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

***Abstract*—Real-time driving behaviors monitoring is a corner stone to improving driving safety. Existing works on driving behaviors monitoring using smartphones only provide detection results after driving behaviors finished, which is not sufficient for avoiding car accidents. In this paper, we leverage audio devices on smartphones to realize early recognition of inattentive driving events including *Fetching Forward*, *Picking Drops*, *Turning Back* and *Eating or Drinking*. Through analyzing driving traces collected in real driving environments, we find that each type of inattentive driving events has unique patterns on the structure of Doppler profiles of the audio signals. Based on the observation, we propose an *Early Recognition* system, *ER*, which can achieve real-time inattentive driving recognition and alert drivers at the early stage of driving events. According to effective features extracted from 4-month driving traces in real driving environments, we utilizes machine learning methods to generate binary classifiers for every pair of inattentive driving events. Then a *modified vote mechanism* is proposed to form a multi-classifier for all inattentive driving events. Afterwards, we turn the multi-classifier model into a *gradient model forest* for early recognition. Through extensive experiments with 8 volunteers driving for another 2 months in real driving environments, ER achieves an average total accuracy of 94.80% for recognition and over 80% inattentive driving events can be recognized before 50% of the finishing time.**

I. Introduction

Inattentive driving [1] has been a significant factor in traffic accidents 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, according to a research [4], since some inattentive driving events are unapparent and easy to be ignored by drivers, most drivers fail to recognize themselves inattentive while driving. Therefore, it is necessary to build an inattentive driving recognition system to alert the drivers in real-time, helping to prevent potential car accidents and correct drivers' bad driving habits.

There has been some existing detection works [5] [6] [7] on abnormal driving behaviors including inattentive driving, drowsy driving and drunk drinking. These works focus on detecting driver's status based on pre-deployed infrastructure, such as cameras, infrared sensors, and EEG devices, which incur high cost. In recent years, with the 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, the existing works on driving behaviors detection using smartphones can only provide a detection result after a specific driving behavior finished, which is less meaningful for drivers to avoid car accidents.

Moving along this direction, we need to consider an approach to recognize inattentive driving at the early stage using smartphone sensors without any additional hardwares. According to the judicial interpretation of inattentive driving [1], there are four of the most commonly occurring inattentive driving events of drivers, i.e. *Fetching Forward*, *Picking Drops*, *Turning Back* and *Eating or Drinking*. Therefore, our goal is to recognize these inattentive driving events and alert drivers as early as possible to prevent drivers from finishing these dangerous driving events. Since human actions may lead to Doppler shifts of signals [11], our work uses the Doppler shifts of audio signals generated by smartphones to recognize different inattentive driving events. To realize the inattentive driving recognition, we face several challenges in practice. Firstly, patterns of inattentive driving events need to be identified from audio devices on smartphones. Secondly, inattentive driving events should be recognized as early as possible under the guarantee of recognition accuracy. Finally, the solution should be effective in real driving environments and computational feasible on smartphones.

In this paper, we first investigate the Doppler shifts of audio signals caused by four inattentive driving events. Through empirical studies of driving traces collected from audio devices on smartphones in real driving environments, we find that each type of inattentive driving events has unique patterns on the structure of Doppler profiles. Based on the observation, we propose an *Early Recognition* system for inattentive driving events, *ER*, which can realize a real-time inattentive driving detection and alert drivers at the early stage. In ER, effective features of audio signals are first extracted through *Principal Components Analysis (PCA)* based on 3-month driving traces in real driving environments involving 8 drivers. 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. Furthermore, for detecting the inattentive driving at the early stage, we analyze the relationship between the completion degree and time duration for each type of inattentive driving events and turn the multi-classifier model into a *Gradient Model Forest*. Additionally, ER is easy to implement and computational feasible on standard smartphones platforms. Our extensive experiments validate the accuracy and the feasibility of using our system in real driving environments.

