# ER: Early Recognition of Inattentive Driving Leveraging Audio Devices on Smartphones

Xiangyu Xu*, Hang Gao*, Jiadi Yu*‡, Yingying Chen†, Yanmin Zhu*, Guangtao Xue*, Minglu Li*
*Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, P.R.China
Email: {chillex,cullen_gao,jiadiyu,yzhu,xue-gt,mlli}@sjtu.edu.cn
†Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, USA
Email: yingying.chen@stevens.edu
‡Corresponding Author

***Abstract*—Real-time driving behavior monitoring is a corner stone to improve driving safety. Most of the existing studies on driving behavior monitoring using smartphones only provide detection results after an abnormal driving behavior is finished, not sufficient for driver alert and avoiding car accidents. In this paper, we leverage existing audio devices on smartphones to realize early recognition of inattentive driving events including *Fetching Forward*, *Picking up Drops*, *Turning Back* and *Eating or Drinking*. Through empirical studies of driving traces collected in real driving environments, we find that each type of inattentive driving event exhibits unique patterns on Doppler profiles of audio signals. This enables us to develop an *E*arly *R*ecognition system, *ER*, which can recognize inattentive driving events at an early stage and alert drivers timely. ER employs machine learning methods to first generate binary classifiers for every pair of inattentive driving events, and then develops a *modified vote mechanism* to form a multi-classifier for all inattentive driving events along with other driving behaviors. It next turns the multi-classifier into a *gradient model forest* to achieve early recognition of inattentive driving. Through extensive experiments with $8$ volunteers driving for about half a year, ER can achieve an average total accuracy of $94.80\%$ for inattentive driving recognition and recognize over $80\%$ inattentive driving events before the event is $50\%$ finished.**

## I. Introduction

Inattentive driving [1] is a significant factor in distracted driving and is associated with a large number of car accidents. According to statistics, in 2014, 3,179 people were killed and 431,000 were injured in the United States alone in motor vehicle crashes involving inattentive drivers [2]. National Highway Traffic Safety Administration(NHTSA) is working to reduce the occurrence of inattentive driving and raise awareness of the dangers of inattentive driving [3]. However, recent research [4] shows that many inattentive driving events are unapparent and thus easy to be ignored by drivers. Most drivers fail to realize themselves as inattentive while driving. Therefore, it is desirable to build an inattentive driving recognition system to alert drivers in real time, helping to prevent potential car accidents and correct drivers' bad driving habits.

There have been existing studies on detecting abnormal driving behaviors [5] [6] [7] including inattentive, drowsy and drunk drinking. These studies detect driver's status based on pre-deployed infrastructure, such as cameras, infrared sensors, and EEG devices, incurring high cost. In recent years, with the increasing popularity of smartphones, more and more smartphone-based applications [8] [9] [10] are developed to detect driving behaviors using sensors embedded in smartphones, such as accelerator, gyroscope, and camera. However, most of these investigations on driving behavior detection using smartphones can only provide a detection result after a specific driving behavior is finished, making it less helpful to alert drivers and avoid car accidents.

Among all kinds of dangerous driving behaviors, inattentive driving is the most common one but also easily to be ignored by drivers. Thus, early recognition of inattentive driving is the key to alert drivers and reduce the possibility of car accidents. Our objective is to build a system for early recognition of inattenive driving using existing smartphone sensors. According to the judicial interpretation of inattentive driving [1], there are four most commonly occurring events of inattentive driving, i.e. *Fetching Forward*, *Picking up Drops*, *Turning Back* and *Eating & Drinking*. Our goal is to recognize these most common inattentive driving events and alert drivers as early as possible to prevent drivers from continuing these behaviors dangerous to driving safety. Our work is grounded on the basic physics pheonomenon that human actions may lead to Dopper shifts of audio signals [11] to recognize different inattentive driving events. To realize the inattentive driving recognition leveraging audio signals, we face several challenges in practice. First, the unique pattern of each type of inattentive driving needs to be distinguished. Second, any inattentive driving event should be recognized as early as possible under the guarantee of a high recognition accuracy. Finally, the solution should be effective in real driving environments and computational feasible on smartphones.

In this paper, we first investigate the patterns of Doppler shifts of audio signals caused by inattentive driving events. Through empirical studies of the driving traces collected from real driving environments, we find that each type of inattentive driving event exhibits an unique pattern on Doppler profiles of audio signals. Based on the observation, we propose an *E*arly *R*ecognition system, *ER*, which aims to recognize inattentive driving events at an early stage and alert drivers in real time for safe driving. In ER, effective features of inattentive driving events on audio signals collected by smartphones are first extracted through *Principal Components Analysis (PCA)*. To improve the recognition accuracy of driving events, we train

these features through a machine learning method to generate binary classifiers for every pair of inattentive driving events, and propose a *modified vote mechanism* to form a multi-classifier for all inattentive driving events based on the binary classifiers. In this work, the training is performed based on 3-month driving traces in real driving environments involving 8 drivers. Furthermore, to detect the inattentive driving at an early stage, we first analyze the relationship between the completion degree and time duration for each type of inattentive driving event, and then exploit the relationships to turn the multi-classifier into a *Gradient Model Forest* for early recognition. Our extensive experiments validate the accuracy and the feasibility of our system in real driving environments.

We highlight our main contributions as follows:

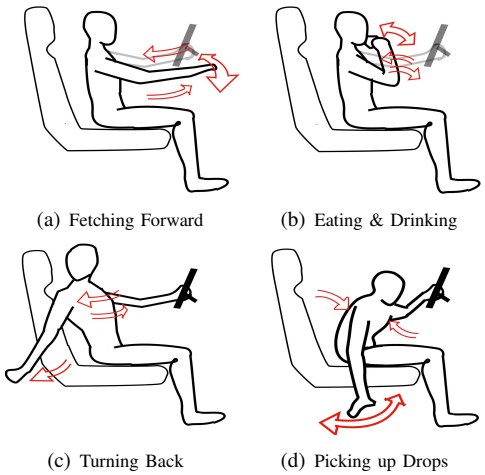
- We design an early recognition system of inattentive driving, ER, leveraging audio devices on smartphones. It aims to recognize inattentive driving behaviors at an early stage and alert drivers in real time for safe driving.
- We find each inattentive driving event presents an unique pattern on Doppler profiles of audio signals. We validate this important finding through empirical analysis of the driving traces collected from real driving environments.
- We propose a modified vote mechanism based on Supporting Vector Machine to generate a highly accurate multi-classifier for inattentive driving recognition, and further present a gradient model forest for early recognition of inattentive driving events.
- We conduct extensive experiments in real driving environments. The results show that ER achieves an average total accuracy of $94.80\%$ for recognition and over $80\%$ inattentive driving events can be recognized before a specific event is half-way done.

