Improving Pedestrian Safety with Consumer Grade Earphones

Siddarth Madala*
smadala2@illinois.edu
University of Illinois at
Urbana-Champaign

Mubashir Anwar*
manwar@illinois.edu
University of Illinois at
Urbana-Champaign

Simon Kato*
sk106@illinois.edu
University of Illinois at
Urbana-Champaign

ABSTRACT

The prevalence of smartphone use while crossing the street leaves pedestrians distracted and at risk of accidents. As a result, while vehicular fatalities have decreased in the past few years, pedestrian fatalities are on the rise [4]. In this paper, we build upon previous wearable systems [11] that use an array of audio sensors to locate the direction of an approaching vehicle and notify the user of impending danger. Specifically, we achieve a comparable detection accuracy of approaching cars while only relying on the built-in microphones and base hardware of commercial headphones (i.e. Apple Airpods). We provide early danger detection in real time with low latency and high accuracy without using any specialized devices on the pedestrian side or have limited applicability due to specific assumptions about the environment. To reduce power consumption, we utilize a low sample rate with input from only one microphone in the headset that alerts our algorithms of potential vehicles, after which we increase the sample rate and use all available microphones (two in headset and one in the smartphone) to produce a confidence estimate of whether a vehicle is approaching and if so, from which direction. Testing is performed in a 10 kilometer walk around the University of Illinois at Urbana-Champaign (UIUC).

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

Embedded systems, pedestrian safety, sound source localization, wearables

ACM Reference Format:
Siddarth Madala, Mubashir Anwar, and Simon Kato. 2022. Improving Pedestrian Safety with Consumer Grade Earphones. In *Proceedings of Smart-X '22: Fake ACM Symposium on Smart cities, homes, phones, and beyond (Smart-X '22).* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Modern technology has substantially changed our day-to-day lives. Today, 97% of Americans have a cellphone [3] while 21% own a smart watch [20]. While the rise of technology has arguably made our lives easier, it has come with some drawbacks. In particular, pedestrian traffic safety is of growing concern because of technological distractions. Distracted pedestrians have led to a rise in accidents [4, 22]. A recent study found that 51% of users talk on the phone while walking, and 36% listen to music [2]. A major challenge today is improving pedestrians' safety in traffic situations.

The problem is aggravated by the use of wireless earphones, which is growing rapidly [13]. Traditionally, the sound of an approaching car is critical for its detection [5, 17], but the widespread adoption of noise-cancelling headphones deprives pedestrians of this vital sense, exposing them to a higher risk of crash related injuries. This issue is only aggravated by simultaneous smartphone use when crossing streets.

To combat distractions, municipalities have begun to put "LOOK" signs on the floor at crosswalks and high-danger areas with inconclusive results. A potential solution to this problem is to use the very technology (e.g. smart phones, smart wearables) that are distracting pedestrians to provide safety instead. Past work used a combination of commercial [12, 18, 21] and custom hardware [11, 14, 23], to detect and alarm users of possible dangers to pedestrians (e.g. approaching vehicles, stepping onto the street from a sidewalk etc.). However, practical limitations exist with these solutions. Custom devices are not widely available, serve no other purpose, and are often visually unattractive. Commercial hardware seems a more attractive alternate, but comes with many challenges as well. For example, past work has proposed using smartphone cameras [21] to detect incoming cars, however, this requires the smartphone to be in the correct orientation at all times (towards the road), which is impractical for the end-user. Other methods rely on specific signals (e.g. wireless signals from smart vehicles) [12] or specific cases (e.g. high frequency sound of electrical vehicles) [18] which does not generalize well. We are interested in utilizing commercially available noise-canceling headphones, which come readily available with microphones, alongside the existing microphone in the smartphone, to monitor and alert pedestrians of approaching vehicles.

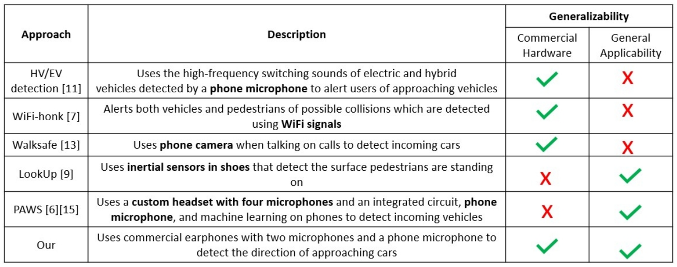
We will be designing a general framework and algorithm for the sound localization of approaching vehicles. Our proposed method is a two-module machine learning model. The first module is a binary vehicle classifier that detects if a moving vehicle is nearby. The second module isolates the direction of the vehicle (front, right, back, or left) relative to the pedestrian.

We aim to collect data from two sources, (i) our own dataset by walking on the streets of Urbana-Champaign, Illinois and (ii) a truncated dataset from PAWS [23]. The PAWS dataset contains seven microphones; we will truncate this data so that only the smartphone's microphones and the microphones located where commercially available headphones have microphones are used.

Our main goal is to evaluate the detection accuracy of approaching vehicles using our approach. We intend to perform this evaluation in the streets of Urbana-Champaign, Illinois by walking at least 3 kilometres around the campus, manually cataloging approaching cars and directions. We will evaluate the sensitivity and

[*] Authors contributed equally to this research.

