

# Acoussist: An Acoustic Assisting Tool for People with Visual Impairments to Cross Uncontrolled Streets

WENQIANG JIN, The University of Texas at Arlington  
MINGYAN XIAO, The University of Texas at Arlington  
HUADI ZHU, The University of Texas at Arlington  
SHUCHISNIGDHA DEB, The University of Texas at Arlington  
CHEN KAN, The University of Texas at Arlington  
MING LI, The University of Texas at Arlington

To cross uncontrolled roadways, where no traffic-halting signal devices are present, pedestrians with visual impairments must rely on their other senses to detect oncoming vehicles and estimate the correct crossing interval in order to avoid potentially fatal collisions. To overcome the limitations of human auditory performance, which can be particularly impacted by weather or background noise, we develop an assisting tool called Acoussist, which relies on acoustic ranging to provide an additional layer of protection for pedestrian safety. The vision impaired can use the tool to double-confirm surrounding traffic conditions before they proceed through a non-signalized crosswalk.

The Acoussist tool is composed of vehicle-mounted external speakers that emit acoustic chirps at a frequency range imperceptible by human ears, but detectable by smartphones operating the Acoussist app. This app would then communicate to the user when it is safe to cross the roadway. Several challenges exist when applying the acoustic ranging to traffic detection, including measuring multiple vehicles' instant velocities and directions with the presence many of them who emit homogeneous signals simultaneously. We address these challenges by leveraging insights from formal analysis on received signals' time-frequency (t-f) profiles. We implement a proof-of-concept of Acoussist using commercial off-the-shelf (COTS) portable speakers and smartphones. Extensive in-field experiments have been conducted to validate the effectiveness of Acoussist in improving mobility for people with visual impairments.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**.

Additional Key Words and Phrases: Pedestrian safety, collision avoidance, acoustic ranging

## ACM Reference Format:

Wenqiang Jin, Mingyan Xiao, Huadi Zhu, Shuchisnigdha Deb, Chen Kan, and Ming Li. 2020. Acoussist: An Acoustic Assisting Tool for People with Visual Impairments to Cross Uncontrolled Streets. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 133 (December 2020), 30 pages. <https://doi.org/10.1145/3432216>

---

Authors' addresses: Wenqiang Jin, The University of Texas at Arlington, [wenqiang.jin@mavs.uta.edu](mailto:wenqiang.jin@mavs.uta.edu); Mingyan Xiao, The University of Texas at Arlington, [mingyan.xiao@mavs.uta.edu](mailto:mingyan.xiao@mavs.uta.edu); Huadi Zhu, The University of Texas at Arlington, [huadi.zhu@mavs.uta.edu](mailto:huadi.zhu@mavs.uta.edu); Shuchisnigdha Deb, The University of Texas at Arlington, [shuchi.deb@uta.edu](mailto:shuchi.deb@uta.edu); Chen Kan, The University of Texas at Arlington, [chen.kan@uta.edu](mailto:chen.kan@uta.edu); Ming Li, The University of Texas at Arlington, [ming.li@uta.edu](mailto:ming.li@uta.edu).

---

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](mailto:permissions@acm.org).

© 2020 Association for Computing Machinery.  
2474-9567/2020/12-ART133 \$15.00  
<https://doi.org/10.1145/3432216>



Figure 1. Some examples of uncontrolled crosswalks.

## 1 INTRODUCTION

**Motivation:** According to the records provided by the World Health Organization in 2018, the global-wide visually impaired people are estimated at 2.2 billion [84]. How to navigate them to cross streets is a long-lasting topic. The state-of-art solution is to install the Accessible Pedestrian Signals (APS) at intersections or crossing sections to assist the visually impaired in determining when it is safe to cross. However, there are even more uncontrolled crosswalks where no traffic control (i.e., traffic signals or APS) is present. These common crossing types occur at non-intersection or midblock locations where they may be marked or not. They typically exist in residential communities, local streets, suburban areas, etc. where sophisticated traffic infrastructures are too expensive to fully deploy around. In these road sections, the visually impaired have to mostly depend on themselves to judge the surrounding traffic condition and decide whether it is safe to proceed to the crosswalks. In practice, the pedestrian leverage hearing to discriminate between traffic sounds that are too far away to pose a hazard to crossing and those that are within close proximity. Nonetheless, hearing based judgment is not always reliable. Hearing capability varies from person to person; young people generally have more sensitive hearing than seniors. Besides, environmental conditions may affect traffic sounds. For example, rain and wind may enhance or distort sounds; snow can muffle sounds; background construction sounds or talking from people nearby may even overwhelm the traffic sounds.

This paper aims to develop a portable tool that assists pedestrians with vision impairments to cross uncontrolled streets. The tool alerts pedestrians with the presence of oncoming vehicles that may cause hazard. To achieve this goal, it is essential to figure out movement status of each vehicle nearby, characterized by, for example, its velocity relative to the pedestrian, direction of arrival (DoA), and its distance to the pedestrian. To measure these parameters, one possible solution is to use radar [68]. Due to its stringent requirements over the received signal quality [101, 122, 124], the existing radar applications mostly operate over licensed spectrum bands. For instance, Federal Communications Commission (FCC) designates the X band frequencies between 10.500-10.550 GHz and the K band frequencies between 24.050-24.250 GHz for police radar gun. In our case, applicable radio frequencies the already over-crowded ISM bands. Besides, given that our problem requires the ranging distance up to 200 ft, the perceived signal-to-noise ratio (SNR) would be too low to achieve meaningful detection. Moreover, we need build our own transmitter or/and receiver by using radar techniques. In contrast, the proposed design can be implemented on commercial off-the-shelf (COTS) devices.

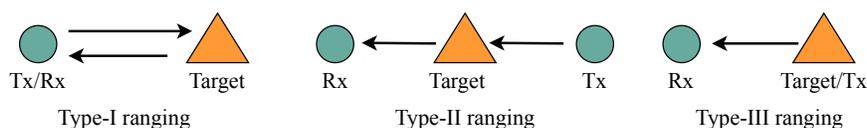


Figure 2. Illustration of three types of ranging.

LiDAR [4, 28, 92], which uses predominantly infrared light from lasers rather than radio waves, is another potential alternative. It has been widely used in autonomous vehicles to monitor surrounding environments to avoid traffic incidents. However, LiDAR is a technology that requires a powerful computation and storage capacity to handle the huge collected dataset. Its cost is another concern.

Given the above analysis, our idea leverages acoustic ranging. It is competent for our implementation in the following aspects. First, acoustic signals can propagate around obstructions through diffraction on their edges or reflection from their surfaces. This capability supports measuring vehicles behind obstructions. Second, its performance does not depend on lighting conditions and is effective even in darkness. Third, we can easily customize transmission signals on commercial COTS speakers and process received signals in smartphones. We propose to utilize ultrasound signals ranging from 17 KHz to 19 KHz. To our knowledge, there is no application with wide deployment operating on this band. Thus, it is less likely to experience cross-application interference.

Motivated by the above observations, we develop Acoussist, an **acoustic based assisting** tool for the visually impaired to cross uncontrolled streets. Acoussist consists of speakers that are mounted on the front of vehicles to emit ultrasonic chirps and an app running on the pedestrian’s smartphone for signal analysis. Whenever a pedestrian senses a clear street and tends to proceed to the crosswalk, she turns on the app to double-confirm her judgment. The app analyzes the received chirps to detect if any oncoming vehicle would cause potential collisions. If yes, an alert is generated. Then, the pedestrian should take precaution and wait at the curb until the street is clear. For vehicles operating in all-electric or hybrid mode, the chirps can be played by their *warning sounds system*<sup>1</sup>. For traditional combustion engine vehicles, chirps are proposed to emit from COTS portable speakers. While pedestrians are “stakeholders” in our scenario, drivers do not necessarily lack the motivation to install the speakers. Department of Public Safety (DPS) [21] regulates that a driver who fails to yield to the pedestrians with vision impairments (regardless of any reasons) is fully/partially liable for any injury caused to the pedestrian. As demonstrated in this work, a participatory vehicle can effectively alert a pedestrian regarding its presence.

**Challenges:** Although the idea of acoustic ranging is not new, turning it into a tool for collision detection is faced with several unique challenges. 1) *Vehicle velocity measurement with mutual interference:* One of the key ingredients to decide if a vehicle causes potential hazard is to figure out its velocity relative to the pedestrian. A straightforward solution is to analyze the Doppler frequency shift of the received chirps at the receiver. This task is easy if only one vehicle is nearby. In our case, oftentimes several vehicles are present. Their emitted chirps overlap, rendering distinguishing among them an extremely challenging task, let alone analyzing the frequency shift for velocity measurement. 2) *Dynamic multi-source localization:* It is also indispensable to figure out the DoA of each vehicle. Acoustic multi-source localization has been studied in the domain of speaker tracking in video conferences [24, 63, 67], indoor localization [23], and noise identification [19]. Existing approaches either assume the number of sources are fixed and *a priori* known, or signals from different sources are of distinct frequency patterns. Thus, none of them is applicable to our problem. It is also worth-mentioning a set of novel ranging-based applications, such as breathing pattern detection [13, 75, 118, 123, 125] and hand and finger gesture

<sup>1</sup>Many countries have approved legislation to enforce “quiet” vehicles install the *warning sounds system*, an array of external speakers that emit artificial engine sounds for pedestrians to be aware of their presence. For example, EU requires all new models of electric and hybrid vehicles developed and sold in EU to equip the system by July 2019 [41].

detection [51, 76, 93, 102, 119]. These applications can be classified as either *type-I ranging*<sup>2</sup> or *type-II ranging*, while our system belongs to *type-III ranging*. For the former two types, the analysis is carried over target-reflected signals. For the latter type, the analysis is over target-emitted signals. Thus, their design rationale and applied techniques are quite different.

**Our Approach:** We find that the received samples associated with each acoustic source generate a series of pulses in the time-frequency (t-f) domain. Lining up these pulses produces a “t-f sweep line” that can identify the corresponding acoustic source due to the unique combination of its signal offset time and the moving speed. Our formal analysis reveals that the slope of each t-f sweep line is a function of the vehicle’s relative velocity. We thus address the first challenge by exploiting this relationship. *To our best knowledge, no existing literature has provided any closed-form formula of a moving source’s t-f sweep line in the expression of its relative velocity, let alone leveraging the relationship for velocity estimation.* We address the second challenge by developing a modified generalized cross correlation method, called *MGCC*. MGCC consists of three major components: 1) extracting the LoS transmission component from the received signal for each acoustic source, 2) applying the generalized cross correlation function over the extracted signals received by two microphones on the smartphone to obtain the time difference of arrival (TDoA), and 3) calculating the DoA for each vehicle based on its TDoA. Comparing with conventional GCC, which is incapable of dealing with association ambiguity caused by coexistence of multiple sources, the proposed MGCC avoids this issue by analyzing t-f profiles of each source extracted from the previous step.

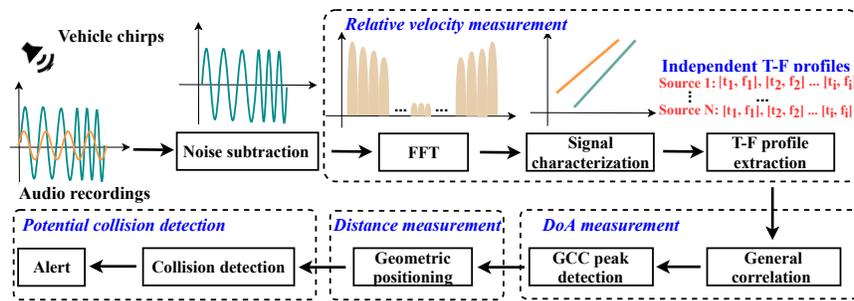


Figure 3. Acoussist system architecture.

The processing flow of Acoussist is summarized in Figure 3. It consists of vehicle-side mounted external speakers that emit acoustic chirps ranging from 17 KHz to 19 KHz and a pedestrian-side detection app. To facilitate the multi-source localization, two microphones at the pedestrian’s smartphone are utilized. Upon receiving acoustic samples, Acoussist applies a high-pass filter to remove low-frequency components in the recorded samples. It then performs the short-time Fourier transform (STFT) over the de-noised samples. A t-f sweep line is identified for each acoustic source. It further estimates the relative velocity for each vehicle by analyzing the slope for each t-f sweep line. Following that, it measures the DoA of acoustic sources via the proposed MGCC method; its inputs are the t-f profiles from each acoustic source obtained from the previous step. By employing the geometric relations between the acoustic source and the pedestrian, it then calculates

<sup>2</sup>By ranging, we mean localizing/detecting objects based on their reflected or emitted electromagnetic (EM) or acoustic signals. According to the relation among the target, transmitter, and receiver in a ranging system, we define the following three types of ranging. For *type-I ranging*, an EM/acoustic wave is emitted from a transmitter and reflects off the target to the receiver. The signal is reflected back to the receiver that picks up the echoed signal. The transmitter and the receiver are collocated. For *type-II ranging*, the signal is also emitted from the transmitter, reflected by the target, and captured by the receiver. The only difference is that the transmitter and the receiver are located differently. They are either cooperative or not. For *type-III ranging*, the target is localized through the analysis over its own emitted signals. Thus, the target is also the transmitter.

each vehicle's moving velocity and its distance to the pedestrian. Finally, the app decides if a pedestrian is safe to proceed to the crosswalk by analyzing the relations among all derived movement parameters, and produces an alert if needed.

As mentioned, Acoussist adopts the framework of *type-III ranging* that analyzes target's self-emitted signal for detection. If we adopt the other two types of ranging that utilizes target's reflected signal, the ranging distance would be significantly reduced due to the high decay coefficient experienced by acoustic signals. As shown in our experiments, the signal is still detectable by a smartphone when the v-p distance is as long as 240 ft under type-III ranging, but it drops to 48 ft under type-I/-II ranging which is unsuitable for moving object collision detection. Besides, Acoussist works for smartphones with more than one microphone. Luckily, most of current smartphones meet this requirement. As a note, Apple equips their devices with even four embedded microphones since iPhone 7 released in 2016.

The key contribution of this paper is summarized as follows:

- We develop an acoustic based collision detection system that assists pedestrians with vision impairments to perceive surrounding traffic conditions before crossing uncontrolled streets. It supplements the conventional hearing based solution. While collision avoidance systems for automobiles have been investigated for more than a decade, human-centered collision detection has rarely been studied.
- We address unique challenges when applying acoustic ranging to collision detection. Two salient technical contributions have been made. First, we propose a novel t-f sweep line based analysis that derives vehicle's relative velocity. With this basis, MGCC is developed to calculate vehicle's DoA according to its TDoA with respect to the smartphone's two mics. Both detection methods are capable of differentiating among multiple vehicles.
- From a generalized point of view, we study a *type-III homogeneous multi-source ranging* problem that has rarely been investigated in prior *ranging* literature.
- We implement Acoussist on COTS speakers and mobiles without involving any central server. We demonstrate the feasibility of Acoussist via extensive in-field testing.

