

# Sensing Driver Phone Use with Acoustic Ranging Through Car Speakers

Jie Yang, *Student Member, IEEE*, Simon Sidhom, Gayathri Chandrasekaran,  
Tam Vu, Hongbo Liu, *Student Member, IEEE*, Nicolae Cecan,  
Yingying Chen, *Senior Member, IEEE*, Marco Gruteser, and Richard P. Martin, *Member, IEEE*

**Abstract**—This work addresses the fundamental problem of distinguishing between a driver and passenger using a mobile phone, which is the critical input to enable numerous safety and interface enhancements. Our detection system leverages the existing car stereo infrastructure, in particular, the speakers and Bluetooth network. Our acoustic approach has the phone send a series of customized high frequency beeps via the car stereo. The beeps are spaced in time across the left, right, and if available, front and rear speakers. After sampling the beeps, we use a sequential change-point detection scheme to time their arrival, and then use a differential approach to estimate the phone's distance from the car's center. From these differences a passenger or driver classification can be made. To validate our approach, we experimented with two kinds of phones and in two different cars. We found that our customized beeps were imperceptible to most users, yet still playable and recordable in both cars. Our customized beeps were also robust to background sounds such as music and wind, and we found the signal processing did not require excessive computational resources. In spite of the cars' heavy multi-path environment, our approach had a classification accuracy of over 90%, and around 95% with some calibrations. We also found we have a low false positive rate, on the order of a few percent.

**Index Terms**—Driving Safety, Driver phone use, Smartphone, Car Speakers, Bluetooth, Acoustic Ranging, Location Classification.

---

## 1 INTRODUCTION

Distinguishing driver and passenger phone use is a building block for a variety of applications but it's greatest promise arguably lies in helping reduce driver distraction. Cell phone distractions have been a factor in high-profile accidents [9] and are generally associated with a large number of automobile accidents. For example a National Highway Traffic Safety Administration study identified cell phone distraction as a factor in crashes that led to 995 fatalities and 24,000 injuries in 2009 [42]. This has led to increasing public attention [8], [32] and the banning of handheld phone use in several US states [4] as well as many countries around the world [1].

Unfortunately, an increasing amount of research suggests that the safety benefits of handsfree phone operation are

marginal at best [16], [41]. The cognitive load of conducting a cell phone conversation seems to increase accident risk, rather than the holding of a phone to the ear. Of course, texting, email, navigation, games, and many other apps on smartphones are also increasingly competing with driver attention and pose additional dangers. This has led to a renewed search for technical approaches to the driver distraction problem. Such approaches run the gamut from improved driving mode user interfaces, which allow quicker access to navigation and other functions commonly used while driving, to apps that actively prevent phone calls. In between these extremes lie more subtle approaches: routing incoming calls to voicemail or delaying incoming text notifications, as also recently advocated by Lindqvist *et al.* [28].

**The Driver-Passenger Challenge.** All of these applications would benefit from and some of them depend on automated mechanisms for determining when a cell phone is used by a driver. Prior research and development has led to a number of techniques that can determine whether a cell phone is in a moving vehicle—for example, based on cell phone hand-offs [23], cell phone signal strength analysis [18], or speed as measured by a Global Positioning System receiver. The latter approach appears to be the most common among apps that block incoming or outgoing calls and texts [3], [10], [11]. That is, the apps determine that the cell phone is in a vehicle and activate blocking policies once speed crosses a threshold. Some apps (e.g., [6]) require the installation of a Bluetooth transmitter module into the vehicle OBD2 port, which then allows blocking calls/text to/from a given phone

- 
- J. Yang, H. Liu and Y. Chen are with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Castle Point on Hudson, Hoboken, NJ 07030.  
E-mail: {jyang, hliu3, yingying.chen}@stevens.edu.
  - S. Sidhom is with the Department of Computer Science, Stevens Institute of Technology, Castle Point on Hudson, Hoboken, NJ 07030 USA.  
E-mail: ssidhom@stevens.edu.
  - G. Chandrasekaran, T. Vu, N. Cecan and M. Gruteser are with the Wireless Information Networks Laboratory (WINLAB), Technology Centre of New Jersey, Rutgers, The State University of New Jersey, 671 Route 1 South, North Brunswick, NJ 08902-3390.  
E-mail: {chandrga, tamvu, gruteser}@winlab.rutgers.edu, cecan@eden.rutgers.edu.
  - R.P. Martin is with the Department of Computer Science, Rutgers University, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019.  
E-mail: rmartin@cs.rutgers.edu.

based on car's speedometer readings and some even rely on a radio jammer [5]. None of these solutions, however, can automatically distinguish a driver's cell phone from a passenger's.

While we have not found any detailed statistics on driver versus passenger cell phone use in vehicles, a federal accident database (FARS) [7] reveals that about 38% of automobile trips include passengers.<sup>1</sup> Not every passenger carries a phone—still this number suggests that the false positive rate when relying only on vehicle detection would be quite high. It would probably be unacceptably high even for simple interventions such as routing incoming calls to voicemail. Distinguishing drivers and passengers is challenging because car and phone usage patterns can differ substantially. Some might carry a phone in a pocket, while others place it on the vehicle console. Since many vehicles are driven mostly by the same driver, the approach of placing a Bluetooth device into the vehicles appears promising. It allows the phone to recognize that the user is in the car by scanning for the device's Bluetooth identifier. Still, this cannot cover cases where one person uses the same vehicle as both driver and passenger, as is frequently the case for family cars. Also, some vehicle occupants might pass their phone to others, to allow them to try out a game, for example.

**An Acoustic Ranging Approach.** In this paper, we introduce and evaluate an acoustic relative-ranging system that classifies on which car seat a phone is being used. The system relies on the assumptions (i) that seat location is one of the most useful discriminators for distinguishing driver and passenger cell phone use and (ii) that most cars will allow phone access to the car audio infrastructure. Indeed, an industry report [39] discloses that more than 8 million built-in Bluetooth systems were sold in 2010 and predicts that 90% of new cars will be equipped in 2016. Our system leverages this Bluetooth access to the audio infrastructure to avoid the need to deploy additional infrastructure in cars. Our classifier's strategy first uses high frequency beeps sent from a smartphone over a Bluetooth connection through the car's stereo system. The beeps are recorded by the phone, and then analyzed to deduce the timing differentials between the left and right speakers (and if possible, front and rear ones). From the timing differentials, the phone can self-determine which side or quadrant of the car it is in. While acoustic localization and ranging have been extensively studied for human speaker localization through microphone arrays, we focus on addressing several unique challenges presented in this system. First, our system uses only a single microphone and multiple speakers, requiring a solution that minimizes interference between the speakers. Second, the small confined space inside a car presents a particularly challenging multipath environment. Third, any sounds emitted should be unobtrusive to minimize distraction. Salient features of our solution that

1. Based on 2 door and 4 door passenger vehicles in 2009. The database only includes vehicle trips ending in a fatal accident, thus it may not be fully representative of all trips.

address these challenges are:

- By exploiting the relatively controlled, symmetric positioning of speakers inside a car, the system can perform seat classification even without the need for calibration, fingerprinting or additional infrastructure.
- To make our approach unobtrusive, we use very high frequency beeps, close to the limits of human perception, at about 18 kHz. Both the number and length of the beeps are relatively short. This exploits that today's cell phone microphones and speakers have a wider frequency response than most peoples' auditory system.
- To address significant multipath and noise in the car environment, we use several signal processing steps including bandpass filtering to remove low-frequency noise. Since the first arriving signal is least likely to stem from multipath, we use a sequential change-point detection technique that can quickly identify the start of this first signal.

By relaxing the problem from full localization to classification of whether the phone is in a driver or passenger seat area, we enable a first generation system through a smartphone app that is practical today in all cars with built-in Bluetooth (provided the phone can connect). This is because left-right classification can be achieved with only stereo audio, and this covers the majority of scenarios (except when the phone is located in the driver-side rear passenger seat, which is occupied in less than 9% of vehicle trips according to FARS). We also show how accuracy can be substantially improved when Bluetooth control over surround sound audio becomes available, or car audio systems provide the function to generate the audio beeps themselves. Given that high-end vehicles are already equipped with sophisticated surround sound systems and more than 15 speakers [2], it is likely that such control will eventually become available.

To validate our approach and demonstrate its generality, we conducted experiments on 2 types of phones in 2 different cars. The results show that audio files played through the car's existing Bluetooth personal area network have sufficient fidelity to extract the timing differentials needed. Our prototype implementation also shows that the Android Developer Phone has adequate computational capabilities to perform the signal processing needed in a standard programming environment. This revised version of our earlier paper [44] includes several presentation updates as well as a new discussion of and initial results for an accelerometer-based discrimination between front and rear seats. This technique can be useful if control over rear speakers is unavailable.

## 2 RELATED WORK

There are active efforts in developing driver distraction detection systems and systems that help managing interruptability caused by hand-held devices. Approaches involving wearing special equipment when driving to detect driver distraction have been developed [14]. Further, Kutila *et al.* [26] proposed a camera vision system. While the system is more suitable

for in-vehicle environments comparing to its predecessors, it did not take the presence of hand-held devices into account. The adverse effects of using a phone on driver’s behavior have been identified [38]. With the increasing number of automobile accidents involved driver cell phone use, more recent contributions are made in the area of reducing driver distraction by allowing mobile users handling their devices with less effort while driving. These systems include Quiet Calls [31], Blind Sight [27], Negotiator [43], and Lindqvist’s systems [28]. They assumed context information of the device and prior knowledge of the phone use by the driver. Our work is different in that we address the fundamental problem of detecting the driver phone use, which can enable numerous safety and interface applications.

Turning to acoustic positioning techniques, Beepbeep [34] proposed an acoustic-based ranging system that can achieve 1 or 2 cm accuracy within a range of 10 meters, which is so far the best result of ranging using off-the-shelf cell phones. It requires application-level communication between two ranging devices. However, in our in-car environment, the head unit is not programmable and only mobile phones are programmable. Cricket [35] and Bat system [24] employed specially designed hardware to compute time difference of arrival or time-of-flight of ultrasonic signal to achieve an accuracy up to several centimeters. ENSBox [22] integrated an ARM processor running Linux to provide high precision clock synchronization for acoustic ranging and achieved an average accuracy of 5 centimeters. WALRUS [15] used the Wi-Fi network and ultrasound to determine location of the mobile devices to room-level accuracy. Sallai *et al.* [37] evaluated acoustic ranging in resource constrained sensor networks by estimating the time-of-flight as the difference of the arrival times of the sound and radio signals.