Smart-X '22, Jan 18–May 05, 2022, Urbana, IL
2022. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn



Approach	**Description**	**Generalizability**	
		Commercial Hardware	General Applicability
HV/EV detection [11]	Uses the high-frequency switching sounds of electric and hybrid vehicles detected by a **phone microphone** to alert users of approaching vehicles	✓	✗
WiFi-honk [7]	Alerts both vehicles and pedestrians of possible collisions which are detected using **WiFi signals**	✓	✗
Walksafe [13]	Uses **phone camera** when talking on calls to detect incoming cars	✓	✗
LookUp [9]	Uses **inertial sensors in shoes** that detect the surface pedestrians are standing on	✗	✓
PAWS [6][15]	Uses a **custom headset with four microphones** and an integrated circuit, **phone microphone**, and machine learning on phones to detect incoming vehicles	✗	✓
Our	Uses commercial earphones with two microphones and a phone microphone to detect the direction of approaching cars	✓	✓

Figure 1: The table denotes the different approaches to pedestrian safety as well as whether they require additional constraints or assumptions.

specificity of both parts of the model. For micro-benchmarking, we will compare multiple popular machine learning models with our own model stratifying on the processed auditory data for detection of danger. Finally, we intend to evaluate the difference in accuracy of detection after varying the frequency of sensing sounds from the microphones. In particular, we are interested in whether or not the binary classifier can have comparable results with only one microphone active at low and high probing frequencies.

To summarize, the major contributions of this work are as follows:

(1) Proposing the use of commercial headphones for detection of approaching vehicles.
(2) Designing, training, evaluating, and comparing the ability of our model to classify an approaching vehicle and its direction.
(3) Experimentally analyzing the performance of the detection algorithm in Urbana-Champaign.

2 RELATED WORK

There have been many studies examining the ability of wearable technology and mobile devices to aid in pedestrian safety in traffic situations. The development of smart phones has yielded the production of many specialized, compact, energy efficient sensors. These sensors are very powerful and can be repurposed to assist in many situations. This has led to an explosion in the breadth of solutions aimed at improving pedestrian traffic safety. These solutions can be divided into two main approaches: those that use special purpose wearables and those that use existing, commercial hardware. Figure 1 details some selected methods with the type of sensors each method requires as well as constraints or assumptions they make. For both commercial and special purpose wearables, we distinguish between methods that use sound for detection and methods that use other sensors for detection.

2.1 Special Purpose Wearables

In this section, we discuss related work on pedestrian safety that use special purpose sensors on wearables for pedestrian safety.

2.1.1 Sound for Vehicle Detection. Headphones have been used in the past to detect incoming cars from all directions using microphones in a custom headset [8, 11, 23]. This approach achieved high detection accuracy by using four microphones in a headset, including one at the back and one at the front in addition to microphones at the left and right ear. Moreover, this approach conserved power by using a custom microcontroller in the headset for initial signal processing before sending data to smartphones for further processing, if necessary. CSafe [24] used a similar approach with a custom wearable with four microphones to be attached to construction workers to detect incoming vehicles.

Microphones have also been used for sound localization for detecting general events. [19] used a microphone array of up to 8 omni-directional microphones mounted on a robot to localize people and interesting events.

2.1.2 Other Sensors. Smart shoes with inertial sensors have been used to conduct surface gradient profiling and step pattern identification [14]. Inertial sensors detect transitions between sidewalks and the street to localize and alarm users if they were on a street. The goal of this work was to direct user attention to the street when they step off from a sidewalk, instead of alarming users to approaching vehicles.

Although specialized wearables are able to produce good detection results, they suffers from a few main limitations:

(1) Custom hardware is expensive.
(2) Specialized hardware is, by nature, not widely accessible.
(3) Such wearables are visually unattractive, bulky, or inconvenient.

2.2 Commercial Hardware

2.2.1 Sound for Vehicle Detection. Even though most sound-based methods on detecting vehicles use custom-built earphones or microphone arrays, there has been some work on using off-the-shelf products for vehicle detection. [15] used the embedded microphone in smartphones to detect incoming vehicles and localize their direction. However, this approach relies on the smartphone being out in the hands of the pedestrian, and would not work when the smartphone is in the pockets or bags, where the microphone does not have 360 degrees coverage of the environment. Similar to our approach, [16] used off-the-shelf earphones to localize events based on sound. However, this approach does not achieve real time detection and would not work for a safety system. In [6], the authors use a single microphone from surveillance cameras for real-time event detection. Similar to our system, the method uses a two-step model for detection. The first model detects if a sound is a vocal or a non-vocal event and the second model further classifies events into excited and normal. Although this approach can detect the sound of vehicles, it does not localize their direction.

2.2.2 Other Sensors. WalkSafe [21] uses the back camera in smartphones coupled with on-device machine learning to alert users of approaching cars. The camera is only active during phone calls, whereby the user is likely to be distracted and the back camera is likely facing the road. However, this approach is only useful for a single direction of a street. Moreover, people oftentimes use their earphones for calls – with their cellphones in their pockets – limiting the effectiveness and usefulness of this solution.

Other approaches proposed the idea of allowing approaching vehicles to send a signal which could be identified by smart phones

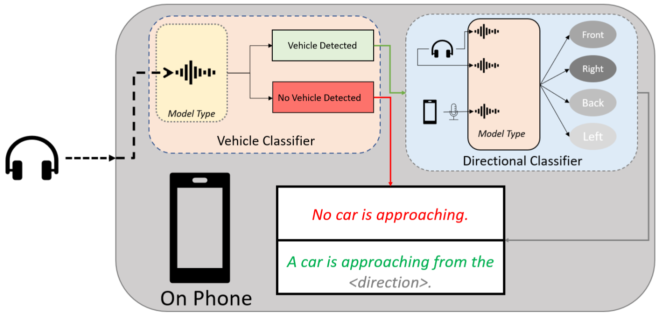


Figure 2: Proposed two-step model. The model consists of a vehicle classifier module and a directional localization classifier. The dashed lines from the headphones to the first module indicate a low frequency of querying. The first module will use a support vector machine to classify whether the audio has a vehicle. If there is not a vehicle, the model restarts once the headphones are queried again. If there is a vehicle, the model will use a much higher frequency (indicated by solid lines) data stream from the headset microphones in addition to the phone microphone to perform directional localization classification. The output of the model will indicate whether a vehicle is approaching; and if there is, the direction relative to the headphones from which it is approaching.

