints are independent of each other, and hence not considered for evaluation.

The results of the experiment are as shown in table 4.4. The overall accuracy of the system is 74%. This is close to the 73.1% reported by video[40], but lower than the 81% reported by audio.

	Positives	Negatives
True	20	N/A
False	7	0

Table 4.4: Accuracy For Simple Combination Of Audio And Video

One interesting observation from this experiment is that the number of false negatives drop to zero. The resulting system did not miss any of the rear approaching vehicle passes. The single case of missed alert from video is due to another rider following the cyclist who occludes the rear approaching vehicle[40]. Audio is unaffected by this and a correct alert is generated. The two cases of false negatives where audio detection missed rear approaching cars were due to sudden road noise and noise created by front approaching cars. Video was unaffected by this and produced correct alerts. The resulting *timeliness* for the combined system is much closer to that observed on video. This is because video generates alerts much earlier than audio for almost all observed cases.

A slight disadvantage to this combination was the increase in the total number of false alerts. This is because false alerts in this type of combination have an adverse cumulative effect, where as true positive cases tend to rescue each others faults.

From this discussion, it becomes clear that there is a strong case towards combining audio and video towards a multi-modal detection system. Overall vehicle detection rate and timeliness of alerts improved, at the cost of consistency. Since the power requirements and efficiency requirements for audio are much low compared to video, the costs of accommodating audio into an existing video based system could not be too prohibitive.

Chapter 5

Future Directions

In this chapter, we present several future directions that can be explored in order to expand the research proposed by this thesis. Firstly, it presents various ideas that can potentially improve timeliness and accuracy of the audio based detection system. This is an important direction to work on, as improved accuracy and timeliness is directly proportional to the effectiveness of the alert system. Subsequently, we present some interesting approaches towards combining audio and video into a single multi-modal detection system.

5.1 Using Multiple Microphones

Section 4.1 suggests that audio based detection mechanism suffers from low potential timeliness. In other words, alerts are generated much later than what is ideally possible, and the cyclist gets much less time to take preventive action before the vehicle passes the bike. This is mostly due to the fact that audio features that are used for prediction have to distinguish between rear and front approaches based on the duration of the event. For example, figure 5.1 indicates the variation of SPC for front approaches and rear approaches. Clearly, the classifier is dependent on the duration of the event to make a positive distinction between the two cases, since the feature shows a similar variation in the beginning of the curve. Any prediction can be made only after ruling out a front approach event. This affects the timeliness of the alert adversely.

To improve this situation, we can distinguish rear and front approaches through other indirect methods. One possible way to do this is to use two microphones, one turned towards rear and another towards front. The front facing microphone must be sufficiently shielded from the wind for this to be effective. Once set up, we can

