SpeedTalker: Automobile Speed Estimation via Mobile Phones

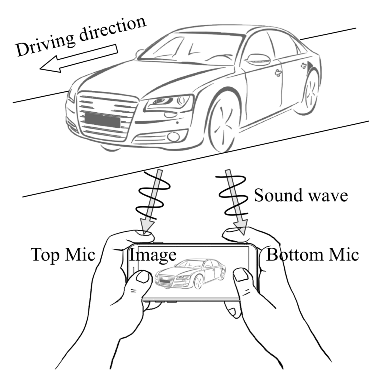
Xinran Lu, Lei Xie, *Member, IEEE,* Yafeng Yin, *Member, IEEE,* Wei Wang, *Member, IEEE,* Yanling Bu, *Member, IEEE,* Qing Guo, and Sanglu Lu, *Member, IEEE*

Abstract—Among all the road accidents, speeding is the most deadly factor. To reduce speeding, it is essential to devise efficient schemes for ubiquitous speed monitoring. Traditional approaches either suffers from using special equipment(e.g., radar speed gun) or special deployment(e.g., position-fixed cameras). In this paper, we propose SpeedTalker, a mobile phone-based approach to perform speed detection on automobiles. By leveraging the built-in microphones and camera from the mobile phone, SpeedTalker estimates the automobile speed by passively sensing the acoustic and image signals. We propose an integrated solution to effectively estimate the automobile's speed based on COTS devices, and provide a platform for every pedestrian to help report the speeding event of automobiles. Specifically, we use the time difference of arrivals (TDOA) model based on acoustic signals to figure out the candidate trajectories of automobile, and use the pin-hole model based on image frames to figure out the vertical distance between the user's position and the automobile's trajectory, thus to estimate the unique trajectory. Combined with the time stamp of the trajectory, the automobile speed can be estimated. Besides, we propose a method to effectively mitigate the influence of the movement jitters of mobile phone. We implemented a system prototype for SpeedTalker and estimated the automobile speed with high accuracy. Experiment results show that in the scenario of single automobile, SpeedTalker can achieve an average estimation error of 6.1% compared to radar speed guns. In the scenario of multiple automobiles, SpeedTalker can achieve an average estimation error of 9.8%, which is acceptable for usage.

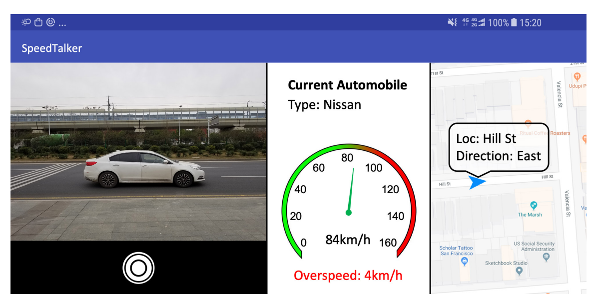
1 Introduction

1.1 Motivation

Nowadays, more and more traffic violations occur due to the increase of the automobile, e.g., in 2016, the number of the road traffic deaths reached 1.35 million. Among all kinds of the traffic violations, speeding is the most deadly factor[1]. Appropriate reductions in speed can reduce fatal and serious crash risk to prevent death and serious injury[2]. To reduce speeding, it is essential to devise efficient schemes for ubiquitous monitoring on traffic. Traditional ways to monitor the traffic are using speed radar or using cameras. However, they are costly and inconvenient since they need wide deployment of special equipment. As a result, a low-cost and mobile solution to measure the speed is needed. It is noted that, the mobile phones embedded with many kinds of sensors, such as cameras and microphones, have become indispensable in daily life. By utilizing the built-in sensors, we can propose a method to measure the automobile speed with mobile phones. Specifically, we can use the microphones and camera to recover the trajectory of the automobile and estimate the speed. IMU sensors are utilized to remove jitters to raise the accuracy of the system. In this way, every pedestrian can help to monitor the traffic condition with his/her mobile phone. Furthermore, all people can participate in the activities of reporting traffic conditions by sufficiently applying the crowdsourcing method [3].



(a) Illustration of the system.



(b) The application of the system.

Fig. 1: Application scenario of SpeedTalker.

A typical scenario of SpeedTalker is as follows. In the speed prone areas, the pedestrians who volunteer to monitor the traffic can arrive at the area in advance and continuously record the acoustic and the visual signals of the au-

- *Xinran Lu, Lei Xie, Yafeng Yin, Wei Wang, Yanling Bu, Qing Guo and Sanglu Lu are with the State Key Laboratory for Novel Software Technology, Nanjing University, China.*
  E-mail: luxinran@smail.nju.edu.cn, lxie@nju.edu.cn, yafeng@nju.edu.cn, ww@nju.edu.cn, yanling@smail.nju.edu.cn, guoqing@smail.nju.edu.cn, sanglu@nju.edu.cn.
- *Lei Xie is the corresponding author.*

tomobiles. The pedestrians only need to use the system for a few minutes to collect the traffic speed information in this period. SpeedTalker estimates the speed of the automobiles and collects the speeding related information. Traffic speed information will be uploaded onto the server of the related department. With the help of volunteers, data from different regions at different time then can be analyzed for traffic control. The distributions of traffic police and equipment can be optimized and the drivers and pedestrians can be warned of danger when moving in this area.

1.2 Limitation of Prior Art

There exist two main approaches to measure the speed of the automobiles. One approach is to use the fixed devices to measure the speed of the automobiles. The cameras and coils are traditional fixed devices for speed detection. They can monitor whether there exist automobiles at two pre-set locations. If the automobile passes the two corresponding locations, the system then records the time interval the automobile uses. Thus the speed of the automobile can be easily estimated. However, if the fixed speed measurement devices are widely deployed to monitor the traffic, the cost is unacceptable. Besides, the drivers can easily figure out whether there exist speed measurement devices since their positions are fixed. Moreover, each speed detection camera needs its own parameters to estimate the speed of the automobiles. The height, gesture and the field of view(FOV) determines the detection region of the camera deployed on the traffic pole. This makes the estimation simple but can only work for the specific camera.

Another approach to measure the speed of the automobile is to use portable devices, such as radar speed gun[4] or lidar[5]. Radar speed guns use *Doppler Effect* to perform speed measurement. They send out a radio signal in a narrow beam, then receive the same signal back after it bounces off the target object. If the object is moving, the frequency of the radio waves change. According to the difference between the reflected radio waves and transmitted waves, the speed of the object can be calculated. However, there exist limitations when using these portable devices. First, special devices are needed to emit the directional modulated electromagnetic waves in certain frequency. This increases the cost of the hardware and prohibits it to be widely used by ordinary people. Second, the electromagnetic wave emitted by the equipment can be easily detected by radar detector in the automobile. Usually this makes them fail to capture the speeding event, since the automobiles may intentionally slow down when they pass by.

Therefore, in order to make every pedestrian become potential speeding inspectors, it is essential to leverage portable daily devices, such as mobile phone, and propose easy-to-use measurements to measure the speed of automobiles. In fact, by sufficiently using the embedded sensors like the microphones and cameras, we can effectively use the mobile phones to measure the automobiles' speed.

1.3 Our Approach

In this paper, we propose SpeedTalker, a mobile phone-based approach to perform speed detection on automobiles. Instead of using special devices, the pedestrian on the sidewalk can utilize mobile phones' built-in microphones and camera to estimate the speed of the automobile. IMU sensors are utilized to compensate the jitters caused by users. Figure 1 illustrates the application scenario of the system. To perform speed detection, the user needs to hold the mobile phone in *landscape orientation* as shown in the figure, i.e., the top microphone and the bottom microphone are placed in a left-and-right manner. When the automobile passes by, both two microphones record the sound of the automobiles. And the camera records the movement of the automobile. According to the measurements from these two kinds of sensors, SpeedTalker estimates the speed of the automobiles. Specifically, during the process when the automobile is passing by, the sound wave reaches the top and bottom microphones at different time, respectively. According to the time difference of arrivals (TDOA) derived from acoustic signals obtained by different microphones, SpeedTalker estimates the *candidate trajectories* of the automobile as a set of hyperbolas. According to the obtained frames from the camera, SpeedTalker estimates the *vertical distance* between the user's position and the automobile's trajectory, by referring to the pin-hole model of the camera. Then, the *trajectory* of the automobile can be determined from the candidates by referring to the unique *vertical distance*. Combined with the temporal information in acoustic signals, SpeedTalker is able to estimate the speed of the automobiles. Besides, since the mobile phones are held in hands, the jitters may cause rotation and translation of the mobile phones. IMU sensors can be used to compensate the translations and rotations and reduce the errors.

1.4 Challenges

There are three main challenges in our work. The first challenge is to propose a passive sensing method to measure the speed of automobile. Passive sensing means the detection system does not actively transmit any detecting signals, such as ultrasonic and flash light. Active sensing has two limitations for the speeding detection. First, the active signals, e.g., the electromagnetic wave, can be easily detected by the radar detectors. Second, an ultrasonic wave or flash light actively generated by the mobile phone will be dramatically attenuated when it is transmitted outdoors. To address this challenges, we propose a passive sensing method to estimate the speed of the automobile, by utilizing two microphones and one camera in the mobile phones. Instead of actively transmitting the modulated signals and receiving the reflected signals, our solution only collects the acoustic signals and the image frames from the automobiles in a passive manner. The trajectory of the automobiles can be estimated by the acoustic signals from the two separated microphones and the image frames from the camera. Combined with the timestamp of the trajectory, the speed of the automobiles can be estimated.

The second challenge is to derive the automobile speed from the complicated acoustic signals. The complication of the acoustic signals comes from two aspects. On one hand, the automobile noises are made up of many parts, including the tire noise, engine noise, exhaust noise, wind noise, etc [6]. These noises are mixed not only in time domain but also in frequency domain. Therefore, it is hard to separate different noises with two built-in microphones of mobile phones. On the other hand, there might be many kinds of noises in the environment, especially for the sound of other automobiles on the road. It is hard to remove the environment noises, since the frequencies of other automobiles mainly lie in very close frequency band with the target automobile. To address this challenge, we consider the acoustic signals at full frequency as a whole. We utilize the cross-correlation of the acoustic signals from the top and bottom microphones to estimate the time difference of arrivals (TDOA). As the automobile is continuously moving, we can obtain a series of time delays through TDOA at different time. The candidate trajectories of the automobile can be estimated as a set of hyperbolas according to the curve of the time delay. Thus the automobile speed can be further estimated.

The third challenge is to estimate the speed of multiple automobiles. We can not separate the sound of multiple automobiles. Therefore, when multiple automobiles pass through the mobile phone, it is challenging to estimate the speed. To address the challenge, we utilize the multiple peaks in the cross-correlation figures between the top and bottom microphones. Then we may recover the delay curve of each automobile and calculate the speed of the automobiles.

1.5 Contributions

This paper makes four contributions: First, this is the first work that estimates the automobile speed via mobile phones through passive sensing of acoustic and image signals. We propose an integrated solution to effectively estimate the automobile's speed based on commercial off-the-shelf(COTS) devices, and provide a platform for every pedestrian to help report the speeding event of automobiles. Second, we use the time difference of arrivals (TDOA) model based on acoustic signals to figure out the candidate trajectories of automobile, and use the pin-hole model based on image frames to figure out the vertical distance, thus to estimate the unique trajectory. Combined with the timestamp of the trajectory, the automobile speed can be estimated. Third, we implemented a system prototype for SpeedTalker and estimated the automobile speed with high accuracy. The system works in the outdoor environment and effectively mitigates the ambient environmental interference. Experiment results show that SpeedTalker can achieve an average estimation error of 6.1% in the scenario of single automobile. In the scenario of multiple automobiles, SpeedTalker can achieve an average estimation error of 9.8%.

