An Acoustic-based Encounter Profiling System

Huanle Zhang, *Member, IEEE,* Wan Du, *Member, IEEE,* Pengfei Zhou, *Member, IEEE,* Mo Li, *Senior Member, IEEE,* and Prasant Mohapatra, *Fellow Member, IEEE*

Abstract—This paper presents DopEnc, an acoustic-based encounter profiling system on commercial off-the-shelf smartphones. DopEnc automatically identifies the persons that users interact with in the context of encountering. DopEnc performs encounter profiling in two major steps: (1) Doppler profiling to detect that two persons approach and stop in front of each other via an *effective* trajectory, and (2) voice profiling to confirm that they are thereafter engaged in an interactive conversation. DopEnc is further extended to support parallel acoustic exploration of many users by incorporating a unique multiple access scheme within the limited inaudible acoustic frequency band. All implementation of DopEnc is based on commodity sensors like speakers, microphones and accelerometers integrated on mainstream smartphones. We evaluate DopEnc with detailed experiments and a real use-case study of 11 participants. Overall DopEnc achieves an accuracy of 6.9% false positive and 9.7% false negative in real usage.

Index Terms—Encounter profiling, Acoustic signals, Doppler effect, Voice profiling, Multiple access.

1 Introduction

PEOPLE often encounter and interact with many persons during a social event or during the day. Encounter profiling aims at identifying the encountered persons and recording the interaction context, e.g., the time and place that a particular person is met, or detailed information of all the persons that the user encountered and interacted with. It is useful in life logging and memory assistance. Existing solutions (e.g., Sony's LifeLog [2] and Google's Keep [3]), however, require user involvement to manually record the person, the time and place, etc., which introduce extra overhead to the users.

It will be very convenient if a smartphone based system can automatically identify the persons that one interacted within a certain duration. For example, Alice encounters Bob in a social event. After a short chat, they say goodbye to each other. Beside Bob, Alice may encounter many other persons. During each interaction, Alice does not need to take any special actions to log the time and place of meeting Bob. An app on her smartphone automatically identifies the persons she meets. When she returns to her place or when needed, Alice is able to retrieve the related information from her smartphone about all the persons she has met during the event. It is desirable for the system to not require any customized devices or pre-deployed infrastructure. To the best of our knowledge there are no proposed solutions to this problem at this moment. Existing techniques for human sensing can detect the handshaking with Skin Potential Level (SPL) sensors [4], identify human groups through trajectory tracking [5], or detect human proximity with short distance communication like Bluetooth Low Energy (BLE) [6] or Near Field Communication (NFC) [7]. Those techniques with their different design purposes cannot truthfully identify the persons that users met. Besides, they all have limitations in usage, e.g., customized SPL devices worn on wrists, infrastructure support for human tracking and detection, extra assumptions on human relationships, movements, ways of interactions, etc.

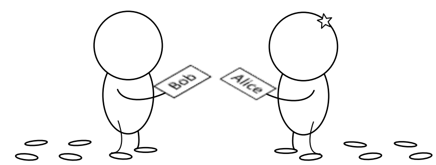


Fig. 1. DopEnc identifies the persons that one has interacted with in two steps: (1) Two persons approach each other in an *effective* trajectory; (2) They are engaged in an interactive conversation

This paper proposes DopEnc, which targets at identifying the persons that users have interacted with by leveraging acoustic signal transmission with common sensors on smartphones (i.e., speaker, microphone and accelerometer). In designing DopEnc, we carefully analyze the normal procedure of human interactions in the context of encountering, and propose to identify the persons that one is interested to interact with in two major steps, namely (1) trajectory analysis, and (2) conversation confirmation. Figure 1 illustrates the DopEnc procedure.