We highlight our main contributions as follows:

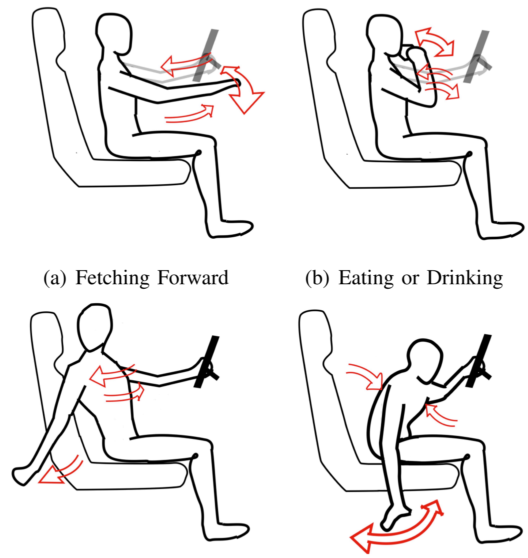
- We capture the unique patterns on the structure of Doppler profiles through empirical analysis of driving traces collected from audio devices on smartphones in real driving environments.
- We develop a system, ER, which can realize a real-time inattentive driving recognition and alert drivers at the early stage using smartphones only.
- We propose a modified vote mechanism based on SVM to generate a high accuracy multi-classifier for inattentive driving events, and further present a gradient model forest for early recognition.
- We conduct extensive experiments in real driving environments. The result shows that ER achieves an average total accuracy of 94.80% for recognition and over 80% inattentive driving events can be recognized before 50% of the finishing time.

The rest of the paper is organized as follows. The related work is reviewed in Section II. Patterns of inattentive driving events from Doppler profiles of audio signals are analyzed in Section III. Section IV presents the design details of ER. Issues and their solution in the implementation of ER are showed 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 the driving events detection by using professional infrastructure, such as EEG [5] and water cluster detector [6], or common sensors such as infrared sensors [12] and cameras [7]. However, the solutions all rely on pre-deployed infrastructures and additional hardwares that incur installation cost. Moreover, those additional hardwares 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 the driving event 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] leverages rear camera of smartphones to monitoring the road condition, and [13] leverages the dual cameras of smartphones to tract road conditions and detect drivers' statement at the same time. However, the accuracy of vision-based approaches depends on weather, lighting and smartphones placement, so it is unstable and infeasible. In sensor-based solutions, [10] leverages accelerators and gyroscopes to detect abnormal driving behaviors, and [14] combines the sensors by using Inertial Measurement Units (IMUs) on smartphones to detect various steering maneuvers of a vehicle. These two solutions can only provide detection results after driving events finished. Besides, [9] uses accelerators of smartphones to determining



(a) Fetching Forward (b) Eating or Drinking

(c) Turning Back (d) Picking Drops

Fig. 1. Illustration of Inattentive driving events.

usage of phones while driving, but this work can not recognize other inattentive driving but usage of phones.

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. Still, none of these works could realize recognition at the early stage of events. Our work, however, achieves early recognition of inattentive driving events, which is more meaningful for safety in real driving environments.

III. Inattentive Driving Events Analysis

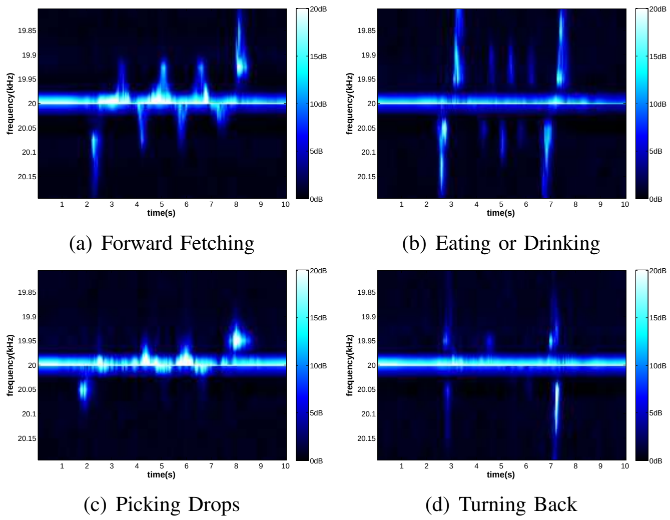
Our work is built on audio devices embedded in smartphones. In this section, we first give a brief introduction to inattentive driving events, and then analyze patterns of the events on Doppler profiles of audio signals.

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.

Fetching Forward: this state refers to the condition where driver fetches out to search widgets, such as phones, car audio consoles, etc.

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



(a) Forward Fetching (b) Eating or Drinking

(c) Picking Drops (d) Turning Back

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

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

Eating or Drinking: drivers eating snacks or replenishing water when driving.