The rest of the paper is organized as follows. The related work is reviewed in Section II. Patterns of inattentive driving events on Doppler profiles of audio signals are analyzed in Section III. Section IV presents the design details of ER. Systems issues and possible solutions in the implementation of ER are presented in Section V. We evaluate the performance of ER and present the results in Section VI. Finally, we give our solution remarks in Section VII.

## II. Related Work

In this section, we review the existing works on driving events detection. Some existing works realize driving events detection by using professional infrastructure including EEG [5] and water cluster detectors [6], or common sensors such as infrared sensors [12] and cameras [7]. However, the solutions all rely on pre-deployed infrastructure and additional hardware that incur installation cost. Moreover, those additional hardware could suffer the difference of day and night, bad weather condition and high maintenance cost.

To overcome the limitations of pre-deployed infrastructure, recent studies put their efforts to exploit smartphones on driving events detection, which can be categorized as vision-based solutions [8] [13] and sensor-based solutions [9] [10] [14]. In vision-based solutions, the build-in cameras are used to capture the graphic information for processing. [8] uses rear cameras of smartphones to monitor road conditions, [13] leverages dual cameras of smartphones to tract road conditions and detect drivers' status at the same time. However, the accuracy of vision-based approaches is unstable depends on weather, lighting and smartphones placement. In sensor-based solutions, [10] leverages accelerators and gyroscopes of smartphones to detect abnormal driving behaviors, and [14] combines sensors by using Inertial Measurement Units on smartphones to detect various steering maneuvers. These two solutions can only provide detection results after driving behaviors finished. Besides, [9] uses accelerators of smartphones to determine usage of phones while driving, but this work can not recognize other driving behaviors but usage of phones.



(a) Fetching Forward (b) Eating & Drinking

(c) Turning Back (d) Picking up Drops

Fig. 1. Illustration of inattentive driving events.

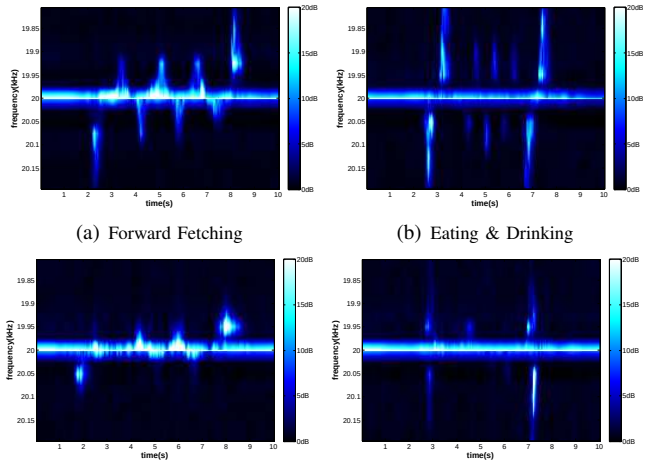
Moreover, there are some works of gesture recognition and behavior monitoring based on the acoustic techniques [15]–[18]. [15] proposes an audio-based system to sense gestures for laptops. As implemented on smartphones, [18] builds an acoustic system leveraging FMCW to detect sleeping condition. [17] realizes a virtual mouse based on audio signals of smartphones. [16] leverages microphones in smartphones as well as car speakers for determining usage of phones. Unlike the above works, our work achieves early recognition of inattentive driving events using acoustic techniques, which is meaningful for safety in real driving environments.

## III. Inattentive Driving Events Analysis

In this section, we first give a brief introduction to inattentive driving events, and then analyze patterns of these events on Doppler profiles of audio signals.

### A. Defining Inattentive Driving Events

Drivers are encountered with a variety of road hazards because of their unawareness of being in negligent driving state, such as eating or picking drops while driving. These inattentive driving events are potentially posing drivers in danger. According to reference [1], there are four types of the most commonly occurring inattentive driving events of drivers themselves, as shown in Fig. 1.



(a) Forward Fetching (b) Eating & Drinking

(c) Picking up Drops (d) Turning Back

Fig. 2. Frequency-time Doppler profiles of inattentive driving events.

*Fetching Forward:* this state refers to the condition where drivers fetch out to search widgets like keys, car audio consoles, etc.

*Eating or Drinking:* drivers eat snacks or replenishing water when driving.

*Turning Back:* drivers intend to take care of their children in rear seat, or turn around searching for bags or packages placed on rear seat.

*Picking Drops:* drivers are likely to pick dropped keys or other objects when driving where their heads temporarily moves away from the front sight.

Through analyzing the above four inattentive driving events, we realize each driving event is not a transient action, but a consecutive action lasts for a time period. For example, Fig. 1(a) shows a Fetching Forward event, which can be demonstrated as stretching out to reach the deck, searching for something, and stretching retrieved to normal condition. Our work is to detect these consecutive inattentive driving events in real time and try to recognize these events at the early stage, so as to alert drivers as early as possible.

### B. Analyzing Patterns of Inattentive Driving Events

We utilize the Doppler shifts of audio signals to recognize inattentive driving events. *Doppler shift* (or Doppler effect) is the change in frequency of waves from observers moving relative to sources. Specifically, a mass point moving at speed $v$ and angle $\theta$ to a speaker brings a frequency change:

$$\Delta f = \frac{2vcos(\theta)}{c} \times f_0, \tag{1}$$

where $c$ and $f_0$ denotes the speed and frequency of the signal.

We recruit five volunteers to perform four inattentive driving events depicted in Fig. 1 while driving in relatively safe area. The experiments are conducted by generating continues pilot tones from speakers and then collect the audio signals from microphones on smartphones.

When selecting the frequency of audio signals to use, we take two factors into consideration, i.e. background noise and unobtrusiveness. According to [16], frequency range from

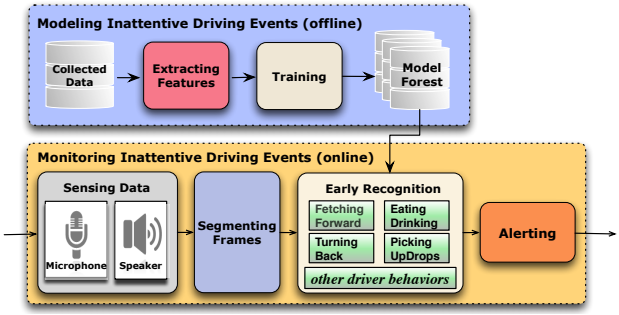



Fig. 3. System architecture and work flows.