[10, 12]. As this method relies on vehicles to communicate with users, the effectiveness of this approach is heavily reliant on manufacturers’ commitment to follow communication protocols. Moreover, this method does not immediately address the issue of pedestrian traffic safety. Even if this method were to become an imposed standard, existing vehicles without early warning functionality would still pose a threat.

Takagi et al. [18] takes advantage of the fact that electric and hybrid engines generate high frequency sounds while in operation. This approach uses smartphone microphones which are sensitive to high frequency sounds to identify and alert users of approaching vehicles. This method does not rely on extra signals from vehicles and proactively uses a smart phone microphone for detection, making it practical and immediately useful. However, regular engines do not produce the high frequency sounds, so this method cannot classify regular cars, limiting its generalizability as a solution.

3 METHOD

We will design a general framework and algorithm for the sound localization of approaching vehicles. To the best of our knowledge, we are the first to attempt detection of incoming vehicles through non-specialized hardware in an unconstrained environment. The framework will consist of two modules, a car classifier and a directional classifier 2. The vehicle classifier will perform binary classification to indicate the presence a vehicle detected in the audio stream. Conditional on there being a vehicle, the directional classifier will output the direction which has the highest probability relative to the end-user.

The algorithm is a two-step process based on the intuition that the resources needed to identify whether a car is present in the audio stream is not the same as what is required to perform directional localization. For example, the first module will only require that only one microphone be queried with some tunable frequency. Since most wireless earphones are configured to have one of the two microphones turned on at all times, this does not waste any extra energy. To perform the directional localization, the microphones need to be streaming data in real-time and might even require the phone’s microphone for better prediction. Furthermore, if car identification can be accomplished with data from only one microphone, the two-step model would let us reduce power consumption.

3.1 Vehicle Detection

The vehicle classifier is a support vector machine pinged at the low-frequency rate previously described. Specifically, we utilize an SVM with a radial basis kernel function to embed our audio data into an infinite dimensional space, theoretically allowing for easier classification. The SVM will split the car based noises from other sounds captured in the audio stream, thereby detecting if any vehicles are approaching the user. If a vehicle is detected, then this result is passed to our directional localization classifier to determine what path a vehicle is approaching from.

3.2 Directional Localization

The (proposed) directional localization classifier is a 4-way Bayesian Filter based approach. At a high level, the Bayesian Filter is split into two steps: predict and (measurement) update. The predict step takes the previous state and an action and outputs a prediction of what it believes the new state will be. The (measurement) update step will incorporate the observation information into the prediction of the state. The reason for using a Bayesian Filter approach is a Bayesian Filter keeps track of a belief of a state when the true state is unknown, like in the case of noisy data. The noisiness of the microphone data suggests the use of a Bayesian Filter approach. Moreover, this method will be able to run in real-time with tunable values.

A particle filter is a type of Bayesian Filter which keeps track of the belief of the state using particles. That is, a number of particles are kept track of, each seemingly representing the state. The prediction and update is applied to each particle at a time to represent how the state would evolve from a range of potential values.

The intuition of the proposed method is that the particle filter is going to be keeping an internal belief of what state it is and will calculate the probability of receiving an observation, we can utilize the internal states to determine which direction is most likely given the four models. Moreover, the directional components of the car will be embedded in the dynamics model as the sound will evolve differently depending on the assumed direction of the car. Then, the observation will inform which of the predicted stated is most in line with reality.

The prediction and the update of a particle filter relies on a dynamics and observation model. The dynamics model informs the dynamics of the system. That is, given a state and an action, the resulting state. The observation model gives the probability of an observation given a state.

To conduct directional localization, we propose to have four particle filters. In specific, we will have four different dynamics and observation models which indicate how the sound will evolve assuming the car is coming from a particular direction, and the probability of an observation. In our proposed approach, the dynamics models are learned from data. The time-domain data will be fit with a polynomial function, and the final polynomial will be the average of all the polynomial fits. Of note, the data is patted with zeros to reach five seconds long to address issues with the tail of the polynomial going to positive or negative infinity. Below, the mathematical derivation of the polynomial fit for m number of data for a particular direction:

$$data_j = data_j + \vec{0}_{[5-length(data_j)]}$$

$$poly(data_j) = a_{0,j} + a_{1,j}x + \cdots + a_{n,j}x^n$$

$$Polynomial\ Fit\ (PF) = \frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{n} a_{i,j}x^i$$

The dynamics model, how the sound evolves, is modelled by the derivative of the polynomial fit. By having the dynamics be the derivative of the polynomial fit, the prediction step of the particle filter can be thought of as essentially being Euler integration with noise. The observation model is assumed to be normally distributed. The mean of the normal distribution is the value of the model fit at the current time, and the standard deviation is the average of the sample standard deviation of all the data.

$$\text{Standard Deviation (SD): } \frac{1}{m}\sum_{j=1}^{m} sample\_std(data_j)$$

$$Dynamics\ Model\ (DM) = \frac{\mathrm{d}}{\mathrm{dx}}\frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{n} a_{i,j}x^i$$

$$Observation\ Model = Pr(X = z_t), \text{ given } X \sim N(PF(t), SD)$$

This process is done for each direction to create four dynamics and observations models.