Through analyzing the above four inattentive driving events, we realize each driving event is a consecutive action, not a transient action. 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. The *Doppler shift* (or Doppler effect) is the change in frequency of a wave (or other periodic event) from an observer moving relative to its source. Specifically, a mass point moving at speed v and angle θ to the speaker, brings a change [19] in frequency, i.e.

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

where c and f_0 denotes the speed and frequency of the audio signals.

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 takes two factors into consideration, i.e. background noise and unobtrusiveness. According to [16], frequency range from $50Hz$ to $15,000Hz$ covers almost all naturally occurring sounds, and human hearing becomes extremely insensitive to frequencies beyond $18kHz$. Therefore we could straightforwardly filter the background noise and eliminate the effects for people by locating our signal above $18kHz$ Furthermore, a higher transmitter 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 giving inattentive driving events at the rate of $44.1kHz$ and then transform it into frequency domain using *N-points Fast Fourier Transform* (*FFT*). 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 consist of positive and negative Doppler shifts, the patterns are different across the four events in frequency amplitude, time duration, relations of the Doppler shifts, etc.

From above analysis, we find that each type of inattentive driving events has unique pattern on the structure of Doppler profiles. Although some existing work, [11] and [10], present human action recognition methods based on the unique patterns of actions already, the action recognition can only be done when the whole specific action finished, which is acceptable for transient actions like a single gesture 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.

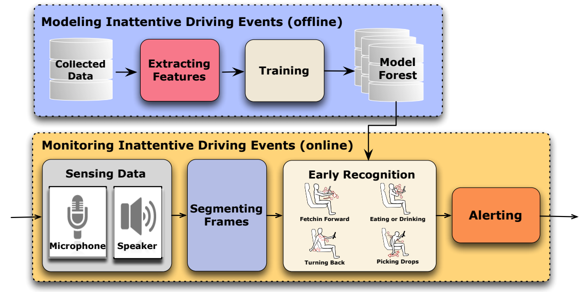


Fig. 3. System architecture and work flows.

IV. SYSTEM DESIGN

In order to monitor inattentive driving events effectively and efficiently, we present an inattentive driving event detection 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 infrastructures and additional hardwares.

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, effective features are extracted from the Doppler profiles of audio signals for different types of inattentive driving events based on traces collected from real driving environments. Then 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 events based on them. Afterwards, the multi-classifier model is turned into a *Gradient Model Forest* for realizing early recognition, which is stored in model database.

In the online part, ER senses real-time audio signals generated by a speaker and received by a microphone. 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. In Early Recognition part, ER extracts features from segments and identifies whether the event are at some early stages based on our 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.

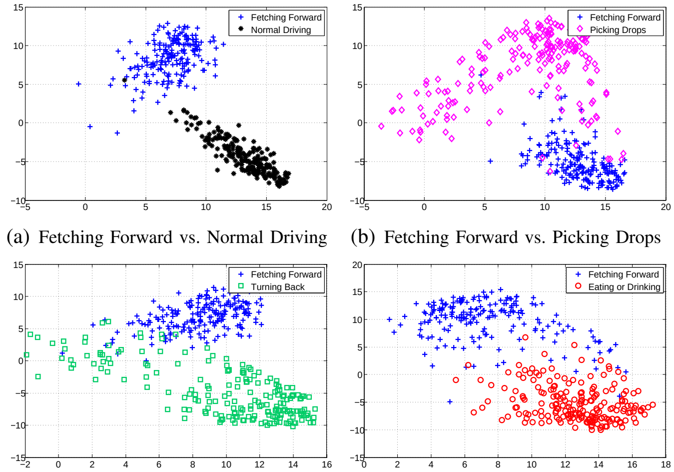
B. Model Training at Offline Stage

1) Establishing Training Dataset: To collect data in real environment, we develop an Android-based program to generate audio signals by the speaker and collect readings from the microphone, and then transform the raw data to the frequency-time Doppler profile of the audio signals.

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 collecting spans from October 2015 to January 2016, during which all the daily driving such as 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 is severed as the ground truth. Afterwards, we combined the Doppler profile of the audio signals of these samples into training dataset X.