Trajectory analysis. Two persons who interact with each other often approach in an *effective* trajectory and stop in front of each other. Not all approaching trajectories are effective, e.g., the trajectory of two persons bypassing each other does not constitute an effective trajectory. To distinguish the effective trajectories from the others, DopEnc does not employ existing trajectory mapping/localization solutions which often incur heavy computation and communication overhead. Instead, DopEnc exploits the Doppler effect of

- *Huanle Zhang and Prasant Mohapatra are with the Department of Computer Science, University of California, Davis, USA 95616. Email: dtczhang@ucdavis.edu; pmohapatra@ucdavis.edu.*
- *Wan Du is with the Department of Electrical Engineering and Computer Science, University of California, Merced, USA 95343. Email: wdu3@ucm.edu.*
- *Pengfei Zhou is with the School of Software, Tsinghua University, China 100084. Email: zhoupf05@tsinghua.edu.cn.*
- *Mo Li is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798. Email: limo@ntu.edu.sg.*

A preliminary version of this work was published in the Proceedings of ACM MobiCom 2016 [1].

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.2017.2776915, IEEE Transactions on Mobile Computing

acoustic signals transmitted between the users' smartphones and derives concise Doppler profiles to tell effective trajectories. DopEnc configures the smartphones to broadcast inaudible acoustic signals and derives from the Doppler effect the relative velocities between two persons. According to the estimated Doppler profile, DopEnc identifies different types of user approach trajectories and finds the effective ones. DopEnc applies a data cleaning technique to the rough acoustic signals obtained in practice that exploits the unique Doppler characteristics during human walkings and essentially improves the accuracy in classifying the trajectories.

Conversation confirmation. An effective trajectory may not always lead to conversations, e.g., when two persons approach the same object (e.g., a poster or booth) but they do not interact with each other. DopEnc therefore performs voice profiling to confirm whether two persons are engaged in conversations. Existing works on voice recognition [8], [9] often require a pre-established voice feature database of all users and incur high processing overhead. They cannot be applied to our scenario where no global voice feature database exists and computation has to be performed on the smartphones. DopEnc adopts a lightweight approach to confirm user conversations in a distributed manner. DopEnc lets each user's smartphone identify its owner's voice only and the smartphones of two approaching users exchange the recognized voice traces. DopEnc calculates the alternativeness ratio and duty ratio of user speeches, and based on that infers whether the two persons are likely engaged in conversations.

After confirming the encounter, DopEnc can thus record the time and the place where the encounter happens. Based on the users' choices, DopEnc can be further configured to assist exchanging some public personal information (e.g., electronic name card) between two users with conventional data exchanging schemes, e.g., Bluetooth or cellular.

DopEnc is further extended to support parallel acoustic exploration of many users with a specially designed multiple access scheme. The inaudible acoustic frequency band is divided into a limited number of channels with careful consideration of channel capacity and possible Doppler offsets due to human walk. Different users may perform Doppler profiling and voice profiling simultaneously over different channels. The multiple access scheme considers the frequency features of Doppler effect and interferences of acoustic signals, and adopts priority-based channel switching to coordinate the channel access of multiple users, which essentially reduces the collisions.

We implement and evaluate DopEnc on Commercial Off-The-Shelf (COTS) smartphones. To the best of our knowledge, DopEnc is the first practical system of its kind that is able to automatically identify people encounters. An experiment with 11 participants in a real event demonstrates that DopEnc can accurately identify the encounter events. The false positive and false negative rates are 6.9% and 9.7% respectively. We emphasize that the goal of DopEnc is not to achieve 100% accuracy. Small false positives and false negatives are tolerable since encounter profiling does not involve sensitive information. Slightest user intervention can greatly reduce the false postives and false negatives. In addition, tunable trade-off can also be set in DopEnc to control the false positives and false negatives so the user can configure more conservative or aggressive schemes for encounter profiling. DopEnc can also be easily extended to other useful applications. For example, an automatic name card exchange system can be realized when name cards are exchanged between the people who have interacted with each other; elderly citizens can use our system as a memory assistant to keep track of daily events.

2 Doppler Profiling

In this section, we describe how DopEnc derives the Doppler profiles of the acoustic signals transmitted between smartphones. Based on the concise Doppler profiles, DopEnc is able to classify the approaching trajectories of people and identify the effective ones that may lead to human interactions. In particular, DopEnc considers the human walking characteristics and devises special techniques to remove noises and errors in practical Doppler measurements.