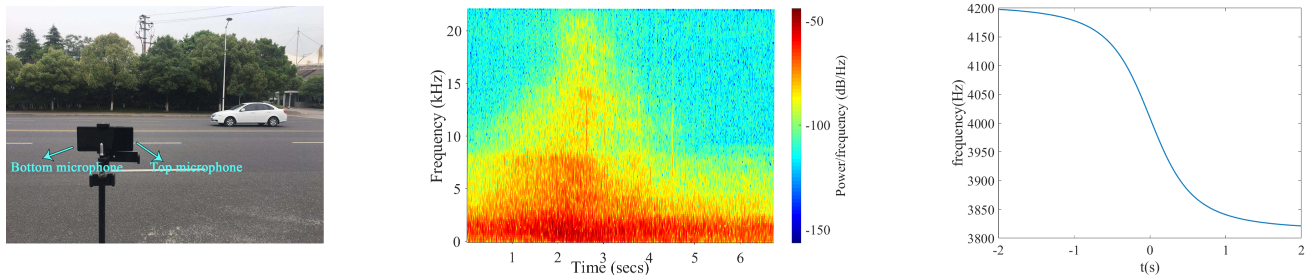
2 Related Work

Automobile detection via visual signals: Traditional approaches utilize cameras to calculate the speed of the automobiles. Kumar[7] and Czajewski[8] use computer vision based technologies to detect automobiles. The cameras are deployed in fixed positions and gestures above the street. As a result, the detection region is known and fixed. That means the moving distance of the automobiles can easily be acquired. Then the speed of the automobiles can be calculated. However, the scenarios of SpeedTalker is different from that of traditional visual approaches. The mobile phones are at the sidewalk and the positions and gestures are unknown. So novel approaches utilizing mobile phones to calculate the speed of the automobiles are needed. To get the relative position information between automobiles and mobile phones, we need to use cameras inside the mobile phones, which is analogous to knowing the position and gestures of the cameras in traditional CV based approaches. Apart from distance calculation, SpeedTalker utilizes acoustic signals to estimate the candidate trajectory of the automobiles. There are two advantages of acoustic signals over the visual signals. Firstly, the detection region of acoustic signals is broader than that of visual signals. Common cameras inside the microphones usually have narrow field of view(FOV). For example, the wide-angle camera of Samsung Galaxy Note 8 has $77°$ field of view. If we utilize the microphones of Samsung Galaxy Note 8 to detect automobiles, the detection field of view is around $160°$ according to the hyperbola model we propose. Secondly, compute complexity of acoustic signals processing is much lower than that of visual signals. If visual signals are utilized to complete the same work, each frame of the videos should be processed. The compute complexity of the processing is unacceptable.

Automobile detection via mobile phones: Automobile detection is an important research area since undetected automobiles are likely to endanger human life. Mobile phones can be utilized to inform the users of the approaching automobiles. There are three approaches to sense the automobiles with mobile phones. The first approach is to install applications both on the automobiles and the mobile phones. Oki Electric Industry Co. Ltd. develops a mobile phone that notifies the users of the presence of the automobiles using DSRC[9]. Car-2-X utilizes ad-hoc and cellular networks to inform the pedestrians of the automobile with the same method[10]. The second approach is to sense the moving automobiles via images. Sivaraman proposed a general active-learning framework for on-road automobiles recognition and tracking based on videos[11]. Wang proposed WalkSafe, a mobile phone application based on the back camera to sense the automobiles[12]. The drawback of these work is that image processing needs huge calculating resources. And the camera of the mobile phone is needed to face the road, which makes the detection inconvenient. The third approach is to utilize acoustic signals to sense the automobiles. Tsuzuki proposed an automobile sound detection system for a mobile phone[13]. Takagi introduced a hybrid and electric vehicles detection system[14], which focused on switching noise of the electric motor. So they failed to detect automobiles other than these types. Li proposed Auto++, a system that detects approaching automobiles for smart phone users to detect all kinds of automobiles via overall acoustic signals[15]. However, all these works can only inform the user of the approach of the automobiles and can not estimate the speed of the automobile.

Sensing via acoustic signals with mobile phones: Sensing with daily equipment is a popular issue. Sound waves can easily be transmitted and received by daily equipment, such as mobile phones and smart watches. Much work based on sound wave has been published. AAMouse measures the Doppler Shift of the sound waves transmitted by a mobile phone to track the phone itself with an accuracy

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.3034354, IEEE Transactions on Mobile Computing



(a) Empirical study setup.

(b) STFT of acoustic signals when the automobile passes by.

(c) S-shaped curve.

Fig. 2: Simple analysis on acoustic signals.

of 1.4cm[16]. Wang proposed a device-free gesture tracking method using acoustic signals[17]. It has a tracking accuracy of 3.5mm and 4.6mm respectively for 1-D hand movement and 2-D drawing in the air. ApenaApp, uses chirp signals to detect the changes in reflected sound that are caused by human breaths[18]. The system applies FFT over the acoustic signals to monitor the periodical movements that have frequency lower than 1Hz. All these works need to transmit active sound wave to sense objects. However they do not work if they are applied outdoors in a long distance with powerful environmental noise. Above all, calculating the speed of the automobile with a mobile phone in ourdoor environment is quite challenging.

Distance perception via cameras: Distance perception is demanded in computer vision technology to optimize the algorithm and enhance the performance. Traditional approaches to estimate the distance between the object and the camera is to use binocular system to calculate the depth. Hartley gives detailed view geometry in computer vision for distance calculating[19]. Tram utilizes two cameras mounted in the automobiles to capture LED light and estimate the distance between vehicles[20]. However, the approach is not suitable for our scenario. Although some mobile phones have multiple cameras at the backside, the cameras have its own roles. Some cameras have wide-angle lens, some have telephoto lens and some have infrared lens. They may not work together at the same time. Moreover some mobile phones only have one camera at the backside. Some other papers use one camera to estimate the distance. Diaz-Cabrera utilizes one camera to estimate the distance between the automobile and the traffic light[21]. They need to know the height of the traffic light and the parameter of the cameras in advance. Rahman utilizes one camera to estimate the distance between the user and the camera[22]. They also need to know the distance between the eyes and the parameters of the cameras. Our approach uses similar view geometry to calculate the distance and we can get the real diameter of the wheel hub through machine learning approaches.

3 Empirical Study and Modeling

3.1 Acoustic Signal Study

3.1.1 Measurement of Acoustic Signals via Mobile Phones

In order to study the relations between the acoustic signals and the speed of the automobile, we need to collect acoustic signals when automobiles pass by. To avoid the influence of jitters from the mobile phone, we deploy the mobile phone on a tripod. As shown in figure 2a, the tripod is set at one side of the road with its camera facing the road. And the mobile phone is in the landscape orientation. The mobile phone is about 1.5m above the ground, and about 8m away from the lane. The mobile phone records the sound when the automobile passes by. The sampling rate f_s of the sound in empirical study is 44.1kHz.

3.1.2 Doppler Effect

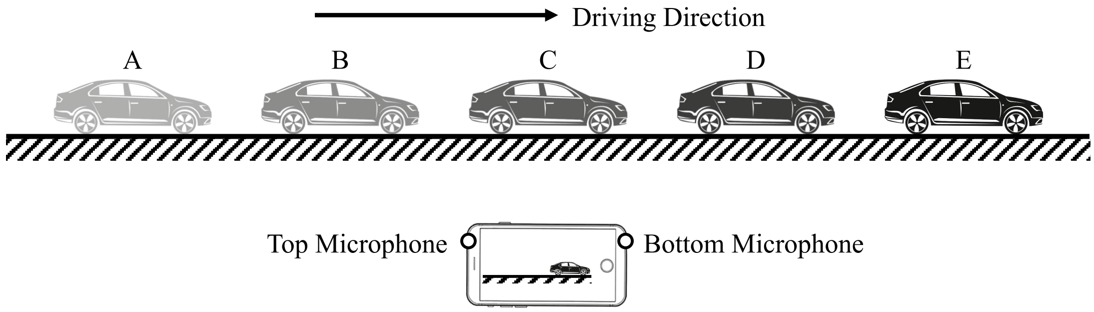
The usual way to estimate the speed of the moving object is to utilize *Doppler Effect*. If we already know the frequency f of the original wave, the frequency f' of real-time wave should be given by:

$$f' = \frac{C^2 f}{C^2 - v^2} \left\{ 1 - \frac{v^2 t}{\sqrt{C^2 v^2 t^2 + l^2 (C^2 - v^2)}} \right\} \quad (1)$$

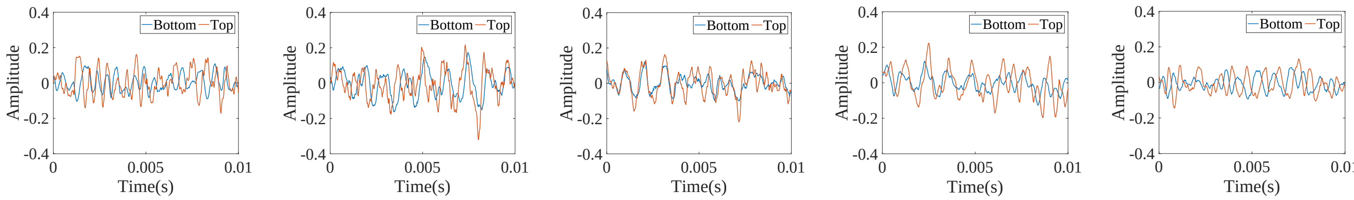
where C is the velocity of sound, v is the velocity of the automobile, l is the closest distance between the mobile phone and the automobile, and t is the time[14]. The distance between the mobile phone and the automobile is shortest at $t = 0$. To calculate the speed of the automobile, one of the problems is to find the original frequency f and real-time frequency f' of a specific sound wave.

First we focus on the original frequency f of the moving automobile. Since active sensing does not work in our scenario, we do not transmit sound wave in specific frequency. As a result, we have to analyze the sound made by the automobile to find the original frequency. In fact, automobile noises include tyre noise, engine noise, wind noise, exhaust noise, wind noise and so on. The frequency of tyre noise is widely distributed. The peak part locates between $315Hz$ and $1000Hz$[23]. The engine noise is dominated by the rotation speed of the engine. The frequency of the engine noise is mainly distributed from $1600Hz$ to $4000Hz$ and the peak part concentrates in the range from $100Hz$ to $400Hz$[24]. The frequency of exhaust noise and wind noise is closely related to the speed of the automobiles. All these noises vary with the type of the automobiles, tyres, engines and so on. This means automobile noise does not have specific frequency and varies with specific automobiles. We cannot find the original frequency f in our scenario.

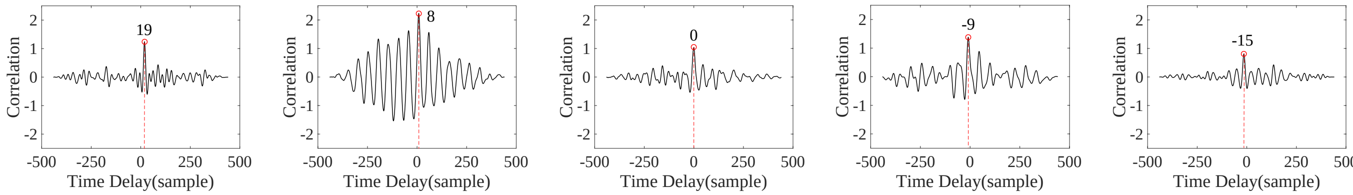
Then we focus on the real-time frequency f'. Figure 2b shows the short-time Fourier transform(STFT) of the process when the automobile passes by. We can see that the power of full-frequency band increases. It is a hard job to focus on a specific frequency to calculate the speed. That is to say, we can hardly know the reason for the increase of the specific frequency power. The increase may be because



(a) Positions of the automobile at different time.



(b) Signals from microphones at position A.
(c) Signals from microphones at position B.
(d) Signals from microphones at position C.
(e) Signals from microphones at position D.
(f) Signals from microphones at position E.



(g) Cross-correlation at position A.
(h) Cross-correlation at position B.
(i) Cross-correlation position C.
(j) Cross-correlation at position D.
(k) Cross-correlation at position E.

Fig. 3: Empirical study.

of the approaching of the automobile, or the shift of the original high-power frequency.

To conclude, if *Doppler Effect* can be utilized to solve the problem in our scenario, we should find some S-shaped curves[14] in the spectrogram. The S-shaped curves show that some specific frequencies shift to the lower frequency in the spectrogram of the acoustic signals, which can be calculated by equation(1). For example, if we let $f = 4000$Hz, $l = 10$m, $v = 20$m/s, $C = 340$m/s, we can get the S-shaped curve as figure 2c shows. We can not find any S-shaped curve in the spectrogram of the acoustic signal. That means *Doppler Effect* cannot be used to estimate the speed in our scenario.

3.1.3 Correlation between the Acoustic Signals

Since frequency domain cannot help us estimate the speed of the automobile, we may look for clues in time domain. To understand how automobile speed affects the acoustic signals from automobiles, it is essential to extract spatial and temporal information from received acoustic signals. Since we have two audio streams recorded at the same time from the top and the bottom microphones, we have the chance to calculate the spatial information.