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

In PCA algorithm, to extract features from data matrix X, a projection matrix W that contains direction 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, Σ is a $m \times n$ matrix



(a) Fetching Forward vs. Normal Driving (b) Fetching Forward vs. Picking Drops (c) Fetching Forward vs. Turning Back (d) Fetching Forward vs. Eating&Drinking

Fig. 4. Recognizing fetching forward events from other inattentive driving events and normal driving in 2-dimensional feature space.

with non-negative singular values on the diagonal. W is a $n \times n$ unitary matrix. 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

$$\min_{d} \quad \frac{\sum_{i=1}^{d} \sigma_i}{\sum_{i=1}^{n} \sigma_i} \geqslant t, \tag{2}$$

where σ_i is the i^th largest singular value of data matrix X, which denotes the importance of the i^th features, and t is the threshold of reconstruction, denoting the remaining percentage of 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 the information of the raw data. Fig.4 shows the distribution of Fetching Forward events versus other three inattentive driving events and normal driving in a 2-dimensional feature space. It is can be seen clearly 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 thus improve recognition accuracy, we train classifier models for inattentive driving events pairwise.

3) Training a Multi-Classifier: After features extracting through PCA, we train a multi-classifier for all inattentive driving events. We first use *Supportive Vector Machine* (*SVM*) to train binary classifiers for all inattentive driving events pairwise. Based on the binary classifier models, a voting mech-

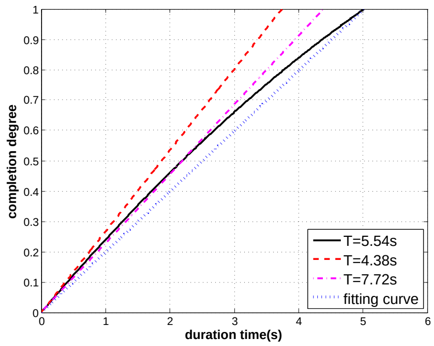


Fig. 5. The relationships between the completion degree α and duration time τ of fetching forward events under different the complete time T.

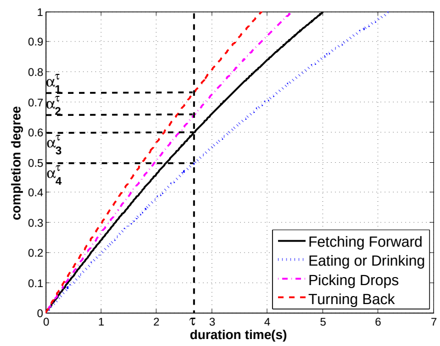


Fig. 6. The relationships between the completion degree α and duration time τ for four kinds of inattentive driving events.

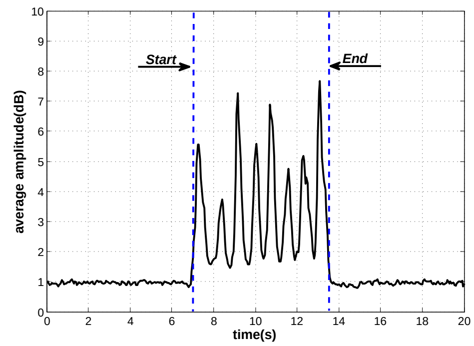


Fig. 7. The average amplitude of frequency bands except the pilot frequency during a 20 seconds driving containing a fetching forward event.

anism is proposed to form a multi-classifier to differentiate all four inattentive driving events.

Given that each binary classifier has one vote $v \in \{0, 1\}$ for the multi-classifier. Considering a binary classifier for separating event a from event b, 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 model 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 votes of a 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 a event which get through the multi-classifier and gets a classification result c from Eq.4, if the event does not satisfy Eq.5, we consider the event as *other driving events*, such as shifting gear, pushing glasses, etc, which are not so danger for drivers.

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 event 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 α of the event 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 τ denotes the time duration of e at time t and T denotes the finishing time duration 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 α_e is as small as possible. As a result, based on different α of inattentive driving events at different τ, we set up a bunch of classifiers for early recognition, i.e. the gradient model forest.

For modeling the complete degree α at different time duration τ of inattentive driving events, the variation of the finishing time duration T among all events should be considered. For different types of events, T is difference because of the nature difference of the events themselves, so we set different models for different types of events. Moreover, for a specific type of inattentive driving events, T is also difference depending on the driving situation. We also need to take this variation into consideration when setting models.

According to statistics of the dataset established in Section IV-B1, the finishing time durations T of 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, over 95% of Fetching Forward events have T from $3.74s$ to $5.02s$. As shown in fig.5, based on the completion degree-duration time relations of Fetching Forward events having T equals to $3.74s$, $4.38s$ and $5.02s$, a quadratic curve is fit to modeling the relation between α and τ of the Fetching Forward event, which starts at the origin, goes through the mid-point of the line of $4.38s$ and ends at the end of the line of $5.02s$. 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. With the similar analysis, ER models a relationship between α and τ for each type of inattentive driving events. Fig.6 shows relationships between α and τ for all four types of inattentive driving events.