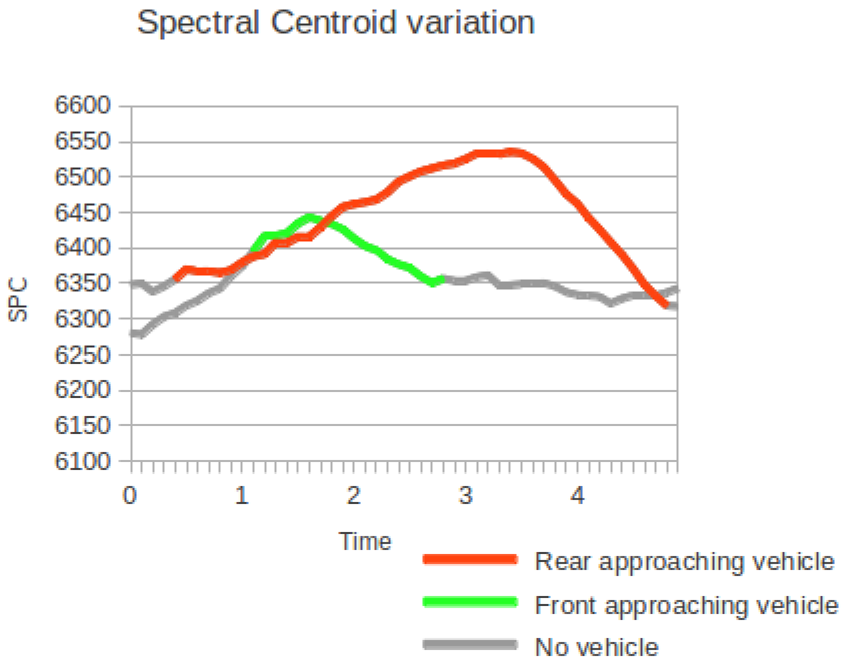

Figure 5.1: Spectral Centroid Variation When Vehicle Passes The Biker.
The rear approach event causes the SPC value to rise gradually and more higher. The front approach event shows a shorter and lower rise in SPC. (This figure is the same as figure 2.3, copied here for the reader's convenience)

differentiate front and rear approaches by comparing the features generated by the two microphones. This works due to the fact that both microphones will register the events slightly differently. A front approaching vehicle will clearly have a bigger impact on the front facing microphone. This method of distinguishing front and rear approaches should work faster than the current approach, since we no longer depend on duration of the event.

The addition of another microphone brings only a small increase in hardware costs. Since audio processing algorithms are relatively light, the additional load on the system will not make much impact.

5.2 Filtering Frequency Bands

Another direction of research, which has the potential to improve the effectiveness of the system, is to filter out the effect of road noise. Instead of focusing on the entire spectrum, it is possible to isolate the bands where the noise due to the road surface are prominent, and filter them out before generating features. Since most road surface noise is low in the frequency, a well designed high-pass filter or a band-pass filter can cut out the impact of the road surface. A high-pass filter attenuates the lower frequencies below a certain threshold frequency, while permitting the higher frequencies to pass. A band-pass filter works similarly, though it can attenuate all frequencies except a given band of frequencies within the spectrum. Once the spectrum is free of bands strongly influenced by noise generated by road surface, the generated features are less affected by the noise.

However, while selecting the frequency bands to avoid, care should be given so that we do not filter out those bands that are strongly influenced by rear approaching vehicle activity. The overall effect of such filtering, if successful, will be an increase in accuracy and consistency. A sudden rough patch on the road will not confuse the detection system or create false alerts.

5.3 Combining Audio And Video

In section 4.2.3, we presented some preliminary results of a simple combination of audio and video, in order to investigate the benefits of a combined multi-modal approach proposed by [40]. The results indicate a slight overall accuracy improvement over video, and most importantly, the resulting system eliminated all false negatives. In other words, the combined system detected all rear approaching vehicles correctly. In this section, we describe possible design directions for a Cyber-Physical Bike that combine audio and video detection systems together. Two approaches are presented in the following sections. Firstly, both the systems can be treated as black boxes and combined at the alert level. A second way of combination involves integrating feature level information from both levels within each other. In the final section, we attempt to tackle the issue of performance, which arises when we combine audio and video together. In order to overcome the performance constraints, we propose a cloud based approach, which pushes costly computation to a central server.

5.3.1 Alert Level Combination

This architecture treats the two detection systems as independent, non interacting systems. The combining mechanism looks at alerts generated from both systems and performs a linear weighted combination to maximize accuracy, timeliness and consistency. A resulting architecture is described by the Figure 5.2. As an addition to this model, the audio and video subsystems can provide accuracy estimates depending on environmental conditions so that the combiner can perform the alert generation using dynamic weights. For example, the system may give more weightage to audio during failing light conditions, under which video accuracy may be low. On the other hand, the system may give more weightage to video under severely noisy conditions where audio accuracy could be inconsistent.