Toward speaker localization for in-car environment, both Swerdlow [12] and Hu [25] proposed to detect the speaker’s location inside a car using the microphone array. Rodriguez-Ascariz *et al.* [36] developed a system for detecting driver use of mobile phones using specialized rectenna. These approaches either require additional hardware infrastructure or involve expensive computation, making them less attractive when distinguishing driver and passenger phone use. Our system leverages the existing car stereo infrastructure to locate smartphones by exploiting only a single microphone and multiple speakers. Our approach is designed to be unobtrusive and computationally feasible on off-the-shelf smartphones. A key contribution is its robustness under heavy multipath and noisy in-car environments.

### 3 SYSTEM DESIGN

To address the driver-passenger challenge, we introduce an acoustic ranging technique that leverages the existing car audio infrastructure. In this section, we discuss in detail design goals, the ranging approach, and the beep design. And in the following section we present beep signal detection and location classification.

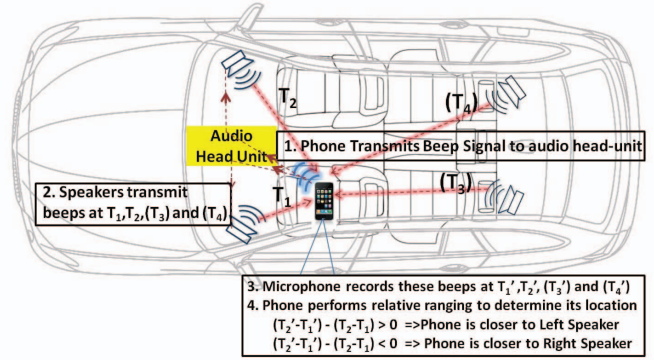


Fig. 1. Illustration of the logical flow in our system.

### 3.1 Challenges and Design Goals

The key goal that led to our acoustic approach was to be able to determine seat location without the need to add dedicated infrastructure to the car. In many cars, the speaker system is already accessible over Bluetooth connections and such systems can be expected to trickle down to most new cars over the next few years. This allows a pure phone software solution. The acoustic approach leads, however, to several additional challenges:

**Unobtrusiveness.** The sounds emitted by the system should not be perceptible to the human ear, so that it does not annoy or distract the vehicle occupants.

**Robustness to Noise and Multipath.** Engine noise, tire and road noise, wind noise, and music or conversations all contribute to a relatively noisy in-car environment. A car is also a relatively small confined space creating a challenging heavy multipath scenario. The acoustic techniques must be robust to these distortions.

**Computational Feasibility on Smartphones.** Standard smartphone platforms should be able to execute the signal processing and detection algorithms with sub-second runtimes.

### 3.2 Acoustic Ranging Overview

The key idea underlying our driver phone use detection system is to perform relative ranging with the car speakers. As illustrated in figure 1, the system, when triggered, say, by an incoming phone call, transmits an audio signal via Bluetooth to the car head unit. This signal is then played through the car speakers. The phone records the emitted sound through its microphone and processes this recorded signal to evaluate propagation delay. Rather than measuring absolute delay, which is affected by unknown processing delays on the phone and in the head unit, the system measures relative delay between the signal from the left and right speaker(s). This is similar in spirit to time-difference-of-arrival localization and does not require clock synchronization. Note, however, that the system does not necessarily perform full localization.



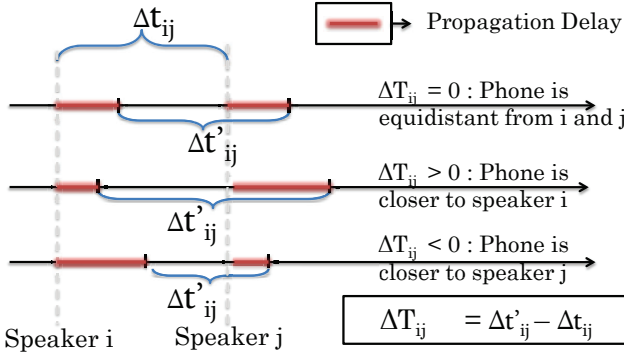


Fig. 2. Relative ranging when applied to a speaker pair  $i$  and  $j$ , for example front-left and front-right.

In virtually all cars, the speakers are placed so that the plane equidistant to the left and right (front) speaker locations separates the driver-side and passenger-side area. This has two benefits. First, for front seats (the most frequently occupied seats), the system can distinguish driver seat and passenger seat by measuring only the relative time difference between the front speakers. Second, the system does not require any fingerprinting or calibration since a time difference of zero always indicates that the phone is located between driver and passenger (on the center console). For these reasons, we refer to this approach as relative ranging.

This basic two-channel approach is practical with current handsfree and A2DP Bluetooth profiles which provide for stereo audio. The concept can be easily extended to four-channel, which promises better accuracy but would require updated surround sound head units and Bluetooth profiles. We will consider both the two and four channel options throughout the remainder of the paper.

Our system differs from typical acoustic human speaker localization, in that we use a single microphone and multiple sound sources rather than a microphone array to detect a single sound source. This means that time differences only need to be measured between signals arriving at the same microphone. This time difference can be estimated simply by counting the number of audio samples between the two beeps. Most modern smartphones offer an audio sampling frequency of 44.1 kHz, which given the speed of sound theoretically provides an accuracy of about 0.8 cm—the resolution under ideal situation, since the signal will be distorted.

Our multi-source approach also raises two new issues, however. First, we have to ensure that the signals from different speakers do not interfere. Second, we need to be able to distinguish the signals emitted from the different speakers. We address both through a time-division multiplexing approach. We let speakers emit sounds at different points in time, with a sufficiently large gap that no interference occurs in the confined in-vehicle space. Since the order of speakers is known to the phone, it can also easily assign the received sounds to the respective speakers.

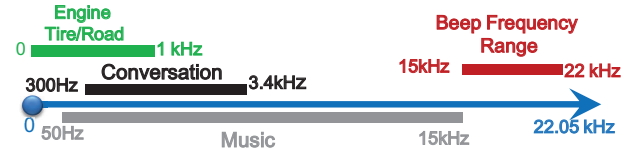


Fig. 3. Frequency distribution of noise and beep signal.

Figure 2 illustrates relative ranging approach for any two speakers  $i$  and  $j$ , for example, front-left and front-right. Assume the fixed time interval between two emitted sounds by a speaker pair  $i$  and  $j$  is  $\Delta t_{ij}$ . Let  $\Delta t'_{ij}$  be the time difference when the microphone records these sounds. The time difference of signal from  $i$ th and  $j$ th speaks arriving at phone is defined as

$$\Delta(T_{ji}) = \Delta t'_{ij} - \Delta t_{ij}, \quad i \neq j \quad i, j = 1, 2, 3, 4. \quad (1)$$

Had the microphone been equidistant from these two speakers, we would have  $\Delta(T_{ji}) = 0$ . If  $\Delta(T_{ji}) < 0$ , the phone is closer to the  $i$ th speaker and if  $\Delta(T_{ji}) > 0$ , it is closer to the  $j$ th speaker.

In our system, the absolute time the sounds emitted by speakers are unknown to the phone, but the phone does know the time difference  $\Delta t_{ij}$ . Similarly, the absolute times the phone records the sounds might be affected by phone processing delays, but the difference  $\Delta t'_{ij}$  can be easily calculated using the sample counting. As can be seen, from the equations above, these two differences are sufficient to determine which speaker is closer.

### 3.3 Beep Signal Design

In designing the beep sound played through the car speakers, we primarily consider two challenges: background noise and unobtrusiveness.

**Frequency Selection.** We choose a high frequency beep at the edge of the phone microphone frequency response curve, since this makes it both easier to filter out noise and renders the signal imperceptible for most, if not all, people. The majority of the typical car noise sources are in lower frequency bands as shown in Figure 3. For example, the noise from the engine, tire/road, and wind are mainly located in the low frequency bands below 1 kHz [17], whereas conversation ranges from approximately 300 Hz to 3400 Hz [40]. Music has a wider range, the FM radio for example spans a frequency range from 50 Hz to 15,000 Hz, which covers almost all naturally occurring sounds. Although separating noise can be difficult in the time domain, we enable straightforward separation in the frequency domain by locating our signal above 15 kHz.

Such high-frequency sounds are also hard to perceive by the human auditory system. Although the frequency range of human hearing is generally considered to be 20 Hz to 20 kHz [21], high frequency sounds must be much louder to be noticeable. 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. Figure 4(a) shows how the ATH varies over frequency, as given in [20]. Note,

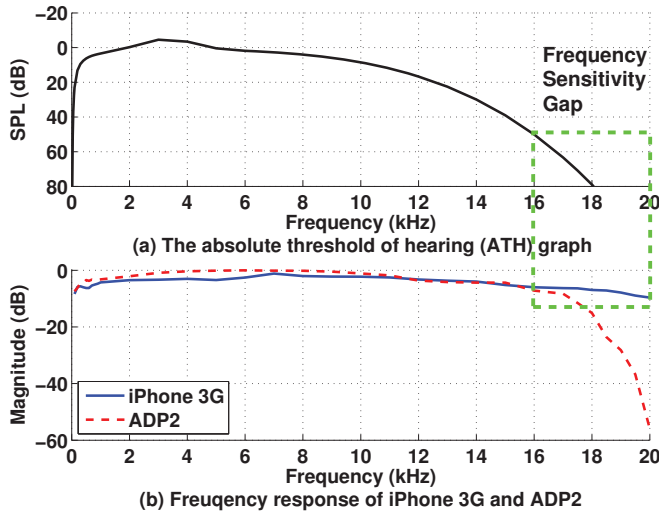


Fig. 4. Frequency sensitivity comparison between the human ear and smartphone.

how the threshold of hearing increases sharply for frequencies over 10 kHz and how human hearing becomes extremely insensitive to frequencies beyond 18 kHz. For example, human ears can detect sounds as low as 0 dB sound pressure level (SPL) at 1 kHz, but require about 80 dB SPL beyond 18 kHz—a 10,000 fold amplitude increase.

Fortunately, the current cell phone microphones are more sensitive to this high-frequency range. We experimented with an iPhone 3G and an Android Developer Phone 2 (ADP2), and plotted their corresponding frequency response curves in Figure 4(b). Although the frequency response also falls off in the high frequency band it is still able to pick up sounds in a wider range than most human ears. We therefore choose frequencies in this high range. Since our frequency response experiments in Figure 4(b) show noticeable difference among phones beyond 18kHz, we chose both the 16-18kHz range on the ADP2 phone and the 18-20kHz range on the iPhone 3G for our experiments. Energy is uniformly distributed over the entire range.