With the relationships between α and τ, for any given time duration τ, ER can gets 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 τ, as shown in Fig.6. According to A^τ, ER segments the Doppler profiles of all types of inattentive driving events

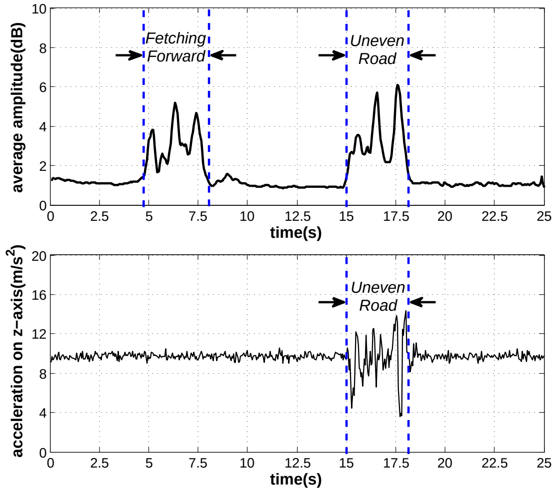


Fig. 8. The average amplitude of audio signals and readings from the accelerator's z-axis in 25-second driving which contains an Fetching Forward event and a driving condition that the vehicle goes across a speed bump.

$X = \{X_1, X_2, \cdots, X_k\}$ and then gets the new input dataset $X^\tau = \{X_1^\tau, X_2^\tau, \cdots, X_k^\tau\}$. Selecting n different τ by gradient, we form a n-element time duration set $T = \{\tau_1, \tau_2, \cdots, \tau_n\}$. ER then segment the Doppler profiles based on T and end 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^τ, patterns of inattentive driving events are guaranteed to be unique, ER can always get a multi-classifier θ^τ. 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, where each of θ^τ is a matrix containing all binary classifiers as

$$\theta^\tau = \begin{bmatrix} -(\theta_1^\tau)^T- \\ -(\theta_2^\tau)^T- \\ \vdots \\ -(\theta_m^\tau)^T- \end{bmatrix}, \tag{7}$$

where $m = C_k^2$ for all pairwise inattentive driving events. Specially, the last multi-classifier model of the model forest, i.e. θ^{τ_n}, is the multi-classifier model for recognizing inattentive driving events after they finished. Finally, we obtain a gradient model forest Θ, which could be used online to realize early recognition for inattentive driving events.

C. Recognizing Inattentive Driving Events at Online Stage

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

As mentioned in Section III-B, all driving events of drivers 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 data collected in real driving environments, we find that when an event begins, the average amplitude of frequency bands except the pilot frequency rises and keeps a relatively

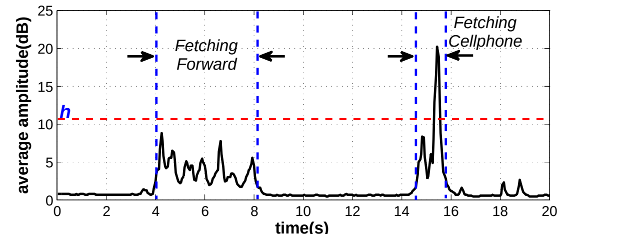


Fig. 9. The average amplitude of audio signals in 20 seconds driving which contains an inattentive driving event and a event of fetching smartphones.

high value until the end of the event. Fig.7 shows that the average amplitude of frequency bands except the pilot frequency during a 20-second driving containing a fetching forward event. From fig.7, it can be seen that the average amplitude pattern for events is much greater than that without events.

ER can employs the average amplitude to capture Doppler shift caused by driving events. Specifically, sliding window method is used here. ER keeps computing the average amplitude of frequency bands 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 segment a frame from the start point to current point and sends the frame to the offline-trained model forest for early driving event recognition at short time intervals, until a recognition result is output or the end point is detected.

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

For a frame e of length τ ($\tau > \tau_i$ and $\tau < \tau_i + 1$), 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 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 may cut the event and use the last classifier of the model forest, θ^{τ_n}, to finish the recognizing and get the corresponding output. For each event, ER records the output recognition result and the result from classifier model θ^{τ_n}. The proposed mechanism can reduce incorrect driving event recognition effectively, and avoid disturbing drivers from false warning.