Pseudocode for the predict in algorithm 1, update in algorithm 2, and to calculate the probability of calculating from which direction the car is coming from will be given below in 3. Moreover, a figure of the general proposed structure can be seen in Fig 3

Algorithm 1 Predict

```
Require: time, particles, control, dt, $\sigma$
  for particle in particles do
      $particle = particle + control(time) * dt + normal(0, sigma)$
  end for
  return particles
```

Algorithm 2 Update

```
Require: particles, $\mu, \sigma$
  $weights = \vec{0}_{particles.size}$
  for i in range(particles.size) do
      $weights[i] = pr(X = particles[i] | X \sim N(\mu, \sigma))$
  end for
  $weights = weights/||weights||_1$
  particles = $Sample(particles, size = particles.size, prob = weights, replace = True)$
  return particles
```

Algorithm 3 Calculate Probability

```
Require: dynamic model, particles, $\sigma$
  $mu = model(time)$
  $probs = \vec{0}_{particles.size}$
  for i in range(particles.size) do
      $probs[i] = Pr(X = particles[i] | X \sim N(\mu, \sigma))$
  end for
  return $\frac{1}{N}\sum_{i=1}^{N} probs[i]$
```

3.3 Preprocessing

The preprocessing of the data for training out SVM consists of converting audio files into *wav* format for easy processing. Once in wav format, the audio signals can easily be loaded and a Fast-Fourier Transform can map the audio signal into the frequency domain. This frequency domain is currently fed in its raw form into the SVM for training along with its label based on the title of the audio file (audio files from cars with have the word car and the direction from which it came). Further processing including smoothing, as well as labeling when car noise begins in a signal would allow us to further refine our model output to relay when exactly a car is heard instead of generalizing over a longer period of audio capture.

For the directional classifier, the preprocessing was to truncate the data to contain only the portion of the vehicle approaching, and to fill in the audio with zeros. The truncation was done to be able to learn how the audio evolves while the car is approaching, the portion of interest of the sound. Moreover, the audio was also filled in with zeros to address the polynomial fit. To elaborate, the truncated data are of varying lengths. Without padding, the polynomial will fit the values at the truncated data and oftentimes, after (and before), will go off to either negative or positive infinity. Because the coefficients are being averaged, this would have the effect of our average model only really being well defined for the length of time of the shortest audio. By padding zeros, the polynomial fits are forced to go to zero after they finish.

3.4 Challenges

There are three main challenges to our approach:

(1) Determining how to identify approaching cars with only two to three microphones.
(2) How to manage the power consumption for in-device processing since unlike PAWS [11], we do not have a specialized, low-power microprocessor.
(3) Manufacturer firmware currently only allows for one microphone to be on at a time due to the difficulties of binaural microphone synchronization.

Given the results of [11], it is known that with a sufficient number of microphones, accurate identification of approaching cars using audio data is feasible. Thus, our intuition is that it should be possible to localize approaching vehicles using just two microphones since

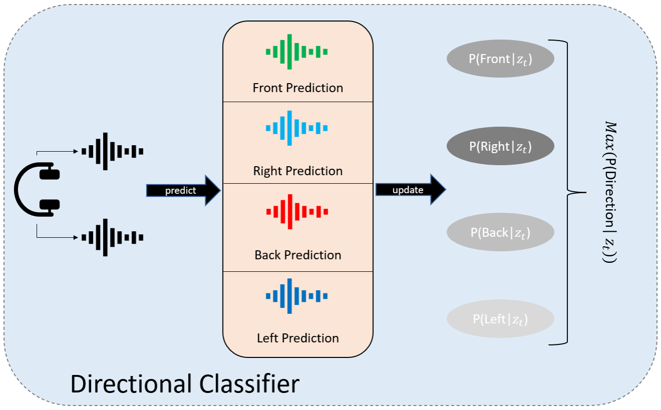


Figure 3: The proposed directional classifier structure. In real time, the audio channel will be fed into the Filter which will call four different predicts based on the direction. These four predictions will then be analyzed by their probability to have occurred given the next observation, z_t. The directional localization will be the maximum of all the probabilities of the directions.

our ears are able to do so. The shape of our ears enable us to differentiate sounds coming from the back from the sounds coming from the front. Since earbuds are very near to our ears, we believe that they should be able to use the same property for localization. Moreover, the addition of the phone's microphone can provide valuable information, especially if the orientation of the phone is known.

For the second challenge, we aim to leverage the two-step model to vary the frequency of sound recording as well as attempt to use only one microphone for vehicle classification, thereby lowering the power consumption. Instead of having always-on microphones that constantly send data to the smartphone, we will condition on the first classifier to detect and send sound data at lower frequencies and increase the frequency during the second classifier when there is a potential approaching car. We intend to manually obtain the actual frequencies required for in-time detection using our tests.

The third challenge is a hardware difficulty. Current manufacturer settings only use one microphone in any setup due to battery concerns and binaural synchronization difficulties. This is an impossible issue to remedy in the short time as pushing changes to proprietary firmware is extremely difficult without extensive negotiation with the company who sells the commercial hardware. However, this does not prevent lab tests/simulations nor detracts from the use of the single microphone SVM. Moreover, utilizing the research of Clearbuds [9], manufacturers have a clear path to add fast, effective binaural microphone synchronization utilizing open-source firmware. We hope that in the near future, our solution will be easily adopted in commercial headphones.

4 DATASETS

To train and test our machine learning models, we use two datasets: (1) PAWS Dataset [11] containing the sound of vehicles, people talking, and ambient noise. (2) Our own approaching vehicle dataset collected on both idle and busy streets.