**Clarifications:** *First*, Acoussist does not intend to overwrite a pedestrian's judgement; instead, it should be treated as an assisting tool that provides an added layer of protection for the visually impaired. That being said, if a conflict occurs between a pedestrian's judgement and the detection result, the pedestrian still holds the responsibility of decision making. A suggestive choice is to wait by the curb when either source indicates a potential collision. *Second*, Acoussist is designed to use at uncontrolled crosswalks existing in residential communities, local streets, and suburban areas, where there are common needs from the visually impaired for daily activities and commute. These venues generally impose relatively conservative vehicle speed limits. In an interview with five visually impaired students in our university, all of them claim that they would never consider crossing any less regulated road sections that allow 45 mph speed limit or higher on their own. *Third*, we stress that the functionality of Acoussist does not require all vehicles to participate. It can perform collision detection only to the participatory ones. In a worst case that no vehicle in the pedestrian's vicinity opt-in, it degrades to the hearing-based judgment scenario. Therefore, Acoussist will not perform worse than the current solution.

## 2 OVERVIEW AND BACKGROUND

### 2.1 Design Rationale

The White Cane Laws give visually impaired pedestrians the right-of-way in crosswalks, whether or not they are marked [81]. The laws require drivers to stop and yield to the blind who is crossing the street. On the other hand, the visually impaired rely on themselves to judge if the street is clear and when to proceed to the crosswalk by mainly referring to hearing. Since hearing is not always reliable, for example, misled by the background noise, the blind may wrongly judge the traffic condition and enter streets even there are oncoming vehicles within close

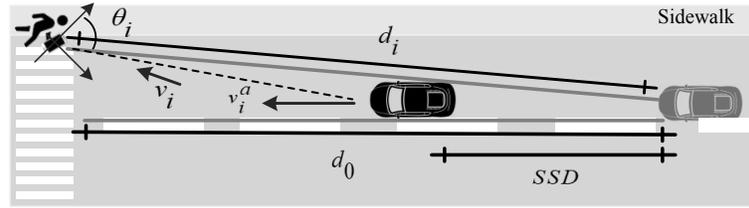


Figure 4. Design rationale of Acoussist.

proximity. To avoid this hazard situation, Acoussist senses surrounding traffic conditions and estimates whether there is sufficient time for nearby drivers to spot the pedestrian and then take reaction to stop cars. If not, an alert is generated at the pedestrian's smartphone, keeping her from proceeding to the crosswalk.

The design of Acoussist relies on a basic assumption that drivers obey the laws; for the collisions that are caused by careless driving are out of the scope of our discussion. Besides, there are some scenarios that people are using headphones/earphones listening to music or talking on the phone when crossing streets without care. They are not the focus of this work too. As shown in Figure 4, denote by  $d_0$  the distance between a vehicle and the crosswalk, when the pedestrian is first spotted entering the crosswalk. Then the driver takes reaction and stops the vehicle. The distance that the vehicle travels before complete stop is called *stopping sight distance* (SSD) [60]. It is a near worst-case distance a driver needs to be able to see in order to have room to stop before colliding with something in the roadway. SSD is also one of several types of sight distance commonly used in road design. If  $d_0 > SSD$ , i.e., the driver can stop the car before hitting into the pedestrian, the pedestrian can safely cross. This inequality can be used as a condition for pedestrian safety. Nonetheless, it is challenging for the pedestrian to measure  $d_0$ . Alternatively, we consider  $d_i$ , the v-p distance. Generally,  $d_i \approx d_0$  when the vehicle is faraway. For example, given  $d_0 = 150$  ft and the street width 30 ft, then  $d_i \leq \sqrt{150^2 + 15^2} = 150.7$  ft. Thus, the pedestrian safety condition can be rewritten as  $d_i > SSD$ .

As discussed later,  $d_i$  is a function of DoA of the vehicle to the pedestrian, denoted by  $\theta_i$ , and SSD is a function of the vehicle's moving velocity, denoted by  $v_i^a$ . Besides,  $v_i^a$  is dependent of  $\theta_i$  and the vehicle's velocity relative to the pedestrian, denoted by  $v_i$ . Eventually, our problem to determine if  $d_i > SSD$  is satisfied is converted to estimate the values of  $v_i$ ,  $\theta_i$ ,  $v_i^a$ , and  $d_i$ .

## 2.2 Acoussist Signal Design

Acoussist uses external speakers to emit acoustic chirps periodically. As shown in Figure 5(a), a chirp's frequency linearly sweeps from the minimum  $f_l$  to the maximum  $f_h$  over time. Chirp signals are widely used in radar applications for its capability of resolving multi-path propagation. In the time domain, the expression for one chirp is

$$s_c(t) = A \cos\left(\pi \frac{B}{T} t^2 + 2\pi f_l t\right) \quad (1)$$

where  $A$  is the amplitude,  $B = f_h - f_l$ ,  $t \in (0, T]$ , and  $T$  is the chirp duration. We choose a high frequency chirp ranging from  $f_l = 17$  KHz to  $f_h = 19$  KHz. Such a range has been adopted by quite a few novel applications, such as biometric sensing [87] and acoustic imaging [71].

Although the frequency range of human hearing is generally considered from 20 Hz to 20 KHz, high frequency sounds must be much louder to be noticeable (including children and young adults) [91]. This is characterized by the absolute threshold of hearing (ATH), which refers to the minimum sound pressure that can be perceived in a quiet environment. According to [103], we depict in Figure 5(b) the ATH with respect to sound frequency. ATH increases sharply for frequencies over 10 KHz. In particular, human ears can detect sounds of 1 KHz at 0 dB sound pressure level (SPL), but above 75 dB SPL for sound beyond 17 KHz, which has about 10,000 fold

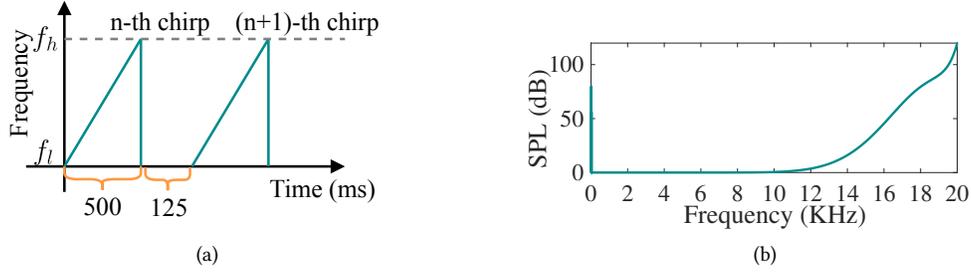


Figure 5. (a) Time-frequency domain representation of the chirp signal. (b) Human hearing threshold.

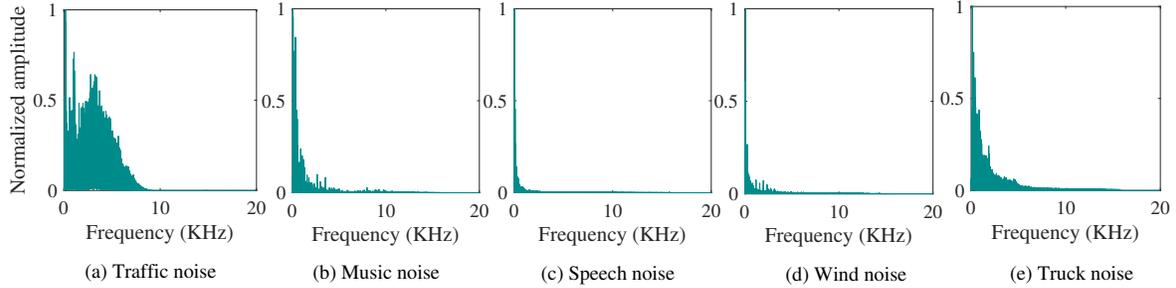


Figure 6. Frequency spreading of different background noises.

amplitude increase. In our implementation, the chirp signal is played at 69.3 dB and thus hardly perceptible by human hearing.

Another concern of applying acoustic ranging is that the signal may be polluted by background noise in outdoor environments. We extract some recordings for commonly seen outdoor noise from the well acknowledged dataset provided by the Google Audioset [36] and analyze their spectrum distribution. We find in Figure 6 that the background noise mainly concentrates on the lower-end of the frequency, mostly lower than 10 KHz. As our chirp signals are above 17 KHz, there is a clear gap between these two. By applying a high-pass filter to the received signal can easily filter out background noise.

The chirp duration and the separation between two consecutive chirps impact the overall performance of Acoussist. Too short a chirp will cause blurs in t-f profiles of received signals. Also, the chirp separation should be large enough to ensure that the main reflected signals of the current chirp are received before the next chirp is transmitted. However, too long a chirp duration or the separation will add delay to the system. As examined in 5.2, we found empirically that a duration of 500 ms and a separation of 125 ms represent a good tradeoff.

### 3 MULTI-VEHICLE SIGNAL CHARACTERIZATION

In this section, we model the received signal with the presence of multiple vehicles<sup>3</sup>. The insight that we develop from this model guides the design of different modules of Acoussist.

As surrounding objects, such as buildings and trees, reflect acoustic signals, a chirp emitted from vehicle  $s_i$  arrives at a microphone  $r_m$  from multiple paths. Denote by  $d_{i,m}^k$  the length of the  $k$ -th path associated with  $s_i$  and  $r_m$ .  $v_a$  is the speed of acoustic signals, which is considered as 340 m/s in our system. Let  $\phi_{i,k}$  be the angle between the DoA of the  $k$ -th path and the line-of-sight (LoS) path respect to mic  $r_m$ . Following [110], the corresponding

<sup>3</sup>In this paper, we use “vehicle” and “source” interchangeably without causing confusion.

time-dependent signal propagation delay is calculated by

$$\tau_{i,m}^k(t) = \frac{d_{i,m}^k - v_i \cos(\phi_{i,k})t}{v_a}.$$

Let  $a_{i,m}^k(t)$  be the attenuation experienced by the acoustic signal transmitted via the  $k$ -th path. Then, the aggregated time-domain channel response between  $s_i$  and  $r_m$  is expressed by

$$h_{i,m}(\tau, t) = \sum_{k=1}^K a_{i,m}^k(t) \delta(\tau - \tau_{i,m}^k(t))$$

i.e., the accumulated pulses arriving at different propagation delays. Given a particular time instance  $t$ , the source signal (1) can be rewritten as  $s_c(t) = A \cos(2\pi f_t t)$ , in which  $f_t = \frac{d(\pi \frac{B}{T} t^2 + 2\pi f_i t)}{2\pi dt} = \frac{B}{T} t + f_i$ . Then, the time-domain expression for mic  $r_m$  received signal coming from vehicle  $s_i$  is

$$y_{i,m}(\tau, t) = \sum_{k=1}^K A a_{i,m}^k(t) \cos(2\pi f_t (t - \tau_{i,m}^k(t))) = \sum_{k=1}^K A a_{i,m}^k(t) \cos(2\pi f_t (1 + \frac{v_i \cos(\phi_{i,k})}{v_a})t - \frac{2\pi f_t d_{i,m}^k}{v_a}).$$

By applying the continuous Fourier transformation over  $y_{i,m}(\tau, t)$ , its t-f representation is

$$Y_{i,m}(f, t) = \frac{A}{2} \sum_{k=1}^K \frac{a_{i,m}^k(t) v_a}{v_a + v_i \cos \phi_{i,k}} e^{-j2\pi f \frac{d_{i,m}^k}{v_a + v_i \cos \phi_{i,k}}} \times [\delta(f - f_t \frac{v_a + v_i \cos \phi_{i,k}}{v_a}) + \delta(f + f_t \frac{v_a + v_i \cos \phi_{i,k}}{v_a})].$$

Then, the t-f representation of the aggregated received signal from all  $N$  vehicles at mic  $r_m$  is written as

$$Y_m(f, t) = \frac{A}{2} \sum_{i=1}^N \sum_{k=1}^K \frac{a_{i,m}^k(t) v_a}{v_a + v_i \cos \phi_{i,k}} e^{-j2\pi f \frac{d_{i,m}^k}{v_a + v_i \cos \phi_{i,k}}} \times [\delta(f - (\frac{B}{T} t + f_i) \frac{v_a + v_i \cos \phi_{i,k}}{v_a}) + \delta(f + (\frac{B}{T} t + f_i) \frac{v_a + v_i \cos \phi_{i,k}}{v_a})]. \quad (2)$$

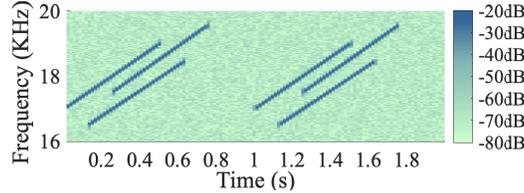


Figure 7. T-F profile of the received signal with the presence of three vehicles.

Figure 7 depicts the t-f profile of the received signal  $Y_m(f, t)$  at a microphone when three vehicles are present, with their relative velocities  $v_i$ 's at 0 mph, 20 mph, and -20 mph, respectively. The darker part in this heat map indicates the components of large power.

**Insight:** We can tell from (2) that the t-f profile of received samples consist of a series of impulses that exist when the corresponding  $f$  and  $t$  satisfy a set of linear equations  $f = (\frac{B}{T} t + f_i) \frac{v_a + v_i \cos \phi_{i,k}}{v_a}$  ( $k \in \{1, \dots, K\}$ ). Besides, according to [11],  $a_{i,m}^k(t) \propto e^{-\gamma d_{i,m}^k(t)}$  where  $\gamma$  is the acoustic amplitude decay coefficient. In the air propagation environment,  $\gamma$  is at least 40.3 dB/100m when the acoustic signal's frequencies are between [17 KHz, 19 KHz] [116]. Thus, the signals coming from the LoS transmission path (with index  $k = 1$ ) out-weight other components from the rest paths. This is also another benefit of employing acoustic signals rather than radio signals. It is more

convenient to extract LoS component from received signals that are mixed with multi-path transmissions. Each “line” in Figure 7 can be specified by

$$f = \left(\frac{B}{T}t + f_l\right) \frac{v_a + v_i \cos \phi_{i,1}}{v_a} = \left(\frac{B}{T}t + f_l\right) \frac{v_a + v_i}{v_a} \quad (3)$$

as  $\phi_{i,1} = 0$ . We call such a line as a *t-f sweep line*. As shown in Figure 7, each vehicle corresponds to a unique t-f sweep line, due to its unique combination of  $v_i$  and the chirp signal offset time. Therefore, the number of different t-f sweep lines that a mic detects implies the number vehicles within its proximity. More importantly, we also notice that the slope of a line contains the information of  $v_i$ . Therefore, we are able to infer the number of nearby vehicles and their relative velocities by analyzing t-f profile of received signals.

While the t-f signal analysis is a common approach for target tracking in a radar system, no existing work has tried to establish an explicit relation between the received signal’s t-f sweep line and the source relative velocity. Such a relation enables velocity estimation when multiple sources with homogeneous signals are present. As discussed later, the analysis also lays the basis for the DoA measurement module. One reason that this idea has not been explored previously is because the generalized problem, *type-III homogeneous multi-source ranging*, is hardly observed in any other real-world ranging-based applications. For example, speaker localization [24, 63, 67] and noise identification [19] can be classified as *type-III heterogeneous multi-source ranging*. Radio-based indoor localization [14] can be treated as *type-III single-source ranging*. Radio-based breathing pattern detection [2], sleep monitoring [64], gesture detection [114], and radar guns all belong to *type-I ranging* or *type-II ranging*.