V. System Implementation

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

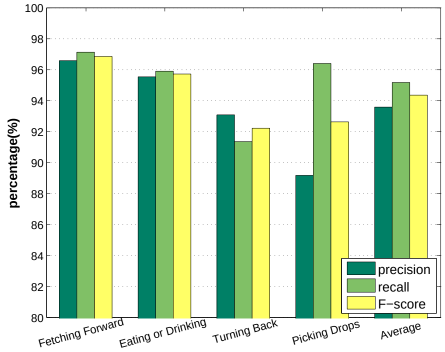


Fig. 10. The Precision, Recall and F-Score for all types of inattentive events.

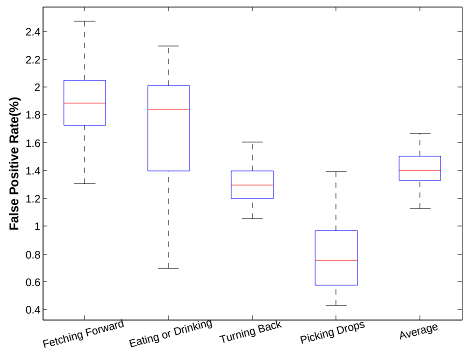


Fig. 11. Box plot of False Positive Rate of all types of inattentive events over 8 drivers.

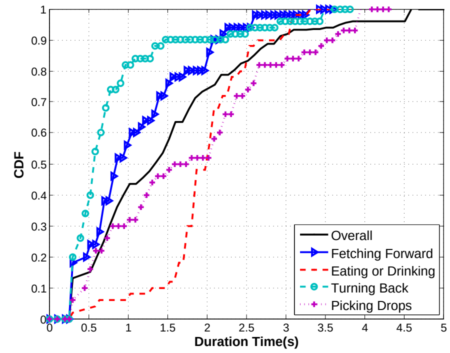


Fig. 12. The CDF of recognition time for all types of inattentive events.

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 condition or other temporary situation. Under this condition 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 [20] presents a result that the data pattern 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 can collects reading from the accelerometer on smartphones to sense stops of a vehicle in real time. Once ER detects that a vehicle is stopped, it does not analyze the audio signals until the vehicle moves again.

B. Filtering Influence of Uneven Road

Uneven road may result in strong vibration on smartphones, which could influent the audio signals collected by the microphone and causes mistaken recognition of ER. It is necessary to separate the Doppler shifts caused by the uneven road from that caused by inattentive driving events. Based on our observation from the collected data in real driving environments, the strong vibration can also be reflected on the acceleration on vehicle's z-axis, so uneven road can be sensed by the readings from accelerator's z-axis. Fig.8 plots the readings from accelerator's z-axis and the average energy amplitude described in fig.7 in a 25 seconds driving which contains an Fetching Forward event and a driving condition that a 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 not influent the acceleration on the z-axis, while an uneven road brings Doppler shifts along with great jitter to the acceleration on the z-axis. Based on the observation, ER could separate the Doppler shifts caused by uneven road from that brought by driver's events.

C. Preventing Usage of Phone while Driving

Using smartphones, such as calling, texting message, browsing webpages, etc, is very danger when driving, so it should be prohibited. Since the usage of smartphones may greatly influent audio signals by changing the position and post of smartphone over time, ER regards usage of phone as a special case of inattentive driving events to detect and alert. Fig.9 shows the average amplitude of audio signals in 20 seconds driving which contains an inattentive driving event and an event of fetching smartphones. From fig.9, it can be seen that when driver fetches a smartphone, there is a remarkable Doppler shift, which the average amplitude is far greater than other driving events. Therefore, ER could recognize usage of smartphones by detecting the event of fetching smartphones, i.e., once Doppler shift of audio signals is greater than a threshold h, ER regards the driving event as usage of smartphones and sends an alert to drivers.

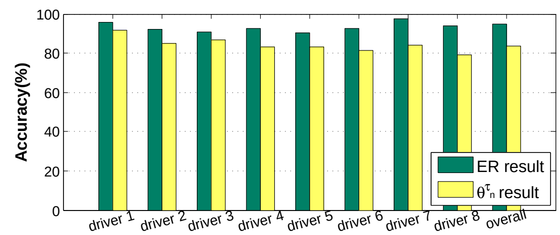


Fig. 16. The total accuracy of ER and classifier model θ^{τ_n} over 8 drivers.