4.1 PAWS Dataset

The data for PAWS [11] was collected using four microphones in five different locations, including parking lots and busy interactions. The data included the sounds of 47 cars, along with sounds of human talking, silence, and ambient noise on the streets. The dataset included 92 recordings. For our purpose, we only used the recordings from one channel (one microphone), since we wanted to use them for our SVM to distinguish between the sound of cars and other sounds. We used this dataset with the permission of the authors.

4.2 Approaching Vehicle Dataset

Although the PAWS dataset provided substantial data to differentiate the sound of approaching vehicles from other sounds, it did not contain information about the direction of the approaching vehicles, which is vital for direction localization. To obtain this data, we conducted tests in the roads of Champaign. We selected Apple Airpods (2nd generation) [1] as representative consumer grade earphones, since Airpods are the most popular wireless brand of earphones in the US [7]. We conducted two sets of experiments, one on an idle road, and another on a busy road. The earphones were worn by one of the authors, who stood at the edge of a cross-walk while the vehicles passed by. For the idle road, two vehicles, driven by the other authors, passed by from four different directions (left, right, back, and front). For each direction, there were two possibilities in the orthogonal direction (e.g. a vehicle coming from the left while the pedestrian is behind the car and a vehicle coming from the left while the car is behind the pedestrian). Thus, a total of 8 recordings were captured, each recording with two approaching cars. For the busy road, one of the authors with the earphones plugged in stood near an intersection, and recorded the sound of cars coming from different directions. We made a video recording of this in order to tag sounds with the direction of the approaching vehicles.

We faced a few challenges while collecting this data. The first challenge was that Apple, along with other popular wireless earphone brands, do not allow stereo recording on their earphones. Thus, only one of the two microphones is active at any given time. This is done to save battery and can only be changed through the firmware. To circumvent this problem, we used two different Airpods (the same version) for left and right ears. This lead to another challenge: how can we synchronize the two microphones when they are not connected together? We synchronized the sound by clapping at the start of each recording, and synchronizing the sounds of the two channel based on when the clap sound was made. Note that this problem is exasperated because we used two different Airpods connected to different phones. In practice, we hypothesize that the synchronization provided by two microphones connected to the same phone would be enough for our purposes, since we treat the recordings from each microphone as a separate channel and do not need a high degree of synchronization for our models to work. The last challenge was figuring out the exact time where a car would collide with a pedestrian. One approach would be to make a video of the approaching vehicle and pedestrian during the experiments

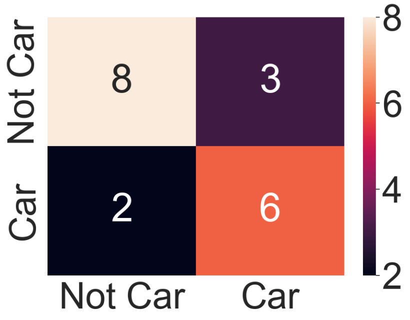


Figure 4: Confusion matrix of our SVM on a test set of 19 differing car positive and negative audio files. X-axis is true label and Y-axis is predicted label.

and then figuring out the time a car is about to hit. However, this would result in synchronization issues of the video with the audio. We use a simple solution to solve this problem: we estimate the time where the car is nearest to the pedestrian as the time where the loudest amplitude is measured. This intuition assumes that the most dominant sound is the sound of the vehicles in the recordings, and this sound has the highest amplitude when the vehicle is about to collide. Thus, in our data pre-processing, we first synchronize the audios using the sound of the clap and then cut the audio at the maximum amplitude (to make sure that we are not using the sound after the vehicle has already passed and possibly collided with the pedestrian). We performed these recordings across two sessions, with 8 recordings in each session. Thus, we collected a total of 16 stereo recordings for our models.

5 EVALUATION

Our main goal is to evaluate the detection accuracy of approaching vehicles using our approach. We intend to perform this evaluation in the streets of Urbana-Champaign, Illinois by walking at least 10 kilometres around the campus, manually cataloging the sensitivity and specificity of our method.

We intend to evaluate the difference in accuracy of detection after varying the frequency of sensing sounds from the microphones. This will be useful for decreasing power consumption, since we do not want the earbuds to continuously send auditory data. We want to observe if we can send data at a lower frequency and increase this frequency only when danger exists.

The evaluation of the SVM will discuss model performance in detecting vehicles as well as the types of noises that induced error in the model. Figure 4 is a confusion matrix of the results from testing our SVM on 19 distinct noises which are around an even mix of car positive and negative audio files. The x-axis represents true classification while the y-axis represents predicted class. We achieve an accuracy rate near 75% with minimal preprocessing. When analyzed, false positive samples were seemingly noisy samples that likely had frequency elements similar to that of an engine or tires on the road. False negative samples were from our dataset and had the car either nearly inaudible or too much random noise from the environment in the file.

The evaluation of the directional classifier will discuss the fits of the polynomial models as well as the results of the directional classifiers. Figure 5 shows the fits of the polynomial fits for the behind, front, left, and right as well as the polynomial fit for the test case that has the sound is coming from behind from the left ear phone, and Figure 6 shows from the right ear phone. The polynomial fit graphs demonstrate the limitations of the temporal-domain fitting. This is because all the models have a spike, but the timing of which are different. Moreover, another clear issue is the magnitude of the spikes. It is expected that the front and behind are similar in timing and amplitude, we expect the left and right to be similar in amplitude but differ slightly in timing. Figure 7 shows the probability of directions of the car while it is approaching from the left earphone, Figure 8 shows from the right ear phone. The figures show results for a 10, 50, and 100 degree polynomials. The probability direction reflect that the directions with the highest probability are those which have a similar polynomial fit. Accordingly, behind and left show a higher probability across the entire time.