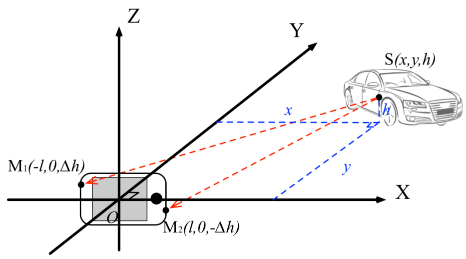
Figure 3a shows five positions of the automobile's trace we choose to study. We record the sound for 0.01s with both top and bottom microphones at each place. Figure 3b to figure 3f show the raw signals at position A to position E. Although the waveforms of the two acoustic signals are different in detail due to the difference of the microphones, they are in the similar shape with certain time delays. To further study the relation between the two signals, we calculate the cross-correlation[25] between the two signals. Figure 3g to figure 3k show the cross-correlation between signals collected by the top and the bottom microphones at different positions. Signals in figure 3b and figure 3c are recorded at the top side of the mobile phone. We can see the signals from top microphone is ahead of the signals from bottom microphone. From figure 3g and figure 3h we can see the time delay can be calculated from the value of cross-correlation between the two signals. Similarly, figure 3e and figure 3f show the signals when the automobile is at the bottom side of the mobile phone. Time delays of position D and position E are -9 and -15.

Since we have the idea that the acoustic signals from the top microphone and the bottom microphone are temporally related, we can split the signals into small segments to study the detailed relationship. This give us the chance to calculate the speed of the automobile.

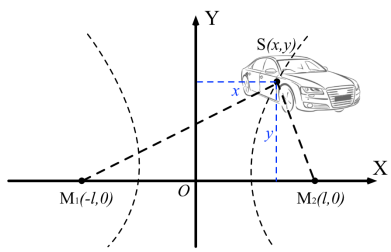
3.2 Modeling Automobile Speed via Microphone and Camera

3.2.1 Build the Coordinate System

We can use three-dimensional coordinate system to describe the scenario, just as figure 4a shows. The origin is located at the midpoint of M_1M_2. M_1 and M_2 are the points representing the two microphones. The x-axis is horizontal and points to the right, the y-axis points towards the outside of the screen face and the z-axis is vertical and points up.

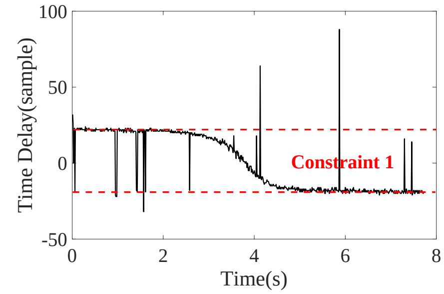


(a) 3-D coordinate system of the scenario.

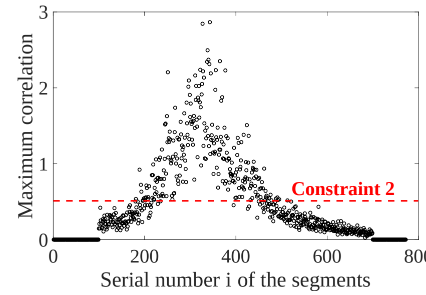


(b) Simplified 2-D coordinate system.

Fig. 4: The model of the scenario.



(a) Time delays at different time.



(b) Maximum correlation distribution.

Fig. 5: Cross-correlation of the corresponding segments.

The coordinate of M_1 and M_2 is $(-l, 0, \Delta h)$ and $(l, 0, \Delta h)$. $2l$ represents the horizontal distance between the two microphones and $2\Delta h$ represents the height difference between the two microphones when the mobile phone is in landscape orientation. $S\,(x, y, h)$ represents the sound source. h of $S\,(x, y, h)$ represents the height difference between the x-y plane and the sound source. Since $\Delta h \ll l$ and $h \ll x$ or y, we can simplify the scenario into a 2-D model as shown in figure 4b which means h and Δh can be ignored.

3.2.2 Preprocessing of the Acoustic Signals

In this section we split the acoustic signals into small segments and calculate the cross-correlation of the corresponding segments to get the time delay.

To further study the time delay between the two acoustic signals, we need to split the signals into segments sorted by time. The size of the segment needs to be discussed. We will get one time delay from one pair of corresponding segments. The more sampling points one segment includes, the more time one segment will last for. As a result, the fewer segment pairs and time delays we will get. This will cause two troubles. First, the automobile will change its position in one segment. If the size is too large, the automobile will drive for a long distance. This makes the time delay inaccurate since the sound source cannot no longer be considered as a point. Second, if the amount of time delays is too small, the time delay curve we draw will be coarse-grained. This influence the estimation precision. However, the fewer sampling points one segment includes, the more easily the segment will be influenced by the environment noise. As a result, we need to choose an appropriate segment size.

In our scenario, we let one segment consist of $n_s = f_s/100 = 441$ samples, which means one segment lasts for 0.01 second. In this case, the signals from the top and the bottom microphones are similar enough to calculate the time delay. Suppose the speed of the automobile is about 50m/s(180km/h), in one segment the automobile moves only 0.5 meter, which means we can approximately consider that the position of the sound source remains unchanged in one segment.

After the segmentation of the acoustic signals, we get two sequences of segments $S_1 = \{W_{11}W_{12}\ldots W_{1n}\}$, $S_2 = \{W_{21}W_{22}\ldots W_{2n}\}$ from the top and the bottom microphones respectively. The following equations calculate the cross-correlations R_i and delays Δd_i, where i represents the serial number of the segment pairs:

$$R_i(n) = \sum_{m=-n_s}^{n_s} W_{1i}(m)W_{2i}(m+n). \tag{2}$$

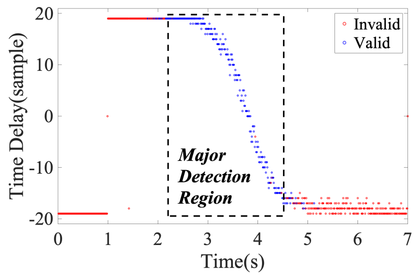
$$\Delta d_i = \arg\max_{t \in N}(R_i(t)). \tag{3}$$

After we get the result of cross-correlation $R_i(n)$, we may find the the largest element $R_i(t)$. And the $\Delta d_i = t$ who makes $R_i(n)$ largest is the time delay of the i-th pair of segments.

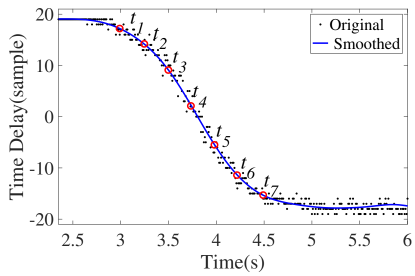
After we get the time delay Δd between the corresponding segments, we want to know whether Δd is suitable for our system. Some points we get from the equation may be erroneous due to different kinds of noises. The time delays with little noise, which are suitable for further calculation should satisfy the the following constraints:

1) The delay Δd should be less than the maximum time delay Δd_m determined by the type of the mobile phone.
2) The correlation of the corresponding segments should exceed a preset threshold R_s.

The upper bound of the valid delay in constraint 1 is inferred from triangle inequality. We can see from figure 4b that the $|\overline{M_1S} - \overline{M_2S}| < \overline{M_1M_2}$, where $|\overline{M_1S} - \overline{M_2S}|$ can be calculated by the time delay and $\overline{M_1M_2}$ is the distance between the two microphones. As a result, the value is mainly determined by the distance between the top and bottom microphones. Suppose the sampling rate is f_s, the

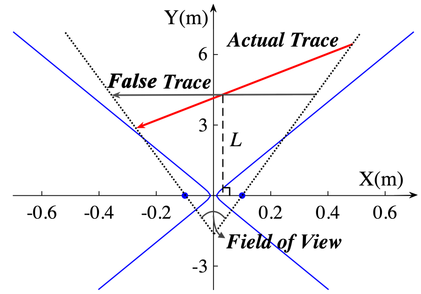


(a) Time delay curve.

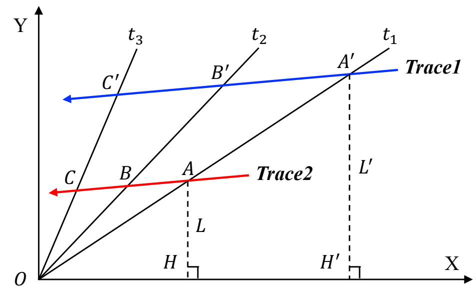


(b) Smoothed time delay curve.

Fig. 6: Generating Time Delay curve



(a) Hyperbolas generated by the time delays.



(b) A simplified model of the asymptotes and the trace.

Fig. 7: Slope Calculating.

maximum valid delay Δd_m between the two signals should be calculated as equation (4):

$$\Delta d_m = \frac{2lf_s}{C}, \tag{4}$$

where $2l$ is the distance between the two microphones and C is the speed of sound. For example, we set $f_s = 44.1KHz$ and $C = 343m/s$, and the distance, without loss of generality, of Samsung note 8(Experiments in Section 5 are based on this type of mobile phone.) $2l = 15cm$, thus Δd_m is $\frac{0.15m\times 44100s^{-1}}{343m/s} = 19.2 \approx 20$ samples. We denote the delay between the segment pair as Δd. According to triangle inequality, the valid delay we get from cross-correlation should be an integer whose absolute value $|\Delta d|$ is less than Δd_m. That means Δd should be an integer ranging from $-\Delta d_m$ to Δd_m, just as figure 5a shows. We define the region where the time delays vary from Δd_m to $-\Delta d_m$(or on the contrary) as *Major Detection Region*. For example, the *Major Detection Region* in figure 5a starts at 2.5s and ends at $5s$.

In constraint 2, the threshold R_s has its physical interpretation. It implies that the automobile should be close enough to the mobile phone, which means the signals from the two corresponding segments should be similar enough. Cross-correlation is a measure of similarity of two signals. The larger the correlation is, the more similar the two signals will be. If the automobile is far from the mobile phone, the sound made by the automobile will be too weak to dominate the signal, which means the signals from the top and the bottom microphones are not similar enough. In this case, the cross-correlations of these segments are quite small. These time delays are not suitable for speed calculation. The threshold will change with different scenarios. And the threshold can be determined with constraint 1. Since we should pay attention to *Major Detection Region*, we can set the maximum cross-correlation of the boundary segments in *Major Detection Region* as the threshold. In other words, according to constraint 1, there must exist a process in which the time delays are around Δd_m. The threshold can be determined by the correlation of these time delays. For example, in figure 5b, the threshold can be set as 0.5. Constraint 2 can help us remove some of the noise appears in *Major Detection Region*. Figure 6a draws the time delay curve with blue points represent the valid time delays and red points represent the invalid time delays.

3.2.3 Candidate Trajectories Estimation

After we get a series of time delays, we need to recover the trace of the automobile. We utilize *Major Detection Region* to estimate candidate trajectories of the automobile. The duration of *Major Detection Region* is less than 3 seconds in most situations. For example, in figure 6a the duration of *Major Detection Region* is 1.5 seconds. Since the duration is short we can assume that the trace of the automobile is a line. It is known that in two dimensions, the linear trace can be represented as:

$$y = mx + b, \tag{5}$$

which means that we need two parameters to determine a line. The parameter m determines the slope of the line and the parameter b determines the vertical distance between the automobile and the mobile phone.

First we try to calculate the parameter m through the time delays curve. If the time delay between the top and the bottom microphones is Δd at time t, the automobile should locate in the hyperbolas whose foci are $M_1\left(-l,0\right)$ and $M_2\left(l,0\right)$ and vertices are $V_1\left(-\frac{1}{2\Delta d},0\right)$ and $V_2\left(\frac{1}{2\Delta d},0\right)$ at this moment. The mathematical expression of the hyperbola is:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \tag{6}$$

where $a = \frac{\Delta d}{2}$ and $b = \sqrt{l^2 - a^2}$.