The advantages of this system is that it is quite simple to build. We can upgrade or replace either of the individual detection systems without affecting the implementation of each other. The only change required is that the the weights associated with each type

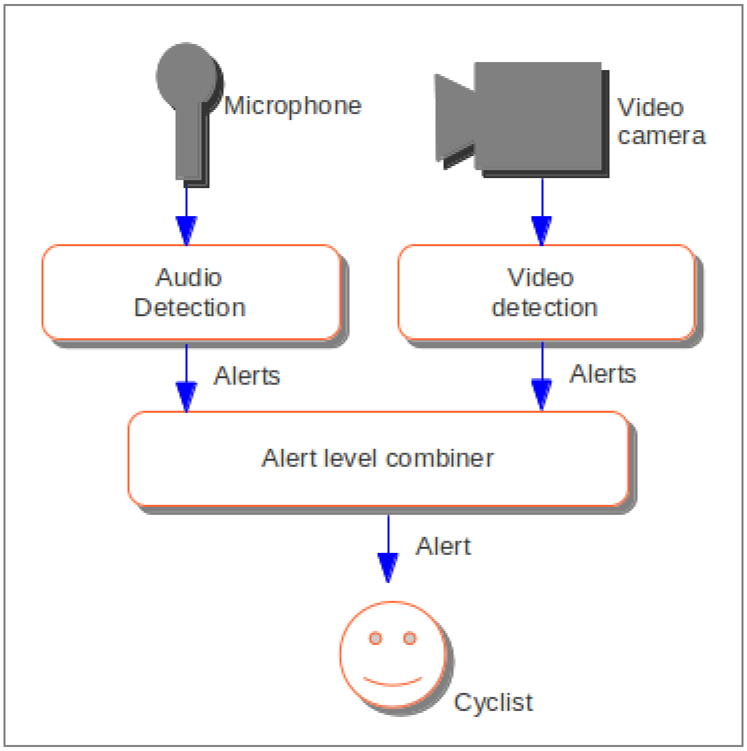

Figure 5.2: Alert Level Combination Architecture

of alert may change. This is also one of its disadvantages, as feature level information from one system can be useful in performing optimizations within the other system, thus improving the overall efficiency.

5.3.2 Feature Level Combination

It is possible to create a truly integrated audio and video detection system that uses each others internal features to boost accuracy or increase efficiency. The architecture can be visualized as each subsystem providing feedbacks to each other to improve accuracy or efficiency. This form of cross-modal feedback to increase the adaptivity of the system is also suggested by [40]. The figure 5.3 represents a feature level combining architecture for audio and video subsystems.

As an example for feedbacks, audio RMS levels can give us an indication of whether the bike is parked or moving. This information can be used by video to switch off its operation while being in parked state, in order to save energy. In a different approach,