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 different drivers with distinct vehicles in real driving environments to collect data 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 the same method described in Section IV-B1, we obtain a test set with 2473 inattentive driving events to evaluate the performance of ER.

A. Metrics

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

- *Accuracy*: The probability that an event is correctly identified for all type of events.

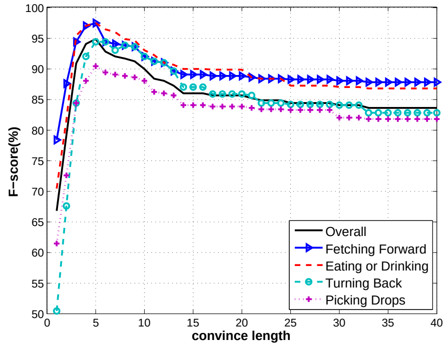


Fig. 13. F-score under different convince length for all types of inattentive events.

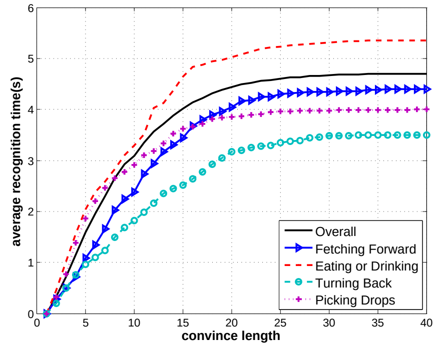


Fig. 14. average recognition time under different convince length for all types of inattentive events.

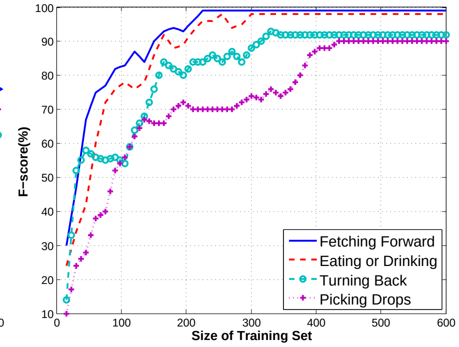


Fig. 15. F-score under different size of training set for all types of inattentive events.

- *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 combines precision and recall($2 \times \frac{Precision \times Recall}{Precision + Recall}$). We use F-score as our metric for evaluate the recognition accuracy for specific types of in attentive events in the following evaluation.

B. Overall Performance

Fig.16 plots the recognition accuracy of ER and the classifier model θ^{τ_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 θ^{τ_n} is 84.78%. Further, ER performs far better than θ^{τ_n} for any of the 8 drivers. The lowest accuracy for ER 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.10. 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%. Further, ER has a better performance with all three metrics above 95% when recognizing *Fetching Forward*and *Eating and drinking* events since their patterns are more distinctive as shown in fig.2.

Moreover, for each of the 8 drivers, we evaluate the FPRs of recognizing specific type of inattentive driving events. fig.11 shows the box-plot of the FPRs for each type of inattentive driving events and the average FPR. We can see from fig.11 that the highest FPR is no more than 2.4% and the average FPR is as low as 1.3% over the four events and eight 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 these events in Fig.12. It can be seen from fig.12 that 50% of all types of inattentive events are recognized before $1.4s$ and 80% can be recognized before $2.3s$, while the average finishing time 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 finishing time of all events. For each specific type of events, The 80%-recognized time is around $2s$, $2.5s$, $1.0s$ and $2.6s$ for Fetching Forward, Eating and drinking, Turning Back and Picking Drops, separately. And the corresponding average finishing time for the four events is $4.3s$, $5.4s$, $3.5s$ 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.13 shows the recognition performance of different types of inattentive driving events at different convince lengths. It can be seen that as convince length increases, the F-score of all 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 result, so that the output result is more accurate, which explains the increasing of F-score. However, as convince length keeps increasing, the output condition becomes too strict for ER to output a recognition result before the end of a event. As a result, more events need to be recognized through classifier θ^{τ_n}, which is less accurate than ER according to fig.16. Fig.14 shows the average recognition time of different types of inattentive driving events at different convince lengths. From fig.14, we observe that the average recognition time of all 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 convince length is greater. ER chooses 5 as the convince length by empirical studies.