$50Hz$ to $15,000Hz$ covers almost all naturally occurring sounds, and human hearing becomes extremely insensitive to frequencies beyond $18kHz$. Thus, we could straightforwardly filter the background noise and eliminate the effects for people by locating our signal above $18kHz$. Furthermore, a higher frequency results in a more discernible Doppler shift confined by Eq. 1, and most phone speaker systems only can product audio signals at up to $20kHz$. Taking all above analysis into account, $f_0 = 20kHz$ is selected as our frequency of pilot tone, through which we sample raw data from given inattentive driving events at the rate of $44.1kHz$, which is the default sampling rate of audio signals under $20kHz$. Then we transform it into frequency domain using *2048-points Fast Fourier Transform* (*FFT*) for appropriate computation complexity and relative high frequency resolution. Fig. 2 shows the structure of Doppler profiles of the four inattentive driving events. From Fig. 2, it can be seen that although the four profiles share the similarity that they all consist of positive and negative Doppler shifts, the patterns are different across the four events in frequency range, energy amplitude, etc.

From above analysis, we find that each type of inattentive driving events has unique patterns on the structure of Doppler profiles. Although some existing works, [11] and [10], present human action recognition methods based on the unique patterns of actions already, the recognition can only be done after actions finished, which is acceptable for transient actions like single gestures in [11], but not good enough for consecutive actions like inattentive driving events here because it is too late to alert drivers after the driving events finished in a driving security warning system. Our goal is to recognize inattentive driving events as early as possible and alert drivers timely.

## IV. System Design

In order to monitor inattentive driving events effectively and efficiently, we present an early recognition system, ER, which can alert drivers as early as possible when they are performing inattentive driving events. ER does not depend on any pre-deployed infrastructure and additional hardware.

### A. System Overview

ER can recognize inattentive driving events through analyzing patterns of Doppler profiles of audio signals over time. The work flow of ER is shown in Fig. 3. The whole system is divided into offline part - *Modeling Inattentive Driving Events*, and online part - *Monitoring Inattentive Driving Events*.

In the offline part, for different types of inattentive driving events, effective features are extracted from the Doppler profiles of audio signals collected in real driving environments. Then, we train these features through machine learning methods to generate binary classifiers for every pair of inattentive driving events, and propose a *modified vote mechanism* to form a multi-classifier for all inattentive driving events along with other driving behaviors based on them. Afterwards, the multi-classifier model is turned into a *gradient model forest* for realizing early recognition, which is stored in the database.

In the online part, ER senses real-time audio signals generated by speakers and received by microphones. The audio signals are first transformed through FFT to Doppler profiles. Then, ER detects the beginning of an event and continuously segments the corresponding frequency-time Doppler profile from the beginning to current time and sends to *Early Recognition* until ER outputs a recognition result. Then in *Early Recognition*, ER extracts features from segments and identifies whether the events are inattentive driving events or other driving behaviors at some early stages based on the trained model forest. Finally, if any of the four inattentive driving events is recognized through the above procedure, ER sends a warning message to alert driver.

### B. Model Training at Offline Stage

*1) Establishing Training Dataset:* To collect data in real environments, we develop an Android-based program to generate and collect audio signals, and then transform the raw sampled signals to the frequency-time Doppler profiles.

We collect these transformed data from 8 drivers with distinct vehicles. 8 smartphones of 4 different types are used, which are HTC Desire G7, ZTE U809, HTC EVO 3D and SAMSUNG Nexus5, two for each type. Meanwhile, all vehicles are equipped with a special camera so that drivers' events can be recorded as the ground truth. Our data collection spans from October 23, 2015 to January 27, 2016, during which all the daily driving including commuting to work, shopping, touring, etc. is recorded. Drivers are not told about our purpose so that they can drive in a natural way. And each of our volunteer has their own driving routes differs from each other. After that, we ask 5 experienced drivers to watch the videos recorded by the cameras and recognize all types of inattentive driving events from the 3-month traces. In total, we obtain 3532 samples of inattentive driving events from the collected traces, which are severed as the ground truth. Afterwards, we combined the collected Doppler profiles of audio signals and their labels into a training dataset $X$.

*2) Extracting Effective Features:* Traditional feature extracting methods extract features by observing the unique patterns manually. Features extracted by these methods usually have redundant information and are poor in robustness. To achieve better features, ER leverages *Principal Components Analysis*(*PCA*) algorithm to the raw data.

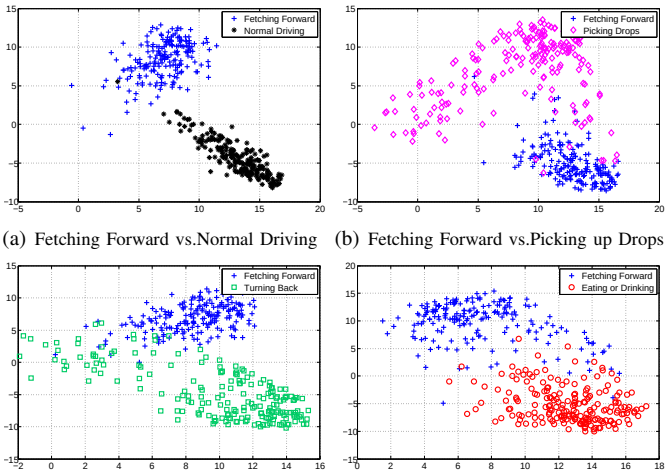
In PCA algorithm, to extract features from training dataset $X$, a projection matrix $W$ that contains features vectors ranked by variance, is calculated using *Singular Value Decomposition* (*SVD*), which is given by $X = U\Sigma W^T$. For $m \times n$ matrix $X$, $U$ is a $m \times m$ unitary matrix, $\Sigma$ is a $m \times n$ matrix with non-negative singular values on the diagonal. $W$ is a $n \times n$ unitary matrix, which has n orthogonal features ranked by importance. Since too many features may bring in the danger of over-fitting, we should select the minimum number of features, $d$, which contains enough information of the raw data. Considering the reconstruction property of PCA, the object function is

$$\min_{d} \quad (\sum_{i=1}^{d} \sigma_i)/(\sum_{i=1}^{n} \sigma_i) \geqslant t \quad t \in [0,1], \tag{2}$$

where $\sigma_i$ is the $i^{th}$ largest singular value of matrix $X$, which denotes the importance of the $i^{th}$ features in $W$, and $t$ is the threshold of reconstruction, denoting the remaining information of the raw data. In ER, $t$ is set to be 0.95 to guarantee the features' validity. For all four inattentive driving events, we have $d = 17$ from Eq. 2, which is slightly large.