2.1 Doppler Profiling on Smartphones

Doppler effect refers to the frequency change of a wave, when an observer is moving relative to the source [10]. The frequency offset is determined by the relative velocity between the source and the observer [10]:

$$\Delta f = \frac{\Delta v}{c} \cdot f_o, \tag{1}$$

where frequency offset $\Delta f = f - f_o$ is the received frequency f subtracted by the emitted frequency f_o; c is the speed of waves. Both f_o and c are known. If f is detected, we can calculate the frequency offset and thus derive the relative velocity by Doppler effect.

DopEnc uses inaudible acoustic signals ($18kHz - 20kHz$[1], details in Section 4.1) to perform Doppler profiling. DopEnc configures the smartphone speakers to broadcast acoustic signals with fixed frequency. The other smartphone records the received signals using its microphone and calculates the frequency offset using Fast-Fourier-Fransform (FFT) analysis. The relative velocities of the senders can thus be derived by the receivers using Equation (1). Figure 2 depicts an example piece of Doppler profile which records the relative velocity measured by two COTS smartphones when one person walks towards the other. Initially, they remain stationary, of zero relative velocity. The relative velocity becomes positive when one walks towards the other and returns to zero when he stops in front of the other.

Two parameters determine the quality of the obtained Doppler profile: *velocity resolution* and *temporal fidelity*. Velocity resolution (Δv_{res}) represents the minimum difference of relative velocities that the measured Doppler effect can tell. Temporal fidelity (Δt_{res}) is the minimum time interval of two consecutive measurements that the system can take. We set the microphone sampling rate to $48kHz$ (supported by most mainstream smartphones) and perform

1. The sound frequency above $15kHz$ is already inaudible for most adults. We set the signal frequency to be above $18kHz$ to safeguard that it will not even affect other groups like young kids and infants who are more sensitive to high frequency sounds. The sounds might be annoying to some animals. The adopted frequency band has much lower noise than audible bands in daily environment.

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.2017.2776915, IEEE Transactions on Mobile Computing

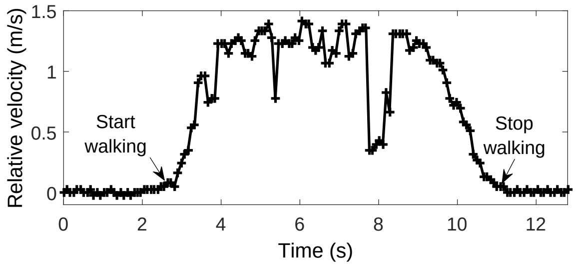


Fig. 2. The Doppler profile that records the relative velocity between two smartphone users who walk towards each other

4096-FFT (i.e., one FFT for 4096 samples, which is affordable by the processing capability of mainstream smartphones). The frequency granularity Δf_{res} is $11.7Hz$ ($48kHz/4096$). The corresponding velocity resolution is only $22.5cm/s$ (refer to Equation (1)) which is not accurate to capture human movements. To improve the velocity resolution of Doppler profiling, DopEnc adopts undersampling technique that translates high-frequency bandpass signals to low-frequency lowpass signals without frequency spectrum distortion. The resultant lowpass signals can be sampled with a much lower undersampling rate. The velocity resolution can be improved by $n\times$ (n is the undersampling factor). Undersampling has been adopted by other works like Spartacus [11] to improve audio sensing precision. DopEnc adopts undersampling factor of 8 (the supported frequency band is illustrated in Section 4.1), corresponding to a velocity resolution of $2.82cm/s$, which is sufficient to accurately capture slight human movements.