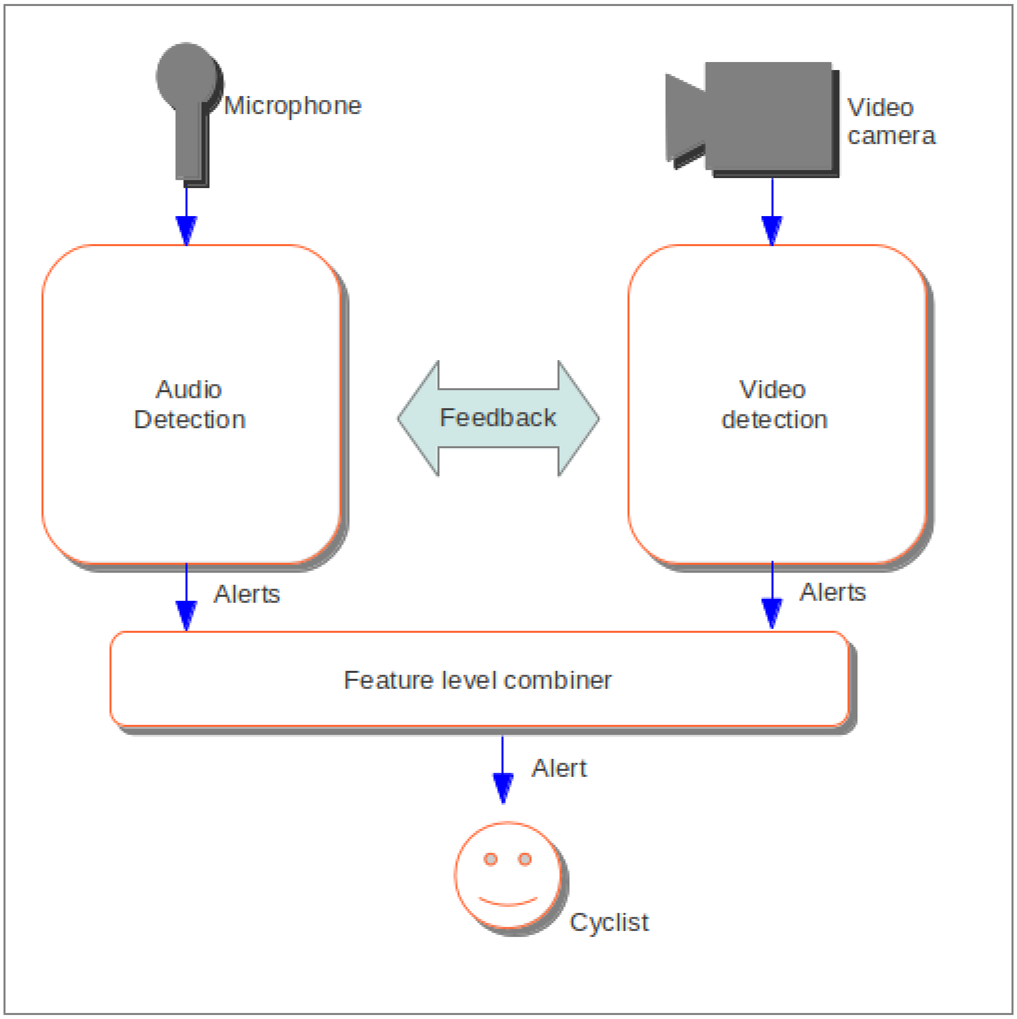

Figure 5.3: Feature Level Combination Architecture With Feedback

video could reduce its input frame rate until audio starts a positive detection, which can save valuable energy. The same trick can be applied during low light conditions, when video detection accuracy may be low. Low light conditions can be determined by the video detection subsystem, or could simply be based on timezones. The system then switches to audio based detection as its primary alert mechanism. The same technique can be applied in the reverse direction; if video accuracy levels are high enough, then audio detection could altogether be switched off to save energy. As an addition to this design, we can build a feature level combiner that uses a truly integrated classifier model based on features generated from audio as well as hints from the video detection system. In future, this sort of architecture helps us to bring in more information from other sensors like the gyroscope, GPS or accelerometer into the system, so that the full power of combined multi-model sensing can be utilized.

The upside to this design is that it can combine audio and video features in unique and advantageous ways, which could result in better performance and accuracy. However, this design is rather complex and both detection subsystems are no longer independent with respect to each other. Any improvement in video or audio detection systems cannot be directly integrated into the system. For example, if a key feedback for the integrated system depends on one of the specific audio features, there may be some difficulty in replacing or removing that feature from the audio system in the future.

5.3.3 Cloud Based Prediction

In a combined system that utilizes both audio and video, the associated performance costs can drive up the costs of the hardware. It is possible that we may need to make trade-offs on accuracy in order to comply with the performance constraints. However, with fast wireless network technologies like 3G/4G available through smartphones nowadays, we can solve this problem by pushing computationally intensive operations to the cloud. In this design, the sensors in the Cyber-Physical Bike capture streams of realtime information that includes video and audio, and send it to a central server.

This server analyses the information, performs costly operations like generation of audio features, optical flow analysis and vehicle tracking, and sends back alerts to the Cyber-Physical Bike. Current mobile wireless technology is capable of supporting such bandwidth and latency requirements, as demonstrated by the introduction of video calling applications like Skype[39], which work over 3G.

A cloud based approach has the advantage that it can afford more computationally complex predictive models and feature generation. This can bring in improvements in accuracy. As an added benefit, the centralized nature of the system encourages more elaborate model construction, using accumulated historical data from multiple users. This could lead to improved models, which can handle more edge-cases. However, a cloud based approach requires consistent network quality through out the duration of the bicycle ride. This issue becomes more relevant to rural bike routes, where network coverage may be incomplete or inconsistent. In addition to this, a cloud based approach also has to deal with network latency issues, which can vary across the bike route. Any delay in the network can potentially delay alert generation and reduce the effectiveness of the alert.