To further reduce $d$, we analyze inattentive driving events pairwise. According to Eq. 2, for any pair of the four inattentive driving events, $d = 2$ is good enough to represent most information of the raw data. Fig. 4 shows the distributions of Fetching Forward events versus other three inattentive driving events and normal driving in 2-dimensional feature spaces. It is can be seen from Fig. 4 that Fetching Forward events can be discriminated from other inattentive driving events along with normal driving using two features extracted by PCA. Similarly, for all pairs for inattentive driving events along with normal driving, this conclusion remains. Therefore, in order to reduce the amount of features and improve recognition accuracy, we extract features for inattentive driving events pairwise.

*3) Training a Multi-Classifier:* After features extracting through PCA, We first use *Supportive Vector Machine* (*SVM*) to train binary classifiers for every pair of inattentive driving events. Based on the binary classifiers, a voting mechanism is proposed to form a multi-classifier to differentiate all four inattentive driving events.




(a) Fetching Forward vs.Normal Driving (b) Fetching Forward vs.Picking up Drops

(c) Fetching Forward vs.Turning Back (d) Fetching Forward vs.Eating & Drinking

Fig. 4. Distributions of Fetching Forward events with other inattentive driving events and normal driving in 2-dimensional feature space.

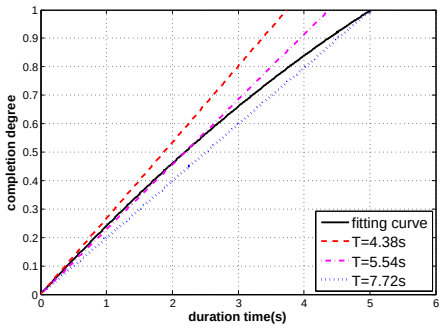



Fig. 5. The relationships between the completion degree $\alpha$ and time duration $\tau$ of fetching forward events under different the complete time $T$.

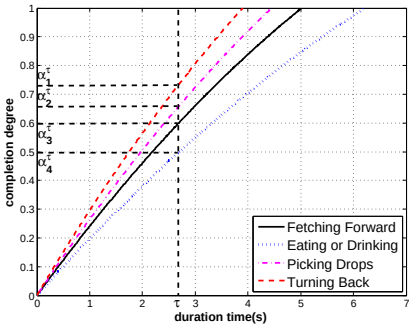



Fig. 6. The relationships between the completion degree $\alpha$ and time duration $\tau$ for four kinds of inattentive driving events.

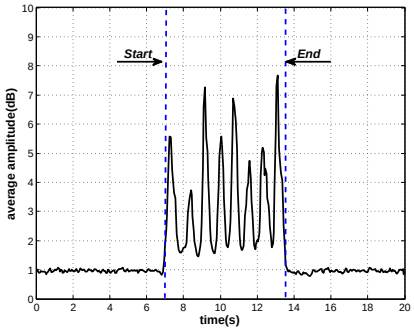



Fig. 7. The average amplitude of frequency bands except the pilot frequency during a 20 seconds driving containing a fetching forward event.

Given that each binary classifier has one vote $v \in \{0, 1\}$ for building the multi-classifier. Considering a binary classifier for separating event $a$ from event $b$, then for a specific event $e$, if the binary classifier identifies $e$ as event $a$, then event $a$ get a vote $v_a = 1$, event $b$ get a vote $v_b = 0$. Assuming an event set $E$ containing $k$ types of events, a classifier group has $C_k^2$ binary classifiers. For the event $e$, the votes of all $C_k^2$ binary classifiers can be denoted as

$$V(e) = \sum_{j \in [1, C_k^2]} v_j, \tag{3}$$

where $v_j$ is a vote vector of $k$ elements that denotes the vote of the $j^{th}$ binary classifier. The event class which get the most votes in $V(e)$, i.e.

$$c = \max_j \ V_j(e) \quad j \in [1, k], \tag{4}$$

is supposed to the classified event of $e$.

Moreover, for a specific type of inattentive driving events, there are exactly $k-1$ binary classifiers directly related to this type of event in all $C_k^2$ binary classifiers. So the votes of the winning event class should satisfies

$$V_c(e) = k - 1. \tag{5}$$

For an event which gets through the multi-classifier and gets a classification result $c$ from Eq. 4, if it does not satisfy Eq. 5, ER considers the event as *other driving behaviors*, such as shifting gear, pushing glasses, etc, which are not so dangerous for drivers as inattentive driving events.

*4) Setting up Gradient Model Forest for Early Recognition:* To approach the goal of recognizing inattentive driving events as early as possible, we propose an early recognition method.

Considering an inattentive driving events set $E$ containing $k$ types of events, $E = \{e_1, e_2, \cdots, e_k\}$. For a given event $e$ started at $t_0$ and finished at $t_1$, the *completion degree* $\alpha$ of the event $e$ at time $t$ is

$$\alpha_e = \frac{t - t_0}{t_1 - t_0} = \frac{\tau}{T} \quad t \in [t_0, t_1], \tag{6}$$

where $\tau$ denotes the *time duration* of $e$ at time $t$ and $T$ denotes the *total length* of $e$. Obviously, $\alpha_e \in [0, 1]$. And when $\alpha_e = 1$, the event $e$ finishes. Eq. 6 shows that the goal to recognize a inattentive driving event $e$ as early as possible is equivalent to finish recognition when $\alpha_e$ is as small as possible. As a result, based on different $\alpha$ of inattentive driving events at different $\tau$, we set up a bunch of classifiers for early recognition, i.e. the gradient model forest.

For modeling the complete degree $\alpha$ at different time duration $\tau$ of inattentive driving events, the variation of the total length $T$ among all events should first be considered. For different types of inattentive driving events, $T$ varies because of the nature differences of the events, thus we set different models for different types of events. Moreover, for a specific type of inattentive driving events, $T$ also varies depending on different drivers and driving situations. We also need to take this variation into consideration when setting models.

