BO-Ear: Unsupervised Car Sound Sensing and Tracking Using Microphones on Smartphones

Sugang Li, Xiaoran Fan, Yanyong Zhang, Wade Trappe, Janne Lindqvist, Richard Howard

WINLAB, Rutgers University, North Brunswick,NJ, USA

1. ABSTRACT

As people become increasingly accustomed to using smartphones while they walk, there will be a corresponding concern about pedestrian safety as smartphone users might become distracted. In this work, we address this concern by developing a system for smartphones that warns users when it detects an oncoming car. In addition to detecting the presence of a vehicle, it can also estimate the vehicle's driving direction. We achieve these goals by processing the acoustic signal captured by microphones embedded in the user's mobile phone. In order to achieve more robust and timely detection, we also explore two novel feature, namely, Power on Certain Frequencies and Top Right frequency.

2. INTRODUCTION

In the US, during the year 2013 (latest available statistics), a pedestrian was killed in a traffic accident every two hours, and injured every 8 minutes [4], totally 4,735 fatalities and 66,000 injuries. In addition to loss of human lives, the economic and societal impacts of these incidents are estimated to be billions USD. As a result, improving the pedestrian safety has long been studied in many areas. Firstly, there are *macro-scale* solutions, including solutions that focus on safety manuals (e.g., World Health Organization [5]) and the solution focus on improving road infrastructure (such as the one proposed by Fitzpatrick et al. [2]). Secondly, there are *car-centric* solutions that focus on equipping the vehicles with the sensors that can detect the presence of the pedestrian. Most of these solution are based on vision or motion sensing, such as discussed in [1]. Thirdly, in the recent few years, *user-centric* solutions have emerged, which focus on equipping the pedestrian users with sensors or mobile systems that can help them become more situation-aware.

In this paper, we set out to devise such a mobile sensing system. Our objective is to give user a warning when there are cars approaching and may reach the user in a few seconds. Figure 1 illustrates two use cases of BO-Ear. In scenario A, the user walks towards the road when the car is approaching and our system warns the user that there is car n seconds away, coming from the front. In scenario B, the user walks along the road (when the sidewalk is too close to the main roads or when there are no sidewalks), and our system warns the user that there is a car n seconds away, coming from the back.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

S3'16, October 03-07, 2016, New York City, NY, USA

© 2016 ACM. ISBN 978-1-4503-4255-1/16/10...$15.00

DOI: http://dx.doi.org/10.1145/2987354.2987362

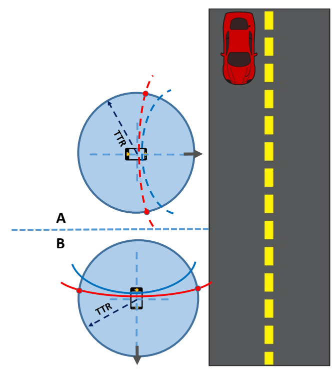

Figure 1: Two use cases for our proposed mobile system: (i) the user is walking towards the road (scenario A), and (2) the user is walking along the road (scenario B). In both cases, our system will send out warnings about the approaching car, with the estimated distance and direction information

Our main idea is to use the microphones that are embedded in the user's mobile device to capture the sound around the user. By analyzing the continuous stream of sound signal, we can detect whether there is a car approaching, estimate its distance from the user, and estimate from which direction the car is coming (as illustrated in Figure 1). Inspired by barn owls [3] who rely on their acute hearing to accurately localize their preys in the dark, we call our system ***Barn-Owl-Ear***, **BO-Ear** in short.

3. SYSTEM OVERVIEW

BO-Ear analyze the acoustic signal captured by built-in microphones to answer the following questions in (1) Is there a vehicle approaching the user? (2) How far is it? (3) What is its direction? The former concerns with the presence of the vehicle, and the later two concern with car tracking. In addition to answering these these questions, we would like to detect the car as early as possible so that the user has enough time to take safety measures. Figure 2 depicts the overview