**Length.** The length of the beep impacts the overall detection time as well as the reliability of recording the beep. Too short a beep is not picked up by the microphone. Too long a beep, will add delay to the system and will be more susceptible to multi-path distortions. We found empirically that a beep length of 400 samples (i.e., 10 ms) represents a good tradeoff.

## 4 DETECTION ALGORITHM

Realizing our approach requires four sub-tasks: *Filtering*, *Signal Detection*, *Relative Ranging* and *Location Classification*. These correspond to the same parts the algorithm shown in Figure 5.

To classify the phone’s location, the specially designed beeps, stored in files, are transmitted to the head-unit and played via the car’s speakers. Just before the beeps are

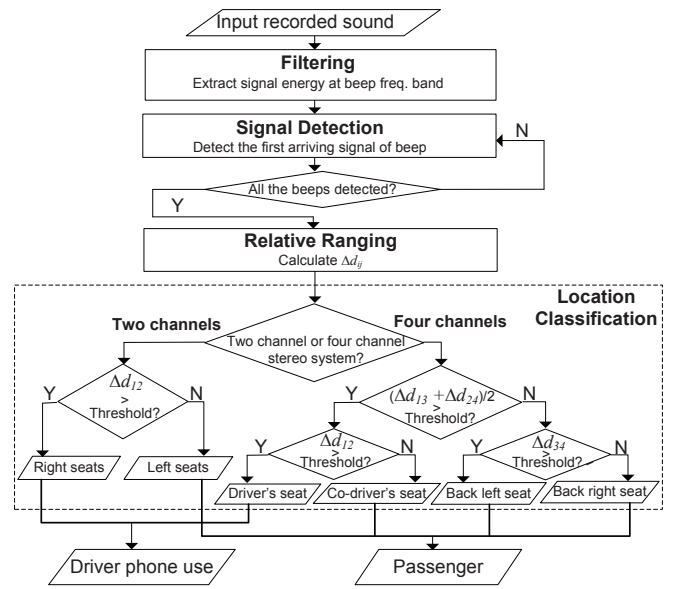


Fig. 5. Flow of the detection algorithm.

transmitted, the microphone is turned on and starts recording. The recorded sound is bandpass filtered around the frequency band of the beep using a short-time Fourier transform (STFT) to remove most background noise. Next, as shown in Figure 5, a *signal detection algorithm* is applied. After each beep sound is detected, its start time is noted and *relative ranging* is performed to obtain the time difference between the two speakers. Given a constant sampling frequency and known speed of sound, the corresponding physical distance is easy to compute. Finally, *location classification* determines the position of the phone in car. Figure 6 shows the walkthrough of the detection system. We next describe the two most important tasks, *beep signal detection* and *ranging and location classification*, in detail.

### 4.1 Detecting Beep Arrival Time

Detecting the beep signal arrival under heavy multipath in-car environments is challenging because the beeps can be distorted due to interference from the multipath components. In particular, the commonly used correlation technique, which detects the point of maximum correlation between a received signal and a known transmitted signal, is susceptible to such distortions [34]. Furthermore, the use of a high frequency beep signal can lead to distortions due to the reduced microphone sensitivity in this range.

For these reasons, we adopt a different approach where we simply detect the first strong signal in our frequency band. This is possible since there is relatively little noise and interference from outside sources in our chosen frequency range. This is known as sequential change-point detection in signal processing. The basic idea is to *identify the first arriving signal* that deviates from the noise after filtering out background noise [13]. Let  $\{X_1, \dots, X_n\}$  be a sequence of recorded audio signal by the mobile phone over  $n$  time point.

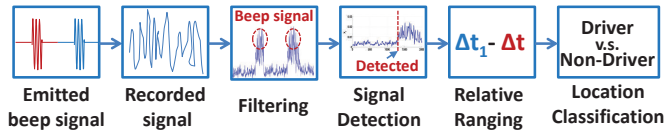


Fig. 6. Walkthrough of the detection system.

Initially, without the beep, the observed signal comes from noise, which follows a distribution with density function  $p_0$ . Later on, at an unknown time  $\tau$ , the distribution changes to density function  $p_1$  due to the transmission of beep signal. Our objective is to identify this time  $\tau$ , and to declare the presence of a beep as quickly as possible to maintain the shortest possible detection delay, which corresponds to ranging accuracy.

To identify  $\tau$ , we can formulate the problem as sequential change-point detection. In particular, at each time point  $t$ , we want to know whether there is a beep signal present and, if so, when the beep signal is present. Since the algorithm runs online, the beep may not yet have occurred. Thus, based on the observed sequence up to time point  $t$   $\{X_1, \dots, X_t\}$ , we distinguish the following two hypotheses and identify  $\tau$ :

$$\mathcal{H}_0 : X_i \text{ follows } p_0, \quad i = 1, \dots, t,$$

$$\mathcal{H}_1 : \begin{cases} X_i \text{ follows } p_0, & i = 1, \dots, \tau - 1, \\ X_i \text{ follows } p_1, & i = \tau, \dots, t. \end{cases}$$

If  $\mathcal{H}_0$  is true, the algorithm repeats once more data samples are available. If the observed signal sequence  $\{X_1, \dots, X_t\}$  includes one beep sound recorded by the microphone, the procedure will reject  $\mathcal{H}_0$  with the stopping time  $t_d$ , at which the presence of the beep signal is declared. A *false alarm* is raised whenever the detection is declared before the change occurs, i.e., when  $t_d < \tau$ . If  $t_d \geq \tau$ , then  $(t_d - \tau)$  is the *detection delay*, which represents the ranging accuracy.

Sequential change-point detection requires that the signal distribution for both noise and the beep is known. This is difficult because the distribution of the beep signal frequently changes due to multipath distortions. Thus, rather than trying to estimate this distribution, we use the cumulative sum of difference to the averaged noise level. This allows first arriving signal detection without knowing the distribution of the first arriving signal. Suppose the cell phone estimates the mean value  $\mu$  of noise starting at time  $t_0$  until  $t_1$ , which is the time that the phone starts transmitting the beep. We want to detect the first arriving signal as the signal that significantly deviates from the noise in the absence of the distribution of the first arriving signal. Therefore, the likelihood that the observed signal  $X_i$  is from the beep can be approximated as

$$l(X_i) = (X_i - \mu), \quad (2)$$

given that the recorded beep signal is stronger than the noise. The likelihood  $l(X_i)$  shows a negative drift if the observed signal  $X_i$  is smaller than the mean value of the noise, and a positive drift after the presence of the beep, i.e.,  $X_i$  stronger

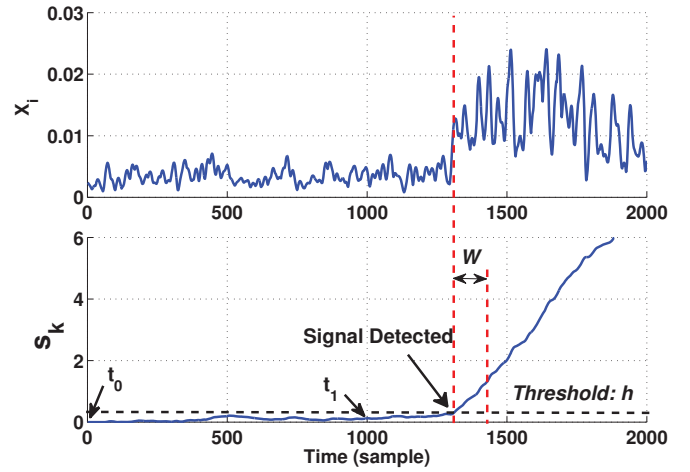


Fig. 7. An illustration of detecting the first arriving signal using our system prototype. The upper plot shows the observed signal energy along time series and the lower plot shows the cumulative sum of the observed signal and the detection results.

than the noise. The stopping time for detecting the presence of the beep is given by

$$t_d = \inf\{k | s_k \geq h\}, \quad \text{satisfy } s_m \geq h, m = k, \dots, k + W, \quad (3)$$

where  $h$  is the threshold,  $W$  is the robust window used to reduce the false alarm, and  $s_k$  is the metric for the observed signal sequence  $\{X_1, \dots, X_k\}$ , which can be calculated recursively:

$$s_k = \max\{s_{k-1} + l(X_k), 0\}, \quad (4)$$

with  $s_0 = 0$ .

Figure 7 shows an illustration of the first arriving signal detection by using our system prototype. The upper plot shows the observed signal energy along time series and the lower plot shows the cumulated sum of the observed signal.

Our approach of cumulative sum of difference to the averaged noise level is inspired by Page's cumulative sum (CUSUM) procedure [33], which was shown to minimize average detection delay when both  $p_0$  and  $p_1$  are known *a priori*. Although the CUSUM algorithm can be generalized as GLR (generalize likelihood ratio) [30] without knowing the distribution of signal, the high computational complexity and large detection delay of GLR make it infeasible in our system design, which requires efficient computation on mobile devices and high accuracy.

**Prototype Considerations.** In our system implementation, we empirically set the threshold as the mean value of  $s_k$  plus three standard deviations of  $s_k$  when  $k$  belongs to  $t_0$  to  $t_1$  (i.e., 99.7% confidence level of noise). The window  $W$  is used to filter out outliers in the cumulative sum sequence due to any sudden change of the noise. We set  $W = 40$  in our implementation. At the time point that the phone starts to emit the beep sound, our algorithm starts to process the recorded signal sequences. Once the first arriving signal of

the first beep is detected, we shifts the precessing window to the approximated time point of the next beep since we know the fixed interval between two adjacent beeps.

## 4.2 Ranging and Location Classification

After the first arriving time of the beeps are detected, the system first calculates the time difference  $\Delta T_{ij} = \frac{S_{ij}}{f}$  between the speakers. Here  $S_{ij}$  is the number of samples that the beeps were apart and  $f$  is the sampling frequency (typically 44.1kHz). In a two-channel system,  $i$  and  $j$  are simply the left speaker (speaker 1) and the right speaker (speaker 2).

The distance difference from the phone to two speakers can be calculated as:

$$\Delta d_{ij} = c \cdot \Delta T_{ij}, \quad (5)$$

where  $i$  and  $j$  represent the  $i^{th}$  and  $j^{th}$  speakers in Figure 1 and  $c$  is the speed of the sound.