According to statistics of the dataset established in Section IV-B1, the total length $T$ for each type of inattentive driving events approximately satisfies a Gaussian distribution. For example, $T$ of Fetching Forward events approximately satisfy a Gaussian distribution of mean value $\mu = 4.38s$ and standard deviation $\sigma = 0.32s$. Since two standard deviations from the mean account for $95.45\%$ data in Gaussian distribution, we can think that about $95\%$ of Fetching Forward events have $T$ from $3.74s$ to $5.02s$. As shown in Fig. 5, based on the completion degree-time duration relations of Fetching Forward events having $T$ equals to $3.74s$, $4.38s$ and $5.02s$, a quadratic curve is fit to model the relationship between $\alpha$ and $\tau$ for Fetching Forward event, which starts at the origin, goes through the mid-point of the line $T = 4.38s$ and ends at the end of the line $T = 5.02s$. For any $\tau > 5.02s$, $\alpha = 1$. The fitting carve can thus represent most Fetching Forward events because it closes to most Fetching Forward events at some time period of completion degree-time duration relations. With the similar analysis, ER models a relationship between $\alpha$ and $\tau$ for each type of inattentive driving events. Fig. 6 shows relationships between $\alpha$ and $\tau$ for all four types of inattentive driving events.

With the relationships between $\alpha$ and $\tau$, for any given time duration $\tau$, ER can get a completion degree set $A^\tau = \{\alpha_1^\tau, \alpha_2^\tau, \cdots, \alpha_k^\tau\}$, which contains the completion degree for each type of inattentive driving events at time duration $\tau$, as shown in Fig. 6. According to $A^\tau$, ER segments the Doppler profiles of all types of inattentive driving events $X = \{X_1, X_2, \cdots, X_k\}$ and then gets the new input dataset $X^\tau = \{X_1^\tau, X_2^\tau, \cdots, X_k^\tau\}$. Selecting n different $\tau$ by gradient,

we form a n-element time duration set $T = \{\tau_1, \tau_2, \cdots, \tau_n\}$. ER then segments the Doppler profiles based on $T$ and ends up with a gradient dataset forest $X = \{X^{\tau_1}, X^{\tau_2}, \cdots, X^{\tau_n}\}$. Afterwards, $X$ is trained through the methods in Section IV-B2 and Section IV-B3. Although for a specific dataset $X^{\tau}$, patterns for parts of inattentive driving events are not guaranteed to be unique, ER can always get a multi-classifier $\theta^{\tau}$. Based on the new input dataset $X$, a gradient model forest $\Theta = \{\theta^{\tau_1}, \theta^{\tau_2}, \cdots, \theta^{\tau_n}\}$ is set up, and each of $\theta^{\tau}$ is a matrix containing all binary classifiers as $\theta^{\tau} = ((\theta_1^{\tau})^T; (\theta_2^{\tau})^T; \ldots; (\theta_m^{\tau})^T)$, where $m = C_k^2$ for all different pairwise inattentive driving events. Specially, the last multi-classifier of the model forest, i.e. $\theta^{\tau_n}$, is a multi-classifier for recognizing inattentive driving events after they finished. Finally, we obtain a gradient model forest $\Theta$, which could be used to realize early recognition of inattentive driving events.

### C. Recognizing Inattentive Driving Events at Online Stage

*1) Segmenting Frames through Sliding Window:* In order to recognize current driving events, ER first needs to determine the time duration by recognizing the beginning and the end of the driving events.

As mentioned in Section III-B, all driving events occur with positive and negative Doppler shifts in the frequency-time Doppler profiles, i.e. energy fluctuation near the pilot frequency ($20kHz$) as shown in Fig. 2. From analyzing the traces collected in real driving environments, we find that when events occur, the average amplitude of frequency bands beside the pilot frequency keeps a relatively high value. Fig. 7 shows the average amplitude of frequency bands beside the pilot frequency during 20 seconds driving containing a fetching forward event. From Fig. 7, it can be seen that the average amplitude for events is much greater than that without events.

Based on patterns of the average amplitude, ER employs the sliding window method to capture Doppler shifts caused by driving events. ER keeps computing the average amplitude within a window and compares with thresholds to determine the start point and end point of an event. The window size and thresholds can be learned from the collected data. When the start point of an event is detected, ER segments a frame from the start point to current point and sends the frame to go through early recognition. ER keeps segmenting and sending at short time intervals until a recognition result is output.

*2) Detecting Inattentive Driving Events at Early Stage:* After getting a frame of a driving event, according to the time duration of the frame, ER inputs the frame into the corresponding classifier in the model forest $\Theta$ to recognize the driving event. For frames with small time durations, the recognition results may not be accurate because these frames contains few information of the events. Thus, ER proposes a mechanism to guarantee the validity of early recognition.

For a frame $e$ of time duration $\tau \in [\tau_i, \tau_{i+1}]$, ER calls the classifier $\theta^{(\tau_i)}$ and $\theta^{(\tau_{i+1})}$ to recognize the driving event. From the two classifiers, ER gets the classification results $c_1$ and $c_2$. Only when $c_1 = c_2$, ER admits the validity of the result and temporarily stores it. After ER detects several continuous

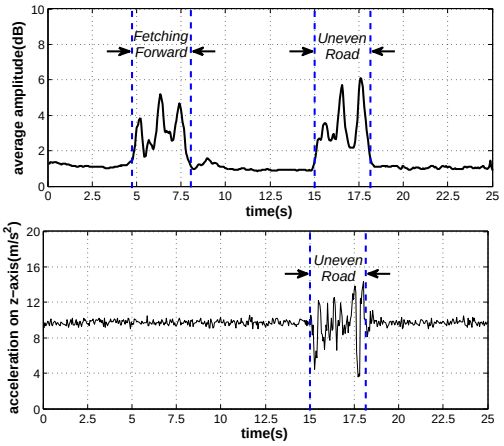



Fig. 8. The average amplitude of audio signals and readings from the accelerator's z-axis in 25-second driving containing a Fetching Forward event and a driving condition that the vehicle goes across a speed bump.

valid results that denote the same inattentive driving event, an alert is sent to the driver until ER detects the end of the event. The number of continuous valid results which output the same is defined as *Convince Length*. If ER does not output a result before recognizing the end of an event, it uses the last classifier of the model forest, $\theta^{\tau_n}$, to finish the recognizing and get the corresponding output. For each event, ER records the output recognition result and the result from the classifier $\theta^{\tau_n}$. The proposed mechanism of ER can reduce mistaken recognition and avoid disturbing drivers from false warning effectively .

## V. System Implementation

In the implementation of ER, we are facing several practical issues as follows.