6 LIMITATIONS AND DISCUSSIONS

- We did not synchronize the audio in real time, since our dataset was pre-processed offline using methods that are not possible in practice (e.g. clapping before an audio). Although our directional localization approach is not highly sensitive to non-synchronized audio, synchronized audio should still be beneficial for the model accuracy. Prior work (Clearbuds [9]) has shown synchronization methods that can be used with our approach. Although these methods use a different hardware, then can be used with out approach if such hardware is commonly adopted by earphone manufacturers.
- The SVM is currently using minimal preprocessing and works at a higher sample rate than desired. In addition, the training data is quite noisy and does not give the model a clear indication of whether a car exists. While we achieve decent accuracy on our test set, there is no indication that our model generalizes beyond the data we collected or the PAWS dataset incorporated into our work. In addition, the cases of false negatives existing provides some worry since it is critical to hear all cars approaching, as even one missed car can lead to an accident. Future work would involve developing efficient preprocessing algorithms, collecting significantly more data at a low sample rate, and testing our solution in more urban environments where distractions other than cars exist.
- The directional localization approach is currently fitting to temporal-domain data. This has shown to have a lot of issues. One issue is that we must train on data that are of varying length will negatively impact the coefficients and capturing the spike. To address the problem of varying length data, padding the data with zeros presents its own problems in synchronization of the model coefficients. Another issue is that the audio quality between recordings is so drastically different that even for the left and right headphone, the audio tracks are very different which can be seen in the polynomial fitted models being so different between every direction.
- A future direction is to use the frequency-domain instead of the temporal-domain. The frequency domain would fit in

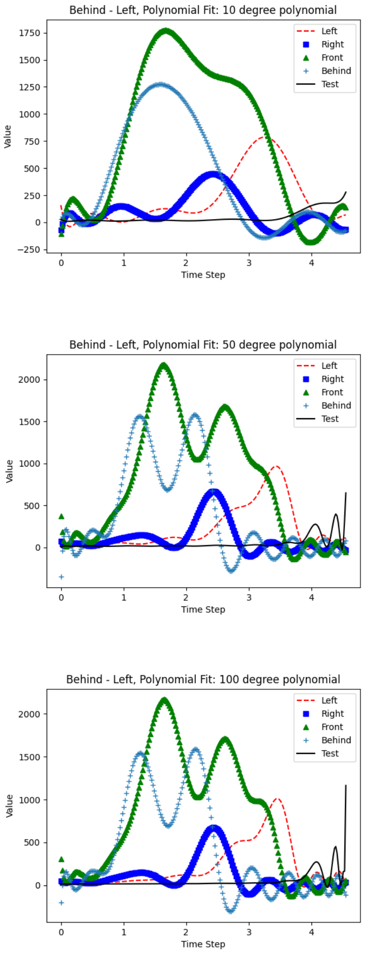


Figure 5: The polynomial fit for the different directions with varying degrees of polynomial. The black line is the test case of an audio coming from behind from the left headphone.

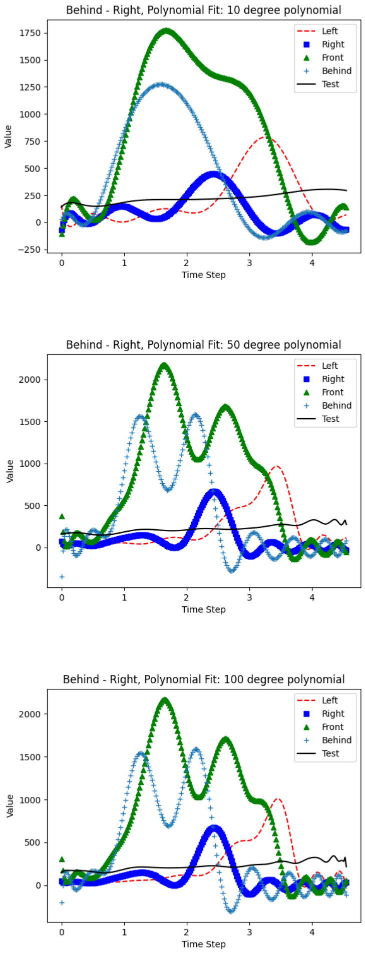


Figure 6: The polynomial fit for the different directions with varying degrees of polynomial. The black line is the test case of an audio coming from behind from the right headphone.

nicely with the learning of the dynamic models; however, the observation model might require some modification. The frequency domain would also benefit greatly from low-pass filter. The best solution would likely be a hybrid solution between the temporal-frequency domains.

- Our collected data had a lot of problems. In our first session. we had different sound settings on the two phones for the microphones, resulting in very different output. This data was unusable, so we conducted the tests again with the same settings. However, due to a problem in one of the Airpods, the recording of one earphone was still distorted. Our results are from only two faithful recordings, which prevented us from obtaining a high accuracy on our models.
- Our work is on predicting approaching vehicles regardless of the location of the pedestrian. For practical deployment, the location of the pedestrian has to be considered to avoid false positives (e.g. an alarm when the pedestrian is on the sidewalk). One potential method would be to collect data where pedestrians are closer to the car vs. when they are a bit far (possibly on a side walk). However, such an approach would still not be able to distinguish the case where a car passes by close to a sidewalk vs a car approaching the pedestrian on the road. The other method would be to use existing approaches that make use of GPS signals to figure out if a pedestrian is on the road or not. Such methods can be integrated with our approach to synthesize a pedestrian safety warning system.
- In the future, we can use multiple microphones for localization. For example, we can use the microphone of the phone while estimating its orientation using gyroscope. This will give another dimension of data, potentially increasing the directional localization accuracy. Similarly, other wearables, like smart watches with different sensors (e.g. microphones, GPS signals) can be used to better localize cars. The key is to