In a two-channel system, the driver-side can then be identified based on the following condition.

$$\Delta d_{12} > TH_{lr}, \quad (6)$$

Here,  $TH_{lr}$  is a threshold that could be chosen as zero, but since drivers are often more likely to place their phone in the car's center console, it often makes sense to assign a negative value of about 5cm.

In a four-channel system, we can first use two pairs of left speakers and right speakers to classify whether the mobile phone is located in the front or back seats. Given a threshold  $TH_{fb}$ , the mobile phone is classified as in the front seat if

$$(\Delta d_{13} + \Delta d_{24})/2 > TH_{fb}, \quad (7)$$

where  $\Delta d_{13}$  represent the distance difference from two left side speakers and  $\Delta d_{24}$  is the distance difference from two right speakers. If the phone is in the front, it will then use the same condition as before to discriminate driver side and passenger side. If the system is in the back, it would use  $\Delta d_{34}$  instead, since the rear speakers are closer.

In order to improve the reliability of the measured distance difference, the median distance difference measured from multiple runs is applied. In our implementation, we used four runs, which is robust up to two outliers. Therefore, there is four beeps in each channel and it takes one second to emit all beeps for two-channel and about two seconds for four-channel systems.

## 5 EVALUATION

We have experimented with this technique in two different cars and on two different phones to evaluate driver-passenger classification accuracy. We also studied how our algorithm compares to correlation-based methods and measured the runtime on the Android Developer Phone 2 platform. The following subsections detail the methodology and results.

### 5.1 Experimental Methodology

**Phones and Cars.** We conducted our experiments with the Android Developer Phone 2 (Phone I) and the iPhone 3G (Phone II). Both phones have a Bluetooth radio and support 16-bit 44.1 kHz sampling from the microphone. The iPhone 3G is equipped with a 256MB RAM and a 600 MHz ARM Cortex A8 processor, while the ADP2 equipped with 192 MB RAM and the slower 528MHz MSM7200A processor.

We created four beep audio files in MATLAB for the two phones, each with 4 beeps for each channel in car's stereo system. Two of these are for two channel operation (one for each phone) and the other two files are designed for four channel operation. To create these files, we first generated a single beep by creating uniformly distributed white noise and then bandpass filtered it to the 16-18kHz for Phone I and 18-20kHz band for Phone II. We then replicated this beep 4 times with a fixed interval of 5,000 samples between each beep so as to avoid interference from two adjacent beeps. This 4 beep sequence is then stored first in the left channel of the stereo file and after a 10,000 sample gap repeated on the right channel of the file.

The accuracy results presented here were obtained while transmitting this audio file from a laptop to the car's head unit via Bluetooth Advanced Audio Distribution and recording it back on one of the phones using an audio recorder application for offline analysis. We subsequently also created an Android prototype implementation that simultaneously streams A2DP audio and records audio from the microphone to confirm feasibility.

We experimented in a Honda Civic Si Coupe (Car I) and an Acura sedan (Car II). Both cars have two front speakers located at two front doors' lower front sides, and two rear speakers in the rear deck. The interior dimensions of Car I are about 175cm (width) by 183cm (length) and about 185cm by 203cm for Car II.

Since both cars are equipped with the two-channel stereo system, the four channel sound system is simulated by using the headunit's fader system. Specifically, we encode a two channel beep sound and play the two channel beep sound first at two front speakers while muting the rear speakers, we then play the two channel beep sound at two rear speakers while muting the front speakers.

**Experimental Scenarios.** We conducted experiments where we placed a phone in various positions that we believe are commonly used. We also varied the number of passengers and the amount and type of background noise. Due to safety reasons (experiments require manual intervention and changing phone positions can be difficult), we restricted the number of experiments while driving and conducted more exhaustive testing in a stationary setting.

We organized our experiments in three representative scenarios:

*Phone I, Car I:* This set of experiments uses the Android Developer Phone 2 in the Honda Civic while stationary. Background noises stem from conversation and an idling



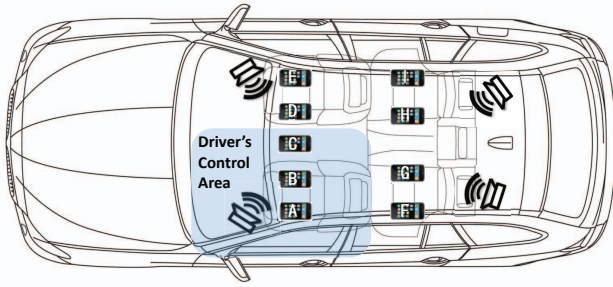


Fig. 8. Illustration of testing positions in *Phone I Car I* scenario and driver's control area.

engine. As illustrated in Figure 8, we placed the phone in nine different locations: Driver's left pant pocket (A), driver's right pant pocket (B), a cupholder on the center console (C), front passenger's left pant pocket (D), front passenger's right pant pocket (E), left rear passenger's left pocket, left rear passenger's right pocket (G), right rear passenger's left pocket (H) and right rear passengers right pocket (I). When the phone was in the 5 front positions, there are two cases: (1) only driver and front passenger were in the car; and (2) driver, front passenger, and left rear passenger were in the car. When the phone was located in the rear positions, the additional rear passenger always occupied the car.

*Phone II, Car II:* These experiments deploy the iPhone 3G in the Acura, again stationary but this time without background noise. We use three occupancy variants: only driver is in the car; driver and co-driver are in the car; driver, co-driver and one passenger are in the car. (1) There are two positions tested in the first case: driver door's handle and cup holder; (2) four positions in the second case: the same two positions as before, plus co-driver's left pant pocket and co-driver door's handle; and (3) six positions in the third case: all four positions from the second case, plus passenger holding the phone at rear left seat and rear left door's handle.

*Highway Driving:* ADP2 is deployed in Car I. The car is driving on highway at the speed of 60MPH with music playing in the car. The four positions tested are: driver's left pant pocket, cup holder, co-driver holding the phone, and co-driver's right pant pocket. We also repeat this set of experiments with both front windows open, as a worst case background noise scenario.

**Metrics.** One of our key evaluation questions is how accurately our technique distinguishes phones that likely are used by the driver from phones likely used by passengers. In this evaluation, we consider all phones in positions that are within easy reach of the driver as phones used by the driver. This includes the driver's left and right pockets, the driver door's handle, and the cup holder. We have marked this as the driver's control area in Fig. 8. We consider all other positions passenger phone positions. To evaluate the performance of our system, we therefore define the following metrics:

*Classification Accuracy (Accuracy).* Classification accuracy is defined as the percentage of the trials that were correctly

Scenario	Threshold	DR	FPR	Accuracy
Two-channel stereo system, phone at front seats				
Highway	Un-calibrated	99%	4%	97.5%
	Calibrated	100%	4%	98%
Phone I, Car I	Un-calibrated	94%	3%	95%
	Calibrated	98%	7%	96%
Phone II, Car II	Un-calibrated	95%	24%	87%
	Calibrated	91%	5%	92%
Four-channel stereo system, phone at all seats				
Phone I, Car I	Un-calibrated	94%	3%	95%
	Calibrated	94%	2%	96%
Phone II, Car II	Un-calibrated	84%	16%	84%
	Calibrated	91%	3%	94%

TABLE 1

Detection rate (DR), false positive rate (FPR) and accuracy when determining the driver phone use under various scenarios.

classified as driver phone use or correctly classified as passenger phone use.

*Detection Rate (DR), False Positive Rate (FPR).* Detection rate is defined as the percentage of trials within the driver control area that are classified as driver phone use. False positive rate is defined as the percentage of passenger phone use that are classified as driver phone use.

*Measurement Error.* Measurement error is defined as the difference between the measured distance difference (i.e.,  $\Delta d_{ij}$ ) and the true distance difference. This metric directly evaluates the performance of relative ranging in our algorithm.

## 5.2 Classification of Driver Phone Use

### 5.2.1 Driver vs. Passenger Phone Use

Table 1 shows the detection rate, false positive rate and accuracy when determining driver phone use using the two channel stereo system. Note that since the 2-channel system cannot distinguish the driver-side passenger seat from the driver seat, we have only tested front phone positions for this experiment. To test the robustness of our system to different types of cars, we distinguish between the *Un-calibrated* system, which uses a default threshold, and the *Calibrated* system, wherein the threshold is determined by taking into the consideration of car's dimensions and speaker layout.

We set the *Un-calibrated* default threshold  $TH_{lr} = -5cm$  for both Car I and Car II. We shift the  $TH_{lr}$  from 0cm to -5cm, because we define the cup holder position within the driver's control area. Recall, that the cup holder is equidistant from both speakers and results in distance difference near zero. For *Calibrated* threshold, it is  $TH_{lr} = -7cm$  and  $TH_{lr} = -2cm$  in Car I and Car II settings respectively.

**Two-channel stereo system.** From Table 1, the important observation in the Highway scenario is that our system can achieve close to 100% detection rate (with a 4% false positive rate), which results in about 98% accuracy, suggesting our system is highly effective in detecting driver phone use while driving. The detection rate for both *Un-calibrated* and *Calibrated* is more than 90% while the false positive rate is around



5% except for Car II setting. This indicates the effectiveness of our detection algorithm. The high false positive rate of Car II setting can be reduced through calibration of the threshold. Although the detection rate is reduced when reducing the false positive rate for Car II, the overall detection accuracy is improved. Further, we observed that the results of Phone I are slightly better than those of Phone II. The difference between the results mainly comes from the different beep files that we used. Specifically, 16-18kHz range has been chosen for Phone I, whereas the 18-20kHz range was chosen for Phone II during our experiments. And the frequency response at around 16kHz for Phone I is comparable or better than that of at 18kHz for Phone II. The energy at higher frequency degrades more easily than that of lower frequency range due to reflection, refraction and path loss, especially in a confined in-car environment where there is no line of sight. We found that using the beep sound at lower frequency band can improve the accuracy of relative ranging; however, beep signals located at lower frequency band will be picked up by human ears easier. Overall, the experimental results show that our system is robust to different types of cars and can provide reasonable accuracy without calibration (although calibration still helps).