Figure 7a shows the hyperbolas generated by different time delays. We can see from the figure that the hyperbolas look like a line. The reason is that in our scenario, $x, y \gg l$, where x, y is the coordinate of the automobile in figure 4b, since l is usually shorter than 10cm and x, y are usually longer than 5 meters. So we can use asymptote of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.3034354, IEEE Transactions on Mobile Computing

the hyperbola instead using the analytic expression of the hyperbola[26] to simplify the calculation. The mathematical expression of the asymptote is:

$$y = \frac{b}{a}x. \tag{7}$$

a and b in the equation (7) are the same as that in equation (6). The exact location of the automobile is impossible to be determined by one single time delay point, but through the time delay curve, we can get a series of hyperbolas the automobile should be located in. As a result, we use a series of time delays in *Major Detection Region* to estimate the slope of the trace.

First we can select n time delays with the same time intervals between the adjacent time delays. For example, we select $n = 7$ time delays from t_1 to t_7 in figure 6b with the time interval equals 0.25s. To illustrate the estimation in detail, we focus on the three adjacent asymptotes in figure 7b. We can simplify the mathematical expression of asymptote t_1 as $y = k_1 x$ where $k_1 = \frac{b_1}{a_1} = \frac{\sqrt{4l^2 - \Delta d^2}}{\Delta d}$. Similarly, the mathematical expressions of asymptote t_2 and t_3 are $y = k_2 x$ and $y = k_3 x$. Combined with equation 5, we can get the coordinates of intersection points A, B, C between the asymptotes and the trace t_1, t_2, t_3. The coordinate of A is $(\frac{b}{k_1 - m}, \frac{k_1 b}{k_1 - m})$, the coordinate of B is $(\frac{b}{k_2 - m}, \frac{k_2 b}{k_2 - m})$ and the coordinate of C is $(\frac{b}{k_3 - m}, \frac{k_3 b}{k_3 - m})$. The distance l_1 between A and B is:

$$l_1 = |\overline{AB}| = b(k_2 - k_1)\frac{\sqrt{m^2 + 1}}{(k_1 - m)(k_2 - m)}. \tag{8}$$

The distance l_2 between B and C is :

$$l_2 = |\overline{BC}| = b(k_3 - k_2)\frac{\sqrt{(m^2 + 1)}}{(k_2 - m)(k_3 - m)}. \tag{9}$$

The duration of automobile moving from A to C is short enough(0.5s). It is reasonable to assume the speed of the automobile remains stable in this period. That means $\frac{l_1}{t_2 - t_1} = \frac{l_2}{t_3 - t_2}$. Then $\frac{l_1}{l_2} = \frac{t_2 - t_1}{t_3 - t_2} = 1$ and $\frac{l_1}{l_2}$ is independent of the parameter b. That means the parameter b can not be determined and the parameter m can be calculated.

However, the parameter m should not be calculated by only three asymptotes. So we use least squares estimation(LSE) to estimate the parameter m. Since the automobile's speed remains unchanged during the period from t_1 to t_3 and $t_2 - t_1 = t_3 - t_2$, the length difference Δl_{12} between l_1 and l_2 should be minimum. The estimation function of m is:

$$\begin{aligned} E(m) &= |l_1 - l_2|^2 \\ &= b^2 \frac{(Am + B)^2(m^2 + 1)}{((k_1 - m)(k_2 - m)(k_3 - m))^2}, \end{aligned} \tag{10}$$

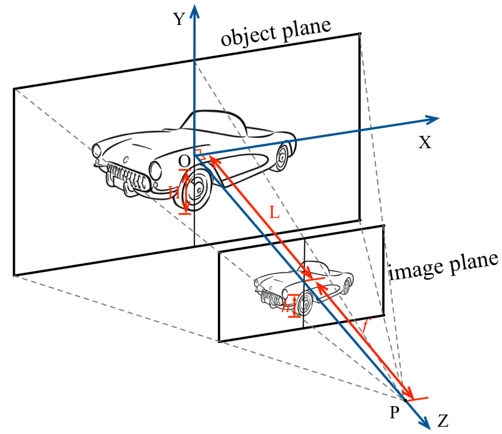
where $A = k_1 + k_3 - 2k_2, B = k_1 k_2 + k_2 k_3 - 2k_1 k_3$ and parameter b should be seen as a constant.

After we get the estimation function of m with three asymptotes, we can modify the estimation function into n asymptotes to find the fittest m. When we take the n time delays we select before, the estimation function is:

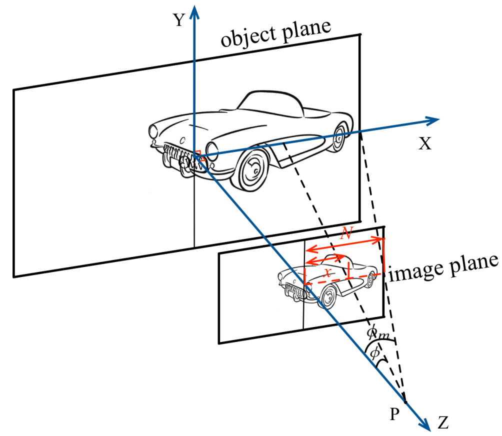
$$E(m) = \sum_{i=1}^{n-2} |l_{i+1} - l_i|^2. \tag{11}$$

l_i can be calculated similarly as equation (8).

The parameter $m = \arg\min_{m} E(m)$. We can let $\frac{\partial E(m)}{\partial m} = 0$ to calculate m.



(a) The pin-hole model of the camera system.



(b) Position estimation.

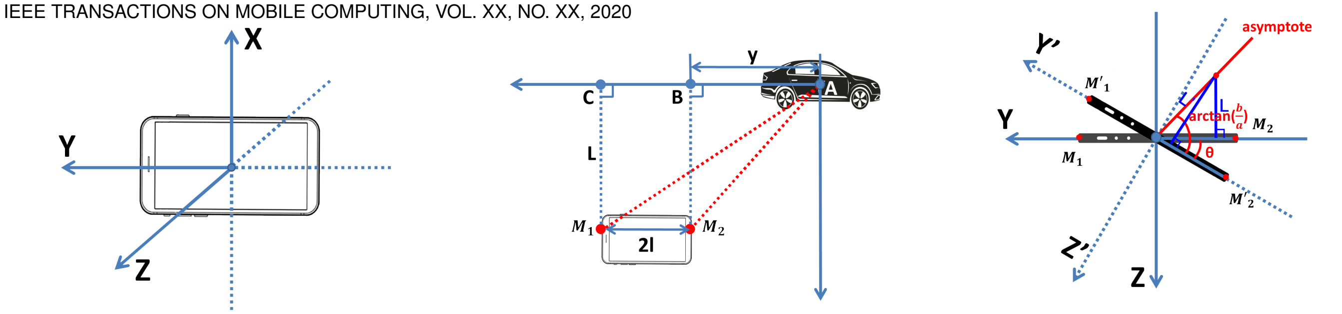
Fig. 8: Image processing.

3.2.4 Estimating Distance through the Image of the Automobile

After we use least square estimation(LSE) to calculate the slope m of the trajectory, we need to calculate the parameter b of the mathematical expression of the trajectory. The parameter b can not be determined by acoustic signals. It is known that everything looks small in the distance and big on the contrary. Similarly, the longer the distance between the object and the camera is, the fewer pixels the object in the image taken from cameras contains. Therefore, we use the camera to estimate the distance between the automobile and the mobile phone. We can continuous record the images of the automobile, and estimate the vertical distance L between the mobile phone and the automobile. In this section, we will illustrate how to estimate the vertical distance from one frame of the video. The increase of frame processing will improve the distance evaluation accuracy and increase the time consumption of the system.

One camera can be simplified to a pin-hole model. In figure 8a and figure 7b, L is the vertical distance between the automobile and the mobile phone. H is the real length of the object, and h represents the length of the object in image plane. Since $\frac{H}{h} = \frac{L+f}{f}$, we can get the distance as follows:

$$L = f\left(\frac{H}{h} - 1\right) \approx f\frac{H}{h}, \tag{12}$$



(a) Coordinate system of IMU on mobile phones.

(b) Distance difference between acoustic signals.

(c) Rotating around x-axis.

Fig. 9: Jitters removing.

where f is the focal length of the camera. f and h are related to the type of the mobile phones. The most ideal situation is that the parameters of mainstream mobile phones are stored in the application. And the approach to calculate the parameters by the system is not complicated. If we take a picture of an object H_c meters height and measure the distance L_c between the object and the camera with a constant resolution. The distance L can be calculated by the equation $L = L_c\frac{Hh_c}{H_c h}$. If resolution of the image remains unchanged, we can use the number of pixels to replace h, the length of the object in image plane.

As a result, we can calculate the distance by the image if we already know the size of some components of the automobile. Wheel hubs have mature standards and are easy to be extracted from the picture. So we extract wheel hubs in the image to estimate the distance. To avoid the stretch of the wheel hubs caused by rolling shutter effect, we calculate the diameter of the wheel hub in the vertical direction. The stretch of the wheel hubs will happen in the direction of automobiles' driving direction.

Usually the automobile will not appear in the middle of the picture. In this situation, the vertical distance L does not equal the parameter b. Just as shown in the figure 8b, we can calculate the offset angle ϕ between the center of the picture and the automobile with the viewing angle ϕ_m and the resolution of the image. First we can locate the automobile and get the offset pixels x from the automobile to the centre of the iamge. Then we can get $\tan\phi = \frac{x}{f}$ and $\tan\phi_m = \frac{N}{f}$. The angle ϕ can be calculated:

$$\phi = \arctan(\frac{x\tan\phi_m}{N}). \tag{13}$$

ϕ_m is determined by the type of the camera.

After we get L and ϕ, the trajectory of the automobile can be determined.

3.2.5 Analysis of the Jitters

Although in our scenario the mobile phone needs to be held still, jitters are unavoidable since the procedure of signals collection lasts for several seconds. The jitters will change the coordinate system and increase the speed detection error. As a result, inertial measurement unit (IMU) is used to compensate the error. In this part we need to study different translations and rotations caused by the jitters in our model. We will analyze the influence of different translations and rotations of mobile phone.

Figure 9a shows the IMU coordinate system, which is different from the coordinate system in Section 3.2.1. If the mobile phone is hold in landscape orientation as shown in the figure, the x-axis is verticle and points up, the y-axis is horizontal and points to the left, and the z-axis points towards the outside of the front face of the screen. We can see jitters can be divided into six categories, the rotation around three axes and the translation along the directions of three axes. As figure 9b shows, with the translation and rotation of the mobile phone, the path length difference $\Delta D = \overline{M_1A} - \overline{M_2A}$ changes even if the automobile is at the same position.

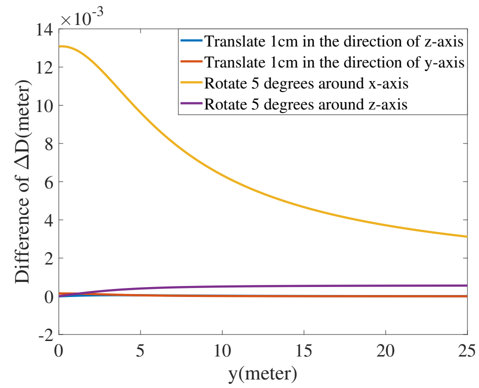


Fig. 10: Overall analysis.

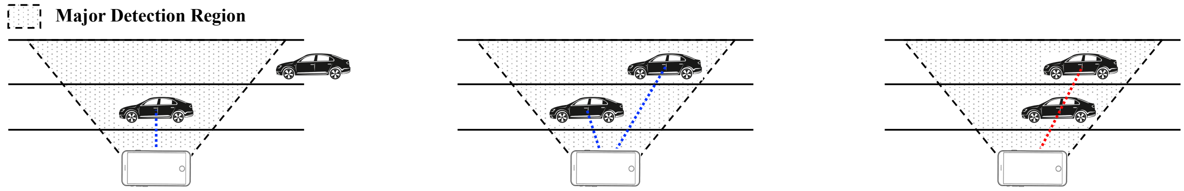
In fact, the path length difference ΔD can be calculated as follows:

$$\Delta D = \sqrt{(y+2l)^2 + L^2} - \sqrt{y^2 + L^2}, \tag{14}$$

where $y = |\overline{AB}|$ represents the displacement between the automobile and the closer microphone projected onto y-axis, $2l$ represents the distance between the two microphones and L represents the closest distance between the automobile and the mobile phone.