### A. Allowing Inattentive Driving Events when Vehicle Stops

When driving in real environments, drivers may stop the vehicles because of red traffic lights, heavy traffic conditions or other temporary situations. In the conditions that the vehicles are stopped, inattentive driving events are not so dangerous and should be allowed to perform. However, ER can not sense stops of vehicles based on audio signals. The work [19] presents a result that the data patterns of the acceleration on vehicle's z-axis for stop is remarkably different from that for moving. Specifically, the standard deviation of the acceleration on z-axis is remarkably low while a vehicle stops. Therefore, ER collects readings from the accelerometers on smartphones to sense stops of vehicles in real time. Once ER detects that a vehicle is stopped, it suspends analysis on audio signals until the vehicle moves again.

### B. Filtering Influence of Uneven Road

Uneven road may result in strong vibrations on smartphones, which could affect audio signals collected by the microphone and cause mistaken recognition of ER. It is necessary to separate Doppler shifts caused by the uneven road from that caused by driving events. Based on our observation from traces collected in real driving environments, the strong vibration can also be reflected on the acceleration on vehicle's z-axis, so

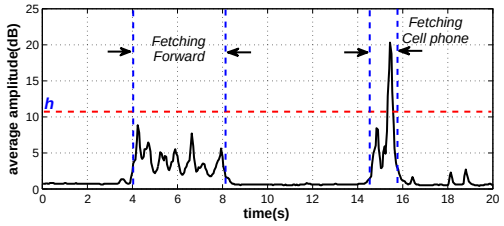



Fig. 9. The average amplitude of audio signals in 20 seconds driving containing a Fetching Forward event and an event of fetching smartphones.

uneven road can be sensed by the readings from accelerator's z-axis. Fig. 8 plots the readings from the accelerator's z-axis and the average energy amplitude described in Fig. 7 in a 25 seconds driving which contains a Fetching Forward event and a driving condition that the vehicle goes across a speed bump. From Fig. 8, it can be seen that when a driver performs an event, it brings Doppler shifts of audio signals but does not influent the acceleration of z-axis, while the condition of uneven road brings Doppler shifts along with great jitter to the acceleration of z-axis. So ER could filter the influence of vibrations by sensing uneven road and ignoring the corresponding Doppler shifts.

### C. Preventing Usage of Phone while Driving

Using smartphones, such as calling, texting messages, browsing webpages, etc, is very dangerous when driving and thus should be prohibited. Since the usage of smartphones may greatly influent audio signals by changing the position and post of smartphones over time, ER regards usage of phones as a special type of inattentive driving events to recognize and alert. Fig. 9 shows the average amplitude of audio signals in 20 seconds driving which contains an Fetching Forward event and an event of fetching smartphone. From Fig. 9, it can be seen that when a driver fetches the smartphone, there is a remarkable Doppler shift, where the average amplitude is far greater than other driving events. Therefore, ER could recognize usage of smartphones by detecting the events of fetching smartphones. Once the average amplitude is greater than a threshold $h$, ER regards the driving event as usage of smartphone and sends an alert to driver.

## VI. Evaluation

In this section, we evaluate the performance of ER in real driving environments. We implement ER as an Android App and install it on smartphones. ER is running by 8 drivers with distinct vehicles in real driving environments to collect traces for evaluation. Drivers are not told about our purpose so that they can drive in a natural way. Meanwhile, each car is implemented with a camera for recording driver's driving behaviors and 5 experienced drivers are asked to recognize inattentive driving events as the ground truth. After data collection from March 11 to May 6, 2016 using method described in Section IV-B1, we obtain a test set with 1473 inattentive driving events to evaluate the performance of ER.

### A. Metrics

To evaluate the performance of ER, we define metrics as follows.

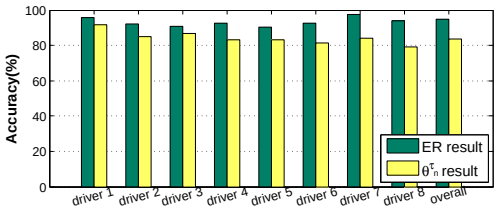



Fig. 10. The total accuracy of ER and classifier model $\theta^{\tau_n}$ over 8 drivers.

- *Accuracy*: The probability that an event is correctly identified for all type of events.
- *Precision*: The probability that the identification for an event A is exactly A in ground truth.
- *Recall*: The probability that an event A in ground truth is identified as A.
- *False Positive Rate(FPR)*: The probability that an event not of type A is identified as A.
- *F-Score*: A metric that combines precision and recall ($2\times \frac{Precision\times Recall}{Precision+Recall}$). We use F-score as our major metric to evaluate the recognition accuracy for specific types of inattentive driving events in the following evaluation.

### B. Overall Performance

Fig. 10 plots the recognition accuracy of ER and the classifier $\theta^{\tau_n}$ for 8 drivers, it can be seen that ER achieves a total accuracy of 94.80% for recognizing all types of inattentive driving events, while the total accuracy for $\theta^{\tau_n}$ is 84.78%. Further, ER performs far better than $\theta^{\tau_n}$ for any of the 8 drivers. The lowest accuracy for ER of the 8 drivers is 91.73%, which validate the effectiveness and stability of ER in real driving environments.

For different types of inattentive driving events, the precision, recall and $F$-score for recognition is showed in Fig. 11. It can be seen that these three metrics is high for every type of inattentive driving events. Specifically, the precision is no less than 89%, while the recall is above 91%, and the F-score is more than 92%.

Moreover, for each of the 8 drivers, we evaluate the FPRs of recognizing specific type of inattentive driving events. Fig. 12 shows the box-plot of the FPRs for each type of inattentive driving events. We can observe from Fig. 12 that the highest FPR is no more than 2.5% and the average FPR is as low as 1.4% over the four events and 8 drivers. It shows that ER could realize inattentive driving events recognition with few false alarms, which is user-friendly for drivers.

We plot the CDF of recognition time for each type of inattentive driving events and the CDF of all types of events in Fig. 13. It can be seen from Fig. 13 that 50% of all inattentive driving events are recognized by ER before $1.4s$ and 80% can be recognized before $2.3s$, while the average total length of all events is around $4.6s$. In another word, more than 80% inattentive driving events can be recognized at the time less than 50% of the average total length of all events. For each specific type of events, The 80%-recognized time are around $2s$, $2.5s$, $1.0s$ and $2.6s$ for Fetching Forward, Eating or drinking, Turning Back and Picking Drops respectively. And

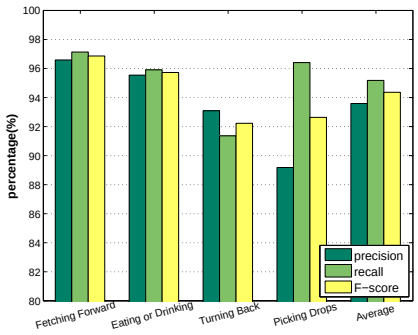



Fig. 11. The Precision, Recall and F-Score for all types of inattentive events.