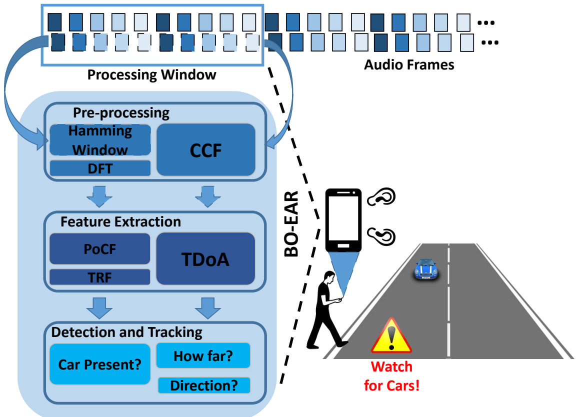

Figure 2: System Architecture

of BO-Ear, which consists of the following components:

1. **Pre-processing:** Before we can use the captured sound signal to detect or track the car, we first need to perform several pre-processing tasks, including segmenting the continuous stream of sound signal into smaller chunks, multiplying each chunk with a suitable window function, and/or conducting required processing on each chunk, such as discrete Fourier transform (DFT) or cross-correlation function (CCF). These steps prepare the data for subsequent signal processing.
2. **Feature Extraction:** Next, we extract features that can be used for car detection and tracking. Unfortunately, we find that those features that are commonly used to localize sound sources are not suitable for our system (explained later this section) as car sound is dominated by tire noise and lacks distinctive spectral and temporal features. In this work, we have proposed new features that are unique to car sound signals and that facilitate accurate, timely, and unsupervised car detection and tracking.
3. **Presence Detection:** Our presence detection algorithm is unsupervised. It can successfully differentiate between the scenarios that involve an approaching car and those that do not involve a car, but with high ambient noise. In addition, it can detect the car and issue a warning to the user early enough so that the user has time to respond to it.
4. **Tracking:** Our unsupervised tracking phase targets at estimating both the car's distance and direction. Considering that cars may move at different speeds, we introduce a new distance metric, Time to Reach (TTR), which denotes the time needed for the car to reach the user. This metric is not affected by the car's speed and is more meaningful to the user than the spatial distance. Further, we can also estimate the driving direction if the mobile device contains two microphones.

System Asumptions: In our approach, we assume that the user carries a mobile device that is equipped with one or two microphones. Today, all smartphones and tablets have at least one microphone, and some have two microphones, such as Google Nexus 6 and Samsung Galaxy Note 5. Further, we assume that the microphones have a similar degree of sensitivity in all directions. Also, we assume that the smartphone is held in hand or armband by the user, not in a pocket or in a bag.

4. KEY FEATURES

In this study, we propose two features that address the unique characteristics of car sound signals, and our features facilitate accurate, timely, and unsupervised car detection and tracking subsequently.

4.1 Power on Certain Frequencies (PoCF)

The first feature we propose is the power level on certain frequencies, PoCF in short. The rationale for PoCF stems from the observation that even though the overall power level is unstable and easily affected by noise in the environment, the power levels on carefully selected frequency ranges are much more stable and robust. In particular, we can choose those frequency ranges that have little overlapping with wind noise. Usually, we observe that the power level goes up when the car gets closer to the user. This is a natural consequence of the sound signal's path loss, $PL = (\frac{4\pi d}{\lambda})^{\alpha} = (\frac{4\pi df}{c})^{\alpha}$, where d is the distance between the transmitter and the receiver, λ is the signal wavelength, f is the signal frequency, c is the speed of sound, and α is a constant that is determined by the propagation model. Fig. 3(a) shows two STDFT snapshots from an approaching car's sound signal. In this example, we focus on frequency 10KHz, and find that the PoCF value continues to increase as the car drives near the user. PoCF serves as a more reliable indicator of the car's presence and distance than the overall SPL.

After selecting a suitable frequency range, BO-Earcalculates the PoCF values for each frame, and uses the moving average over 100 frames to generate smooth results.

4.2 Top-Right Frequency (TRF): Maximum Frequency Whose Power Reaches a Certain Threshold

We discover our second feature after carefully examining the car sound signal's STDFT results. Fig. 3(b) plots the STDFT result of the sound signal from a 2007 Toyota Camry when it drove 100 meters to reach the user. We collected the sound samples using a Nexus 6 smartphone. From the figure, we observe that when the car gets closer to the receiver, the energy on each frequency range increases. Further, when we consider the highest frequency component whose power level reaches a certain level at different time instances, the corresponding frequencies at different times form the red line in Figure 3(b), which shows that, as the car moves closer, this frequency continues to increases, and its increase becomes much sharper when the car is really close to the user. This observation suggests that this particular frequency can be used as a new feature to infer whether the car is present, and how far it is from the user.