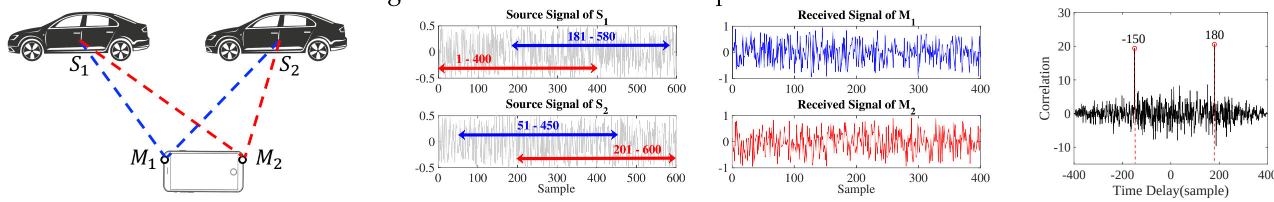
Then, we will evaluate the influence of translation and rotation. The translation in the direction of x-axis can be ignored. The reason is that we have already ignored the height difference when modeling the system. And the translation in the direction of x-axis is much shorter than the height difference between the mobile phone and the sound source. As a result, the translation in the direction of x-axis can be ignored. Similarly, the rotation around the y-axis does not affect the positions of microphones on y-axis. The rotation around y-axis can also be ignored.

Figure 10 shows the influences of remaining rotations and translations. We can take the translation in the direction of z-axis as example. The translation in the direction of z-axis influences the distance L we estimate between the mobile phone and the automobile in figure 9b. The path length difference ΔD can be modified as:



(a) Case 1. (b) Case 2. (c) Case 3.

Fig. 11: Different cases of multiple automobiles.



(a) Illustration of multiple sound source. (b) Original signals from different sources. (c) Received signals of different receivers. (d) Correlation between the received signals.

Fig. 12: Multiple peaks analysis.

$$\Delta D_z = \sqrt{(y+2l)^2 + (L+\Delta l)^2} - \sqrt{y^2 + (L+\Delta l)^2}, \quad (15)$$

where Δl represents the translation in the direction of z-axis. The other rotations and translation can be calculated in the similar way. From the figure we can see that the influence of rotation around x-axis is much more than that of others. As a result, we just need to analyze the rotation around x-axis to remove the jitters.

3.2.6 Multiple Automobiles Detection

While when multiple automobiles pass by the mobile phone, the model presented before can only estimate the speed of part of the automobiles. The problem is caused by multiple sound sources. When multiple automobiles pass by the mobile phone, the sound made by different automobiles interferes with each other. And the sounds of the automobiles share comparable intensity. The time delays we calculate from cross-correlation between the top and bottom microphones cannot always form a complete time delay curve.

According to the empirical study, it is hard to separate the specific automobile sounds from microphones of mobile phones. However, it is possible to decide whose sound decides dominate the acoustic signals collected by the microphones. In this section we define the automobile whose sound dominates the acoustic signals as *Major Detection Object*. Then we classify the scenario of multiple automobiles into three situations as shown in Figure 11.

We classify the scenarios of multiple automobiles into three cases as shown in figure 11. *Major Detection Region* refers to the areas between the asymptote corresponding to the time delay Δd_m and $-\Delta d_m$ as shown in figure 11a. The three cases can be concluded as:

1. Only one automobile is moving in *Major Detection Region*.

2. Multiple automobiles become *Major Detection Object* in turn.

3. Only one of the automobiles become *Major Detection Object* when multiple automobiles go through the *Major Detection Region*.

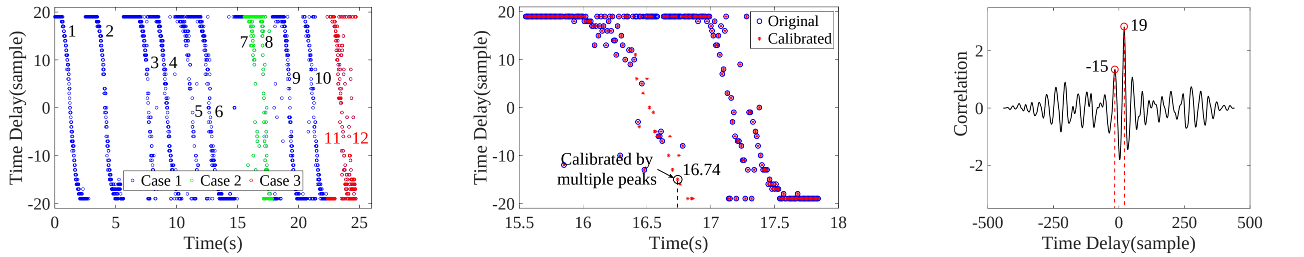
Figure 13a is the time delay curve of 12 automobiles in the real environment drawn by means of Section 3.2.2. All three cases appear in the figure. The solutions to the cases are as follows.

Case 1: Only one automobile is moving in Major Detection Region. In this case, before one automobile finishes its movement in its major detection region, no other automobile will enter the major detection region. The automobile in the major detection region will become the major detection object. The acoustic signals we collect from the automobile in this case is similar to the signals from single automobile. The blue circles in figure 13a form the delay curve of the corresponding automobiles. The speed of the automobiles can be estimated through the algorithm we propose. And according to our experiments, when it is not at the peak time, this is the most common case for multiple automobiles.

Case 2: Multiple automobiles become Major Detection Object in turn.

In this case, we should analyze how multiple automobiles affect the correlation we calculate through the acoustic signals. Cross-correlation measures the similarity between the two signals. The time delay between the two signals can be calculated with the correlation values. First we can analyze the mixed signals with different delays. Figure 12a illustrates the scenario of multiple signal sources and multiple receivers. Figure 12b shows the signal waves of S_1 and S_2. Since $|S_1M_1| < |S_1M_2|$, the signals from S_1 will arrive at M_1 earlier than arrive at M_2. Similarly, the signals from S_2 will arrive at M_1 later than arrive at M_2. In figure 12c, the signals R_1 received by M_1 are the superposition of sample 181-580 in S_1 and sample 51-450 in S_2. The signals R_2 received by M_2 are the superposition of sample 1-400 in S_1 and sample 201-600 in S_2. That means the signals of S_1 involved in R_1 is 180 samples earlier than in R_2 and S_2 involved in R_1 is 150 samples later than in S_2. The correlation between the received signals is shown in figure 12d. We have the chance to calculate the time delay of different sources.

We can take automobile 7 and 8 in figure 13a as example. The time delay curve of automobile 7 is not complete. The reason is that automobile 7 and 8 are coming at the same time. And the sound made by automobile 8 is louder than that made by automobile 7. Then we may use multiple peaks



(a) Time delays of microphones with 12 automobiles.

(b) Time delays when 7th and 8th automobiles pass by.

(c) Cross-correlation curve when t = 16.74s.

Fig. 13: Processing of acoustic signals of multiple automobiles.

in the correlation figures to recover the time delay curve. Then we focus on the correlation between the corresponding segments at time 16.74s. In figure 13a, we have time delay $\Delta d = 10$ around $t = 16.4s$ and time delay $\Delta d = -5$ around $t = 16.6s$. According to these existing points, we can estimate the range of the time delay when $t = 16.74s$. Figure 13c shows the cross-correlation of the corresponding segments. Several peaks appear in the figure. We calculate top three peaks in correlation figure. We can see the second highest peak is located on time delay -15 and is suitable for the time delay curve. That means the highest peak at $\Delta d = 19$ is caused by automobile 8 and the second highest peak at $\Delta d = -15$ is caused by automobile 7. At last if no time delay satisfy the range, we will abandon the corresponding segments.

Case 3: Only one of the automobiles becomes Major Detection Object when multiple automobiles ao through the Major Detection Region.

The situation is shown in figure 11c. In this case the line of sight(LOS) path between the automobiles overlap within the major detection region. That means the delay curve of these automobiles can not be separated from each other. The speed of the automobiles cannot be estimated in this case. In this case, the sound collected by microphones is too complicated to analyze. As a result, we conduct experiments on multiple automobiles and the result shows that if SpeedTalker is not utilized at morning or evening peak, the detection success rate is about 92%. In figure 13a, the curve in red of the 11th and 12th automobiles cannot be distinguished. The two delay curve highly overlap and the two acoustic signals of the automobiles are severely affected by each other. Only the speed of the automobile which can be detected by the cameras can be estimated.

4 System Design

4.1 System Overview

The system architecture is shown in figure 14. There are three main components in SpeedTalker, i.e., *Acoustic Signal Processing*, *Computer Vision Based Processing* and *Speed Extraction*. *Acoustic Signal Processing* first filters the high-frequency signals. Then signals from the top and the bottom microphones will be split into small segments sorted by time. The time delays between the segment pairs can be calculated through cross-correlation. We modify time delays to remove the influence caused by the jitters with inner measurement unit(IMU) in the mobile phone. After that the time delays form a time delay curve. The time delay curve can be smoothed with Gaussian smoothing. *Computer Vision Based Processing* first extracts the automobile in the image with yolo[27]. The automobile extracted from the image will be put into a deep learning network to recognize the type of the automobile and know the diameter of its wheel hub in reality. Then we extract the wheel hub through Hough Transform, get the pixel diameter of the wheel hub in the image and estimate the vertical distance between the automobile and the mobile phone. At last, *Speed Extraction* recovers the trajectory of the automobile and estimates the speed of the automobile.

4.2 Acoustic Signal Processing

The acoustic signals we collect usually contain many noises. We can see from figure 2b that the main energy of the sound made by the automobile is distributed in the low-frequency area. To make the time delay calculated through cross-correlation more accurate, we have let the sound of automobile dominate the acoustic signal. As a result, we use a low-pass filter to remove the high-frequency noises. Different types of automobiles have different noise frequency distribution, but we know that the automobile noise is mainly distributed in the frequencies below 4 kHz. So we use a low-pass filter with the cutoff frequency of 4 kHz.

Next, we need to split the acoustic signals into segments sorted by time. As mentioned before, the size of the segments we design is $f_s/100$. After getting two series of audio segments, we calculate the cross-correlation between the segment pairs to make out the time delay between the corresponding segments. Then we get a series of time delays with timestamp, which can be used to draw the time delay curve. The illustration of the process is shown in figure 15.

At last, we need to smooth the time delay curve. In Section 3.2.2, we propose two constraints to filter the time delays. We can divide the whole curve into two parts: Minor Detection Region and Major Detection Region. Minor Detection Region refers to the region where the time delay remains unchanged, with $\Delta d = \Delta d_m$. The definition of Major Detection Region is in Section 3.2.2. When the acoustic signals is in Minor Detection Region, the automobile is far away from the mobile phone and the time delay is fixed. This region is of little importance to the speed estimation. So we focus on Major Detection Region as shown in the figure 6a. In Major Detection Region, there exist some invalid time delays influenced by environmental noise. There are two steps to smooth the time delay curve. The first step is to replace the invalid delay with some reasonable values. All the time delays with $|\Delta d| > \Delta d_m$ should be

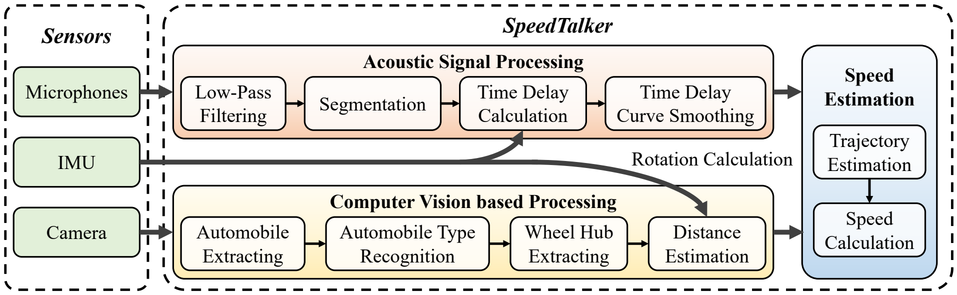


Fig. 14: System Design.

modified to Δd_m or $-\Delta d_m$. Then, the time delays in Major Detection Region should not change rapidly. That means the difference between the adjacent time delays should not exceed a threshold. We can see from figure 6a that the largest gradient of the curve is at the time when $\Delta d = 0$. Without loss of generality, we suppose the vertical distance between the automobile and the mobile phone L is 5 m, the distance between the two microphones $2l$ is 0.2 m. If the difference between adjacent time delays is 3, the speed of the automobile is 75 m/s(270 km/h), which is rarely seen in daily life. As a result, the difference between adjacent time delays should less than 3 in ideal situations. We relax the threshold to 5. So we can exclude extremely abnormal time delays which may influence the traditional smoothing method and retain other time delays. The value of extremely abnormal time delays can be modified to that of its neighbors.