## 4 DESIGN DETAILS OF ACOUSSIST

Next, we discuss the design details of Acoussist’s modules and describe how they interact to perform multi-vehicle detection.

### 4.1 Measurement of Relative Velocity

The objective of this module is to estimate each vehicle’s velocity relative to the pedestrian,  $v_i$ . A straightforward solution is to analyze the Doppler frequency shift of the received chirps at the receiver. This task is easy if only one source is nearby or source signals are heterogeneous, e.g., different combinations of frequency components. In our case, oftentimes several vehicles are present. Additionally, they emit homogeneous chirps. These chirps overlap at the receiver, rendering distinguishing among them an extremely challenging task, let alone analyzing the frequency shift for velocity measurement. To address this issue, in the previous section, we formulate the received signal into a generalized expression that quantifies the effect of source movements by jointly characterizing the received signal’s time and frequency properties. More importantly, we model the t-f profile of a source as a closed-form expression of its relative velocity. Specifically, the slope of a t-f sweep line, denoted as  $\kappa$ , is unique and dependent of  $v_i$ , i.e.,  $\kappa = \frac{B}{T} \frac{v_a + v_i}{v_a}$ . Hence,  $v_i$  is calculated as  $v_i = v_a(\kappa T/B - 1)$ . As  $v_a$ ,  $T$  and  $B$  are known values, the remaining task is to find out  $\kappa$  of each t-f sweep line.

The app takes the denoised-stream at runtime as input and continuously slides a window of short time-width over it to get t-f profile by applying short-time Fourier transform (STFT) at each window. Consider a sliding window indexed by  $l$ ; the window size is  $\Delta t$ . Figure 8(a) depicts all frequency components contained in this window. As a note, Figure 8(a) is actually a slice of Figure 7 at window  $l$ . Each peak exists at the frequency  $f_i = \left(\frac{B}{T}t + f_l\right) \frac{v_a + v_i}{v_a}$  with  $t = l \cdot \Delta t$ . The rest components are the signals from multi-path propagation or ambient noise that may consist of sound of sudden wind or machinery in a construction site or their harmonics in the higher frequency range. We then identify all the peaks in window  $l$  by applying the *peak detection algorithm* [8]. Denote by  $(l, f_i)$  an index-frequency pair of window  $l$ . We are able to identify three such pairs in Figure 8(a), indicating three detectable vehicles.

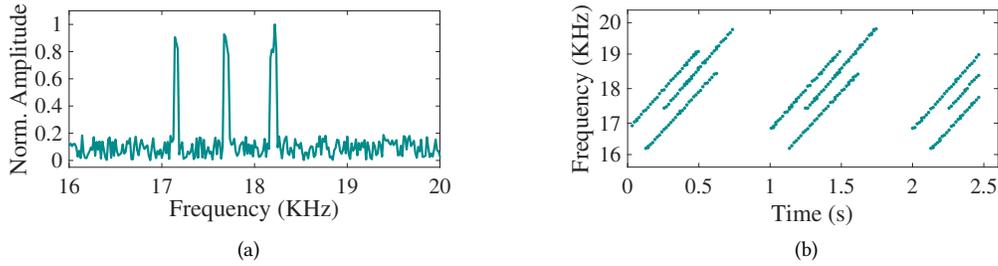


Figure 8. (a) Normalized amplitude of all frequency components in a window. (b) All detected peaks in the received signal's t-f profile.

Figure 8(b) plots all the index-frequency pairs obtained at all windows of the received signal's t-f profile; each dot is associated with one pair. To extract the t-f sweep lines from a set of discrete dots, we employ the *Hough transformation* technique [6], which is a classic scheme in detecting edges, including lines, circles and ellipses, from a digital image. In our scenario, by treating the collected index-frequency pairs as the entire dataset, the t-f sweep lines are then detected as the edges by applying Hough transmissions over the dataset. Since there are substantial prior discussions on Hough transformation, we omit its implementation details here. So far, we are able to extract the t-f sweep lines and thus the slope  $\kappa$  for each of them. Then the number of lines is exactly the number of detectable vehicles. Each relative velocity is computed by  $v_i = v_a(\kappa T/B - 1)$ .

One critical issue in STFT is to decide the frequency resolution ( $\Delta f$ ) and time resolution, i.e., sliding window size ( $\Delta t$ ). These two parameters determine whether frequency components close together can be separated and the time at which frequencies change. A properly selected set of  $\Delta f$  and  $\Delta t$  produces concentrated, rather than blurred, t-f sweep lines, which are essential to measure relative velocity accurately. Suppose the microphone's sampling rate is 64 KHz. Here, 64KHz is a conservative value. Many smartphones, such as Razer phone 2 and Pixel XL, even support a sampling rate of 192KHz.  $\Delta t$  and  $\Delta f$  are computed as  $\Delta t = N/64$  KHz and  $\Delta f = 64$  KHz/ $N$ , respectively, where  $N$  is the number of samples taken from a window. To strike a balance between these two,  $N$  is set to 2048 in our system. Given the chirp duration as 500 ms, the time resolution  $\Delta t = 2048/64$  KHz = 32 ms is precise enough to capture the variance of received chirps in the time domain caused by multi-path propagation. Consider that most of vehicles are not traveling with speed lower than 10 mph, i.e., 4.5 m/s. According to (3), its spectrum offset is  $f_i \frac{v_i}{v_a}$ , which ranges from 225 Hz ( $f_i = 17$  KHz) to 251 Hz ( $f_i = 19$  KHz). Thus, the frequency resolution  $\Delta f = 64$  KHz/2048 = 31 Hz is satisfactory to measure the spectrum offset.

#### 4.2 Measurement of DoA

This module estimates the vehicle's DoA with respect to the pedestrian,  $\theta_i$ . As shown in Figure 9, it is the angle between the LoS transmission and the line connecting two mics on the phone. Its measurement is achieved by analyzing the time difference of arrival (TDoA), denoted by  $\tau_i$ , at the two mics. Since the v-p distance is much larger than the inter-mic distance, denoted by  $D$ , LoS propagation paths to the two mics are deemed parallel with each other. Then,  $\theta_i$  is calculated as  $\theta_i = \arccos(\frac{\tau_i v_a}{D})$ . As  $v_a$  and  $D$  are available values, the remaining task is to find out  $\tau_i$ . The inter-distance of two microphones of a mobile phone is available in open source dataset like [37]. This value can be instantiated during the initiation of the app.

**A naive approach:** It examines the inter-channel phase difference (ICPD) to derive DoA [16, 127]. Specifically, the ICPD observed by two mics can be

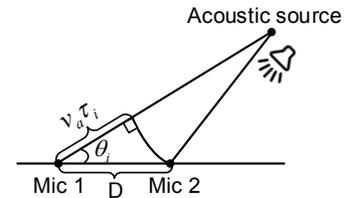


Figure 9. Illustration of vehicle DoA.

expressed as

$$\psi_i(f) = \angle \frac{Y_1(f, l)}{Y_2(f, l)} = 2\pi f \tau_i + 2\pi p_f \quad (4)$$

where  $Y_m(f, l)$  ( $m = 1$  or  $2$ ) is the T-F representation of the signal received at mic  $r_m$  under discrete time. The wrapping factor  $p_f$  is a frequency-dependent integer and  $2\pi p_f$  represents possible phase wrapping. If  $p_f = 0$ , then  $\tau_i$  is calculated by  $\angle \frac{Y_1(f, l)}{Y_2(f, l)} / 2\pi f$  and thus our problem is solved. In theory,  $p_f = 0$  when  $D$  is smaller than half the wavelength [16]. In our system, the longest wavelength is about 2cm ( $\frac{340\text{m/s}}{17\text{kHz}}$ ). Its halve, i.e., 1cm, is apparently smaller than the inter-mic distance for many smartphones, e.g., the ones with one mic located on bottom and the other on top. Therefore, the ICPD based TDOA measurement is unreliable in our case.

**The proposed approach:** We start from a conventional approach, called generalized cross correlation (GCC), to identify TDoA [15, 113]. Consider a function

$$R(\tau) = \left| \sum_l \sum_f \frac{Y_1^*(f, l) Y_2(f, l)}{|Y_1(f, l) Y_2(f, l)|} e^{-j2\pi f \tau} \right| = \left| \sum_i \sum_l \sum_f e^{-j\psi_i(f)} e^{-j2\pi f \tau} \right|$$

where  $Y_m(f, l)$  ( $m = 1$  or  $2$ ) follows the definition above.  $Y_m^*(f, l)$  stands for the conjugate of  $Y_m(f, l)$ .  $\psi_i(f, l) = 2\pi f \tau_i$  is the phase difference between  $Y_1$  and  $Y_2$ . Ideally,  $R(\tau)$  shows a peak at  $\tau = \tau_i$ . However, due to the presence of multiple vehicles, the strong interference from multiple acoustic sources generates multiple peaks, as shown in Figure 10(a).

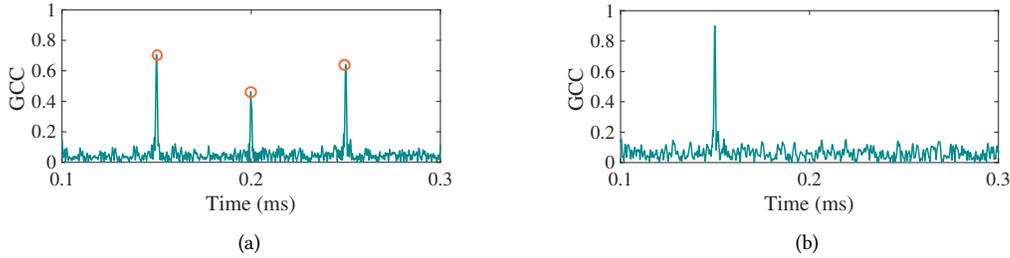


Figure 10. (a) Association ambiguity caused by multi-source and multi-path interference. (b) Eliminate association ambiguity by calculating GCC of LoS signals from a separated single source.

To avoid the association ambiguity in the conventional GCC approach, we adapt it to multi-source scenarios. Recall that in Section 4.1 we are able to extract the discrete index-frequency pairs for each source  $s_i$ , and thus the t-f profile of the received signal from  $s_i$  via the LoS path, denoted as  $Y_{i,m,1}(f, l)$ . Here, “1” means the index of the LoS path. Then, a modified GCC (MGCC) function that associated with a particular source  $s_i$  is expressed as

$$\begin{aligned} R_i(\tau) &= \left| \sum_l \sum_f \frac{Y_{i,1,1}^*(f, l) Y_{i,2,1}(f, l)}{|Y_{i,1,1}(f, l) Y_{i,2,1}(f, l)|} e^{-j2\pi f \tau} \right| \\ &= \left| \sum_l \sum_f e^{-j\psi_i(f)} e^{-j2\pi f \tau} \right|. \end{aligned} \quad (5)$$

Since  $Y_{i,1,1}$  and  $Y_{i,2,1}$  only contain the signals that are transmitted via the LoS paths from source  $s_i$ , there is only one peak at  $\tau = \tau_i$ , as shown in Figure 10(b). Thus, the TDoA  $\tau_i$  of source  $r_i$  can be calculated by  $\tau_i = \arg \max R_i(\tau) \ i \in [1, N]$ . Then,  $\theta_i$  is derived accordingly.

### 4.3 Measurement of Vehicle Velocity and V-P Distance

**Vehicle velocity:** The measurement of vehicle  $s_i$ 's velocity  $v_i^a$  relies on the estimation results of its relative velocity  $v_i$  and DoA  $\theta_i$  derived in Section 4.1 and 4.2, respectively.

Denote by  $v_i^a(t_1)$  and  $v_i^a(t_2)$  the vehicle's instance velocities at  $t_1$  and  $t_2$ , respectively. Let  $\delta_t = t_2 - t_1$  be the measurement interval. In the implementation,  $\delta_t$  is set to 500 ms. Thus,  $v_i^a(t_1)$  and  $v_i^a(t_2)$  are deemed equal. As shown in Figure 11, denote by  $\alpha$  the angle between the vehicle's moving direction  $AB$  and the line connecting two mics  $O_1O_2$ . Besides, let  $\beta$  be the angle between the vehicle's velocity  $v_i^a$  and its velocity relative to the pedestrian  $v_i$ . Then we have the following system of equations

$$\begin{cases} v_i^a(t_1) \cos \beta(t_1) = v_i(t_1), & v_i^a(t_2) \cos \beta(t_2) = v_i(t_2) \\ \alpha + \beta(t_1) = \theta_i(t_1), & \alpha + \beta(t_2) = \theta_i(t_2), & v_i^a(t_1) = v_i^a(t_2) \end{cases}$$

The first two equations are from the relation between  $v_i^a$  and  $v_i$ . Regarding the third equation,  $\theta(t_1) = \alpha + \angle OAO_1$  due to the application of exterior angle theorem in the triangle  $\Delta AO_1B$  in Figure 11. Similarly, the fourth equation holds. Since there are five variables ( $\alpha, \beta(t_1), \beta(t_2), v_i^a(t_1), v_i^a(t_2)$ ) and five uncorrelated equations above, we can derive the closed form expression for  $v_i^a(t)$  as

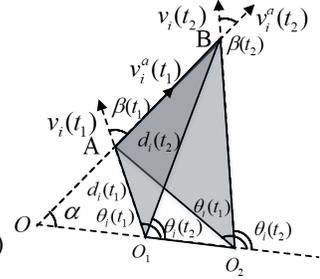


Figure 11. Geometric relation illustration.

$$v_i^a(t) = v_i(t_2) \left( \cos \left( \arctan \left( \frac{\cos(\theta(t_1) - \theta(t_2)) - \frac{v_i(t_1)}{v_i(t_2)}}{\sin(\theta(t_1) - \theta(t_2))} \right) \right) \right)^{-1}. \quad (6)$$

**V-P distance:** The v-p distance at  $t$ , i.e.,  $d_i(t)$ , is calculated based on the knowledge of  $v_i^a(t)$ ,  $\theta_i(t)$ , and  $\tau_i(t)$  which are all known values by now. Consider two triangles  $\Delta AO_1B$  and  $\Delta AO_2B$  in Figure 11. Since v-p distance is significantly larger than the inter-mic distance, vehicle's DoAs with respect to the two mics are deemed the same, denoted by  $\theta_i(t)$ . Thus,  $\angle AO_1B = \theta_i(t_1) - \theta_i(t_2)$  and  $\angle AO_2B = \theta_i(t_1) - \theta_i(t_2)$ . Due to the law of cosines in trigonometry, we have the following relation

$$\begin{cases} \cos(\theta_i(t_1) - \theta_i(t_2)) = \frac{d_i^2(t_1) + d_i^2(t_2) - (v_i^a \Delta t)^2}{2d_i(t_1)d_i(t_2)} \\ \cos(\theta_i(t_1) - \theta_i(t_2)) = \frac{(d_i(t_1) + v_a \tau_i(t_1))^2 + (d_i(t_2) + v_a \tau_i(t_2))^2 - (v_i^a \Delta t)^2}{2(d_i(t_1) + v_a \tau_i(t_1))(d_i(t_2) + v_a \tau_i(t_2))} \end{cases} \quad (7)$$

which is a system of two quadratic equations of two variables,  $d_i(t_1)$  and  $d_i(t_2)$ . Thus, it is not difficult to solve it by some existing libraries. In the implementation, we use the GSL [33] that provides a library to compute the root of polynomials.