Chapter 6
Conclusion

In this thesis, we introduced the design of an audio based Cyber-Physical Bicycle to improve bicycle safety. We proposed a system that augments the ordinary road bicycle with a microphone and computational capabilities, in order to detect rear approaching vehicles automatically and alert the cyclist. The aim of this system is to reduce the cognitive load on the biker while riding on roads amidst other vehicular traffic. The design of the system involved generation of audio features that correlate well when a vehicle approaches the biker from behind. A classification model was then constructed based on these features which predicts a rear approaching vehicle.

Through evaluation of our prototype system, we determined the feasibility and effectiveness of the system, and also compared it with the existing video based detection system proposed by the Cyber-Physical Bike project[40]. In addition to this, we investigated the potential for a combined multi-modal detection system comprising both video and audio, by evaluating a simple combination of the audio detection system with the results of the existing video based detection.

The key conclusions of this thesis are as follows:

- The evaluation of our prototype clearly suggests that an audio based detection system for the Cyber-Physical Bicycle is feasible.
- The system performed with high level of accuracy (81%) during our evaluation.
- The average amount of time provided to the cyclist to take preventive action after an alert was 1.4 seconds. The lowest was 0.6 seconds. This time forms a low percentage (47%) of the duration for which the vehicle can be heard before it passes the biker. This is a limitation to the system, which requires improvement.

- The system can function and produce alerts in realtime, while running at the required frame rate without any buffering delay. The system does not utilize the hardware to its full, and indicates that it can be realized on much slower hardware.

- The system can function with very low power requirements, and presented a battery discharge rate of 8.5 Watts. This discharge rate is very close to the idling requirements for the hardware we used, which indicates that the system can run for several hours without requiring a recharge. Additionally, the system can be realized with much lower capacity batteries.

- The three audio features presented in the design, Spectral Centroid, Spectral Entropy and Root-mean square amplitude display good correlation with a rear approaching vehicle event. Other features including Zero Crossing rate, Spectral Roll-off and Spectral Flux together reinforce the accuracy of prediction.

- The system is susceptible to the road surface noise as well as noise caused by front approaching vehicles.

- When compared to the video based detection system, the audio based system performed with a comparable accuracy. (81% vs 73% displayed by video)

- When it came to *timeliness* of alerts, which indicates the time provided to the cyclist to take preventive action, the video based detection system performed much better when compared to audio. (3.5 seconds when compared to 1.4 seconds by audio).

- The audio based system performed with much better end-to-end performance and significantly lower battery requirements when compared with video.

- When we combined both detection systems trivially, the resulting system performed with slightly higher (74%) overall accuracy than video and eliminated all false negatives. In other words, the combined system generated alerts for all rear approaching events without miss. The system could also generate alerts much more earlier than audio. This indicates that there is a strong case for building a multi-modal detection system by combining video and audio together.

We found that the audio features that we currently use to distinguish rear and front approaching vehicles are dependent on the duration of these events. This is causing the alerts to be delayed. In order to improve this, we suggest the use of multiple microphones, one facing the rear and another facing the front, as a possible future way of distinguishing front and rear approaches. In order to reduce the effect of road surface noise, an approach based on filtering out noisy frequency bands is suggested. The thesis also suggests future design directions for a combined multi-modal approach based on a simple weighted alert level combination, as well as a more integrated, feedback based feature level combination.

References

[1] 11 most bike friendly cities in the world. http://www.virgin-vacations.com/site_vv/11-most-bike-friendly-cities.asp.

[2] Bikely. http://www.bikely.com/.

[3] Cyclopath. http://cyclopath.org/.

[4] Mapmyride. http://www.mapmyride.com/.

[5] veloroutes.org. http://veloroutes.org/.