The second step is to smooth the sequence with traditional smoothing methods, e.g., Gaussian smoothing. Then we will get the smoothed time delay curve as figure 6b shows.

4.3 Computer Vision Based Processing

After we smooth the delay curve and get the Major Detection Region, we may get a series of candidate trajectories with the same slope m. However, the other parameter b of the trajectory cannot be estimated by acoustic signals. So we use the image of the automobile to complete the estimation of the trajectory. From equation 12 we can see that we need to calculate the pixel diameter h of the wheel hub and the real diameter H of the wheel hub. The key point of this section is to extract the wheel hub information in the image and in reality. Figure 16 shows the process of image processing, including *Automobile Extraction*, *Automoblie Type Recognition*, *Hough Transform* and *Diameter Estimation*.

First we need to extract the automobile in the image. The image we get from the camera contains too many objects, which makes it difficult to recognize the type of the automobile. Besides, too many pixels in the image increase the complexity of image processing. We just want to pay attention to the automobile itself. As a result, we utilize yolo[27], which applies a single neural network to the full image. The network divides the image into regions and predicts bounding boxes and probabilities for each region. With Yolo, we can extract the automobile and get the position of the automobile in the image. Some empirical approaches are utilized to solve the problem. The pixel locations of the car should be a rectangle whose length-width ratio is more than 2. And the automobile should not be static in different frames.

The second step is to recognize the type of the automobile so as to get the type of the wheel hub. Each type of automobile has its well-matched wheel hub size[28]. If we recognize the type of the automobile, we can get the real diameter H of the wheel hub. There are several computer vision tools based on machine learning can recognize the type, such as DeepVision[29], OrpixVision[30] and BaiduAI[31]. The method using computer vision gives APIs for our system to recognize the type of the automobile as figure 16b shows.

After we get the real parameter H of the wheel hub, we need to calculate the pixel parameter h of the wheel hub in the image. The shape of the wheel hub is a circle. Due to the movement of the automobile, the wheel hubs in the image are not standard circles. However, the movement of the automobile is in the horizontal direction, which has no effect on the diameter in the upright direction. As a result, we focus on the wheel hub in the upright direction. The method to detect the rough circle is using Hough transform. Hough transform can be utilized to arbitrary shapes[32]. First we detect the edge of the automobile with Canny edge detection algorithm[33]. To achieve better performance, we cut out the lower right corner of the automobile extracted from the figure since the wheel hub of the automobile are mostly appear at the bottom of the image. We may find the wheel hub by Hough transform. The vertical distance L can be calculated through equation (12).

With yolo we also get the position of the automobile in the image, with equation (13), the offset angle can also be calculated.

4.4 Jitters Removing

Since we get the approach of time delay modification in Section 3.2.5, the following part of this section is to calculate the rotation angle around x-axis. We define the moment when we start to record the sound of the automobile as T_0. The coordinate system at T_0 is denoted as C_{T_0}. We need to estimate the speed of the automobile in a stable reference system. As a result, we transform the time delay Δd calculated by cross-correlation in altered coordinate system C_{T_n} into the time delay $\Delta d'$ in C_{T_0}. We have already analyze how different translations and rotations will influence the time delay. The influence of the jitters can be resolved into these translations and rotations. And we just need to achieve the rotation around x-axis. We utilize rotation vector in android system[34] to calculate the rotation. The rotation vector represents the orientation of the device as a combination of an angle and an axis, in which the device has rotated through

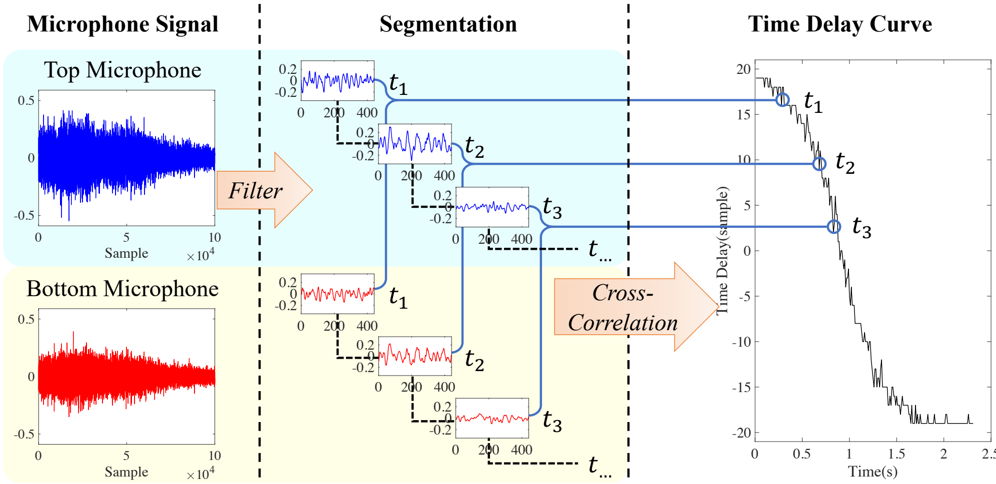
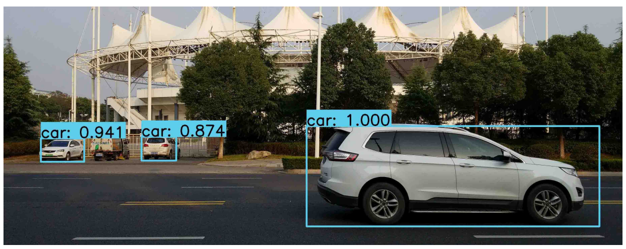
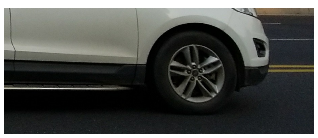


Fig. 15: Illustration of generating time delay curve.

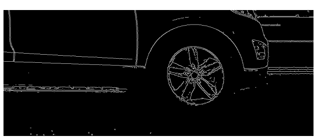


(a) Automobile recognition.

(b) Automobile extraction.



(c) Narrow the range.



(d) Edge detection.

(e) Circle detection.

Fig. 16: Wheel hub extraction

an angle θ around an axis $< x, y, z >$. The three elements of the rotation vector are $< x\sin\frac{\theta}{2}, y\sin\frac{\theta}{2}, z\sin\frac{\theta}{2} >$. The three elements are equal to the last three components of a unit quaternion $< \cos\frac{\theta}{2}, -x\sin\frac{\theta}{2}, -y\sin\frac{\theta}{2}, -z\sin\frac{\theta}{2} >$. ${}^{A}_{B}\boldsymbol{q}$ describes the orientation of coordinate system C_B relative to coordinate system C_A and $[x, y, z]$ is a vector described in C_A. Since there do not exist rapid rotations in our scenario, the Euler angle can be approximately seen as the rotations around the three axes.

4.5 Speed Estimation

In Section 3.2.3, we propose the least square estimation(LSE) to estimate the slope m of the trajectory. The co-ordinates($L\tan\phi, L$) of the point can be given through image processing, since L is the distance between the automobile and the camera and ϕ is the offset angle from the mid line of the camera. We can recover the real trajectory of the automobile. That means we get the spatial information of the automobile. Moreover, the points on the trajectory have its own timestamp since the trajectory is calculated by the time delay curve. We can choose two asymptotes from the smoothed delay curve. From the two asymptotes and the trace we can get the distance the automobile moves at a period of time so that the speed can be estimated. We can choose two asymptotes with time delay Δd_i and Δd_k and calculate the slope m_i and m_k. To reduce the error, we can make $|\Delta d_i - \Delta d_k|$ as large as possible. After choosing the asymptotes, we can get the distance between the two intersection points of the asymptotes and the trace. At last, the speed of the automobile can be estimated.

TABLE 1: Different parameters of mobile phones.

Type	Horizontal Distance $2l$	Equivalent Focal Length
Samsung S5	0.142 meter	31mm
Iphone 6	0.132 meter	29mm
Honor 7	0.14 meter	28mm
Mi 6	0.145 meter	27mm
Note 8	0.15 meter	26mm

5 Performance Evaluation

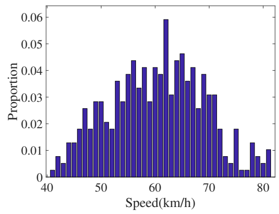
5.1 Experimental Setup and Methodology

We run SpeedTalker on different types of mobile phones, including Samsung S5, Iphone 6, HUAWEI Honor 7, Mi 6 and Samsung Note 8. The parameters of different types of mobile phones are shown in Table 1. The angle of view can be calculated from equivalent focal length. The user stands on the side of the road and hold the mobile phone in the landscape orientation. When the automobile passes by, the mobile phone continuously collect the acoustic and visual signals to estimate the speed of the automobile. In our experiments, the default parameter set is that recording the sound at 44.1 kHz, the sound is first arrive at bottom microphone and the direction is 0 degree. Default resolution of the visual signals is 1280 $\times$ 720. Default segmentation size is 441, which means each segment lasts for 0.01s. Default FPS is 30 fps, and other camera settings including shutter time, aperture, ISO condition(the sensitivity to light) and focusing are automatically determined by the camera system itself. SpeedTalker is set on these mobile phones with default settings to evaluate the effectiveness of the system.

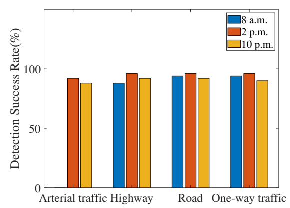
Besides the type of mobile phones, we test SpeedTalker in different usage scenarios with different parameter sets using Samsung Note 8. There are several parameters in the experimental setup: 1) Sampling rate: we set the sampling rate when recording the sound of the automobile

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.3034354, IEEE Transactions on Mobile Computing

(a) Experiment setup.



(b) Speed distribution of multiple automobile.



(c) Detection success rate in different locations at different time.

Fig. 17: Preparation of the Experiment.

as 8 kHz, 22.05 kHz, 44.1 kHz and 192 kHz. 2) Resolution of the image: we set the resolution of the image as 640×480, 1280×720, 1920×1080 and 4032×3024. 3) Direction: we change the angles between the road and the mobile phone from -30 degrees to 30 degrees. 4) Automobile type: we evaluate SpeedTalker on three types of automobiles: cars, buses and trucks. 5)Light condition: we record the signals in five different light conditions including dawn, daytime, dusk, night when it is sunny and daytime when it is rainy. 6)FPS(frame per second) condition: we set the fps conditions of the videos as 30 fps, 60 fps, 120fps, 240fps. 7)Different size of the segments: we change the size of the window in segmentation. The time of duration is 0.005s (221 samples in one segment), 0.01s, 0.05s, 0.1s. Different sets of signals may overlap other sets.

At last, we use relative speed error, which first calculates the speed difference between the estimated speed and ground truth, and then divede the ground truth, to evaluate SpeedTalker. The ground truth is collected by the radar speed gun, BushNell Speedster III as shown in figure 17a. The error associated with radar speed gun is within 1 MPH(1.6km/h) according to the instruction.

5.2 Speed Estimation with Different Setups

In different parameter settings, experimental results show that SpeedTalker achieves an average relative error of 6.1% compared with the groundtruth in the scenario of single automobile. In the scenario of multiple automobiles, the situation is much more complicated. We have analyzed different situations of multiple automobiles in Section 3.2.6. In some situations multiple automobiles cannot be distinguished by SpeedTalker. We test SpeedTalker in four different locations at different time with default parameter setting. The four locations include arterial traffic with six lanes(three lanes in each direction), highway with six lanes, road with four lanes and one-way traffic with one lane. The time includes 8 a.m. when the traffic is very busy, 2 p.m. when the traffic is not so busy in the daytime and 10 p.m. when the traffic is not so busy in the night. Each of the experiments includes about 50 samples. The detection success rate is shown in figure 17c. First of all, SpeedTalker cannot handle the situation of morning peak since all the lanes of arterial traffic are full of automobiles. The acoustic signals are too complicated for analysis. We will further discuss this situation in Discussion Section. Luckily in this situation, speeding detection is not necessary since the automobiles cannot move fast. If we exclude this situation and one-way traffic with one lane, the detection success rate is around 92%. Among the automobiles can be distinguished from each other, the average relative error is 9.8%.