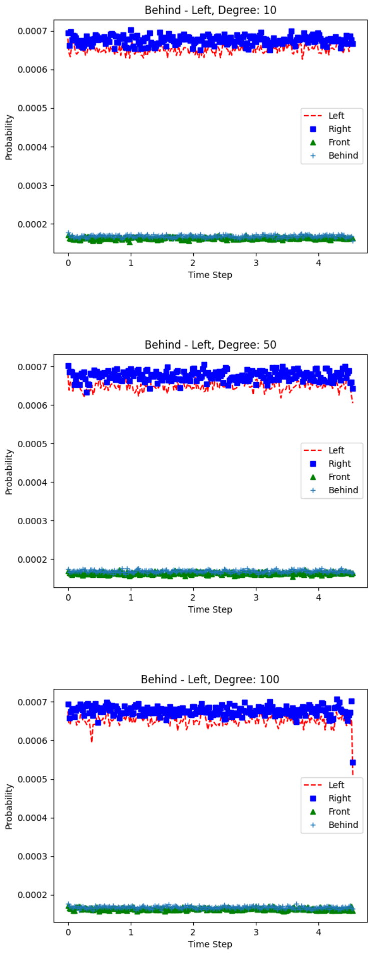
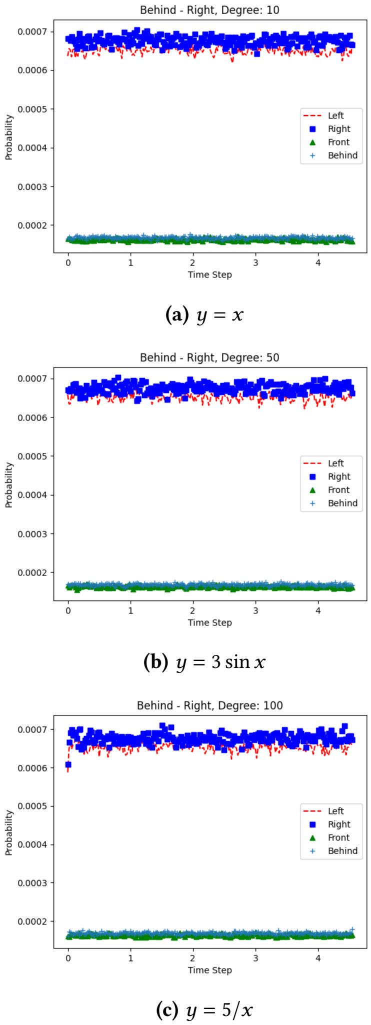


Figure 7: The probabilities as a function of time for the test cause of audio coming from behind recorded from the left headphone. Results are shown for the polynomial fits for multiple degrees.

utilize common wearables, which have a myriad of sensors, for pedestrian safety rather than using specialized hardware, which is harder to adopt.

7 CONCLUSION

This project concerns itself identifying when and from which direction a car, if any, is approaching. The proposed method is a two-step model: a vehicle classification and a direction localization module. The vehicle classification module utilizes an SVM that ... The direction localization module keep track of four beliefs of the audio state assuming the direction of the car utilizing particle filters, and return the direction most in line with the observation. The strength of the two-step method is the ability to alter the frequency of running the different modules and allows improving each individual module independently. (The strength of utilizing an SVM is in quickly being able to do inference?) The particle filter is a real-time method which operates in a closed-loop fashion for better accuracy, has tunable frequency and number of particles for adapting to computation power available as well as battery, and is a robust method of dealing with noisy sensors. However, the model's directional localization is trained on a temporal-domain but future works will investigate how the frequency-domain can be used to improve the dynamics models. Moreover, incorporating the microphone of the phone and sensors on smart wearables can be investigated.



(a) $y = x$

(b) $y = 3 \sin x$

(c) $y = 5/x$

Figure 8: The probabilities as a function of time for the test cause of audio coming from behind recorded from the right headphone. Results are shown for the polynomial fits for multiple degrees.

REFERENCES

[1] [n. d.]. Airpods (2nd generation). https://www.apple.com/airpods-2nd-generation/

[2] 2013. New study shows three out of five pedestrians prioritize smartphones over safety when Crossing Streets. https://www.libertymutualgroup.com/about-lm/news/articles/new-study-shows-three-out-five-pedestrians-prioritize-

smartphones-over-safety-when-crossing-streets

[3] 2021. Mobile fact sheet. https://www.pewresearch.org/internet/fact-sheet/mobile/

[4] 2022. New projection: U.S. pedestrian fatalities reach highest level in 40 years. https://www.ghsa.org/resources/news-releases/GHSA/Ped-Spotlight-Full-Report22

[5] Admin. 2016. New sound rule designed to minimize pedestrian injuries and accidents. https://jonesclifford.com/new-sound-rule-designed-minimize-pedestrian-injuries-accidents/?doing_wp_cron=1666918227.8242299556732177734375

[6] P.K. Atrey, N.C. Maddage, and M.S. Kankanhalli. 2006. Audio Based Event Detection for Multimedia Surveillance. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, Vol. 5. V–V. https://doi.org/10.1109/ICASSP.2006.1661400

[7] Sanuj Bhatia. 2022. No surprises here: Airpods amp; beats are the most popular wireless earphones in the US market. https://pocketnow.com/airpods-beats-most-popular-wireless-earphones-us-market/