[6] A. Amditis, M. Bimpas, G. Thomaidis, M. Tsogas, M. Netto, S. Mammar, A. Beutner, N. Mo andhler, T. Wirthgen, S. Zipser, A. Etemad, M. Da Lio, and R. Cicilloni. A situation-adaptive lane-keeping support system: Overview of the safelane approach. *Intelligent Transportation Systems, IEEE Transactions on*, 11(3):617 –629, sept. 2010.

[7] C. Bahlmann, Y. Zhu, Visvanathan Ramesh, M. Pellkofer, and T. Koehler. A system for traffic sign detection, tracking, and recognition using color, shape, and motion information. In *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*, pages 255 – 260, june 2005.

[8] N. Barnes and A. Zelinsky. Real-time radial symmetry for speed sign detection. In *Intelligent Vehicles Symposium, 2004 IEEE*, pages 566 – 571, june 2004.

[9] Biking Bis. 19 states require 3-foot clearance for bicycles. http://www.bikingbis.com/blog/_archives/2008/3/5/3549263.html.

[10] M.S. Brandstein, J.E. Adcock, and H.F. Silverman. A closed-form location estimator for use with room environment microphone arrays. *Speech and Audio Processing, IEEE Transactions on*, 5(1):45 –50, jan 1997.

[11] Cerevellum. Cerevellum hindsight 35 cyclometer with digital rear-view mirror. http://www.cerevellum.com.

[12] DARPA. Urban challenge, 2007. http://archive.darpa.mil/grandchallenge/index.asp.

[13] A. de la Escalera, L.E. Moreno, M.A. Salichs, and J.M. Armingol. Road traffic sign detection and classification. *Industrial Electronics, IEEE Transactions on*, 44(6):848 –859, dec 1997.

[14] Shane B. Eisenman, Emiliano Miluzzo, Nicholas D. Lane, Ronald A. Peterson, Gahng-Seop Ahn, and Andrew T. Campbell. Bikenet: A mobile sensing system for cyclist experience mapping. *ACM Trans. Sen. Netw.*, 6, January 2010.

[15] M. Enzweiler and D.M. Gavrila. Monocular pedestrian detection: Survey and experiments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(12):2179 –2195, dec. 2009.

[16] US DOT Federal Highway Administration. National household travel survey, 2009.

[17] Jonathan Foote. An overview of audio information retrieval. *ACM Multimedia Systems*, 7:2–10, 1998.

[18] Insurance Institute for Highway Safety. Fatality facts for bicycles, 2008. http://www.iihs.org/research/fatality_facts_2008/bicycles.html.

[19] D.M. Gavrila and V. Philomin. Real-time object detection for ldquo;smart rdquo; vehicles. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, pages 87 –93 vol.1, 1999.

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

[21] Bicycle Helmet Safety Institute. Helmet laws for bicycle riders. http://www.bhsi.org/mandator.htm.

[22] C. Komanoff. Killed by automobile. death in the streets in newyork city 1994-1997. right of way, 1999.

[23] Dongge Li, Ishwar K. Sethi, Nevenka Dimitrova, and Tom McGee. Classification of general audio data for content-based retrieval. *Pattern Recogn. Lett.*, 22:533–544, April 2001.

[24] Chin-Teng Lin, Ruei-Cheng Wu, Sheng-Fu Liang, Wen-Hung Chao, Yu-Jie Chen, and Tzyy-Ping Jung. Eeg-based drowsiness estimation for safety driving using independent component analysis. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 52(12):2726 – 2738, dec. 2005.

[25] H. Lu, W. Pan, N. D. Lane, T. Choudhury, and A. T. Campbell. Soundsense: scalable sound sensing for people-centric applications on mobile phones. In *MobiSys 09: Proceedings of the 7th international conference on Mobile systems, applications, and services*, June 2009.

[26] Ling Ma, Ben Milner, and Dan Smith. Acoustic environment classification. *ACM Trans. Speech Lang. Process.*, 3:1–22, July 2006.

[27] Ling Ma, Dan Smith, and Ben Milner. Environmental noise classification for context-aware applications. In *In Proc. EuroSpeech-2003*, pages 2237–2240, 2003.