D. Impact on Training Set Size

According to Section IV-B1, we collect 3532 inattentive driving events for training. Fig.15 shows the impact of training set size to the recognition performance of ER. From the plot, it can be seen that the F-score rises as training set size increases and after a certain training set size for each inattentive driving events, the F-score goes stable. Specifically, to get a stable F-score, ER needs 220 training examples for Fetching Forward, 300 examples for Eating or Drinking, 340

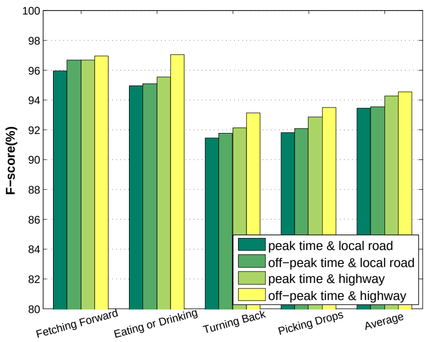


Fig. 17. F-score under different traffic condition and road types for all types of inattentive events.

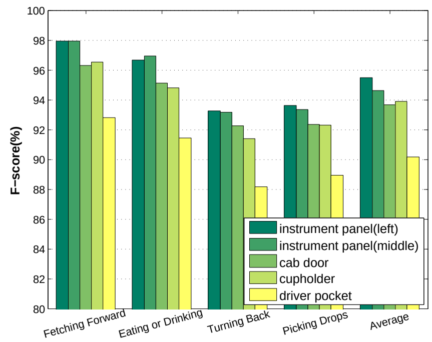


Fig. 18. F-score under different smartphone placement for all types of inattentive events.

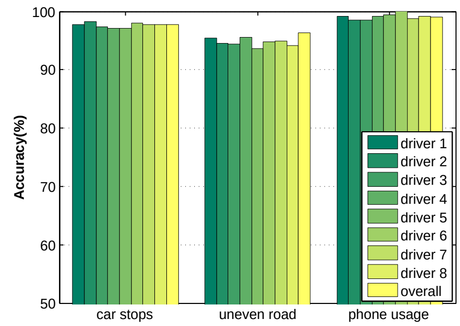


Fig. 19. Recognition accuracy for car stops, uneven road and phone usage over 8 drivers.

examples for Turning Back and 450 examples 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 evaluate the collected driving traces of different road types(local road and highway) and different traffic conditions(during peak time and off-peak time), separately. Fig.17 shows the F-score of recognizing different types of inattentive driving events at all four combinations of road types and traffic conditions. It can be seen that ER achieves fairly good F-score for recognition at any combination of road types and traffic conditions. In addition, during peak time, the F-score of ER is slightly lower than the F-score during off-peak time because heavy traffic condition may bring more stops for vehicle and more driving behaviors such as shifting gears, which brings more chances for ER to make wrong recognitions. Further, the F-score of ER when driving on highway is slightly higher than the F-score 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 experiment of 8 drivers, smartphones are placed on instrument panel(left side), instrument panel(middle part), panel near cab door and cup-holder, or placed in drivers' pockets. Fig.18 shows that ER can achieve fairly good F-score for recognition under different smartphone placement. Specifically, smartphones placed on instrument panel achieve best recognition result 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 smartphone. 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 smartphone has the recognition accuracy over 98% for all 8 drivers and thus ER can effectively prevent drivers from using smartphones while driving. For recognizing car stops and uneven road, the accuracy of ER is above 96% and 94%, respectively.

VII. Conclusion

In this paper, we address the problem of recognizing inattentive driving events at the early stage to improve driving safety. In particular, we propose an system, ER, to recognize different inattentive driving events as early as possible leveraging build-in audio devices, speakers and microphones, of smartphones. To recognize specific inattentive driving events, ER first extract effective features using PCA based on the patterns of specific types of inattentive driving events on Doppler profiles of audio signals. Then ER leverages SVM and a modified vote mechanism to form achieve a multi-classifier model and turns the multi-classifier model into a gradient model fores. We train our gradient model forest based on the dataset collected from 3-month driving traces in real environment. The extensive experiments in real environment of another 2 months show that ER achieves high accuracy for recognition and realizes recognizing at early stage. However, ER is not an approach without limitations, events from the person in passenger's seat may confuse ER and result in mistaken recognition results.

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] W. C. R. G. R. Majewski *et al.*, "Spotlight synthetic aperture radar: Signal processing algorithms," *Norwood, MA: Artech House*, 1995.

[20] 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.