### 4.4 Piecing All Components Together

The design rationale of Acoussist is to estimate whether nearby drivers have sufficient time to spot the blind pedestrian and stop their vehicles when the pedestrian tends to enter the crosswalk.

As discussed, it is equivalent to have driver's SSD larger than the v-p distance  $d_i$ . In practice, we should further take into account the processing latency of the system, denoted by  $t_{dl}$ . We thus adopt the following conservative pedestrian safety condition

$$d_i > SSD_i + v_i^a \times t_{dl}. \quad (8)$$

Acoussist generates an alarm as long as any detectable vehicle  $s_i$  violates the above condition. In our implementation,  $t_{dl}$  is instantiated with a device-dependent value that is associated with 90% confidence level. To obtain this value, app runs on the smartphone dozens of times prior the usage. More details will be discussed in Section 5.2.

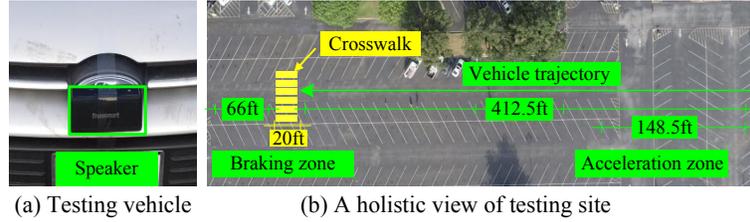


Figure 12. In-field testing setup.

Since  $d_i$  and  $v_i^a$  are all known, the remaining task is to find out a particular vehicle  $i$ 's SSD. As recommended by design standard of American Association of State Highway and Transportation Officials (AASHTO) [1], SSD is estimated by

$$SSD = 1.47 \times v_i^a t_{pr} + 1.075(v_i^a)^2 / a \quad (9)$$

where  $v_i^a$  is the instant vehicle velocity that is derived in Section 4.3.  $t_{pr}$  and  $a$  stand for the driver's perception-reaction time and acceleration rate, respectively. AASHTO allows 2.5 seconds for  $t_{pr}$  and  $11.2 \text{ ft/s}^2$  for  $a$  to accommodate approximately 90% of all drivers when confronted with simple to moderately complex road situations. SSD is the sum of two distances: 1) brake reaction distance (i.e., the distance traversed by the vehicle from the instant the driver sights an object necessitating a stop to the instant the brakes are applied); and 2) braking distance (i.e., the distance needed to stop the vehicle from the instant brake application begins). According to AASHTO, in the above expression for SSD, conservative parameters are used, including a generous amount of time given for the perception-reaction process, and a fairly low rate of deceleration, such that it allows a below-average driver to stop in time to avoid a collision in most cases.

It is noteworthy that our system does not impose any requirement on the position/orientation of the phone during usage. Even though the calculation of DoA  $\theta_i$  is dependent on the phone orientation, the pedestrian safety condition is related to  $v_i^a$  and  $d_i$  (8). As shown in (6) and (7),  $v_i^a$  and  $d_i$  are relevant to  $\theta_i(t_1) - \theta_i(t_2)$  which is independent of the phone orientation.

## 5 IMPLEMENTATION AND IN-FIELD TESTING

### 5.1 Implementation Setup

**Implementation:** As a proof-of-concept implementation, we develop the prototype of Acoussist on four Tron-smart portable speakers, around \$40 each, and three Android smartphones, Google Pixel XL, Galaxy S8 and Nexus 2. Four vehicles, Ford Focus 2014, Ford escape 2019, Toyota corolla 2017 and Honda Accord 2016, are used in the testing. A speaker is mounted in front of the vehicle, shown in Figure 12(a), playing the pre-loaded chirps that sweep from 17 KHz to 19 KHz at 69.3 dB<sup>4</sup>. We use the two built-in microphones at the smartphones to receive signals. An Android app is developed to process received signals and generate alarm when needed. We use NDK [35] to implement the STFT operations, and GNU Scientific Library (GSL) for other mathematical operations in our design.

**In-field testing setup:** All testings are conducted at the campus parking lot as shown in Figure 12(b) during weekends when the space is relatively empty. A pedestrian stands at the end of the crosswalk and records the performance. In each testing round, a driver accelerates the vehicle to a target speed. Meanwhile, the pedestrian activates the app to sense the environment. If no alarm is generated, the pedestrian waves a flag, indicating the action of street crossing. Otherwise, she keeps the flag down, indicating waiting at the curb. Upon noticing a waving flag, the driver takes reaction and stops the car. The reason we use flag signals instead of having a

<sup>4</sup>This value is measured at the speaker directly, which equals to  $8.5 \times 10^{-7} \text{ mW/cm}^2$  [115]. As a reference, the FDA regulation over preamendments diagnostic ultrasound equipment is  $\leq 94 \text{ mW/cm}^2$  [29]. It restricts the maximal intense level of ultrasound exposed when people take medical evaluations such as peripheral vessel detection, cardiac diagnostic, and fetal imaging.

pedestrian physically proceed to the crosswalk is for safety consideration. Besides, as the driver needs be signaled with the pedestrian's action of street crossing by waving a flag, the pedestrian cannot be simulated by a stand mounted with a smartphone. A test is viewed success, if a) the flag is not waved, since a potential collision is detected, or b) the flag is waved while the vehicle stops completely before reaching the crosswalk.

Acoussist requires users to hold their smartphones steady for about 1 second to have an accurate detection of oncoming vehicles. It is also the time duration between the time point that a user activates the app and the time point that he/she decides to wave the flag or not. Beyond the 1 second time limit, the user can choose to wave the flag at any time, as long as no alert is observed. The 1 second is attributed from two aspects, the duration of two consecutive measures to derive the v-p distance ( $\delta_t = 500$  ms) and the processing delay (the 90-percentile value of  $t_{dl} = 220.7$  ms as shown in Section 5.2).

**Evaluation metrics:** The performance of our system is evaluated via the following metrics: ranging distance, warning distance, miss detection ratio (MDR), and false alarm ratio (FAR). Particularly, ranging distance is the v-p distance at which the app is able to measure this value for the first time. It implies the largest detectable range of our system. Warning distance is the v-p distance at which the pedestrian safety condition (8) is violated for the first time. MDR is the probability that a vehicle which has violated (8) but not detected. FAR is the probability that a vehicle satisfies (8) but wrongly reported. Traffic cones are placed along the vehicle trajectory. To measure the ranging distance, the pedestrian records the parking slot number, where the pedestrian stands denotes the first slot. With the assistance of traffic cones, the v-p distance becomes measurable. Then, the ranging distance is approximated by multiplying the slot number with the width of each slot, 10.7 ft in our case. Warning distance is obtained similarly.

## 5.2 Micro Benchmark

**Impact of parameter settings:** We first examine the impact of two most crucial parameters of our system,  $\Delta t$  and  $\delta_t$ . Recall that  $\Delta t$  is the STFT window size and  $\delta_t$  is the time interval between two consecutive measures. Figure 13(a) shows the accuracy performance of Acoussist with various  $\Delta t$ . The best performance 93.6% exists when  $\Delta t = 2048$ . As discussed in Section 4.1, the value of  $\Delta t$  strikes a trade-off balance between frequency resolution  $\Delta f$  and time resolution  $\Delta t$  of STFT. Figure 13(b) shows the accuracy performance with various  $\delta_t$ . The best performance is achieved for  $\delta_t = 500$  ms. On one hand, a large  $\delta_t$  and thus an apparent difference between  $\theta_i(t_1)$  and  $\theta_i(t_2)$  is beneficial for deriving an accurate  $v_i^a$ . On the other hand, it inevitably leads to a long processing delay which impacts the detection accuracy.  $\Delta t$  and  $\delta_t$  are set to 2048 and 500 ms, respectively, in the rest experiments.

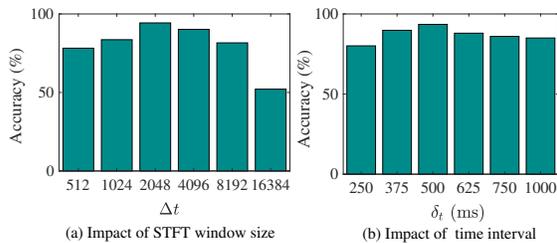


Figure 13. Impact of parameters.

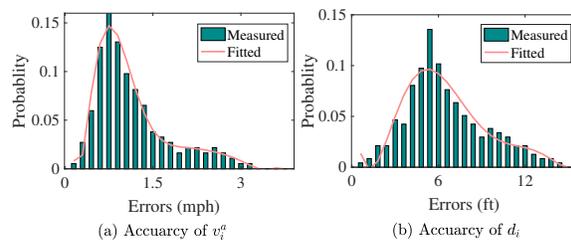


Figure 14. Accuracy of measurements.

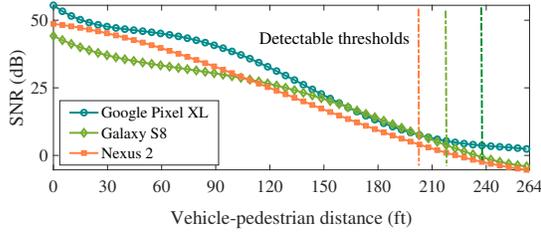


Figure 15. The received signal SNR with different v-p distance.

Table 1. Detection performance of different devices.

Device	Pixel XL	Galaxy S8	Nexus 2
MDR	3.4%	4.1%	3.6%
FAR	3.2%	3.7%	4.8%

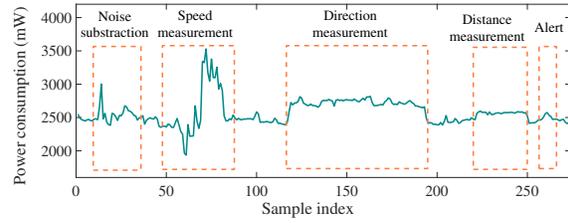
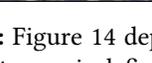
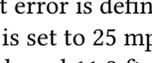


Figure 16. Instant power readings.

Table 2. Impact of phone orientation.

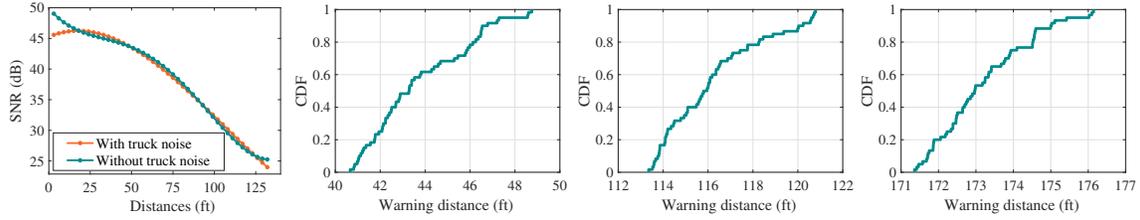
Orientation	Ranging dist.	Warning dist.
	P1 189.8 ± 10.1 ft	157.6 ± 10.4 ft
	P2 181.1 ± 13.2 ft	155.3 ± 15.3 ft
	P3 186.9 ± 11.8 ft	150.7 ± 10.5 ft
	P4 178.1 ± 14.4 ft	159.2 ± 10.5 ft

**Measurement performance of motion parameters:** Figure 14 depicts the distribution of measurement errors of velocity  $v_i^a$  and v-p distance  $d_i$ . The measurement error is defined as the difference between measures and the ground truth. Here, the ground truth of  $v_i^a$  and  $d_i$  is set to 25 mph and 120 ft, respectively. We observe that 90% of errors for these parameters are within 2.4 mph and 11.8 ft, respectively, which are acceptable for implementation.

**Ranging distance and warning distance:** Figure 19(a) evaluates the ranging distance of our system with respect to the vehicle speed. The ranging distance decreases from 208.1 ft to 165.5 ft on average when the speed changes from 5 mph to 45 mph. This is because the vehicle travels a longer distance at a higher speed given  $\delta_t$  and thus perceives a shorter v-p distance when this value is first obtained. As shown in Figure 19(b), the warning distance increases almost linearly as the speed grows from 5 mph to 30 mph. It reaches 175.2 ft when  $v_i^a = 30$  mph. The result meets our expectation; when a vehicle moves faster, the driver needs longer distance to react and stops the vehicle which corresponds to a larger warning distance. However, as the speed continues to increase, the warning distance experiences slight decrease. This is because the warning distance is capped by the ranging distance. As the latter decreases, it also brings down the former.

**Impact of different devices:** Figure 15 shows the smartphone's received SNRs of chirps with frequency 17 KHz-19 KHz at different v-p distances. Three devices exhibit different sensitivity responding to high-frequency signals. Particularly, the detectable threshold, defined as the maximum distance within which the received signal are perceivable from the background noise, is 237.6 ft, 221.1 ft and 207.9 ft for Google Pixel, Galaxy S8, and Nexus 2, respectively. They bring about various detection performances as shown in Table 1. Combining the results of Figure 15 and Table 1, we observe a positive correlation between a device's detectable threshold and its detection accuracy, and Pixel XL has the best performance among the three.

**Impact of phone orientation:** We also examine if the phone orientation in usage impacts the detection performance. We test four different positions, the combinations of the screen facing above/aside and the head pointing up/down, as shown in Table 2. We find that the ranging distance and the warning distance are almost the same for all four positions. Thus, the performance of Acoussist is independent of how the pedestrian holds the phone. It meets our discussion in Section 4.4.



(a) Received signal SNR. (b) Vehicle speed  $v_i^a = 10$  mph. (c) Vehicle speed  $v_i^a = 20$  mph. (d) Vehicle speed  $v_i^a = 30$  mph.

Figure 17. Impact of the background noise.

**Impact of background noise:** Among commonly observed background noise in streets, truck sound is typically the most powerful one. We thus evaluate its impact to the performance of Acoussist. First of all, as shown in Figure 6 (e), the signal frequency components are mainly concentrated on the lower-end of the frequency. Particularly, 88.7% of them reside lower than 10 KHz. Recall that the acoustic chirp signal used by Acoussist ranges between 17 KHz and 19 KHz. Thus, there is a clear gap between the truck sound and the acoustic chirp signal. In our design, a high-pass filter, with a cutting frequency of 10 KHz, is then applied to get rid of most background noise, including traffic noise, music noise, speech noise, construction noise, as well as truck sound.

On the other hand, we notice that truck noise does have frequency components above 10 KHz. To examine its impact to f-t analysis of our system, we first add truck sound as background noise to the chirp signals recorded at different distances by using Matlab audio toolbox [72]. Then SNR is measured after passing the received signal through all noise removal modules. The relation of SNR versus the v-p distance is plotted in Figure 17(a). As a comparison, we also show the SNR without truck noise. We observe that the two curves are quite similar to each other, except when the pedestrian is very close to the noise source, i.e., within 25 ft. We further depict in Figure 17(b)-(d) the CDF of warning distance under different vehicle speeds, from 10 mph to 30 mph. Recall that warning distance is the v-p distance at which the alert is triggered. Take Figure 17(b) as an illustration. When the vehicle speed is at 10 mph, all alerts are generated when the v-p distance is between 40.6 ft and 48.7 ft. Thus, vehicles are all detected even at distances much longer than 25 ft away from the pedestrian. Combining the observations above, we can infer that a truck will trigger the alert even it is larger than 25 ft away from the pedestrian. Besides, its detection performance should be similar to regular vehicles. If a truck is within 25 ft from the pedestrian, its presence can be easily picked up by human ears. In this scenario, the visually impaired pedestrians can simply rely on their hearings to detect the potential hazard.