Recall that in this experiment we only considered front phone positions since the two-channel stereo system can only distinguish between driver-side and passenger-side positions. With phone positions on the back seats, particularly the driver-side rear passenger seat, the detection accuracy will be degraded, although the detection rate remains the same. Real life accuracy will depend on where drivers place their phones in the car and how often passengers use their phone from other seats. Unfortunately, we were unable to gather this information. We did however find information on passenger seat occupancy in the FARS 2009 database [7]. Encouragingly it shows that the two front seats are the most frequently occupied seats. In particular, according to FARS 2009 database, 83.5% of vehicles are only occupied by driver and possibly one front passenger, whereas only about 16.5% of trips occur with back seat passengers. More specifically, only 8.7% of the trips include a passenger sitting behind driver seat—the situation that would increase our false positive rate.

If we weigh the phone locations by these probabilities, the false positive rate only increases by about 8.7% even with the two channel system. The overall accuracy of detecting driver phone use remains at about 90% for all three experimental scenarios in our system. This is very encouraging as it indicates our system can successfully produce high detection accuracy even with the systems limited to two-channel stereo in today’s cars.

**Four-channel stereo system.** We now consider the four-channel system to study how accuracy could be improved when surround sound is available. The results of using four-channel system under both *Un-calibrated* and *Calibrated* thresholds is shown in Tables 1. The *un-calibrated* thresholds are  $TH_{fb} = 0cm$  and  $TH_{lr} = -5cm$  for both Car I and Car II. The *calibrated* thresholds are  $TH_{fb} = 15cm$  and

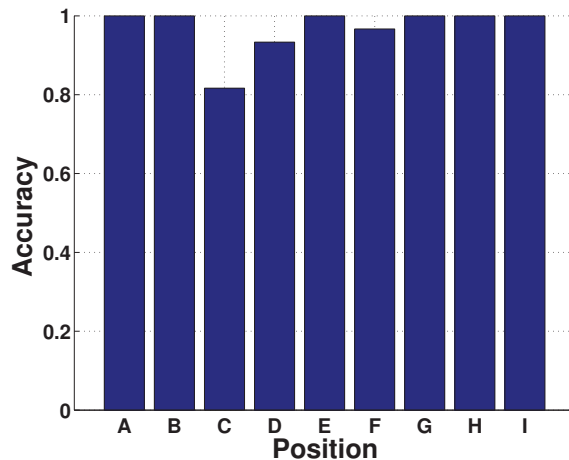


Fig. 9. Accuracy of detecting driver phone use at each position in Car I (i.e., positions plotted in Figure 8) under calibrated thresholds with four-channel stereo system.

$TH_{lr} = -5cm$  for Car I, whereas they are  $TH_{fb} = -24cm$  and  $TH_{lr} = -2cm$  for Car II. We found that with the *calibrated* thresholds, the detection rate is above 90% and the accuracy is around 95% for both settings. This shows that the four-channel system can improve the detection performance, compared to that of the two-channel stereo system. In addition, the performance under *un-calibrated* thresholds is similar to that under *calibrated* thresholds for Car I setting, however, it is much worse than that of *calibrated* thresholds for Car II settings. This suggests that calibration is more important for distinguishing the rear area, because the seat locations vary more in the front-back dimension across cars (and due to manual seat adjustment).

### 5.2.2 Position Accuracy and Seat Classification

We next evaluate our algorithm accuracy at different positions and seats within the car. Figure 9 shows the accuracy of detecting driver phone use for different positions in Car I setting under calibrated thresholds. We observed that we can correctly classify all the trials at the positions A,B,E,G,H,I as denoted in Figure 8, whereas the detection accuracy decreases to 93% for position D (i.e., co-driver’s left pocket) and 82% for position C (i.e., cup holder). Additionally, we tested doors’

	Driver	Co-driver	Rear Left	Rear Right
Phone I, Car I				
Un-calibrated	95%	95%	99%	99%
Calibrated	96%	95%	99%	99%
Phone II, Car II				
Un-calibrated	84%	88%	94%	N/A
Calibrated	94%	94%	98%	N/A

TABLE 2  
 Accuracy of determining the phone at each seat with four-channel stereo system.

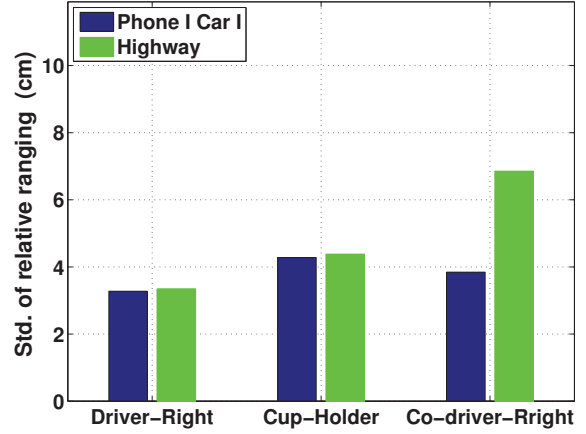
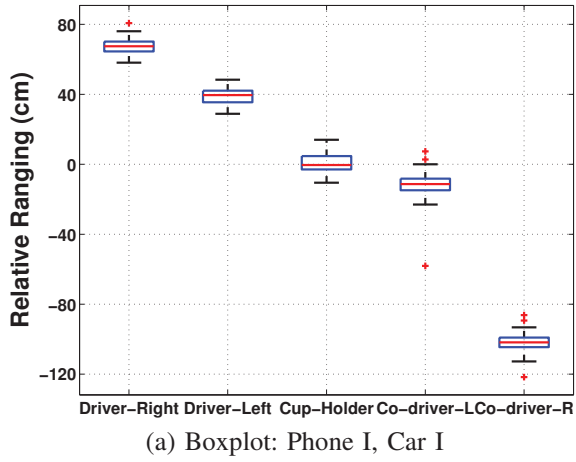


Fig. 11. Stability study of relative ranging between high-way driving and stationary scenarios.

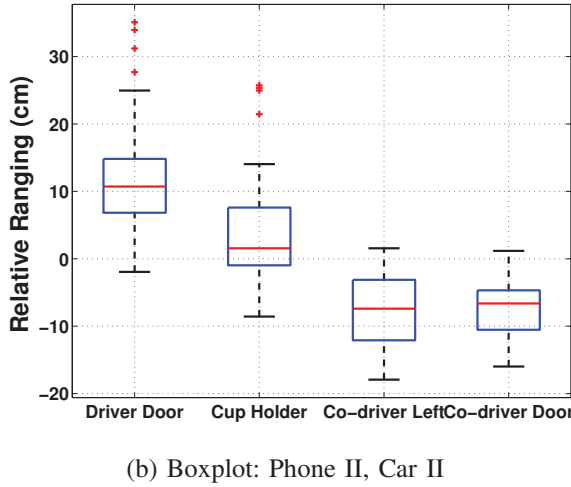


Fig. 10. Boxplot of the measured  $\Delta d_{12}$  for all front positions in two-channel stereo system.

handle positions in Car II setting and found the accuracy for driver’s door handle is 99%, and 97% for co-driver’s door handle. These results provide a better understanding of our algorithm’s performance at different positions in car.

We further derive seat classification results. Table 2 shows the accuracy when determining the phone at each seat under *Un-calibrated* and *Calibrated* thresholds using the four-channel stereo system. We found that the accuracy of the back seats is much higher than that of front seats. Because there is a cup holder position tested in the front. It is hard to classify the cup holder and co-driver’s left position since they are physically close to each other.

### 5.2.3 Left vs. Right Classification

Figure 10 illustrates the boxplot of the measured  $\Delta d_{12}$  at different tested positions. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points. We note that the scale of y-axis in Figure 10 (a) is different from that in Figure 10 (b). We found that these boxes are clearly separated from each other showing that we obtained

different relative ranging values at different positions. And these positions can be perfectly identified by examining the measured values from relative ranging except Cup holder and Co-driver’s left positions for both Car I and Car II settings. By comparing Figure 10 (a) and (b), we found that the relative ranging results of driver’s and co-driver’s doors are much smaller than that of driver’s left and co-driver’s right pockets, which is conflict with the groundtruth. This is mainly because the shortest path that the signal travels to reach the phone is significantly longer than the actual distance between the phone and the nearby speaker when putting the phone at door’s handle since there is no direct path between the phone and speaker, i.e., the nearby speaker is facing the opposite side of the phone.

To compare the stability of our ranging results under the Highway driving scenario to the stationary one, we plotted the standard deviation of relative ranging results at different positions in Figure 11. We observed the encouraging results that our algorithm produces the similar stability of detection when car is driving on highway to that when car is parked. We note that at the co-driver’s right position (i.e., Co-driver-R), the relative ranging results of Highway driving scenario still achieves 7cm of standard deviation, although it is not as stable as that of Phone I Car I setting due to the movement of the co-driver’s body caused by moving car.

### 5.2.4 Front vs. Back Classification

In front and back classification, the detection rate is defined as the percentage of the trials on front seats that are classified as front seats. False positive rate is defined as the percentage of back seat trials that are classified as front seats. Figure 12 plotted Receiver Operating Curve (ROC) of detecting the phone at front seats in Car I setting. We found that our system achieved over 98% detection rate with less than 2% false positive rate. These results demonstrate that it is relatively easier to classify front and back seats than that of left and right seats since the distance between the front and back seats

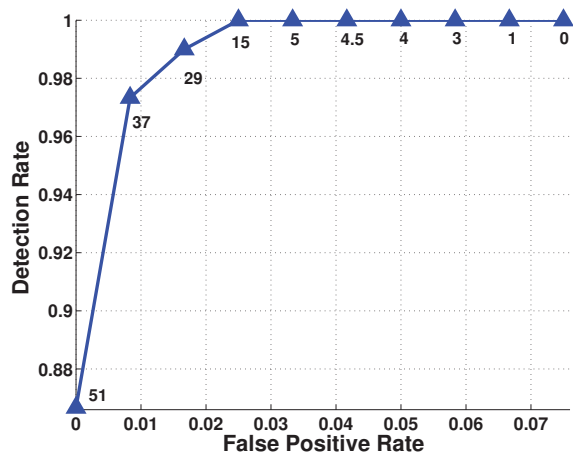


Fig. 12. ROC curve of detecting the phone at front seats for Phone I, Car I scenario.