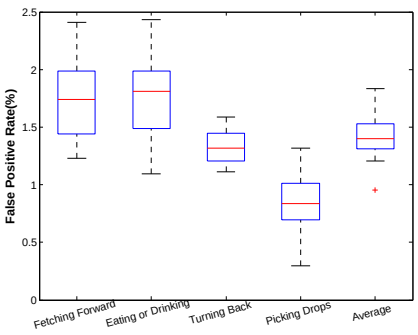



Fig. 12. Box plot of False Positive Rate of all types of inattentive events over 8 drivers.

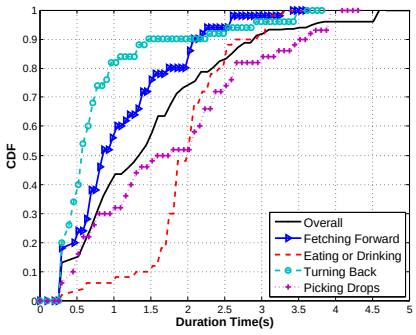



Fig. 13. The CDF of recognition time for all types of inattentive events.

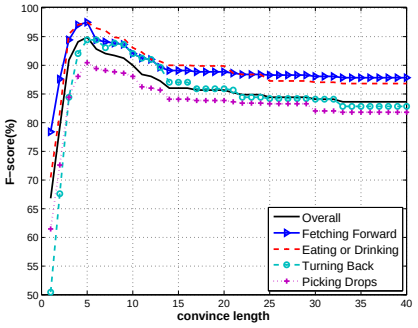



Fig. 14. F-score under different convince length for all types of inattentive events.

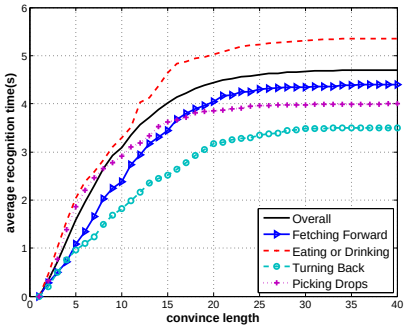



Fig. 15. Average recognition time under different convince length for all types of inattentive events.

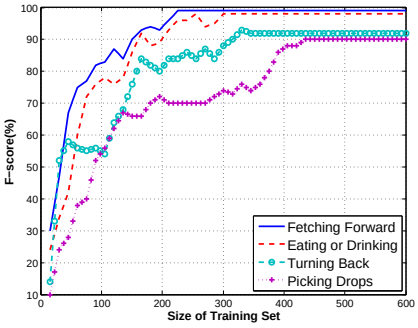



Fig. 16. F-score under different size of training set for all types of inattentive events.

the corresponding average total length for the four events are $4.3s$, $5.4s$, $3.5s$ and $4.0s$.

### C. Impact on Convince Length

Convince length, as defined in Section IV-C2, is the requiring number of continuous valid results of same value for ER to output. Fig. 14 shows the recognition performance for different types of inattentive driving events at different convince lengths. It can be seen that as convince length increases, F-scores of all different types of inattentive driving events first increase to peak value rapidly, then decrease slowly and finally converge to some constants. According to the definition of convince length, greater convince length brings more strictly condition for ER to output the recognition results, so that the output results is more accurate, which explains the increasing of F-scores. However, as convince length keeps increasing, the output condition becomes too strict for ER to output recognition results before the end of events. As a result, more events need to be recognized through classifier $\theta^{\tau_n}$, which is less accurate than ER according to Fig. 10. Fig. 15 shows the average recognition time of different types of inattentive driving events at different convince lengths. We can see that the average recognition time of all different types of inattentive driving events keeps increasing and finally converges to some constants as convince length increases because it always takes longer for a result to output when output condition is stricter. ER chooses 5 as the convince length through empirical studies.

### D. Impact on Training Set Size

According to Section IV-B1, we collect 3532 inattentive driving events for training. Fig. 16 shows the impact of training set size to the recognition performance of ER. From the figure, it can be seen that the F-score rises as training set size increases and goes stable after a certain size for each inattentive driving events. Specifically, to get a stable F-score, ER needs 220 training samples for Fetching Forward, 300 samples for Eating or Drinking, 340 samples for Turning Back and 450 samples for Picking Drops. We use as much training examples as we can get to guarantee the performance of ER.

### E. Impact on Road Types and Traffic Conditions

Different road types and traffic conditions may influent drivers' driving behaviors and vehicle conditions, thus may have an impact on the performance of ER. We analyze the collected traces of different road types (local road and highway) and different traffic conditions (during peak time and off-peak time), respectively. Fig. 17 shows the result. It can be seen that ER achieves fairly good F-scores for recognition at any combination of road types and traffic conditions. In addition, during peak time, the F-scores of ER is slightly lower than the F-scores during off-peak time because heavy traffic condition may bring more stops for vehicle and more driving behaviors such as shifting gears, which may result in more mistaken recognitions. Further, the F-scores of ER when driving on highway is slightly higher than the F-scores on local road since drivers are more concentrate when driving at high speed and the road on highway is more smooth, which brings less influence to ER.

### F. Impact on Smartphone Placement

In our experiments, each driver place the smartphone randomly on instrument panel (left side), instrument panel

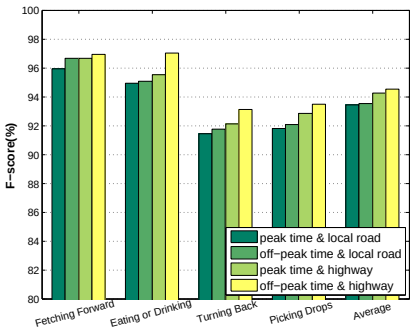



Fig. 17. F-score under different traffic condition and road types for all types of inattentive events.

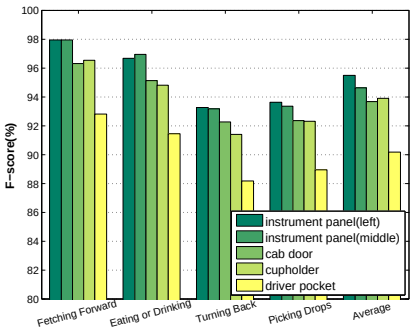



Fig. 18. F-score under different smartphone placement for all types of inattentive events.