**Energy consumption:** A dedicated hardware, Monsoon power monitor [73], is applied to measure the energy consumption of mobile phones for running our app. During the measurement, we keep other components, e.g., WiFi and Bluetooth, offline. Figure 16 shows the instant power reading via the power monitor when executing one detection. We clearly specify the part dedicated to each module. We can tell that the measurement of relative velocity and DoA consumes a larger amount of power among all modules, which is about 2967.2 mW and 2656.8 mW on average, separately. As a note, the average power consumption of some common smartphone tasks, such as video call, map service, and web browsing take 3351.6 mW, 2642.8 mW, and 1732.7 mW, respectively. Besides, our app is only activated when a pedestrian tends to cross streets and thus offline most of the time. Thus, the power consumption of our app is practically acceptable.

**Processing latency:** Figure 18(a) gives the stacked computation time of each system module. The module for DoA measurement incurs the largest delay, which is about 137.2 ms on average. This is because it involves an exhaustive search for the solution of the MGCC function. Figure 18(b) further illustrates the cumulative distribution function (CDF) of the total processing latency of the app. The average value is 186.3ms, with 90% of

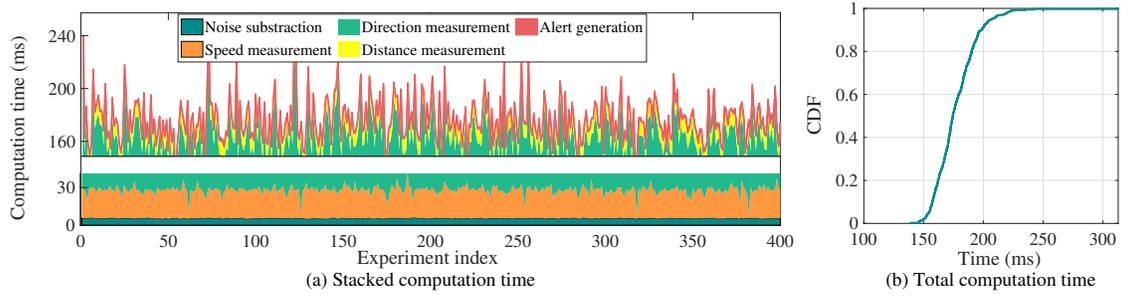


Figure 18. Processing latency.

Table 3. Detection performance when a vehicle is at different speeds.

Speed (mph)	5	10	15	20	25	30	35	40	45
MDR	11.3%	9.9%	6.4%	5.4%	4.7%	3.4%	5.8%	6.3%	8.2%
FAR	10.0%	9.5%	7.2%	6.5%	5.3%	3.2%	2.8%	2.5%	2.1%

measurements lower than 220.7 ms. We thus instantiate  $t_{dl}$  with 220.7 ms for the implementation (8). We also believe that by leveraging the parallelization, we can further bring down the processing latency.

### 5.3 System Benchmark

**Impact of vehicle speeds:** Table 3 gives the detection accuracy of Acoussist toward a vehicle in a wider range of speeds. Interestingly, both MDR and FAR experience significant decrease when the speed increases from 5 mph to 45 mph. This is because a higher speed generates a larger slope of the t-f sweep line as revealed in (3). As a result, the difference between  $f(t + \Delta t)$  and  $f(t)$  will be more apparent to tolerate errors caused by insufficient frequency resolution  $\Delta f$ . Therefore, when a target vehicle moves in a higher speed, its t-f profile tends to more accurate which leads to a better detection accuracy. MDR grows as the speed continues to increase from 30 mph to 45 mph. This is because the ranging distance becomes close or even shorter than the the warning distance when a vehicle is at a high speed. As a result, some hazard situations are missed in the detection.

Table 4. Detection performance with the presence of multiple vehicles.

Speed (mph)	5	10	15	20	25	30	35	40	45
MDR	8.5%	6.8%	4.3%	3.7%	2.6%	1.9%	4.0%	5.5%	6.2%
FAR	12.0%	11.6%	10.1%	8.4%	7.3%	6.7%	5.0%	4.5%	3.2%

**Impact of multiple vehicles:** We examine the detection performance of Acoussist with the presence of four vehicles in Table 4. Compared with Table 3, we notice that FAR slightly increases with the presence of more cars. This is because a false alarm is generated by the system when any of the four vehicles is falsely reported to incur a potential collision. In contrast, MDR becomes smaller when there are more vehicles. This is because a collision is correctly forecast, when any one of the vehicles triggers the alarm. While FAR experiences a slight increase, it does not impact the performance of Acoussist much. A visually impaired pedestrian uses Acoussist to double-confirm the situation when sensing a clear street with hearing. Thus, MDR is more crucial than FAR in practical usage.

Figure 20 shows the ranging distance and warning distance when vehicles move in the same/opposite direction(s). While the average measures are closely the same, the variance associated with opposite directions is

smaller than the same direction. This is because t-f sweep lines of the four vehicles are better separated and easier to extract in the former case.

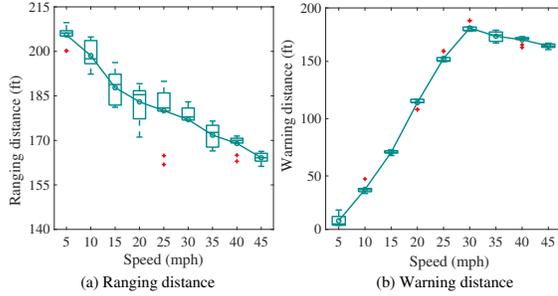


Figure 19. Impact of vehicle speeds.

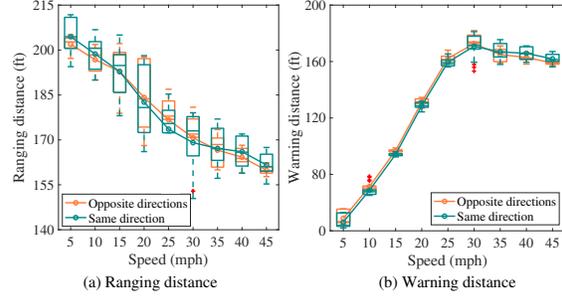
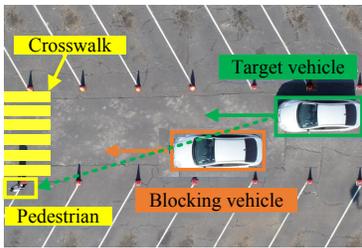


Figure 20. Impact of multiple vehicles.

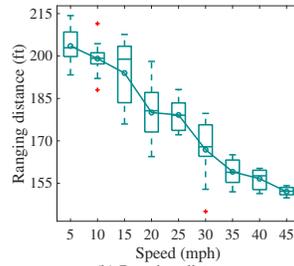
**Impact of vehicles in asynchronous speeds:** In the experiment, two vehicles move toward the same direction at asynchronous speeds of  $v_1$  and  $v_2$ , separately. We find in Table 5 that the ranging distance is mainly determined by  $\min(v_1, v_2)$ , while the warning distance is determined by  $\max(v_1, v_2)$ . For example, the warning distance under the setting ( $v_1 = 10\text{mph}, v_2 = 25\text{mph}$ ) equals to  $161.45 \pm 15.21\text{ft}$ , which is similar to the one measured under ( $v_1 = 25\text{mph}, v_2 = 15\text{mph}$ ), i.e.,  $158.82 \pm 13.36\text{ft}$ . This is because the high-speed vehicle is more easily to break the safety condition. In terms of ranging distance, the low-speed vehicle covers a shorter distance within Acoussist’s detection/processing delay. Since the ranging distance is the v-p distance when vehicles are first detected, it is determined by  $\min(v_1, v_2)$ .

Table 5. Detection performance when two vehicles are in asynchronous speeds.

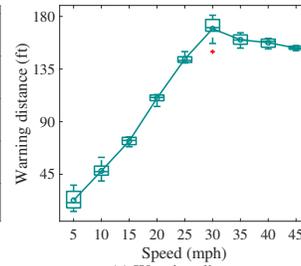
Speeds ( $v_1, v_2$ )	(5 mph, 30 mph)	(10 mph, 25 mph)	(15 mph, 20 mph)	(25 mph, 15 mph)	(30 mph, 10 mph)
Ranging dist.	$210.24 \pm 15.36\text{ft}$	$198.63 \pm 9.81\text{ft}$	$185.72 \pm 10.26\text{ft}$	$187.53 \pm 13.26\text{ft}$	$201.05 \pm 14.12\text{ft}$
Warning dist.	$168.45 \pm 11.21\text{ft}$	$161.45 \pm 15.21\text{ft}$	$136.61 \pm 14.25\text{ft}$	$158.82 \pm 13.36\text{ft}$	$168.63 \pm 10.26\text{ft}$



(a) Testing scenario



(b) Ranging distance



(c) Warning distance

Figure 21. Impact of nearby objects.

**Impact of nearby objects:** To evaluate the impact of nearby objects, we place a second vehicle to partially block the line of sight between the target vehicle and the pedestrian (as shown in Figure 21(a)). The ranging distance and the warning distance toward the target vehicle is shown in Figure 21(b) and 21(c), respectively. The trend of these two distances with respect to the vehicle speed is very similar to that in Figure 19(a) and 19(b),

the performance without any blocking object. However, the two distances exhibit a larger variance with the existence of a blocking object. This is because the blocking object absorbs a portion of energy of chirps in the LoS path and thus slightly impacts the detection performance.

**Impact of time/weather of usage:** Table 6 shows the detection accuracy of Acoussist at different time of a day. The performance is relatively stable. Acoussist performs well at evenings when there are typically lack of visible light. Note that some existing systems for pedestrian safety, such as WalkSafe [117], rely on back camera of mobile phones to detect hazard vehicles. Thus, their performance is largely impacted by the time of usage. Table 7 further compares the detection performance under different weather conditions. We notice that both MDR and FAR experience slight increase in rainy days due to higher loss of acoustic signals when propagating in saturated air.

Table 6. Detection performance at different time of a day.

Usage time	Morning	Noon	Afternoon	Evening
<b>MDR</b>	3.8%	4.5%	4.7%	5.2%
<b>FAR</b>	6.2%	4.9%	5.3%	4.8%

Table 7. Detection performance under different weather conditions.

Usage weather	Sunny	Windy	Cloudy	Rainy
<b>MDR</b>	3.7%	3.5%	4.2%	4.4%
<b>FAR</b>	4.8%	5.3%	4.8%	5.2%

## 6 USER STUDY

**Procedures:** To evaluate the effectiveness and usability of Acoussist from the perspective of visually impaired people, we conduct user study with this demographic group. For recruitment, we advertised our research study through a local mailing list of people with visual impairments. Six volunteers are invited, either totally blind or can barely see to test the system. Their demographics are provided in Table 8. At the end of experiments, they are asked to fill the questionnaire which serves as the basis of our user study.

Table 8. Demographic information of vision impaired participants.

ID	P1	P2	P3	P4	P5	P6
Gender	F	F	M	F	M	F
Age	65	50	46	25	32	58
Eyesight	Blind	Blind	Blind	Barely see	Barely see	Blind

Before the experiment, a short training session (15 - 30 minutes) is organized. An overview of the study and the usage instructions of Acoussist are presented. Volunteers are left with sufficient time to get familiar with the Acoussist app. The field testing is conducted at the campus parking lot as shown in Figure 12(b) during weekends when the space is relatively empty. The pedestrian (i.e., volunteer) stands at the end of the crosswalk and records the performance. In each testing round, a driver accelerates the vehicle to a target speed. Meanwhile, the pedestrian activates the app to sense the environment. If no alert is generated, the pedestrian waves a flag, indicating the action of street crossing. Otherwise, she keeps the flag down, indicating waiting at the curb. Upon noticing a waving flag, the driver takes reaction and stops the car. A test is viewed success, if a) the flag is not

waved, since a potential collision is detected, or b) the flag is waved while the vehicle stops completely before reaching the crosswalk. As the pedestrian cannot observe the test result by him/herself, the result is informed by a second volunteer, i.e., a student researcher in this project, with normal sight vision who stands right next to the visually impaired pedestrian. Besides, the second volunteer watches out for the visually impaired pedestrian's safety in case of any unforeseen situation. To examine which alert modality is better perceived by the visually impaired pedestrians, we have the smartphone to either emit beeping sound or vibration once a potential hazard is detected.

**Experiment results:** Table 9 shows the detection accuracy of Acoussist when used by visually impaired pedestrians. First of all, both MDR and FAR experience significant decrease when the speed increases from 5 mph to 30 mph. This trend coincides with the result of Table 3, derived from experiments involving normal users. Besides, we are not able to identify noticeable differences on the detection accuracy between the two groups of people. The main reason is that Acoussist involves rather limited user operation, simply turning on the app, during the entire process. Thus, its performance is irrelevant to who the user is.

Table 9. Acoussist detection performances when used by visually impaired pedestrians.

Speed (mph)	5	10	15	20	25	30
MDR	11.5%	9.8%	6.3%	5.1%	4.5%	3.6%
FAR	10.5%	9.2%	7.5%	6.2%	5.8%	3.5%

Table 10. Survey results (1: strongly disagree, 3: neutral, 5: strongly agree).

No.	Questions	P1	P2	P3	P4	P5	P6	Mean	Variance
Q1	Traffic sound is insufficient to make all the judgements needed to cross streets safely.	4	5	3	5	4	4	4.2	0.6
Q2	Acoussist provides additional information to make judgements needed to cross streets safely.	5	4	3	3	4	4	3.8	0.6
Q3	Acoussist helps to make better decisions to cross streets safely.	4	4	3	4	3	5	3.8	0.6
Q4	False alarm produced by Acoussist discourages me to adopt it.	2	2	2	1	3	2	2	0.4
Q5	I prefer vibration than beeping sound as alerts.	4	5	4	4	3	4	4	0.4
Q6	I prefer to receive more information of surrounding traffic via Acoussist to cross streets safely.	3	4	2	3	2	3	2.8	0.6
Q7	Acoussist is easy to use.	5	4	5	3	5	4	4.3	0.7
Q8	I am willing to adopt Acoussist during daily commutes.	3	4	5	3	5	4	4	0.8

**Survey results:** After the experiment, questionnaires are distributed among volunteers. They are asked to rate Acoussist from the perspectives of effectiveness and usability. Questions are list in Table 10. From 5-point Likert scale (1 = strongly disagree, 3 = neutral, 5 = strongly agree), volunteers pick a point that they deem proper. Survey results are shown in Table 10.