[28] Michael Montemerlo, Sebastian Thrun, Hendrik Dahlkamp, and David Stavens. Winning the darpa grand challenge with an ai robot. In *In Proceedings of the AAAI National Conference on Artificial Intelligence*, pages 17–20, 2006.

[29] US DOT National Highway Travel Safety Authority. Travel safety facts 1996, 1996.

[30] US DOT National Highway Travel Safety Authority. National survey of pedestrian and bicyclist attitudes and behaviors, 2002.

[31] US DOT National Highway Travel Safety Authority. Travel safety facts 2008, 2008.

[32] US DOT National Highway Travel Safety Authority. Travel safety facts 2007, 2009.

[33] US DOT National Highway Travel Safety Authority. Travel safety facts 2009, 2009.

[34] The official Google blog. What we are driving at, 2010. http://googleblog.blogspot.com/2010/10/what-were-driving-at.html.

[35] C. OUTRAM, C. RATTI, and A. BIDERMAN. The copenhagen wheel: An innovative electric bicycle system that harnesses the power of real-time information an crowd sourcing. *EVER Monaco International Exhibition & Conference on Ecologic Vehicles & Renewable Energies*, 2010.

[36] Sasank Reddy, Katie Shilton, Gleb Denisov, Christian Cenizal, Deborah Estrin, and Mani Srivastava. Biketastic: sensing and mapping for better biking. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 1817–1820, 2010.

[37] J. Saunders. Real-time discrimination of broadcast speech/music. In *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, volume 2, pages 993 –996 vol. 2, may 1996.

[38] E. Scheirer and M. Slaney. Construction and evaluation of a robust multifeature speech/music discriminator. In *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, volume 2, pages 1331 –1334 vol.2, apr 1997.

[39] skype.com. Skype brings video calling to iphones. http://about.skype.com/press/2010/12/iphone_video_calls.html.

[40] S. Smaldone, C. Tonde, V. Ananthanarayanan, A. Elgammal, and L. Iftode. The cyber-physical bike: A step towards safer green transportation. In *12th Workshop on Mobile Computing Systems and Applications (HotMobile)*, March 2010.

[41] New York State. New york state vehicle and traffic laws. https://www.dot.ny.gov/display/programs/bicycle/safety_laws/laws.

[42] H. Ueno, M. Kaneda, and M. Tsukino. Development of drowsiness detection system. In *Vehicle Navigation and Information Systems Conference, 1994. Proceedings., 1994*, pages 15 –20, aug-2 sep 1994.

[43] Osiris Development website. Battery bar for windows 7, 2010. http://osirisdevelopment.com/BatteryBar/.

[44] Ian H. Witten and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Second edition, 2005.

[45] www.bicyclinginfo.org. Bicycle crash facts. http://www.bicyclinginfo.org/facts/crash-facts.cfm.

[46] L. Zhao and C.E. Thorpe. Stereo- and neural network-based pedestrian detection. *Intelligent Transportation Systems, IEEE Transactions on*, 1(3):148 –154, sep 2000.

[47] E. Zilberg, D. Burton, Zheng Ming Xu, M. Karrar, and S. Lal. Methodology and initial analysis results for development of non-invasive and hybrid driver drowsiness detection systems. In *Wireless Broadband and Ultra Wideband Communications, 2007. AusWireless 2007. The 2nd International Conference on*, page 16, aug. 2007.

Vita

Vancheswaran Koduvayur Ananthanarayanan

2009-2012 M.S. in Computer Science, Rutgers University

1999-2003 B. Tech. in Computer Science from NIT Calicut, India

2011-2012 Android Developer, Graduate School of Education, Rutgers University

2010-2011 Teaching assistant, Department of Computer Science, Rutgers University

2006-2009 Technical Lead, Pathpartner Ltd., Bangalore.

2004-2006 Sr. Software Engineer, Emuzed Ltd., Bangalore.

2003-2004 Software Engineer, Wipro Technologies, Bangalore.