Let us look at two STDFT snapshots in Fig. 3(c). For each STDFT snapshot, we observe that the signal power drops as we go from lower frequency components to higher frequency components. This is because higher frequency components experience higher path loss. Let us first choose a specific power threshold value, e.g, $-58dB$, and then find the corresponding highest frequency component whose power level is

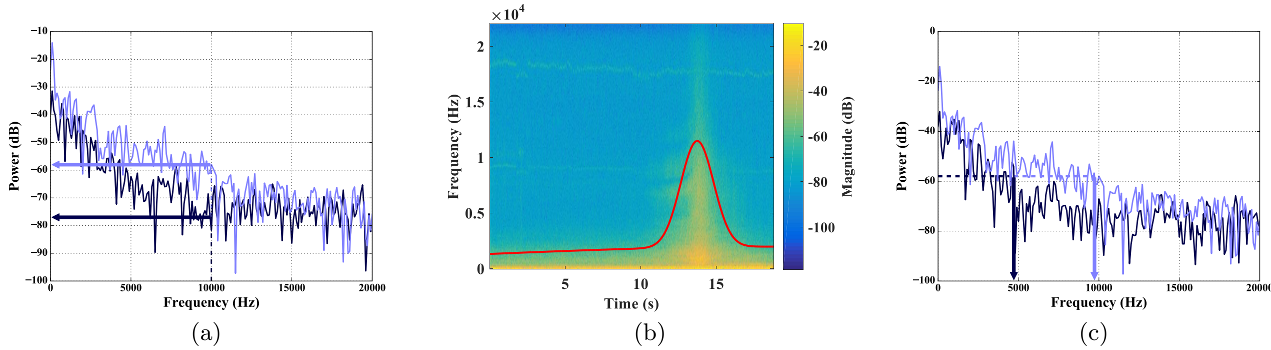

Figure 3: (a) Two STDFT snapshots taken at different timestamps and we see higher PoCF when the car is closer. (b) The STDFT result of the sound signal from a 2007 Toyota Camry which started 100 meters away from the user and ended a few meters past the user. The red line marks the highest frequency component whose power level reaches a certain threshold at different time instances. (c) Two STDFT snapshots taken at different times. The Top-Right frequency (TRF) is the maximum frequency component whose power level reaches a given power level threshold. The TRF is higher when the car is closer to the user

at or above this threshold on Fig. 3(c). We refer to this frequency value as the *Top-Right* frequency, or *TRF* in short, because if we draw a rectangle whose height is given by the power threshold value, the frequency of the top right corner of this rectangle is the frequency we are looking for. When the car is closer to the user, the signal's TRF value is higher. After selecting a suitable power threshold value, BO-Earcalculates the TRF values within each slot, and uses the moving average over 100 frames to smooth the results.

5. CONCLUSION & FUTURE WORK

In this paper, we have presented the overall design of BO-Ear, an unsupervised approach for detecting approaching cars. Also, we have explored two novel features, Power on Certain Frequencies (PoCF) and Top-Right Frequency (TRF), which provide more robust and timely detection than SPL. In the future, we will conduct extensive experiment to validate our system in ideal scenario and field test.

6. REFERENCES

[1] F. Bu and C.-Y. Chan. Pedestrian detection in transit bus application: sensing technologies and safety solutions. In *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE.* IEEE, 2005.

[2] K. Fitzpatrick, S. Turner, and M. A. Brewer. Improving pedestrian safety at unsignalized intersections. *Institute of Transportation Engineers. ITE Journal*, 77(5):34, 2007.

[3] M. Konishi. How the owl tracks its prey: experiments with trained barn owls reveal how their acute sense of hearing enables them to catch prey in the dark. *American Scientist*, 61(4):414–424, 1973.

[4] NHTSA. Traffic safety facts 2013. DOT HS 812 139.

[5] W. H. Organization et al. *Pedestrian safety: A road safety manual for decision-makers and practitioners.* World Health Organization, 2013.