Almost everyone agrees that the sound of traffic is no longer sufficient to make all the determinations and judgments needed to cross streets safely (Q1). For example, P2 and P6 mentioned that these situations take place when there is loud background noise, such as sound of construction and rain. They also mentioned that some "quiet vehicles", such as electrical/hybrid vehicles, cannot be heard until they are too close. In the response to Q2, 4 out of 6 agree that Acoussist provides additional information to cross streets safely. P2 and P5 specifically mentioned that they feel more informed about the surrounding traffic conditions and less panic when making a decision. The survey result of Q3 shows that Acoussist helps the visually impaired to make better decisions to

cross streets safely. P4 and P6 said that choosing the right timing to cross a street can be hard, as hearing an approaching vehicle is one thing but estimating its distance and speed is another. Acoussist is able to effectively alleviate this detection load. When volunteers are asked by Q4 of if false alarm produced by Acoussist discourages them from using it, 5 out of 6 said no while 1 held a neutral opinion. Some of them expressed that they are unlikely to fully rely on Acoussist to decide when to cross streets. Instead, it can be treated as an assistive tool for their judgement, especially that Acoussist is merely a prototype rather than a commercial product. Q5-Q8 are designed to evaluate the usability of Acoussist. In response to Q5, most of the volunteers prefer vibration over beeping sound as the alert modality. This is because beeping sound can cause interference to their own hearings. In the response of Q6, there is no clear preference regarding whether to get more traffic information, such as how many detected vehicles nearby and their distances, out of Acoussist in addition to the current collision avoidance alert. Some of them said more information would distract them from sensing the traffic condition, while others feel additional information help to make more accurate and timely decisions. In Q7, almost everyone agrees that Acoussist is easy to use. In Q8, 4 out of 6 expressed their willingness to adopt Acoussist for daily commute.

**Open questions:** After the in-field experiments, the visually impaired participants are encouraged to share their usage experience. They are also provided with the opportunity to discuss with each other and exchange their thoughts. They are asked by the following two questions.

“**Q1:** *What is the possible role of Acoussist in your daily commute?*”

“**Q2:** *What are possible aspects to improve Acoussist?*”

In response to Q1, three participants mentioned that Acoussist should be used as a supplemental method for estimating the surrounding traffic condition in addition to their hearings. If they intend to cross an uncontrolled street, they will use Acoussist to double-confirm their judgement. P3 suggested that Acoussist should be used in a conservative way. If he hears any oncoming vehicle, he would wait to hear that the vehicle begins to slow down and eventually comes to full stop before crossing the street, even if Acoussist produces a safety sign at the beginning. After discussion with P3, we find that such situation happens when the pedestrian has a keen sense of hearing to detect a vehicle far away. The vehicle does not trigger the alert in the first a few seconds because the driver is deemed to have sufficient time to spot the pedestrian and stop the car, given its instant distance and moving speed. All volunteers agreed that Acoussist should be positioned as an assistive tool that supplements the existing assistive methods, including accessible pedestrian signals and dog guidance. Due to the nature of complexity of street traffics, such as distracted drivers or other unforeseen accidents, it is desirable to provide with them comprehensive information of surrounding traffics that assist them to make accurate judgement.

In response to Q2, all six participants viewed Acoussist as a helpful tool that enhances their mobility experiences outdoors. They also provided valuable comments on its potential improvements. Based on our thematic grouping of their comments, the following themes emerged. **Notify the driver:** P3, P4, and P6 agreed that the drivers should be notified if their vehicles present potential hazard to pedestrians nearby. The alert may allow drivers to react quickly, for example, slowing down the car. **Alert modalities:** P2 and P5 suggested to provide options of different alert modalities, e.g., vibration, beeping sound, and human voice. Besides, Acoussist only provides binary detection results. P1 and P6 suggested to include more diverse information, such as the number of oncoming vehicles and their v-p distances. These information assists with better traffic condition estimation. **Remove the speaker:** Some participants mentioned that the usability of the system would be improved if the installment of external speakers is no longer needed. For example, P5 noted “*people may be lazy or reluctant to install the speakers.*” In future work, we plan to explore the Type-I ranging and utilize reflected signals to perform nearby vehicle detection.

## 7 RELATED WORK

### 7.1 Pedestrian Safety

There is a growing interest in using V2X [31, 105] for pedestrian safety. The idea is to exchange positioning information between vehicles and pedestrians for collision avoidance. In a pedestrian safety project initiated by Honda [121], they use the dedicated short-range communications (DSRC) with the basis of IEEE 802.11p as the module for V2X communications. However, so far, no existing commercial smartphone is installed with IEEE 802.11p module. There neither exists any clear road map regarding its implementation. In parallel with the DSRC-based approach, WiFi is an alternative channel for information exchange between vehicles and pedestrians [20, 47, 65]. However, the WiFi association involved therein introduces extra processing delay and load to the vehicle, which prevents it from large-scale deployment. There are several mobile apps, such as WalkSafe [117], which use the back camera of the mobile phone to detect potentially fatal collisions. However, its performance is subject to lightening conditions.

### 7.2 Collision Avoidance Systems

A collision avoidance system (CAS) is an automobile safety system designed to prevent or reduce the severity of a collision. Since the demonstration of the first modern version in 1995, it has attracted massive attention in the past decades. The most common approaches use radar [68, 86], laser [22, 77, 78], LiDAR [4, 28, 92] and sonar [56, 57] to detect an imminent crash. Nonetheless, CAS is primarily designed for driver safety. It is unclear whether it can be directly applied to our problem. Besides, some of these solutions, such as laser and LiDAR, require the equipment of expensive specialized sensors, which are impractical to install in smartphones. For radar and sonar, as they belong to *type-I/II ranging* when employed for CAS, their techniques are different from ours.

### 7.3 Parameter Estimation of Moving Targets

The estimation of target movement parameters, including speed and DoA, exploits the acoustic energy radiated by a target for its detection and tracking.

**Speed estimation:** Doppler models have been widely used in previous literature that characterize the relation between a target's instant moving speed and the frequency shift incurred at the observer [26, 34, 49, 50, 89]. These models work well for single-target scenarios or multiple targets that emit heterogeneous signals. Specifically, to perform multi-source speed estimation [26, 34, 49] adopt signature signal modeling. Central to their success is to accurately identify each target, for example, by analyzing engine humming sounds from vessels or tire noises made by vehicles. With targets' pre-measured signature signals, source separation models like maximal likelihood approach [12, 69] and nonlinear least squares method [66] are applied to estimate each target's speed. However, to capture a target's sound, these designs rely on ground-mounted sensors; the estimation is conducted when the target passes through them closely, say 3-7 ft, which is too short for vehicle collision avoidance. Besides, the requirement of pre-measurement for each vehicle's signal signature is impractical. More importantly, since vehicles emit homogeneous acoustic chirps in our design, the above-mentioned models are inapplicable.

**DoA estimation:** Another research that is relevant to this work is DoA estimation, a task of identifying the relative position of sound sources with respect to the microphone. It forms an integral part of speech enhancement [106], multichannel sound source separation [46] and spatial audio coding [79]. Popular approaches to DoA estimation are based on time-delay-of-arrival (TDoA) [17, 25, 52], the steered response power (SRP) [18, 45, 109], the generic cross correlation with phase transform (GCC-PHAT) [15, 54, 58], or on subspace methods such as multiple signal classification [55, 96]. While these approaches can estimate DoAs from all acoustic sources, they are mainly for single-source tracking. Clearly, they are inapplicable to Acoussist. To conduct multi-source tracking, the key ingredient is to address the data association problem. For this, probabilistic models like Gaussian mixture model [59, 128] and Laplacian mixture model [16] are employed over the signals' t-f profile to compute

the DoA and map them to each source by a histogram or clustering. These methods require the inter-mic distance smaller than half wavelength of acoustic signals to avoid the spatial aliasing effect. It is practical in generic scenarios where mic arrays can be arbitrarily arranged. In our system, two mics are fixed in smartphones. Besides, the longest wavelength is about  $2\text{cm}$  ( $\frac{340\text{m/s}}{17\text{KHz}}$ ). Its half, i.e.,  $1\text{cm}$ , is apparently smaller than the inter-mic distance for many smartphones, e.g., the ones with one mic located on bottom and the other on top.

#### 7.4 Assistive Navigation Technologies

To improve blind people's navigation experience, researchers have augmented normal white canes with sensors to acquire conditions of surrounding environments, especially the obstacles that are out of the reach of a normal cane. By detecting and analyzing the information about obstacles, such as distance [40, 70, 98, 126] and shape [88, 100, 130], the smart cane decides if the environment is hazardous to users; if so, it provides audio or vibration feedback. More obstacle avoidance and navigation systems are built with various sensing modalities, such as cameras [10, 85, 107, 108], ultrasonic [32, 83, 104, 112], RF sensing [3, 5, 39, 94], depth sensing [48, 61, 90, 99], or fusion of several of them [9, 53, 62]. However, the ranging distance of these systems are mostly bounded within 30 feet. Besides, while they are able to detect static or slow moving obstacles, e.g., walls, tables, trees, and other pedestrians nearby, the detection and avoidance of collisions with moving vehicles in a relatively high speed, is still unexplored in the literature.

There are also other related works aiming to improve travel safety for the visually impaired. For example, [43] collects more rich information of the environmental conditions to assist the blind pedestrians in making appropriate movement decisions via crowdsourcing. Some others customize the traffic lights [7, 74] or crosswalk infrastructures [44] to provide street-crossing guidance. Another line of research [38, 97, 120, 129] builds simulated outdoor environments to help blind pedestrians gain prior navigation experiences in new locations. [82] models the evolution of user expertise throughout repetitions of a navigation task with a smartphone-based turn-by-turn navigation guidance interface. [42] examines the information needs of visually impaired pedestrians at intersections, which may present a specific cause of stress when navigating in unfamiliar locations. However, none of them is about assisting blind pedestrians to cross uncontrolled streets. It is worth mentioning a prior work [95] that shares the author's experience of teaching the blind how to assess when it is safe to cross uncontrolled streets. For the first time, we try to tackle this challenge via novel acoustic sensing techniques.

## 8 DISCUSSIONS

**Impact of ultrasonic signals to animals:** Acoussist operates over the frequency between 17 KHz and 19 KHz, which falls into the hearing range of some animals, such as dogs and bats. While ultrasonic sound with low or medium transmission power might impact animals to some extent, research shows that only extremely loud noise ( $\geq 85$  dB) is harmful to animals [80, 111]. In our system, commercial portable speakers are employed running at 69.3 dB. While the emitted ultrasonic chirps are not harmful, they are perceptible by some animals when they are nearby. To roughly estimate the perceptible range, we measure the chirp sound with respect to the speaker-receiver distance. The result shows that the chirp sound drops to 30 dB, when the distance is about 4.6 meters. As a reference, whisper is typically at the level of 30 dB, while human conversations are at 60 dB [30]. Thus, we claim that Acoussist would cause rather limited interference to animal's hearings even with the existence of multiple opt-in vehicles nearby, as close as 4.6 meters. Besides, ultrasonic sensing has been widely used in many application scenarios, such as collision avoidance detection on autonomous vehicles and underwater navigation. More recently, acoustic sensing has found its novel applications in biometric sensing [87], acoustic imaging [71], and indoor localization [27]. Our work can be treated as another application of acoustic sensing. Lastly, the authors do plan to work with researchers from the Biology Department to carry out formal study regarding the potential impact of our system to commonly found mammals, such as dogs and cats.

**Usage requirements:** Acoussist requires users to hold their smartphones steady for about 1 second to have an accurate detection of oncoming vehicles. This is because two measures of the vehicle's instant velocities at two time instances  $t_1$  and  $t_2$  are needed to derive the v-p distance. A too short  $|t_2 - t_1|$  will result in low accuracy of measuring the v-p distance, while a too large value will cause long waiting duration and postpone the collision detection. To strive a balance between these two, a proper value of  $|t_2 - t_1|$  is critical. During the testing, the best overall performance exists when  $|t_2 - t_1|$  is 500ms. Besides, we have also discussed in Section 5.2 that the 90 percentile signal processing delay of our system is 220.7 ms. By jointly considering these two factors, users need to hold the smartphone for about 1 second to receive a detection result. While we believe this requirement is practical to execute for most of human beings, it would be more convenient without such a restriction. We plan to further look into it in our future study. It is noteworthy that our system does not impose any requirement on the position/orientation of the phone during usage.

**Applicable scenarios:** In the experiment, we find that vehicle's speed has a direct impact on the detection performance of Acoussist. Specifically, as the speed increases from 5 mph to 45 mph, MDR first experiences a drop and then increase, while FAR keeps on decreasing. The higher a vehicle's speed is, the higher chance it is miss-detected, the lower chance a false alarm is generated though. Since MDR is more critical than FAR in our case, Acoussist is deemed to work better to detect low-/medium-speed traffic. Due the imperfect performance, we do not intend to replace human judgement with the detection result of Acoussist. Instead, it is designed as an assistive tool that provides an added layer of protection to the visually impaired when they sense a clear street to cross using hearing. Acoussist is designed to use at uncontrolled crosswalks existing in residential communities, local streets, and suburban areas, where there are common needs from the visually impaired for daily activities and commute. To extend the usage scenarios of our system, we plan to investigate how to improve detection accuracy especially for high-speed vehicles. More sophisticated ranging and estimation algorithms will be developed.

**Feedback design:** In the user study that involves six visually impaired volunteers, we tested two alert modalities, vibration and beeping sound. Smartphones generated either one of the two kinds of alert signals once any potential collision is detected. After the experiments, volunteers are asked for their preference over these two approaches. The result shows that most of them, 5 out of 6, prefer vibration over beeping sound, as the latter will interfere their hearing-based judgement to some extent. In the real application, Acoussist can set vibration as the alert modality by default, while leave users the freedom to switch to beeping sound in the setting menu. So far, Acoussist only provides binary information to users regarding whether a potential collision is detected or not. In the user study, volunteers are asked if they prefer to receive additional information, such as how many detected vehicles nearby and their distances. No consensus is reached; some of them said more information would distract them from sensing the traffic condition with hearing, while others feel additional information help to make more accurate and timely decisions. We plan to implement the optional function of additional information provisioning via Acoussist and carry out a wider user study regarding its feasibility.

## 9 CONCLUSIONS

In this paper, we propose Acoussist, a theoretical grounded acoustic ranging based system that assists visually impaired pedestrians to cross uncontrolled crosswalks. The key novelty of Acoussist lies in the idea of analyzing the t-f sweep line embedded in the received signals to estimate the relative velocity of each vehicle even with the presence of multiple of them who all play homogeneous acoustic chirps. With this basis, we successfully address the association ambiguity issue by proposing MGCC when measuring vehicle's DoA. We explore the geometric relations among system entities to derive important vehicle movement parameters, such as vehicle velocity and v-p distance. From a generalized point of view, we study a *type-III homogeneous multi-source ranging problem* that has not been investigated before. The theoretical results and technical design may shed light to future studies

on this topic. As another contribution of this work, we implement Acoussist using COTS portable speakers and smartphones. Extensive in-field experiments show that the detection accuracy of our system can reach 93.3%.

Acoussist is designed as an assisting tool that provides an added layer of protection to the visually impaired when they sense a clear street to cross using hearing. Due to the same reason, the functionality of Acoussist does not require all vehicles to participate, which is also impractical in reality. Still, Acoussist can effectively detect opt-in vehicles and alert the pedestrian their presence. In a worst case that no vehicle in the pedestrian's vicinity participates the program, it degrades to the conventional hearing-based judgment scenario. Therefore, Acoussist will not perform worse than the current solution.