We present the distribution of the speed groundtruth in figure 17b. The speed of the automobiles we collect approximately obeys Gaussian distribution whose average speed is about 60 km/h. We ignore the automobiles whose speeds are less than 40 km/h(usually meet speed requirement) or more than 80 km/h(lack of experiment data). The signals we collect are sound enough to evaluate SpeedTalker.

5.3 Performance in Different Speed Conditions

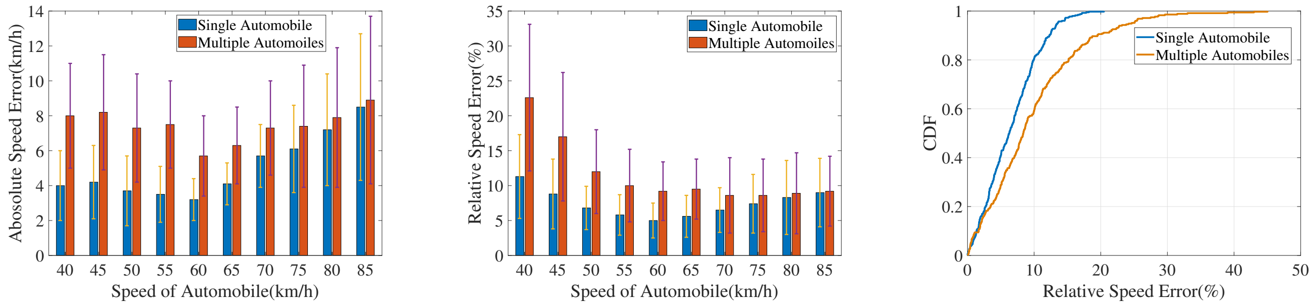
Figure 18a and figure 18b show the absolute and relative speed errors of 10 speed intervals. The interval labeled by 40 means the speed of autmobiles in this section is between 40 km/h and 45 km/h.

In the scenario of single automobile, the value of absolute speed errors does not have the trend to increase where the speed of the automobiles is below 60 km/h. Then the absolute speed errors begin to increase with the increase of the automobiles' speed. The automobiles' moving distance in the duration of one segment increases with the increase of automobile speed, which may brings errors to the cross-correlation between the segment pairs. This makes the absolute speed errors increase. When the speed of the automobiles is below 60 km/h, the absolute speed error does not increase rapidly with the increase of speed. The reason for that is the sound made by the automobiles are tightly related with the speed of the automobiles. When the speed of the automobiles is not fast, the amplitude of the acoustic signals may influence the absolute speed error. However, the relative speed errors shown in figure 18b differ from absolute speed errors. Since groundtruth also increases, the relative speed errors are acceptable if the speed of automobiles is high.

Then we present the CDF figure of the relative speed error in figure 18c. The average error based on the experiments is 6.1%. According to the results, more than 80% of our measurements have an accuracy of 90.2%. SpeedTalker can achieve good accuracy in estimating the speed.

In the scenario of multiple automobiles, we find that the performance of SpeedTalker is worse than that of single automobile. With the increase of the speed, the differences of performances are tending to decrease. The reason is that we need to apply multiple peak model to recover the delay curve of each automobile. The recovered delay curve cannot achieve the same accuracy as the original delay curve. However, the automobiles with higher speed will make louder sound, which will become major detection object for longer time. As a result, the estimated speed in multiple automobiles scenario is close to that in single automobile. The average relative speed error in this scenario is 9.8%. We can see that SpeedTalker can achieve high accuracy in daily use.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.3034354, IEEE Transactions on Mobile Computing



(a) Average absolute speed error VS the speed of automobile.

(b) Average relative speed error VS the speed of automobile.

(c) CDF of relative speed error.

Fig. 18: Evaluate the accuracy of speed with different speed conditions.

5.4 Performance in Different Parameter Conditions

Next we change the parameters in the experiment setup, i.e., sampling rate, resolution, direction, orientation, FPS conditions and the size of the segments. In figure 19a we can see that in both single automobile and multiple automobiles scenarios, the higher the sampling rate is, the smaller the average relative speed error will be. The reason is that the lower the sampling rate is, the fewer sample points one segment will contain. As a result, the time delay calculated from cross-correlation may be inaccurate, which makes the estimation of automobiles' speed inaccurate. However, even if the sampling rate equals 8 kHz, we can still achieve an average relative speed error of 85.3% in single automobile scenario and 75.1% in multiple automobiles scenario.

In figure 19b we can see the higher the resolution of the image is, the smaller the average relative speed error will be. The resolution of the image influences the precision of the distance estimation. If we enhance the resolution of the image, the pixel height of the wheel hub will be more precise. This will enhance the performance of SpeedTalker. In figure 19c we can see that the fps condition does not apparently influence the performance of SpeedTalker. Actually the automobile information collected by the visual signals is influenced by resolution and exposure time of the camera in our model. If the exposure time is too long, the figure of the automobile will be blurred, which will make the estimation inaccurate. The shutter time is the direct factor of the clarity of the frames instead of FPS conditions. We evaluate performance with different shutter time. The results are shown in figure 19d. Different shutter time does not influence the performance dramatically, either. In digital cameras, the exposure time is shorter than the electronic shutter time. Even if the electronic shutter time is 1/30 s, the exposure time is usually less than 1/60 s. The movement of the automobiles will not influence the quality of the figures severely. In figure19e, we can see that in the scenario of single automobile, the performances of SpeedTalker are almost the same. If one segment lasts for 0.05s, the system has the best performance. However, when it comes to multiple automobiles, the methods whose segments lasting for 0.05s and 0.01s have better performance. The reason is that when multiple automobiles pass by, the delay curve of the automobiles may not be complete. If the one segment have too many samples, the less time delay points we have. Then the time delay curve we recover may have more errors.

Figure 19f shows SpeedTalker achieve good performance in different orientations. If we rotate the mobile phone, the relative speed error will increase. But due to our trajectory estimation model, SpeedTalker still performs well. To conclude, if we enhance the resolution of the images and sampling rate of the audio, SpeedTalker performs better.

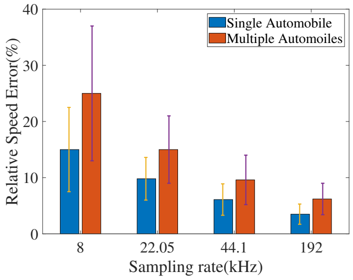
5.5 Performance in Different External Conditions

The external conditions may influence the accuracy of SpeedTalker. And SpeedTalker works in real outdoor environment, the conditions can not be easily controlled. We choose three typical external factors to analyze the impact. First we can see that trucks achieve the best performance while buses perform worst among the three types of automobiles in both scenarios. The reason is that the trucks make the biggest noise, which can dominate the acoustic signals. The noise of buses includes the noise of hydraulic machines, the noise of wobbling and so on, which is much more complicated than the cars. Together with bigger size, the sound source of buses is wider than cars. And buses move at medium speed, usually no more than 50 km/h. With the same absolute speed errors, the relative speed error will be larger than that of cars. Another interesting phenomenon is that SpeedTalker works better on trucks in multiple automobile scenarios since the acoustic signals made by the trucks is loud enough to make the trucks major detection objects. In cases of cars and buses, the performance of SpeedTalker get worse, especially there exist trucks at the same time.

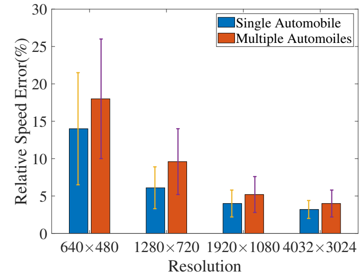
The light condition influences the visual signals we collect. However, it is unlikely to control the light condition outdoors beside the streets. As a result, we choose four typical time, including dawn, daytime, dusk and night(with street lamps) in a day when it is sunny to change the light condition. Besides we also evaluate SpeedTalker in rainy days. Figure 20b shows SpeedTalker performs better in daytime when the light is sufficient in both scenarios. The images we take at dawn, dusk or night are more blurry than that in daytime, which makes the estimation error of distance increase. Although we cannot achieve the same performance as in sunny weather in the daytime, we can utilize SpeedTalker to estimate automobiles' speed in the condition that the wheel hub of the automobiles can be recorded either. If the street lamps are too dark to recognize the wheel, SpeedTalker will not work.

Figure 20c shows the performance of SpeedTalker on different types of mobile phones. We can see from Table 1 that

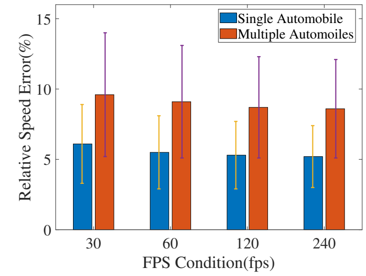
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2020.3034354, IEEE Transactions on Mobile Computing



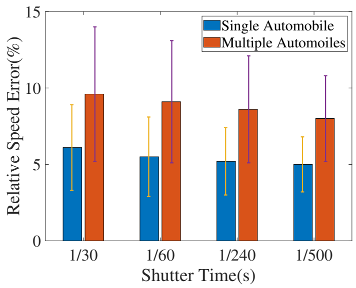
(a) Average relative speed error VS sampling rate.



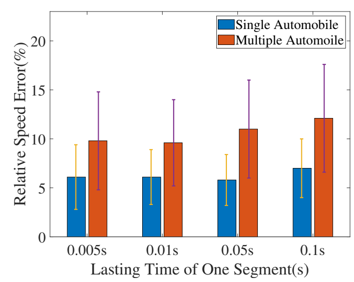
(b) Average relative speed error VS resolution of the image.



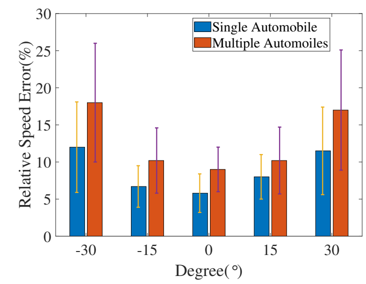
(c) Average relative speed error VS fps condition.



(d) Average relative speed error VS shutter time.



(e) Average relative speed error VS segment lasting time.



(f) Average relative speed error VS degree.

Fig. 19: Evaluate the accuracy of speed with different parameters.

different types of mobile phones have different size. That means the horizontal distance $2l$ and angle of view will vary with different mobile phones. We change the corresponding parameters in our model, and conduct experiments on them. We can see from the figure that different types of mobile phones have similar performance. The mobile phone with longer length have tiny advantage over that with shorter length.

5.6 Cost Evaluation of SpeedTalker

To evaluate the cost of SpeedTalker, we run the system with the default parameter settings. The acoustic signals processing and computer vision based processing can work on concurrent threads since they are not dependent on each other. However, the process of speed estimation needs the results of both acoustic signals processing and computer vision based processing. Acoustic signals processing perform low-pass filtering, cross-correlation calculating, and smoothing, which takes a very short time. Computer vision based processing contains two AI networks and circle detection. First yolo is run to line out the automobiles and the system can get the outline coordinate. After that the extracted automobiles can be upload to another machine learning network (we use Baidu AI here) to get the automobile information. The extracted automobile pictures are upload to the ML network on the server and the results are sent to our system. Hough transform is performed to recognize the wheel hub at the same time. Hough transform does not take a long time since the input images are the extracted automobiles. As a result, the bottleneck of SpeedTalker is utilizing ML network. Table 2 shows the detailed information of time consumption of SpeedTalker. Computer vision based processing are performed on one frame(not all the frames need to processed, one frame per second is sufficient for speed detection). The energy consumption of the system is also shown in the Table. We can see from the table that time consumption of different mobile phones is acceptable for practical usage.

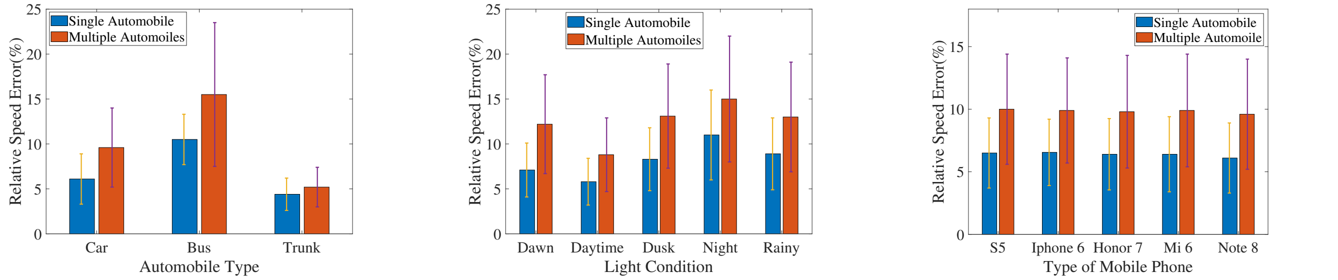
6 Discussions

6.1 Limitation of Our Work

Though SpeedTalker can achieve accurate speed estimation with mobile phones, there exist several limitations in real environments.