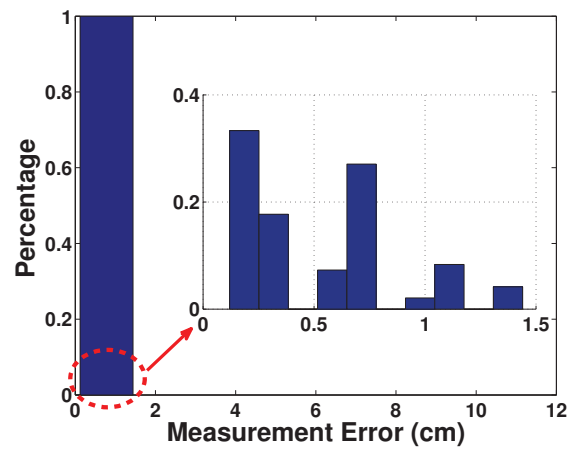
is relatively larger. Our algorithm can perfectly classify front seats and back seats with only a few exceptions.

### 5.3 Results of Relative Ranging

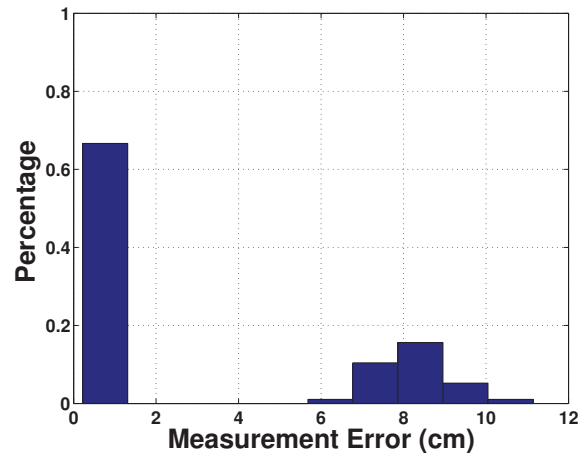
We next present the measurement error of our relative ranging mechanism and compare it to the previous work using chirp signal and correlation signal detection method with multipath mitigation mechanism, which achieved high accuracy for acoustic ranging using off-the-shelf mobile devices [34].

**Correlation-Based Method.** To be resistant to ambient noise, the correlation method uses the chirp signal as beep sound. To perform signal detection, this method correlates the chirp sound with the recorded signal using  $L_2$ -norm cross-correlation, and picks the time point when the correlation value is the maximum as the time signal detected. To mitigate the multipath, instead of using the maximum correlation value, the earliest sharp peak in the correlation values is suggested as the signal detected time [34]. We refer this approach as correlation method with mitigation mechanism.

**Strategy for Comparison.** To investigate the effect of multipath in an enclosed in-car environment and the resistance of beep signals to background noise, we designed experiments by putting ADP2 in car I at three different positions with line-of-sight (LoS) to two front speakers. At each position, we calculated 32 measurement errors to obtain a statistical result. To evaluate multipath effects, we simply measured the TDOA values of our method and correlation method with mitigation mechanism. To test the robustness under background noise, we played music in car at different sound pressure levels, which are 60dB and 80dB, representing moderate noise (e.g., people talking in car) and heavy noise (e.g., traffic on a busy road), respectively. The chirp sound used for correlation method is taken from previous work [34], which is a 50 millisecond length of 2-6kHz linear chirp signal at 80dB SPL and is proven to be a good compromise between multipath effects suppressing and noise resistance. We also found that chirp signals under high frequency band (e.g. beyond 15kHz) does



(a) Our method



(b) Correlation method

Fig. 13. Measurement error of relative ranging. Our method has all the measurement errors within 2cm, whereas more than 30% of the measurement errors of correlation-based method are larger than 2cm.

not perform as well as those under low frequency band. The main reason is that the frequency response of the phone's microphone for the chirp sound at the frequency range 2-6kHz is the best. Once the chirp frequency went very high, the recorded chirp signal suffered large distortion, making the correlation between the original chirp signal and the recorded one very weak.

#### 5.3.1 Impact of Multipath

Figure 13 shows the histogram of measurement error in car for both our method and correlation method with multipath mitigation mechanism. We observed that all the measurement errors of our method are within 2cm, whereas more than 30% of the measurement errors of correlation-based method are larger than 2cm. Specifically, by examining the zoomed in histogram in Figure 13(a), we found that our method has most of the cases with measurement errors within 1cm (i.e.,



1 sample), whereas about 30% cases at around 8cm (i.e., 10 samples) for correlation-based method. The results show that our algorithm outperforms the correlation-based method in mitigating multipath effect in an in-car environment since our signal detection method detects the first arriving signal, not affected by the subsequent arriving signal through different paths.

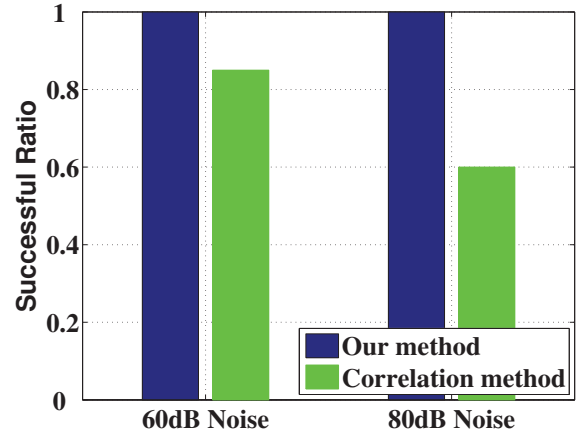
### 5.3.2 Impact of Background Noise

Figure 14 analyzes the impact of background noise. Figure 14(a) illustrates the comparison of successful ratio defined as the percentage of measurement errors within 10cm for two methods. Our method successfully achieves within 10cm measurement error for all the trials under both moderate and heavy noises, whereas the correlation-based method with multipath mitigation scheme achieves 85% for moderate noise and 60% for heavy noise over all the trials, respectively. Figure 14(b) shows the measurement error CDF of our method. The median error of our method is only 0.66cm under moderate noise and it is 1.05cm under heavy noise. We also tested both methods in a room environment (with people chatting at the background) using computer speakers, and found both methods exhibit comparable performance.

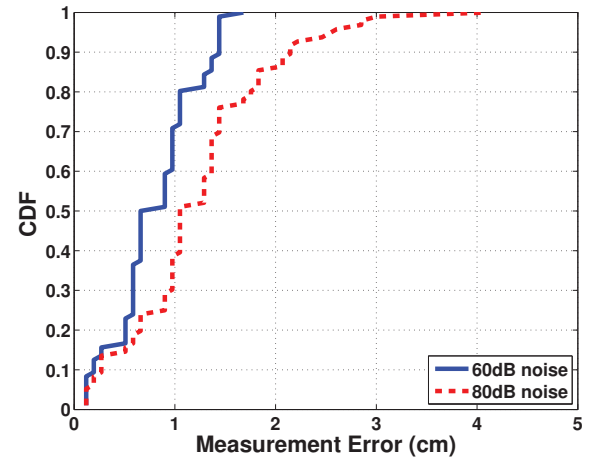
### 5.4 Computational Complexity

Our algorithm complexity is bounded by the length of the audio signal needed for analysis. In order to keep the resolution at one sample and perform noise filtering, we extract the energy within each  $m$  samples moving window at the targeted frequency band (i.e., 16-18kHz for ADP2 and 18-20kHz for iPhone) using a short-time Fourier transform (STFT). Given  $n$  recorded samples and a moving window size  $m$ , the computational cost for energy extraction at the targeted frequency band is  $O(nm \log m)$ . In our implementation, we set the window size as 32 samples. After filtering, the computational cost of signal detection is  $O(n)$ .

**Run Time.** Since the STFT is the most expensive processing step, our implementation limits processing to a 1000 sample window that the beep signal is estimated to fall into (it is chosen wide enough for worst-case propagation delays in the car environment). We then detect the exact time point of the first arriving signal within these 1000 samples. Once the time point of the first arriving signal is determined, our algorithm shifts the preprocessing window to the next beep sound since we know the fixed interval between two adjacent beeps. Thus, the computational time for one beep is approximately equivalent to process 1000 samples. We implemented this step on the ADP2 with JTransforms library for STFT and measured the average processing time of our detection algorithm as about 0.5 second for the two-channel system and about 1 second for the four-channel system. The windowing implementation has significantly reduced the processing time of our algorithm and further optimizations are likely possible.



(a) Successful ratio



(b) Error CDF of our method

Fig. 14. Impact of background noise. Successful ratio is defined as the percentage of measurement errors within 10cm. Our method has 100% successful ratio for both 60dB and 80dB noises. Correlation-based method has 85% for 60dB and 60% for 80dB, respectively.

## 6 DISCUSSION

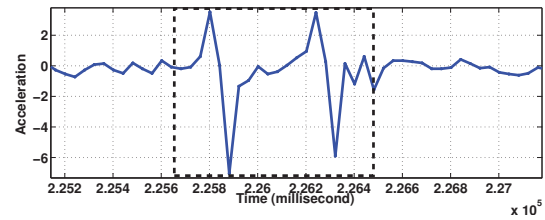
**Bluetooth issues.** We have assumed that a Bluetooth connection is already established. We believe that this is a reasonable assumption for people who (usually) drive a given car. People are likely to pair their phone with the in-car Bluetooth system and after the first pairing, connections are usually automatically established when the phone comes in range of the car. It is not common practice, however, for occasional passengers who are never drivers. There seem to be several possible approaches to address this issue: (i) having phones listen for beeps transmitted by other phones at regular known times, (ii) standardizing a Bluetooth profile for such beep transmission which allows auto-pairing, (iii) building the beep transmissions into car audio systems, so that phones only need to listen. The Bluetooth connection could also be in use for playing music

using the A2DP profile. In this case, the phone should be able to insert the beeps into the music stream.

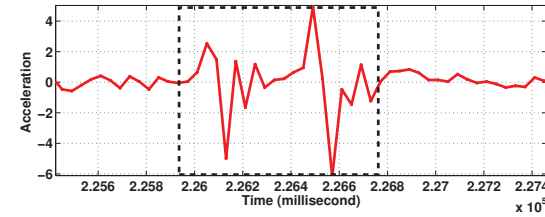
**Driver’s Seat v.s. Driver-side Rear Seat.** Our acoustic ranging based detection system is practical with current handsfree and A2DP Bluetooth profiles which provide for stereo audio. This is because the left-right classification can be achieved with stereo audio, and this covers the majority of scenarios. For the case that the phone is located in the driver-side rear passenger seat (less than 9% of vehicle trips according to FARS), one solution is to explore using smartphone’s built-in sensor to further determine the phone’s location after left-right classification. For example, the smartphone’s built-in accelerometer has been used in Trafficsense [29] and the pothole patrol [19] to identify speed bumps, potholes and other severe road surface anomalies. Indeed, the accelerometer readings on smartphones can also be utilized to distinguish the position of the phone is at driver’s seat from driver-side rear passenger seat when the car passing over the speed bumps or potholes. When passing over a speed bump, the front wheels will hit the bump first and then the rear wheels. Since the driver’s seat is closer to the front wheels whereas the driver-side rear seat is closer to the back wheels, passing over the bump will produce different sensor reading patterns on the phone located at driver’s seat from those in the driver-side rear seat.