## ACKNOWLEDGMENTS

We sincerely thank the anonymous reviewers for their insightful comments and suggestions. We are also grateful to NSF (CNS-1943509, ECCS-1849860) and DoT for partially funding this research.

## REFERENCES

- [1] AASHTO. 2011. A Policy on Geometric Design of Highway and Streets. [https://www.academia.edu/33524500/AASHTO\\_Green\\_Book\\_2011.PDF](https://www.academia.edu/33524500/AASHTO_Green_Book_2011.PDF).
- [2] Heba Abdelnasser, Khaled A Harras, and Moustafa Youssef. 2015. UbiBreathe: A ubiquitous non-invasive WiFi-based breathing estimator. In *Proceedings of the International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*. 277–286.
- [3] Iyad Abu Doush, Sawsan Alshatnawi, Abdel-Karim Al-Tamimi, Bushra Alhasan, and Safaa Hamasha. 2017. ISAB: integrated indoor navigation system for the blind. *Interacting with Computers* 29, 2 (2017), 181–202.
- [4] Alireza Asvadi, Cristiano Premebida, Paulo Peixoto, and Urbano Nunes. 2016. 3D Lidar-based static and moving obstacle detection in driving environments: An approach based on voxels and multi-region ground planes. *Robotics and Autonomous Systems* 83 (2016), 299–311.
- [5] Marco Baglietto, Antonio Sgorbissa, Damiano Verda, and Renato Zaccaria. 2011. Human navigation and mapping with a 6DOF IMU and a laser scanner. *Robotics and Autonomous Systems* 59, 12 (2011), 1060–1069.
- [6] Dana H Ballard. 1981. Generalizing the Hough transform to detect arbitrary shapes. *Pattern recognition* 13, 2 (1981), 111–122.
- [7] Nur Hasnifa Hasan Baseri, Ee Yeng Ng, Alireza Safdari, Mahmoud Moghavvemi, and Noraisyah Mohamed Shah. 2017. A Low Cost Street Crossing Electronic Aid for the Deaf and Blind. In *Proceedings of the International Conference for Innovation in Biomedical Engineering and Life Sciences*. 73–78.
- [8] Eli Billauer. 2019. Peak detection. <http://billauer.co.il/peakdet.html>.
- [9] Nicolae Botezatu, Simona Caraiman, Dariusz Rzeszotarski, and Pawel Strumillo. 2017. Development of a versatile assistive system for the visually impaired based on sensor fusion. In *Proceedings of the International Conference on System Theory, Control and Computing*. IEEE, 540–547.
- [10] Simona Caraiman, Anca Morar, Mateusz Owczarek, Adrian Burlacu, Dariusz Rzeszotarski, Nicolae Botezatu, Paul Hergelegiu, Florica Moldoveanu, Pawel Strumillo, and Alin Moldoveanu. 2017. Computer vision for the visually impaired: the sound of vision system. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 1480–1489.
- [11] NDT Research Center. 2020. Attenuation of Sound Waves. <https://www.nde-ed.org/EducationResources/CommunityCollege/Ultrasonics/Physics/attenuation.htm>.
- [12] Volkan Cevher, Rama Chellappa, and James H McClellan. 2008. Vehicle speed estimation using acoustic wave patterns. *IEEE Transactions on signal processing* 57, 1 (2008), 30–47.
- [13] Jagmohan Chauhan, Yining Hu, Suranga Seneviratne, Archan Misra, Aruna Seneviratne, and Youngki Lee. 2017. BreathPrint: Breathing acoustics-based user authentication. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services*. 278–291.
- [14] Sakmongkon Chumkamon, Peranitti Tuvaphanthaphiphat, and Phongsak Keeratiwintakorn. 2008. A blind navigation system using RFID for indoor environments. In *Proceedings of the International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, Vol. 2. 765–768.
- [15] Alice Clifford and Joshua Reiss. 2010. Calculating time delays of multiple active sources in live sound. In *Proceedings of the Audio Engineering Society Convention*. 1–8.
- [16] Maximo Cobos, Jose J Lopez, and David Martinez. 2011. Two-microphone multi-speaker localization based on a Laplacian mixture model. *Digital Signal Processing* 21, 1 (2011), 66–76.
- [17] Xunxue Cui, Kegen Yu, and Songsheng Lu. 2016. Direction finding for transient acoustic source based on biased TDOA measurement. *IEEE Transactions on Instrumentation and Measurement* 65, 11 (2016), 2442–2453.

- [18] Symeon Delikaris-Manias, Despoina Pavlidi, Ville Pulkki, and Athanasios Mouchtaris. 2016. 3D localization of multiple audio sources utilizing 2D DOA histograms. In *Proceedings of the European Signal Processing Conference*. 1473–1477.
- [19] Nilanjan Dey and Amira S Ashour. 2018. Applied examples and applications of localization and tracking problem of multiple speech sources. In *Direction of arrival estimation and localization of multi-speech sources*. Springer, 35–48.
- [20] 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 *Proceedings of the IEEE Vehicular Technology Conference*. IEEE, 1–5.
- [21] DPS. 2019. Department of Public Safety Guide Book. <http://www.dps.texas.gov/internetforms/forms/dl-7.pdf>.
- [22] Hossam Elsayed, Bassem A Abdullah, and Gamal Aly. 2018. Fuzzy Logic Based Collision Avoidance System for Autonomous Navigation Vehicle. In *Proceedings of the IEEE International Conference on Computer Engineering and Systems*. 469–474.
- [23] Viktor Erdélyi, Trung-Kien Le, Bobby Bhattacharjee, Peter Druschel, and Nobutaka Ono. 2018. Sonoloc: Scalable positioning of commodity mobile devices. In *Proceedings of the Annual International Conference on Mobile Systems, Applications, and Services*. 136–149.
- [24] Christine Evers, Alastair H Moore, and Patrick A Naylor. 2016. Acoustic simultaneous localization and mapping (a-SLAM) of a moving microphone array and its surrounding speakers. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 6–10.
- [25] Mojtaba Farmani, Michael Syskind Pedersen, Zheng-Hua Tan, and Jesper Jensen. 2015. Informed TDoA-based direction of arrival estimation for hearing aid applications. In *Proceedings of the IEEE Global Conference on Signal and Information Processing*. 953–957.
- [26] Brian G Ferguson. 2016. Source parameter estimation of aero-acoustic emitters using non-linear least squares and conventional methods. *IET Radar, Sonar & Navigation* 10, 9 (2016), 1552–1560.
- [27] Viacheslav Filonenko, Charlie Cullen, and James D Carswell. 2012. Asynchronous ultrasonic trilateration for indoor positioning of mobile phones. In *Proceedings of the International Symposium on Web and Wireless Geographical Information Systems*. 33–46.
- [28] Carlos Flores, Pierre Merdrignac, Raoul de Charette, Francisco Navas, Vicente Milanés, and Fawzi Nashashibi. 2018. A cooperative car-following/emergency braking system with prediction-based pedestrian avoidance capabilities. *IEEE Transactions on Intelligent Transportation Systems* 20, 99 (2018), 1–10.
- [29] Food, Drug Administration, et al. 2008. Guidance for industry and FDA staff information for manufacturers seeking marketing clearance of diagnostic ultrasound systems and transducers. *Rockville, MD: FDA* (2008).
- [30] Centers for Disease Control and Prevention. 2020. What Noises Cause Hearing Loss? [https://www.cdc.gov/nceh/hearing\\_loss/what\\_noises\\_cause\\_hearing\\_loss.html](https://www.cdc.gov/nceh/hearing_loss/what_noises_cause_hearing_loss.html)
- [31] Sukru Yaren Gelbal, Sibel Arslan, Haoan Wang, Bilin Aksun-Guvenc, and Levent Guvenc. 2017. Elastic band based pedestrian collision avoidance using V2X communication. In *Proceedings of the IEEE Intelligent Vehicles Symposium*. 270–276.
- [32] W Gelmuda and A Kos. 2013. Multichannel ultrasonic range finder for blind people navigation. *Bulletin of the Polish Academy of Sciences. Technical Sciences* 61, 3 (2013).
- [33] GNU. 2019. GNU Scientific Library. <https://www.gnu.org/software/gsl/>.
- [34] Hüseyin Göksu. 2018. Vehicle speed measurement by on-board acoustic signal processing. *Measurement and Control* 51, 5-6 (2018), 138–149.
- [35] Google. 2019. Android NDK. <https://developer.android.com/ndk>.
- [36] Google. 2020. Audioset. <https://research.google.com/audioset/>
- [37] Google. 2020. Phone Device Metrics. <https://material.io/resources/devices/>.
- [38] João Guerreiro, Dragan Ahmetovic, Kris M Kitani, and Chioko Asakawa. 2017. Virtual navigation for blind people: Building sequential representations of the real-world. In *Proceedings of the International ACM SIGACCESS Conference on computers and accessibility*. 280–289.
- [39] João Guerreiro, Dragan Ahmetovic, Daisuke Sato, Kris Kitani, and Chioko Asakawa. 2019. Airport accessibility and navigation assistance for people with visual impairments. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–14.
- [40] John Castillo Guerrero, Cristhian Quezada-V, and Diego Chacon-Troya. 2018. Design and implementation of an intelligent cane, with proximity sensors, gps localization and gsm feedback. In *Proceedings of the IEEE Canadian Conference on Electrical & Computer Engineering*. 1–4.
- [41] Jack Guy. 2019. EU requires carmakers to add fake engine noises to electric cars. <https://www.cnn.com/2019/07/01/business/electric-vehicles-warning-noises-scli-intl-gbr/index.html>.
- [42] Richard Guy and Khai Truong. 2012. CrossingGuard: exploring information content in navigation aids for visually impaired pedestrians. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 405–414.
- [43] Kotaro Hara, Shiri Azenkot, Megan Campbell, Cynthia L Bennett, Vicki Le, Sean Pannella, Robert Moore, Kelly Minckler, Rochelle H Ng, and Jon E Froehlich. 2015. Improving public transit accessibility for blind riders by crowdsourcing bus stop landmark locations with google street view: An extended analysis. *ACM Transactions on Accessible Computing* 6, 2 (2015), 1–23.
- [44] Satoshi Hashino and Ramin Ghurchian. 2010. A blind guidance system for street crossings based on ultrasonic sensors. In *Proceedings of the IEEE International Conference on Information and Automation*. 476–481.
- [45] Hongsen He, Xueyuan Wang, Yingyue Zhou, and Tao Yang. 2018. A steered response power approach with trade-off prewhitening for acoustic source localization. *The Journal of the Acoustical Society of America* 143, 2 (2018), 1003–1007.

- [46] Takuya Higuchi and Hirokazu Kameoka. 2015. Unified approach for audio source separation with multichannel factorial HMM and DOA mixture model. In *Proceedings of the IEEE European Signal Processing Conference*. 2043–2047.
- [47] Ping-Fan Ho and Jyh-Cheng Chen. 2017. Wisafe: Wi-fi pedestrian collision avoidance system. *IEEE Transactions on Vehicular Technology* 66, 6 (2017), 4564–4578.
- [48] Hsieh-Chang Huang, Ching-Tang Hsieh, and Cheng-Hsiang Yeh. 2015. An indoor obstacle detection system using depth information and region growth. *Sensors* 15, 10 (2015), 27116–27141.
- [49] Jingchang Huang, Xin Zhang, Qianwei Zhou, Enliang Song, and Baoqing Li. 2013. A practical fundamental frequency extraction algorithm for motion parameters estimation of moving targets. *IEEE Transactions on Instrumentation and Measurement* 63, 2 (2013), 267–276.
- [50] Wenchao Huang, Yan Xiong, Xiang-Yang Li, Hao Lin, Xufei Mao, Panlong Yang, and Yunhao Liu. 2014. Shake and walk: Acoustic direction finding and fine-grained indoor localization using smartphones. In *Proceedings of the IEEE Conference on Computer Communications*. 370–378.
- [51] Yasha Irvantchi, Mayank Goel, and Chris Harrison. 2019. BeamBand: Hand Gesture Sensing with Ultrasonic Beamforming. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 15.
- [52] Florian Jacob, Joerg Schmalenstroeer, and Reinhold Haeb-Umbach. 2013. DOA-based microphone array position self-calibration using circular statistics. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 116–120.
- [53] Slim Kammoun, Gaétan Parseihian, Olivier Gutierrez, Adrien Brilhault, Antonio Serpa, Mathieu Raynal, Bernard Oriola, MJ-M Macé, Malika Auvray, Michel Denis, et al. 2012. Navigation and space perception assistance for the visually impaired: The NAVIG project. *Irbm* 33, 2 (2012), 182–189.
- [54] Masanori Kato, Yuza Senda, and Reishi Kondo. 2017. Tdoa estimation based on phase-voting cross correlation and circular standard deviation. In *Proceedings of the IEEE European Signal Processing Conference*. 1230–1234.
- [55] Dima Khaykin and Boaz Rafaely. 2012. Acoustic analysis by spherical microphone array processing of room impulse responses. *The Journal of the Acoustical Society of America* 132, 1 (2012), 261–270.
- [56] SamYong Kim, JeongKwan Kang, Se-Young Oh, YeongWoo Ryu, Kwangsoo Kim, SangCheol Park, and Jinwon Kim. 2008. An intelligent and integrated driver assistance system for increased safety and convenience based on all-around sensing. *Journal of Intelligent and Robotic Systems* 51, 3 (2008), 261–287.
- [57] SamYong Kim, SeYoung Oh, JeongKwan Kang, YoungWoo Ryu, Kwangsoo Kim, SangCheol Park, and KyongHa Park. 2005. Front and rear vehicle detection and tracking in the day and night times using vision and sonar sensor fusion. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2173–2178.
- [58] Charles Knapp and Clifford Carter. 1976. The generalized correlation method for estimation of time delay. *IEEE/ACM Transactions on Acoustics, Speech, and Signal processing* 24, 4 (1976), 320–327.
- [59] Dionyssos Kounades-Bastian, Laurent Girin, Xavier Alameda-Pineda, Sharon Gannot, and Radu Horaud. 2017. An EM algorithm for joint source separation and diarisation of multichannel convolutive speech mixtures. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 16–20.
- [60] Robert Layton and Karen Dixon. 2012. Stopping sight distance. <https://cce.oregonstate.edu/sites/cce.oregonstate.edu/files/12-2-stopping-sight-distance.pdf>.
- [61] Bing Li, J Pablo Munoz, Xuejian Rong, Jizhong Xiao, Yingli Tian, and Aries Ardit. 2016. ISANA: wearable context-aware indoor assistive navigation with obstacle avoidance for the blind. In *Proceedings of the European Conference on Computer Vision*. 448–462.
- [62] Qing Lin and Youngjoon Han. 2014. A context-aware-based audio guidance system for blind people using a multimodal profile model. *Sensors* 14, 10 (2014), 18670–18700.
- [63] Qingju Liu, Wenwu Wang, Teofilo de Campos, Philip JB Jackson, and Adrian Hilton. 2017. Multiple speaker tracking in spatial audio via PHD filtering and depth-audio fusion. *IEEE Transactions on Multimedia* 20, 7 (2017), 1767–1780.
- [64] Xuefeng Liu, Jiannong Cao, Shaojie Tang, and Jiaqi Wen. 2014. Wi-Sleep: Contactless sleep monitoring via WiFi signals. In *Proceedings of the IEEE Real-Time Systems Symposium*. 346–355.
- [65] Zhenyu Liu, Lin Pu, Zhen Meng, Xinyang Yang, Konglin Zhu, and Lin Zhang. 2015. POFS: A novel pedestrian-oriented forewarning system for vulnerable pedestrian safety. In *Proceedings of the IEEE International Conference on Connected Vehicles and Expo*. 100–105.
- [66] Kam W Lo and Brian G Ferguson. 2000. Broadband passive acoustic technique for target motion parameter estimation. *IEEE Trans. Aerospace Electron. Systems* 36, 1 (2000), 163–175.
- [67] Heinrich W Löllmann, Christine Evers, Alexander Schmidt, Heinrich Mellmann, Hendrik Barfuss, Patrick A Naylor, and Walter Kellermann. 2018. The LOCATA challenge data corpus for acoustic source localization and tracking. In *Proceedings of the Sensor Array and Multichannel Signal Processing Workshop (SAM)*. 410–414.
- [68] Ningbo Long, Kaiwei Wang, Ruiqi Cheng, Kailun Yang, and Jian Bai. 2018. Fusion of millimeter wave radar and RGB-depth sensors for assisted navigation of the visually impaired. In *Proceedings of the Millimetre Wave and Terahertz Sensors and Technology XI*. 1–8.
- [69] Roberto López-Valcarce, Carlos Mosquera, and Fernando Pérez-González. 2004. Estimation of road vehicle speed using two omnidirectional microphones: a maximum likelihood approach. *EURASIP Journal on Applied Signal Processing* 2004 (2004), 1059–1077.