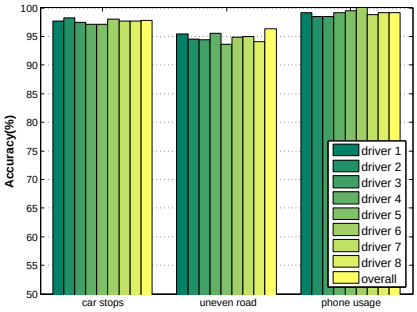



Fig. 19. Recognition accuracy for car stops, uneven road and phone usage over 8 drivers.

(middle part), panel near cab door and cup-holder, or in driver's pocket for daily driving. Fig. 18 shows that ER can achieve fairly good F-scores for recognitions under different smartphone placements. Specifically, smartphones placed on instrument panel achieve best recognition results as the audio devices of smartphones is directly face to drivers. And smartphones placed in drivers' pockets achieves lower F-score than others because the movement of drivers may bring influence to ER. But the F-score for any smartphone placement and any inattentive driving events is above 88%, which is acceptable for using ER in real driving environments.

### G. Evaluations for Implementation

In the implementation of ER, we solve several practical issues by recognizing car stops, uneven road and the event of fetching cell phones. Fig. 19 shows the recognition accuracy for all three situations over the 8 drivers. It can be seen from Fig. 19 that most of these situations are correctly recognized by ER. Among these three situations, the event of fetching cell phones has the recognition accuracy over 98% for all 8 drivers and thus ER can effectively prevent drivers from using phones while driving. For recognizing car stops and uneven road, the accuracy is above 96.5% and 94%, respectively.

## VII. Conclusion

In this paper, we address the problem of recognizing inattentive driving as early as possible to improve driving safety. In particular, we propose an early recognition system, ER, to recognize different inattentive driving events at the early stage leveraging build-in audio devices on smartphones. To achieve the recognition, ER first extracts effective features using PCA based on the patterns of different inattentive driving events on Doppler profiles of audio signals. Then ER leverages SVM and a modified vote mechanism to form a multi-classifier and set up a gradient model forest based on the multi-classifier for early recognition. We train our gradient model forest based on traces collected in real driving environments. The extensive experiments in real driving environments show that ER achieves high accuracy for inattentive driving recognition and realizes recognizing events at early stage.

## Acknowledgment

Research was sponsored by NSFC (No.61170238, 61420106010, 61472254, 61170238), 863 Program (No.2015AA015303), 973 Program (No.2014CB340303), STCSM (No.14511107500 and 15DZ1100305) and SZSTI (No.JCYJ20160407160609492).

## References

[1] USLegal, "Inattentive driving law and legal definition." [Online]. Available: http:// definitions.uslegal.com/i/inattentive-driving, 2016.

[2] U. D. of Transportation, "Traffic safety facts research note, distracted driving 2014." [Online]. Available: https:// crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812260, 2016.

[3] U. D. of Transportation, "Faces of distracted driving." [Online]. Available: http:// www.distraction.gov/faces/, 2016.

[4] C. C. Liu, S. G. Hosking, and M. G. Lenné, "Predicting driver drowsiness using vehicle measures: Recent insights and future challenges," *Journal of safety research*, vol. 40, no. 4, pp. 239–245, 2009.

[5] M. V. Yeo, X. Li, K. Shen, and E. P. Wilder-Smith, "Can svm be used for automatic eeg detection of drowsiness during car driving?," *Safety Science*, vol. 47, no. 1, pp. 115–124, 2009.

[6] M. Sakairi and M. Togami, "Use of water cluster detector for preventing drunk and drowsy driving," in *Proc. Sensors'10*, IEEE, 2010.

[7] S. Al-Sultan, A. H. Al-Bayatti, *et al.*, "Context-aware driver behavior detection system in intelligent transportation systems," *IEEE transactions on vehicular technology*, vol. 62, no. 9, pp. 4264–4275, 2013.

[8] H. Dahlkamp, A. Kaehler, D. Stavens, S. Thrun, and G. R. Bradski, "Self-supervised monocular road detection in desert terrain," in *Robotics: science and systems*, Philadelphia, 2006.

[9] Y. Wang, J. Yang, H. Liu, Y. Chen, M. Gruteser, and R. P. Martin, "Sensing vehicle dynamics for determining driver phone use," in *Proc. Mobisys'13*, ACM, 2013.

[10] Z. Chen, J. Yu, Y. Zhu, Y. Chen, and M. Li, "D3: Abnormal driving behaviors detection and identification using smartphone sensors," in *Proc. SECON'15*, IEEE, 2015.

[11] Q. Pu, S. Gupta, S. Gollakota, *et al.*, "Whole-home gesture recognition using wireless signals," in *Proc. MOBICOM'13*, ACM, 2013.

[12] D. Lee, S. Oh, *et al.*, "Drowsy driving detection based on the driver's head movement using infrared sensors," in *Proc. ISUC'08*, IEEE, 2008.

[13] C.-W. You, N. D. Lane, F. Chen, *et al.*, "Carsafe app: alerting drowsy and distracted drivers using dual cameras on smartphones," in *Proc. Mobisys'13*, ACM, 2013.

[14] D. Chen, K.-T. Cho, S. Han, Z. Jin, and K. G. Shin, "Invisible sensing of vehicle steering with smartphones," in *Proc. Mobisys'15*, ACM, 2015.

[15] S. Gupta, D. Morris, S. Patel, and D. Tan, "Soundwave: using the doppler effect to sense gestures," in *Proc. SIGCHI'12*, ACM, 2012.

[16] J. Yang, S. Sidhom, G. Chandrasekaran, T. Vu, H. Liu, N. Cecan, Y. Chen, M. Gruteser, and R. P. Martin, "Detecting driver phone use leveraging car speakers," in *Proc. MOBICOM'11*, ACM, 2011.

[17] S. Yun, Y.-C. Chen, and L. Qiu, "Turning a mobile device into a mouse in the air," in *Proc. Mobisys'15*, ACM, 2015.

[18] R. Nandakumar, S. Gollakota, and N. Watson, "Contactless sleep apnea detection on smartphones," in *Proc. Mobisys'15*, ACM, 2015.

[19] H. Han, J. Yu, H. Zhu, Y. Chen, J. Yang, Y. Zhu, G. Xue, and M. Li, "Senspeed: Sensing driving conditions to estimate vehicle speed in urban environments," in *Proc. INFOCOM'13*, IEEE, 2014.