Figure 15 shows the accelerometer readings at vertical axis by passing over the speed bump (i.e., highlighted by dashed rectangle) when the phones located at driver’s seat and at driver-side rear passenger seat, respectively. We observed two peaks for both phones (the first peak is produced when the front wheel passing over the bump and the second one is produced by the rear wheels). Furthermore, the first peak is slightly stronger than the second peak when the phone is located at driver’s seat, whereas the strength order of the peaks is reversed when the phone is placed at the driver-side rear seat. Although this pattern cannot be observed throughout all the trials when passing over the bumps, we found that the ratio between the first peak and the second peak when the phone locating at driver’s seat is larger than that of the phone at driver-side rear seat for majority of the cases. Therefore, by examining the ratio of these two peaks, we should be able to determine the phone is more likely, if not absolutely, located at the driver’s seat or the rear seat. Additionally, by accumulating the sensor readings when passing over a sequence of bumps, we can improve the probability of correctly identifying whether the phone is located at driver’s or rear seats.

Figure 16 depicts the mean and standard deviation of the ratio between the first and second peaks when combining the sensor readings from several bumps. Totally, we collected the accelerometer readings for passing over 40 speed bumps at various speeds ranging from 5 to 20 mph. Given the number of combined bumps, the ratio is computed as the ratio of the summation of all the first peaks to the summation of all the second peaks. From Figure 16, we observed that the mean value of the ratio is very stable when the number of the



(a) Driver’s seat



(b) Driver-side rear passenger seat

Fig. 15. Accelerometer readings when passing over the speed bump.

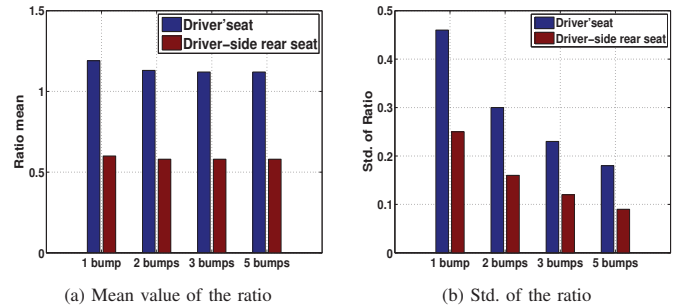


Fig. 16. Mean and standard deviation of the ratio when combining several bumps.

combined bumps varies from 1 to 5. Specifically, the ratio mean when the phone is placed at the driver’s seat is about 1.1, which is much larger than that of at the rear seat (i.e., 0.6). Moreover, the standard deviation of the ratio decreases when increasing the number of combined bumps. This indicates that it is easier to distinguish the phone’s position in driver’s seat from rear seat when the number of combined bumps increases, given there are about 0.5 difference in the mean value of the ratio. Figure 17 presents the receiver operating characteristic (ROC) curves of detecting the phone at driver’s seat under different number of combined bumps ranging from 1 to 5. We found that the curves shift to the left indicating higher detection rate when the number of combined bumps increases. Specifically, when setting the threshold at around 0.75, we can achieve 83% accuracy using the accelerometer readings from one bump, 90% accuracy when combining two bumps, 94% accuracy for combining three bumps, and 98% accuracy when combining the readings from five bumps.

**Limitations.** Even with access to four audio channels, the system might not accurately distinguish driver and passenger

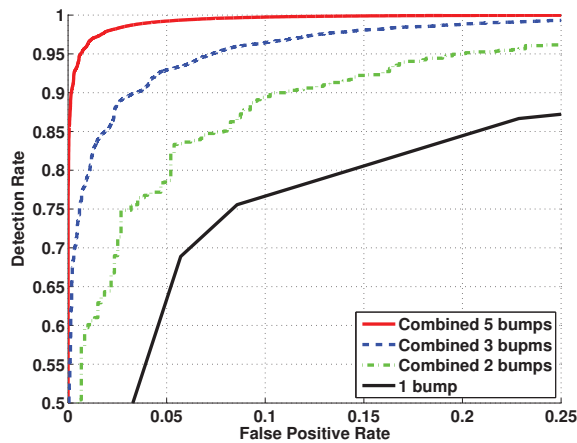


Fig. 17. ROC curves of detecting the phone at driver's seat under different number of combined bumps.

for several reasons. First, if the phone is placed under a heavy winter coat or inside a full bag, the beep sounds might be too muffled to be accurately detected. Second, if the driver places the phone on an empty passenger seat, the system might correctly detect the seat, but an incoming call could still distract the driver. Still, we believe the accuracy of this system will be a significant improvement over current systems that only seek to determine whether the phone is used inside a vehicle. We have also left buses, trains, and other vehicles outside the scope of this work—phones could identify such vehicles by comparing GPS traces with known routes. In these vehicles it is also more cost-efficient to add a device to the driver cabin. A more fundamental limitation is the probabilistic nature of our approach. We can not place hard boundaries on accuracy because of many environmental unknowns, some of which are described above. This means that our approach is less suitable for applications depending on perfect accuracy. Rather, they will serve to enhance the user experience and nudge drivers towards safer behavior. Finally, this system is not intended for continuously tracking phone position, since its energy consumption would be quite substantial. Rather, we envision that this technique would be sporadically triggered, for example, by an incoming phone call or when entering the vehicle (upon Bluetooth connect).

**Applications.** In this paper, we have concentrated only on distinguishing drivers and passengers, a complete system should also include cellphone-based speed detection techniques to determine whether the car is driving. As alluded to in the introduction, there are several applications of this driver phone use detection system: (i) it could automatically bring up less distracting driver user interfaces; (ii) the beeps might only be transmitted when a call or text is coming in, to determine whether the phone should ring or whether the call should go to voicemail; (iii) the 'driving' status might be displayed in friends dialer applications to discourage them from calling. Integration with vehicle controls, is another dimension that could be explored. Perhaps a driver chatting on the phone

should increase the responsiveness of a vehicle's braking system, since this driver is more likely to break late. It could also affect the level of intrusiveness of lane-departure warning and other driver assist systems. Finally, the information could be used to lock the phone to prevent the driver from calling—we note, however, that the system is not secure against a user intentionally trying to fool it. Thus, it is less suitable for such enforcement actions.

## 7 CONCLUSIONS

We developed a driver mobile phone use detection system that requires only software changes on smartphones. It achieves this by leveraging the existing infrastructure of car speakers for ranging via Bluetooth. The proposed system detects driver phone use by estimating the range between the phone and car's speakers. To estimate range, we developed an acoustic based relative ranging technique in which the phone plays and records a specially designed acoustic signal through car's speakers. Our specially designed acoustic signal is unobtrusive as well as robust to background noise when driving. Our algorithm achieves high accuracy under heavy multipath in-car environments by using sequential change-point detection to identify the first arriving signal.

We further demonstrated the viability of distinguishing between driver's and passenger's phone use working within the confines of the existing handsfree audio infrastructure. Our prototype showed the generality of our approach, as we applied it to two different phone types and two different cars under various scenarios. Our system can achieve over 90% of detection rates as well as accuracy, with low false positive rate.

## 8 ACKNOWLEDGMENTS

We thank Adrian Perrig for suggesting the use of accelerometer data. This work is supported in part by the National Science Foundation Grants CNS-0954020, CNS-1016303, CNS-1040735 and CNS-0845896.

## REFERENCES

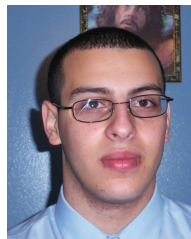
- [1] <http://tinyurl.com/an5y1>.
- [2] <http://www.bang-olufsen.com/sound-system-audi-r8>.
- [3] Drivesmart plus. <http://tinyurl.com/4v7oygy>.
- [4] Governors highway safety association. <http://tinyurl.com/39hfe7>.
- [5] Guardian angel vehicle platform. <http://www.trinitynoble.com/>.
- [6] Key2safedriving app. <http://www.key2safedriving.com/>.
- [7] National highway traffic safety administration: Fatality analysis reporting system. <http://tinyurl.com/24h2t7>.
- [8] The new york times. <http://tinyurl.com/4etj1ja>.
- [9] Reuters. <http://tinyurl.com/4uoxefh>.
- [10] Textecution. <http://www.textecution.com/>.
- [11] txtblocker. <http://www.txtblocker.com/>.
- [12] B. K. Alexej Swerdlow, Timo Machmer and K. Kroschel. Speaker position estimation in vehicles by means of acoustic analysis. In *34th Deutsche Jahrestagung für Akustik (DAGA '07)*, Stuttgart, Germany, 2007. SpringerVerlag.
- [13] M. Basseville and I. Nikiforov. *Detection of abrupt changes: theory and application*, volume 10. Prentice Hall, 1993.
- [14] L. Bergasa, J. Nuevo, M. Sotelo, R. Barea, and M. Lopez. Real-time system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems*, 7(1):63–77, 2006.