- [70] Shachar Maidenbaum, Shlomi Hanassy, Sami Abboud, Galit Buchs, Daniel-Robert Chebat, Shelly Levy-Tzedek, and Amir Amedi. 2014. The EyeCane, a new electronic travel aid for the blind: Technology, behavior & swift learning. *Restorative Neurology and Neuroscience* 32, 6 (2014), 813–824.
- [71] Wenguang Mao, Mei Wang, and Lili Qiu. 2018. Aim: acoustic imaging on a mobile. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services (Mobisys)*. 468–481.
- [72] MathWorks. 2020. MATLAB Audio Toolbox. <https://www.mathworks.com/products/audio.html>
- [73] Monsoon. 2019. Power monitor. <https://www.msoon.com/online-store>.
- [74] A Montanha, MJ Escalon, FJ Domínguez-Mayo, and AM Polidorio. 2016. A technological innovation to safely aid in the spatial orientation of blind people in a complex urban environment. In *Proceedings of the International Conference on Image, Vision and Computing*. 102–107.
- [75] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. 2015. Contactless sleep apnea detection on smartphones. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services*. 45–57.
- [76] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. 2016. Fingerio: Using active sonar for fine-grained finger tracking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1515–1525.
- [77] Fawzi Nashashibi and Alexandre Bargeton. 2008. Laser-based vehicles tracking and classification using occlusion reasoning and confidence estimation. In *Proceedings of the IEEE Intelligent Vehicles Symposium*. 847–852.
- [78] Matthias Nieuwenhuisen, David Droeschel, Marius Beul, and Sven Behnke. 2016. Autonomous navigation for micro aerial vehicles in complex gnss-denied environments. *Journal of Intelligent & Robotic Systems* 84, 1-4 (2016), 199–216.
- [79] Joonas Nikunen and Tuomas Virtanen. 2014. Multichannel audio separation by direction of arrival based spatial covariance model and non-negative matrix factorization. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 6677–6681.
- [80] National Institutes of Health et al. 1990. Noise and Hearing Loss Consensus Conference. *JAMA* 263 (1990), 3185–3190.
- [81] American Council of the Blind. 2019. White Cane Law. <https://www.acb.org/whitecane>.
- [82] Eshed Ohn-Bar, João Guerreiro, Dragan Ahmetovic, Kris M Kitani, and Chieko Asakawa. 2018. Modeling expertise in assistive navigation interfaces for blind people. In *Proceedings of the International Conference on Intelligent User Interfaces*. 403–407.
- [83] Olakanmi Oladayo. 2014. A multidimensional walking aid for visually impaired using ultrasonic sensors network with voice guidance. *International Journal of Intelligent Systems and Applications* 6, 8 (2014), 53.
- [84] World Health Organization. 2018. Blindness and vision impairment report. <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>.
- [85] En Peng, Patrick Peursum, Ling Li, and Svetha Venkatesh. 2010. A smartphone-based obstacle sensor for the visually impaired. In *Proceedings of the International Conference on Ubiquitous Intelligence and Computing*. 590–604.
- [86] Riccardo Polvara, Sanjay Sharma, Jian Wan, Andrew Manning, and Robert Sutton. 2018. Obstacle avoidance approaches for autonomous navigation of unmanned surface vehicles. *The Journal of Navigation* 71, 1 (2018), 241–256.
- [87] Kun Qian, Chenshu Wu, Fu Xiao, Yue Zheng, Yi Zhang, Zheng Yang, and Yunhao Liu. 2018. Acousticcardiogram: Monitoring heartbeats using acoustic signals on smart devices. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*. 1574–1582.
- [88] X Qian and C Ye. 2013. NCC-RANSAC: A fast plane extraction method for navigating a smart cane for the visually impaired. In *Proceedings of the IEEE International Conference on Automation Science and Engineering*. 261–267.
- [89] BG Quinn. 1995. Doppler speed and range estimation using frequency and amplitude estimates. *The Journal of the Acoustical Society of America* 98, 5 (1995), 2560–2566.
- [90] Alberto Rodríguez, J Javier Yebe, Pablo F Alcantarilla, Luis M Bergasa, Javier Almazán, and Andrés Cela. 2012. Assisting the visually impaired: obstacle detection and warning system by acoustic feedback. *Sensors* 12, 12 (2012), 17476–17496.
- [91] A Rodríguez Valiente, A Trinidad, JR García Berrocal, C Górriz, and R Ramírez Camacho. 2014. Extended high-frequency (9–20 kHz) audiometry reference thresholds in 645 healthy subjects. *International Journal of Audiology* 53, 8 (2014), 531–545.
- [92] Zoltan Rozsa and Tamas Sziranyi. 2018. Obstacle prediction for automated guided vehicles based on point clouds measured by a tilted LIDAR sensor. *IEEE Transactions on Intelligent Transportation Systems* 19, 8 (2018), 2708–2720.
- [93] Wenjie Ruan, Quan Z Sheng, Lei Yang, Tao Gu, Peipei Xu, and Longfei Shangguan. 2016. AudioGest: enabling fine-grained hand gesture detection by decoding echo signal. In *Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing*. 474–485.
- [94] Daisuke Sato, Uran Oh, Kakuya Naito, Hironobu Takagi, Kris Kitani, and Chieko Asakawa. 2017. Navcog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment. In *Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility*. 270–279.
- [95] Dana Sauerburger. 1995. Safety awareness for crossing streets with no traffic control. *Journal of Visual Impairment & Blindness* 89, 5 (1995), 423–431.
- [96] Ralph Schmidt. 1986. Multiple emitter location and signal parameter estimation. *IEEE transactions on antennas and propagation* 34, 3 (1986), 276–280.

- [97] Jochen Schneider and Thomas Strothotte. 2000. Constructive exploration of spatial information by blind users. In *Proceedings of the international ACM conference on Assistive technologies*. 188–192.
- [98] Sharang Sharma, Manind Gupta, Amit Kumar, Meenakshi Tripathi, and Manoj Singh Gaur. 2017. Multiple distance sensors based smart stick for visually impaired people. In *Proceedings of the IEEE Annual Computing and Communication Workshop and Conference*. 1–5.
- [99] Tarun Sharma, JHM Apoorva, Ramananathan Lakshmanan, Prakruti Gogia, and Manoj Kondapaka. 2016. NAVI: Navigation aid for the visually impaired. In *Proceedings of the International Conference on Computing, Communication and Automation*. IEEE, 971–976.
- [100] Tushar Sharma, Tarun Nalwa, Tanupriya Choudhury, Suresh Chand Satapathy, and Praveen Kumar. 2017. Smart Cane: Better Walking Experience for Blind People. In *Proceedings of the IEEE International Conference on Computational Intelligence and Networks*. 22–26.
- [101] Bridget Smith and David Sandwell. 2003. Accuracy and resolution of shuttle radar topography mission data. *Geophysical Research Letters* 30, 9 (2003).
- [102] Ke Sun, Ting Zhao, Wei Wang, and Lei Xie. 2018. Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals. In *Proceedings of the International Conference on Mobile Computing and Networking*. 591–605.
- [103] Yoiti Suzuki and Hisashi Takeshima. 2004. Equal-loudness-level contours for pure tones. *The Journal of the Acoustical Society of America* 116, 2 (2004), 918–933.
- [104] AA Tahat. 2009. A wireless ranging system for the blind long-cane utilizing a smart-phone. In *Proceedings of the IEEE International Conference on Telecommunications*. 111–117.
- [105] Amin Tahmasbi-Sarvestani, Hossein Nourkhiz Mahjoub, Yaser P Fallah, Ehsan Moradi-Pari, and Oubada Abuhaar. 2017. Implementation and evaluation of a cooperative vehicle-to-pedestrian safety application. *IEEE Intelligent Transportation Systems Magazine* 9, 4 (2017), 62–75.
- [106] Yu Takahashi, Tomoya Takatani, Keiichi Osako, Hiroshi Saruwatari, and Kiyohiro Shikano. 2009. Blind spatial subtraction array for speech enhancement in noisy environment. *IEEE Transactions on Audio, Speech, and Language Processing* 17, 4 (2009), 650–664.
- [107] Ruxandra Tapu, Bogdan Mocanu, Andrei Bursuc, and Titus Zaharia. 2013. A smartphone-based obstacle detection and classification system for assisting visually impaired people. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 444–451.
- [108] YingLi Tian, Xiaodong Yang, Chucai Yi, and Aries Arditi. 2013. Toward a computer vision-based wayfinding aid for blind persons to access unfamiliar indoor environments. *Machine vision and applications* 24, 3 (2013), 521–535.
- [109] Johannes Traa, David Wingate, Noah D Stein, and Paris Smaragdīs. 2015. Robust source localization and enhancement with a probabilistic steered response power model. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24, 3 (2015), 493–503.
- [110] David Tse and Pramod Viswanath. 2005. *Fundamentals of wireless communication*. Cambridge university press.
- [111] Jeremy G Turner, Jennifer L Parrish, Larry F Hughes, Linda A Toth, and Donald M Caspary. 2005. Hearing in laboratory animals: strain differences and nonauditory effects of noise. *Comparative medicine* 55, 1 (2005), 12–23.
- [112] Iwan Ulrich and Johann Borenstein. 2001. The GuideCane-applying mobile robot technologies to assist the visually impaired. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 31, 2 (2001), 131–136.
- [113] Jose Velasco, Mohammad J Taghizadeh, Afsaneh Asaei, Hervé Bourlard, Carlos J Martín-Arguedas, Javier Macias-Guarasa, and Daniel Pizarro. 2015. Novel GCC-PHAT model in diffuse sound field for microphone array pairwise distance based calibration. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. 2669–2673.
- [114] Raghav H Venkatnarayan, Griffin Page, and Muhammad Shahzad. 2018. Multi-user gesture recognition using WiFi. In *Proceedings of the ACM International Conference on Mobile Systems, Applications, and Services*. 401–413.
- [115] Deutsche Version. 2020. Conversion of sound units. <http://www.sengpielaudio.com/calculator-soundlevel.htm>.
- [116] Deutsche Version. 2020. Damping of Air at High Frequencies. <http://www.sengpielaudio.com/calculator-air.htm>.
- [117] 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 ACM Workshop on Mobile Computing Systems & Applications*. 5–10.
- [118] Tianben Wang, Daqing Zhang, Yuanqing Zheng, Tao Gu, Xingshe Zhou, and Bernadette Dorizzi. 2018. C-FMCW based contactless respiration detection using acoustic signal. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 170.
- [119] Wei Wang, Alex X Liu, and Ke Sun. 2016. Device-free gesture tracking using acoustic signals. In *Proceedings of the ACM Annual International Conference on Mobile Computing and Networking*. 82–94.
- [120] Martin Weiss, Simon Chamorro, Roger Girgis, Margaux Luck, Samira E Kahou, Joseph P Cohen, Derek Nowrouzezahrai, Doina Precup, Florian Golemo, and Chris Pal. 2020. Navigation agents for the visually impaired: A sidewalk simulator and experiments. In *Proceedings of the Conference on Robot Learning*. 1314–1327.
- [121] Xinzhou Wu, Radovan Miucic, Sichao Yang, Samir Al-Stouhi, James Misener, Sue Bai, and Wai-hoi Chan. 2014. Cars talk to phones: A DSRC based vehicle-pedestrian safety system. In *Proceedings of the IEEE Vehicular Technology Conference*. IEEE, 1–7.
- [122] Li Xi, Liu Guosui, and Jinlin Ni. 1999. Autofocusing of ISAR images based on entropy minimization. *IEEE Trans. Aerospace Electron. Systems* 35, 4 (1999), 1240–1252.

- [123] Yanming Xiao, Jenshan Lin, Olga Boric-Lubecke, and M Lubecke. 2006. Frequency-tuning technique for remote detection of heartbeat and respiration using low-power double-sideband transmission in the Ka-band. *IEEE Transactions on Microwave Theory and Techniques* 54, 5 (2006), 2023–2032.
- [124] Jia Xu, Xiang-Gen Xia, Shi-Bao Peng, Ji Yu, Ying-Ning Peng, and Li-Chang Qian. 2012. Radar maneuvering target motion estimation based on generalized Radon-Fourier transform. *IEEE Transactions on Signal Processing* 60, 12 (2012), 6190–6201.
- [125] Yanni Yang, Jiannong Cao, Xiulong Liu, and Xuefeng Liu. 2019. Multi-Breath: Separate Respiration Monitoring for Multiple Persons with UWB Radar. In *Proceedings of the IEEE Computer Software and Applications Conference*. 840–849.
- [126] Cang Ye, Soonhac Hong, Xiangfei Qian, and Wei Wu. 2016. Co-robotic cane: A new robotic navigation aid for the visually impaired. *IEEE Systems, Man, and Cybernetics Magazine* 2, 2 (2016), 33–42.
- [127] Wenyi Zhang and Bhaskar D Rao. 2010. A two microphone-based approach for source localization of multiple speech sources. *IEEE Transactions on Audio, Speech, and Language Processing* 18, 8 (2010), 1913–1928.
- [128] Xiaojia Zhao, Yuxuan Wang, and DeLiang Wang. 2014. Robust speaker identification in noisy and reverberant conditions. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22, 4 (2014), 836–845.
- [129] Yuhang Zhao, Cynthia L Bennett, Hrvoje Benko, Edward Cutrell, Christian Holz, Meredith Ringel Morris, and Mike Sinclair. 2018. Enabling People with Visual Impairments to Navigate Virtual Reality with a Haptic and Auditory Cane Simulation. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. 116–130.
- [130] Shifeng Zhou. 2018. A Smart Cane to Help the Blind People Walk Confidently. *Materials Science and Engineering* 439, 3 (2018), 032121.