[8] Rishikanth Chandrasekaran, Daniel de Godoy, Stephen Xia, Md Tamzeed Islam, Bashima Islam, Shahriar Nirjon, Peter Kinget, and Xiaofan Jiang. 2016. SEUS: A Wearable Multi-Channel Acoustic Headset Platform to Improve Pedestrian Safety: Demo Abstract. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM* (Stanford, CA, USA) *(SenSys '16)*. Association for Computing Machinery, New York, NY, USA, 330–331. https://doi.org/10.1145/2994551.2996547

[9] Ishan Chatterjee, Maruchi Kim, Vivek Jayaram, Shyamnath Gollakota, Ira Kemelmacher, Shwetak Patel, and Steven M. Seitz. 2022. ClearBuds. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*. ACM. https://doi.org/10.1145/3498361.3538933

[10] Klaus David and Alexander Flach. 2010. CAR-2-X and Pedestrian Safety. *IEEE Vehicular Technology Magazine* 5, 1 (2010), 70–76. https://doi.org/10.1109/MVT.2009.935536

[11] Daniel de Godoy, Bashima Islam, Stephen Xia, Md Tamzeed Islam, Rishikanth Chandrasekaran, Yen-Chun Chen, Shahriar Nirjon, Peter R. Kinget, and Xiaofan Jiang. 2018. PAWS: A Wearable Acoustic System for Pedestrian Safety. In *2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI)*. 237–248. https://doi.org/10.1109/IoTDI.2018.00031

[12] Kaustubh Dhondge, Sejun Song, Baek-Young Choi, and Hyungbae Park. 2014. WiFiHonk: Smartphone-Based Beacon Stuffed WiFi Car2X-Communication System for Vulnerable Road User Safety. In *2014 IEEE 79th Vehicular Technology Conference (VTC Spring)*. 1–5. https://doi.org/10.1109/VTCSpring.2014.7023146

[13] Market Research Future. 2022. Wireless earphone market to surpass USD 32.24 billion with a CAGR of 36.10% by 2030 - report by Market Research Future (MRFR). https://www.globenewswire.com/en/news-release/2022/08/04/2492577/0/en/Wireless-Earphone-Market-To-Surpass-USD-32-24-Billion-with-a-CAGR-of-36-10-by-2030-Report-by-Market-Research-Future-MRFR.html

[14] Shubham Jain, Carlo Borgiattino, Yanzhi Ren, Marco Gruteser, Yingying Chen, and Carla Fabiana Chiasserini. 2015. LookUp: Enabling Pedestrian Safety Services via Shoe Sensing. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services* (Florence, Italy) *(MobiSys '15)*. Association for Computing Machinery, New York, NY, USA, 257–271. https://doi.org/10.1145/2742647.2742669

[15] Sugang Li, Xiaoran Fan, Yanyong Zhang, Wade Trappe, Janne Lindqvist, and Richard E. Howard. 2017. Auto++: Detecting Cars Using Embedded Microphones in Real-Time. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 70 (sep 2017), 20 pages. https://doi.org/10.1145/3130938

[16] Annamaria Mesaros, Toni Heittola, Antti Eronen, and Tuomas Virtanen. 2010. Acoustic event detection in real life recordings. In *2010 18th European Signal Processing Conference*. 1267–1271.

[17] María Carmen Pardo-Ferreira, Juan Antonio Torrecilla-García, Carlos de las Heras-Rosas, and Juan Carlos Rubio-Romero. 2020. New Risk Situations Related to Low Noise from Electric Vehicles: Perception of Workers as Pedestrians and Other Vehicle Drivers. *International Journal of Environmental Research and Public Health* 17, 18 (2020). https://doi.org/10.3390/ijerph17186701

[18] Masaru Takagi, Kosuke Fujimoto, Yoshihiro Kawahara, and Tohru Asami. 2014. Detecting Hybrid and Electric Vehicles Using a Smartphone. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. Association for Computing Machinery, New York, NY, USA, 267–275. https://doi.org/10.1145/2632048.2632088

[19] Jean-Marc Valin, François Michaud, and Jean Rouat. 2007. Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering. *Robotics and Autonomous Systems* 55, 3 (2007), 216–228. https://doi.org/10.1016/j.robot.2006.08.004

[20] Emily A. Vogels. 2020. About one-in-five Americans use a smart watch or fitness tracker. https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/

[21] Tianyu Wang, Giuseppe Cardone, Antonio Corradi, Lorenzo Torresani, and Andrew T. Campbell. 2012. WalkSafe: A Pedestrian Safety App for Mobile Phone Users Who Walk and Talk While Crossing Roads. In *Proceedings of the Twelfth Workshop on Mobile Computing Systems Applications* (San Diego, California) *(HotMobile '12)*. Association for Computing Machinery, New York, NY, USA, Article 5, 6 pages. https://doi.org/10.1145/2162081.2162089

[22] Steven Wilson. 2020. Road Safety Annual Report 2013. https://www.itf-oecd.org/road-safety-annual-report-2013

[23] Stephen Xia, Daniel de Godoy Peixoto, Bashima Islam, Md Tamzeed Islam, Shahriar Nirjon, Peter R. Kinget, and Xiaofan Jiang. 2019. Improving Pedestrian Safety in Cities Using Intelligent Wearable Systems. *IEEE Internet of Things Journal* 6, 5 (2019), 7497–7514. https://doi.org/10.1109/JIOT.2019.2903519

[24] Stephen Xia, Jingping Nie, and Xiaofan Jiang. 2021. CSafe: An Intelligent Audio Wearable Platform for Improving Construction Worker Safety in Urban Environments. In *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (Co-Located with CPS-IoT Week 2021)* (Nashville, TN, USA) *(IPSN '21)*. Association for Computing Machinery, New York, NY, USA, 207–221. https://doi.org/10.1145/3412382.3458267