- [15] G. Borriello, A. Liu, T. Offer, C. Palistrant, and R. Sharp. Walrus: wireless acoustic location with room-level resolution using ultrasound. In *Proceedings of the 3rd international Conference on Mobile Systems, Applications, and Services (MobiSys '05)*, pages 191–203, New York, NY, 2005.
- [16] J. Caird, C. Willness, P. Steel, and C. Scialfa. A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention*, 40(4):1282–1293, 2008.
- [17] G. Cerrato. Automotive Sound Quality—Powertrain, Road and Wind Noise. *Sound & vibration*, pages 16–24, 2009.
- [18] G. Chandrasekaran, T. Vu, A. Varshavsky, M. Gruteser, R. P. Martin, J. Yang, and Y. Chen. Tracking vehicular speed variations by warping mobile phone signal strengths. In *Proceedings of International Conference on Pervasive Computing and Communication (PerCom '11)*, Seattle, WA, 2011.
- [19] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan. The pothole patrol: using a mobile sensor network for road surface monitoring. In *Proceeding of the 6th international conference on Mobile systems, applications, and services*, pages 29–39. ACM, 2008.
- [20] H. Fastl and E. Zwicker. *Psychoacoustics: facts and models*. Springer-Verlag New York Inc, 2007.
- [21] S. Gelfand and H. Levitt. *Hearing: An introduction to psychological and physiological acoustics*, volume 2. Marcel Dekker New York, 2004.
- [22] L. Girod, M. Lukac, V. Trifa, and D. Estrin. The design and implementation of a self-calibrating distributed acoustic sensing platform. In *Proceedings of the 4th international conference on Embedded networked sensor systems (Sensys '06)*, pages 71–84. ACM, 2006.
- [23] D. Gundlegard and J. Karlsson. Handover location accuracy for travel time estimation in gsm and umts. In *IEEE Intelligent Transportation Systems Conference (ITSC '09)*, March 2009.
- [24] A. Harter, A. Hopper, P. Steggle, A. Ward, and P. Webster. The anatomy of a context-aware application. In *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking (Mobicom '99)*, pages 59–68, 1999.
- [25] J. Hu, C. Cheng, and W. Liu. A robust statistical-based speaker's location detection algorithm in a vehicular environment. *EURASIP Journal on Applied Signal Processing*, 2007(1):181, 2007.
- [26] M. Kuttila, M. Jokela, G. Markkula, and M. Rue. Driver distraction detection with a camera vision system. In *IEEE International Conference on Image Processing (ICIP '07)*, 16 2007.
- [27] K. A. Li, P. Baudisch, and K. Hinckley. Blindsight: eyes-free access to mobile phones. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems (CHI '08)*, CHI '08, pages 1389–1398, New York, NY, 2008.
- [28] J. Lindqvist and J. Hong. Undistracted driving: A mobile phone that doesn't distract. In *12th Workshop on Mobile Computing Systems and Applications (HotMobile '11)*, 2011.
- [29] P. Mohan, V. Padmanabhan, and R. Ramjee. Trafficsense: Rich monitoring of road and traffic conditions using mobile smartphones. In *Proceedings of the Conference on Embedded Networked Sensor Systems (SenSys 2008)*, pages 323–336, 2008.
- [30] G. Moustakides. Optimal procedures for detecting changes in distributions. *Annals of Statistics*, 14(1379-1387), 1986.
- [31] L. Nelson, S. Bly, and T. Sokoler. Quiet calls: talking silently on mobile phones. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI'01)*, pages 174–181, New York, NY, USA, 2001.
- [32] U. D. of Transport. Faces of distracted driving. <http://www.distraction.gov/faces/>.
- [33] E. Page. Continuous inspection schemes. *Biometrika*, 41(1-2):100, 1954.
- [34] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan. BeepBeep: a high accuracy acoustic ranging system using COTS mobile devices. In *Proceedings of the 5th international conference on Embedded networked sensor systems (Sensys '07)*, pages 1–14, 2007.
- [35] N. Priyantha, A. Chakraborty, and H. Balakrishnan. The cricket location-support system. In *Proceedings of the 6th annual international conference on Mobile computing and networking (Mobicom '00)*, pages 32–43, Aug 2000.
- [36] J. Rodriguez-Ascariz, L. Boquete, J. Cantos, and S. Ortega. Automatic system for detecting driver use of mobile phones. *Transportation Research Part C: Emerging Technologies*, 2010.
- [37] J. Sallai, G. Balogh, M. Maróti, Á. Lédeczi, and B. Kusy. Acoustic ranging in resource-constrained sensor networks. In *Proceedings of ICWN '04*, 2004.
- [38] D. D. Salvucci. Predicting the effects of in-car interfaces on driver behavior using a cognitive architecture. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '01)*, pages 120–127, Seattle, Washington, 2001.
- [39] C. Schreiner. Driver distraction: Relevant research and implications for public policy. <http://tinyurl.com/35gw7k6>.
- [40] I. Titze and D. Martin. Principles of voice production. *Acoustical Society of America Journal*, 104:1148, 1998.
- [41] P. Treffner and R. Barrett. Hands-free mobile phone speech while driving degrades coordination and control. *Transportation Research Part F: Traffic Psychology and Behaviour*, 7(4-5):229–246, 2004.
- [42] U.S. Department of Transportation - National Highway Traffic Safety Administration. Distracted driving 2009. In *Traffic Safety Facts Research Note*, pages 1–9, Sept.
- [43] M. Wiberg and S. Whittaker. Managing availability: Supporting lightweight negotiations to handle interruptions. *ACM Transactions on Computer-Human Interaction*, 12:356–387, 2005.
- [44] J. Yang, S. Sidhom, G. Chandrasekaran, T. Vu, H. Liu, N. Cecan, Y. Chen, M. Gruteser, and R. Martin. Detecting driver phone use leveraging car speakers. In *Proceedings of the 17th annual international conference on Mobile computing and networking*, pages 97–108. ACM, 2011.



(MobiCom) 2011 and the Outstanding Research Award in 2009 from Stevens Institute of Technology. His research has received wide press coverage including MIT Technology Review, The Wall Street Journal, CNET News, and Yahoo News.



**Jie Yang** received his Ph.D. degree in Computer Engineering from Stevens Institute of Technology in 2011. He is currently a postdoctoral fellow in the Data Analysis and Information Security (DAISY) laboratory at Stevens Institute of Technology. His research interests include cyber security and privacy, mobile and pervasive computing, wireless localization systems, and mobile social networks. He is the recipient of the Best Paper Award from ACM International Conference on Mobile Computing and Networking (MobiCom) 2011 and the Outstanding Research Award in 2009 from Stevens Institute of Technology. His research has received wide press coverage including MIT Technology Review, The Wall Street Journal, CNET News, and Yahoo News.

**Simon S. Sidhom** received a BS degree in Cybersecurity as well as a MS degree in computer science both from Stevens Institute of Technology. Graduate certifications include: system administration, cloud computing, service oriented computing, enterprise security and information assurance, and computer systems. His research interests include acoustic localization, mobile systems, and computer security and privacy.



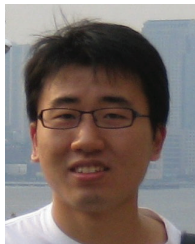
**Gayathri Chandrasekaran** received her M.S and Ph.D. degrees in Computer from Rutgers, The State University of NJ and her bachelors degree in Information Systems from BITS(Pilani) India. She is currently employed with Verizon Wireless. Her research interests include Context Awareness in Mobile Systems, Mobile and Pervasive Computing and its applications in real-world and Wireless device localization. She is a recipient of the best paper award at Mobicom 2011, University Fellowship at the Ohio State

University between the years 2004-2006 and the JNCASR summer research fellowship in 2003 and 2004.



**Tam Vu** is a Ph.D Candidate in Computer Science at Rutgers University and a research assistant in the Wireless Information Networking Laboratory (WINLAB) since 2009. His research in security and privacy focuses on computer networks and distributed systems, with emphasis on existing and future Internet architectures. He also interested in mobile and ubiquitous computing applications. He earned his M.S. in Computer Science at Rutgers University and B.Engineer in Hanoi University of Technology,

Vietnam in 2006.



**Hongbo Liu** is a Ph.D. candidate of the Electrical and Computer Engineering Department at Stevens Institute of Technology. His research interests include information security & privacy, wireless localization and location based services (LBS), wireless and sensor networks. He is currently working in the Data Analysis and Information Security (DAISY) Lab with Prof. Yingying Chen. He got his Master degree in communication engineering from Department of Communication and Information Engineering of University

of Electronic Science and Technology of China in 2008. He received his Bachelor's degree from Department of Communication and Information Engineering of University of Electronic Science and Technology of China, China, in 2005.



**Nicolae Cecan** graduated *summa cum laude* and received a BS in Electrical and Computer Engineering from Rutgers University. He is the recipient of the Steven Petrucelli Endowed Scholarship Award (2010) and the Noe, James & Edna Engineering Scholarship Award (2011). He participated in several projects at the Winlab Rutgers research facility. He is currently working as a Systems Integration Consulting Analyst at Accenture.



**Yingying (Jennifer) Chen** is an assistant professor in the Department of Electrical and Computer Engineering at Stevens Institute of Technology. Her research interests include cyber security and privacy, wireless embedded systems, wireless and sensor networks, mobile social networks and pervasive computing. She received her Ph.D. degree in Computer Science from Rutgers University. She has coauthored the book *Securing Emerging Wireless Systems* and published extensively in journal and conference

papers. Prior to joining Stevens Institute of Technology, she was with Alcatel-Lucent. She received the IEEE Outstanding Contribution Award from IEEE New Jersey Coast Section each year 2005-2009. She is the recipient of the NSF CAREER award 2010. She is the recipient of Google Research Award 2010 and the Best Paper Award from ACM International Conference on Mobile Computing and Networking (MobiCom) 2011. She received Stevens Board of Trustees Award for Scholarly Excellence. She is also the recipient of the Best Paper Award from the International Conference on Wireless On-demand Network Systems and Services (WONS) 2009, as well as the Best Technological Innovation Award from the International TinyOS Technology Exchange 2006.



**Marco Gruteser** is an Associate Professor of Electrical and Computer Engineering at Rutgers University and a member of the Wireless Information Network Laboratory (WINLAB). He is a pioneer in the area of location privacy and also recognized for his work on connected vehicle applications. Beyond these topics, his 80+ peer-reviewed articles and patents span a wide range of wireless, mobile systems, and pervasive computing issues. He received his MS and PhD degrees from the University of Colorado in 2000

and 2004, respectively, and has held research and visiting positions at the IBM T. J. Watson Research Center and Carnegie Mellon University. His recognitions include an NSF CAREER award, MobiCom and MobiSys best paper awards, and a Rutgers Board of Trustees Research Fellowship for Scholarly Excellence. His work has been featured in numerous media outlets including NPR, the New York Times, and CNN TV.



**Richard P. Martin** received the B.A. degree from Rutgers, The State University of New Jersey, Piscataway, and the M.S. and Ph.D. degrees in computer science from the University of California, Berkeley.

He is currently an Associate Professor of computer science with Rutgers, The State University of New Jersey, where he is a Member of the Wireless Network Information Laboratory. His research interests include wireless device localization and human factors in dependable computing. Dr. Martin received of the Best Paper Award at the 2004 IEEE Conference on Sensor and Ad Hoc Communication Networks and a Faculty Early Career Development (CAREER) Award from the U.S. National Science Foundation (NSF). He has served as an Investigator on grants from the Defense Advanced Research Projects Agency, the NSF, and IBM.