First, SpeedTalker does not work in several situations. a) If the automobile is turning a corner, the system cannot work properly. This is because the estimation on candidate trajectories are based on the assumption that the trajectory of the automobile is a line. Meanwhile, the mobile phones nowadays are equipped with more than two microphones, which may give us chances to improve the system. Without more sensors to help sense the automobile, it is difficult to deal with the crooked trajectories. b) If the automobiles are coming from multiple direction, the system cannot work. This situation can be avoided since automobiles coming from all directions is quite a dangerous scenario for the users. c) If the automobile cannot be detected by the camera, the system cannot work. This situation may happen if the system is used in the evening and there exists no street lamp around the users or the detection object is blocked by the barriers. d) As shown in Section 5 Performance Evaluation, SpeedTalker cannot work when using at morning or evening peak. In fact in this situation, due to the traffic jam, the speed of the automobiles cannot be very fast.

Second, based on our model, we need to know the parameters,i.e., the distance between the top and the bottom distance and the view angle of the camera, of the mobile phone which is running SpeedTalker. Moreover, the speed of sound varies with geographical factors, such as elevation, humidity and so on. That means SpeedTalker needs to be adapted to all types of mobile phones with unique



(a) Average relative speed error VS types of automobiles.

(b) Average relative speed error VS light condition.

(c) Average relative speed error VS type of mobile phone.

Fig. 20: Evaluate the accuracy of speed with different environmental conditions.

Type	Acoustic Processing(ms)	YOLO(ms)	Hough Transform(ms)	Baidu AI(s)	Energy Consumption per min(%)
Samsung S5	105	210	46	0.8	0.7
Iphone 6	98	190	45	0.8	0.9
Honor 7	113	224	66	0.8	0.6
Mi 6	96	178	45	0.8	0.7
Note 8	88	152	40	0.8	0.5

TABLE 2: Different parameters of mobile phones.

parameter. Also, we need to take geographical factors into consideration. To handle these two limitations, we need pre-knowledge of the mobile phones and geographical position. SpeedTalker can work with more sensors, such as GPS to improve the performance.

6.2 Improvements and Future Work

a) **Utilizing advanced cameras:** More and more mobile phones have multiple cameras at the back side. Some cameras even have specific sensors to calculate the depth of scenes. Besides, two cameras of the same type can estimate the depth with computer vision based approaches. However, these approaches have their limitations. First, the depth sensors have its own detection region. It will not detect the objects far from the sensors. For example, Kinect can only sense the depth of human body within 3 meters. Second, most depth sensors cannot detect the objects with high speed. Third cameras on the on the mobile phones are normally not the same type, which may not suitable for depth calculating. As a result, the depth sensors cannot work properly in our scenarios.
b) **Future work:** We have conducted experiments on several environmental parameters, such as light condition, sunny or rainy days. However, there exist much more situation that may influence the performance of SpeedTalker. For example, wind and insects may influence the acoustic signals we collect and dusty or foggy weather will influence the cameras. However, it is a hard job for us to control or quantize environmental parameters in a short time. We will conduct experiments on different environmental parameters to further study the characters of acoustic and visual signals influenced by outdoor environments.

7 Conclusion

In this paper, we propose SpeedTalker that leverages mobile phones to estimate the speed of the automobiles. The users just need to stand by the side of the road, hold the mobile phone in landscape orientation and run the application to collect acoustic and image signals. The speed of the automobile can be estimated. We use the time difference of arrivals(TDOA) model based on acoustic signals to estimate the candidate trajectories of the automobile. The images we take from cameras help us determine the trajectory of the automobile. Combined with the timestamp on the trajectory we can estimate the speed. An inner measurement unit(IMU) based jitters removing method is proposed to improve the performance. In our experiment, we measure the speed of automobiles with an average estimation error of 6.1% in the scenario of single automobile and 9.8% in the scenario of multiple automobiles compared to radar speed guns. Our approach demonstrates a portable and low-cost solution to provide accurate speed monitoring via mobile phones.

Acknowledgments

This work is supported in part by National Natural Science Foundation of China under Grant Nos. 61872174, 61832008, 61802169, 61872173, 61902175; JiangSu Natural Science Foundation under Grant Nos. BK20180325, BK20190293. This work is partially supported by Collaborative Innovation Center of Novel Software Technology and Industrialization. Lei Xie is the corresponding author.

References

[1] "Global status report on road safety 2018," https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/.
[2] "Managing speed," https://www.who.int/violence_injury_prevention/publications/road_traffic/managing-speed/en/.
[3] E. Estellés-Arolas and F. González-Ladrón-De-Guevara, "Towards an integrated crowdsourcing definition," *Journal of Information science*, vol. 38, no. 2, pp. 189–200, 2012.
[4] "Radar gun," https://en.wikipedia.org/wiki/Radar_gun, 2018.
[5] C. Weitkamp, *Lidar: range-resolved optical remote sensing of the atmosphere*. Springer Science & Business, 2006, vol. 102.
[6] A. R. George, "Automobile aerodynamic noise," *SAE transactions*, pp. 434–457, 1990.
[7] T. Kumar and D. S. Kushwaha, "An efficient approach for detection and speed estimation of moving vehicles," *Procedia Computer Science*, vol. 89, pp. 726–731, 2016.
[8] W. Czajewski and M. Iwanowski, "Vision-based vehicle speed measurement method," in *International Conference on Computer Vision and Graphics*. Springer, 2010, pp. 308–315.

[9] "Oki succeeds in trial production of world's first "safety mobile phone" to improve pedestrian safety," https://www.einpresswire.com/article/56319627/, 2007.

[10] K. David and A. Flach, "Car-2-x and pedestrian safety," *IEEE Vehicular Technology Magazine*, vol. 5, no. 1, pp. 70–76, 2010.

[11] S. Sivaraman and M. M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 267–276, 2010.

[12] T. Wang, G. Cardone, A. Corradi, L. Torresani, and A. T. Campbell, "Walksafe: a pedestrian safety app for mobile phone users who walk and talk while crossing roads," in *Proceedings of the twelfth workshop on mobile computing systems & applications*. ACM, 2012, p. 5.

[13] H. Tsuzuki, M. Kugler, S. Kuroyanagi, and A. Iwata, "A novel approach for sound approaching detection," in *International Conference on Neural Information Processing*. Springer, 2010, pp. 407–414.

[14] M. Takagi, K. Fujimoto, Y. Kawahara, and T. Asami, "Detecting hybrid and electric vehicles using a smartphone," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014, pp. 267–275.

[15] S. Li, X. Fan, Y. Zhang, W. Trappe, J. Lindqvist, and R. E. Howard, "Auto++: Detecting cars using embedded microphones in real-time," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, p. 70, 2017.

[16] S. Yun, Y.-C. Chen, and L. Qiu, "Turning a mobile device into a mouse in the air," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2015, pp. 15–29.

[17] W. Wang, A. X. Liu, and K. Sun, "Device-free gesture tracking using acoustic signals," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. ACM, 2016, pp. 82–94.

[18] R. Nandakumar, S. Gollakota, and N. Watson, "Contactless sleep apnea detection on smartphones," in *Proceedings of the 13th annual international conference on mobile systems, applications, and services*. ACM, 2015, pp. 45–57.

[19] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[20] V. T. B. Tram and M. Yoo, "Vehicle-to-vehicle distance estimation using a low-resolution camera based on visible light communications," *IEEE Access*, vol. 6, pp. 4521–4527, 2018.

[21] M. Diaz-Cabrera, P. Cerri, and P. Medici, "Robust real-time traffic light detection and distance estimation using a single camera," *Expert Systems with Applications*, vol. 42, no. 8, pp. 3911–3923, 2015.

[22] K. A. Rahman, M. S. Hossain, M. A.-A. Bhuiyan, T. Zhang, M. Hasanuzzaman, and H. Ueno, "Person to camera distance measurement based on eye-distance," in *2009 Third International Conference on Multimedia and Ubiquitous Engineering*. IEEE, 2009, pp. 137–141.

[23] G. Dubois, J. Cesbron, H. Yin, F. Anfosso-Lédée, and D. Duhamel, "Statistical estimation of low frequency tyre/road noise from numerical contact forces," *Applied Acoustics*, vol. 74, no. 9, pp. 1085–1093, 2013.

[24] Y. Liu, Y. B. Jia, X. J. Zhang, Z. C. Liu, Y. C. Ren, and B. Yang, "Noise test and analysis of automobile engine," in *Applied Mechanics and Materials*, vol. 307. Trans Tech Publ, 2013, pp. 196–199.

[25] L. R. Rabiner and B. Gold, "Theory and application of digital signal processing," *Englewood Cliffs, NJ, Prentice-Hall, Inc., 1975. 777 p.*, 1975.

[26] E. W. Swokowski, *Calculus with analytic geometry*. Taylor & Francis, 1979.

[27] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[28] "Wheel sizing," https://en.wikipedia.org/wiki/Wheel_sizing#Wheel_size, 2019.

[29] "Deepvision," http://visualintelligence.deepvisionai.com, 2018.

[30] "Vehicle make and model recognition," http://mmr-demo.orpix-inc.com/detection/main/, 2016.

[31] "Baidu ai," http://ai.baidu.com/tech/imagerecognition/fine_grained, 2018.

[32] D. H. Ballard, "Generalizing the hough transform to detect arbitrary shapes," *Pattern recognition*, vol. 13, no. 2, pp. 111–122, 1981.

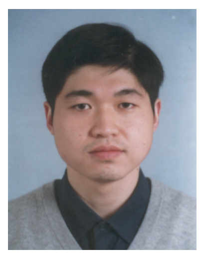
[33] W. Rong, Z. Li, W. Zhang, and L. Sun, "An improved canny edge detection algorithm," in *2014 IEEE International Conference on Mechatronics and Automation*. IEEE, 2014, pp. 577–582.

[34] "Rotation vector," https://developer.android.google.cn/reference/android/hardware/SensorEvent.html#values, 2019.

Xinran Lu received his B.S. degree in computer science from Nanjing University, China in 2017. He is currently a Ph.D. student in the Department of Computer Science and Technology at Nanjing University. Her research interests includes sensing via acoustic signals.

Lei Xie received his B.S. and Ph.D. degrees from Nanjing University, China in 2004 and 2010, respectively, all in computer science. He is currently an associate professor in the Department of Computer Science and Technology at Nanjing University. He has published over 80 papers in IEEE Transactions on Mobile Computing, ACM/IEEE Transactions on Networking, IEEE Transactions on Parallel and Distributed Systems, ACM Transactions on Sensor Networks, ACM MobiCom, ACM UbiComp, ACM MobiHoc, IEEE INFOCOM, IEEE ICNP, IEEE ICDCS, etc.

Yafeng Yin received her Ph.D. degree in computer science from Nanjing University, China in 2017. She is currently an assistant professor in the Department of Computer Science and Technology at Nanjing University. Her research interests include human activity recognition, mobile sensing, wearable computing, etc.



Wei Wang received his MS and Ph.D. degree from the ESE department of Nanjing University and the ECE department of National University of Singapore, in 2000 and 2008 respectively. He is currently an Associate Professor in the CS department of Nanjing University. His research interests are in the area of wireless networks, including Device-free Sensing, Cellular Network Measurements, and Software Defined Radio systems.

Yanling Bu received her M.S. degree in computer technology from Nanjing University, China in 2018. She is currently a Ph.D. student in the Department of Computer Science and Technology at Nanjing University. Her research interests include RFID sensing and localization.

Qing Guo received his B.S. degree from Nanjing University in 2017. He is currently a Ph.D. student in the Department of Computer Science and Technology at Nanjing University. His research interests include optical communication and passive sensing.

Sanglu Lu received her B.S., M.S. and Ph.D. degrees from Nanjing University, China in 1992, 1995 and 1997, respectively, all in computer science. She is currently a professor in the Department of Computer Science and Technology at Nanjing University. Her research interests include distributed computing and pervasive computing.