On the other hand, undersampling reduces temporal fidelity by $n\times$, because the interval between two adjacent samples for FFT are enlarged. To provide high temporal fidelity, DopEnc further adopts overlapping technique [12] which reuses the past sampling data and constitutes a sliding window for FFT. Therefore, less waiting time is needed before performing one FFT. With an overlapping ratio of 87.5%, the temporal fidelity can be improved by $8\times$. We also test higher overlapping ratios (e.g., 88.9% for $9\times$, 90.0% for $10\times$) but do not observe obvious performance improvement. In summary, by incorporating undersampling and overlapping techniques, DopEnc achieves velocity resolution of $2.82cm/s$ and temporal fidelity of $0.085s$ (i.e., $11.7Hz$).

There are other methods to calculate Doppler shifts with different target applications. Frequency Modulated Continuous Waveform (FMCW) is used by [13] to estimate the absolute distance between a transmitter and a receiver, and the corresponding distance change. The authors in [14] propose to directly track signal phases to infer gestures. Both methods have good accuracies but high complexities. On contrary, Goertzel algorithm [15] is lightweight but can only analyze one selectable frequency component. None of these methods can be easily applied to encounter profiling which requires simultaneous transmission among multiple users (refer to Section 4 for more details).

Acoustic Signals vs. Radio Signals. Radio signals, e.g., Wi-Fi and RFID, have been widely used to track human movements [16], [17]. Unlike radio signals which are electromagnetic waves in essence, acoustic signals are mechanical waves that have much slower transmission speeds in the air. Compared to radio signals, acoustic signals are much more sensitive in sensing movement. Equation (1) represents the frequency offset due to the relative movements. Assume that the relative velocity Δv between two persons is $1m/s$. Consider two scenarios: 1) acoustic signal, signal frequency fo is $20kHz$, speed of acoustic signal c is $346m/s$, then the resulting frequency offset Δf of the acoustic signal is $57.8Hz$; 2) radio signal, signal frequency fo is $2.4GHz$, speed of radio signal c is $3\times10^8 m/s$, then the corresponding frequency offset Δf of the radio signal is only $0.008Hz$. The frequency offset of acoustic signals is approximate 7225 (57.8/0.008) times that of radio signals. We released the Android code at https://github.com/dtczhl/doppler-illustrator.

2.2 Classifying Trajectories

A relative trajectory is effective if two persons approach (directly or indirectly) and stop in front of each other. An effective trajectory may lead to interactions. To identify effective trajectories, DopEnc classifies all relative trajectories into five categories as illustrated in the first row in Figure 3: (1) Direct approach where two persons walk directly towards each other (Figure 3(a)). Direct approach leads to an effective trajectory, since the two persons approach and stop in front of each other; (2) Roundabout approach where two persons walk towards each other in a roundabout trajectory (Figure 3(b)). Roundabout approach also gives an effective trajectory. In some cases, the direct path between two persons may be blocked, so they may need to bypass the obstacle and eventually approach each other; (3) Slant approach where two persons slantly walk closer but do not meet each other (Figure 3(c)). Slant approach does not give an effective trajectory, because the two persons do not eventually meet each other; (4) Departing where two persons depart from each other (Figure 3(d)), and (5) Passing by where two persons pass by each other (Figure 3(e)). Departing and passing by scenarios are obviously not effective, as the two persons eventually get away from each other.

The second row in Figure 3 summarizes the theoretical Doppler profiles for the above categories. The relative velocity between two persons is determined by their absolute walking speeds and their relative heading direction. When people walk with targets in mind, their walking speeds remain stable [18], [19] and their relative velocity can thus indicate the relative heading direction between them. Such information tells the effective trajectories. From Figure 3, we can see that each category has unique features in its relative velocity trace. The relative velocity in direct approach (Figure 3(a)) remains steady all the time. In roundabout approach (Figure 3(b)), after obstacles are bypassed, the two persons are relatively heading closer to each other and their relative velocity increases. In slant approach (Figure 3(c)), however, the relative heading direction between the two persons keeps shifting away, which leads to gradually decreasing relative velocities. The relative velocities in departing (Figure 3(d)) and passing by (Figure 3(e)) drop below zero when the two persons walk away from each other. To summarize, the trace of relative velocity recorded

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.2017.2776915, IEEE Transactions on Mobile Computing

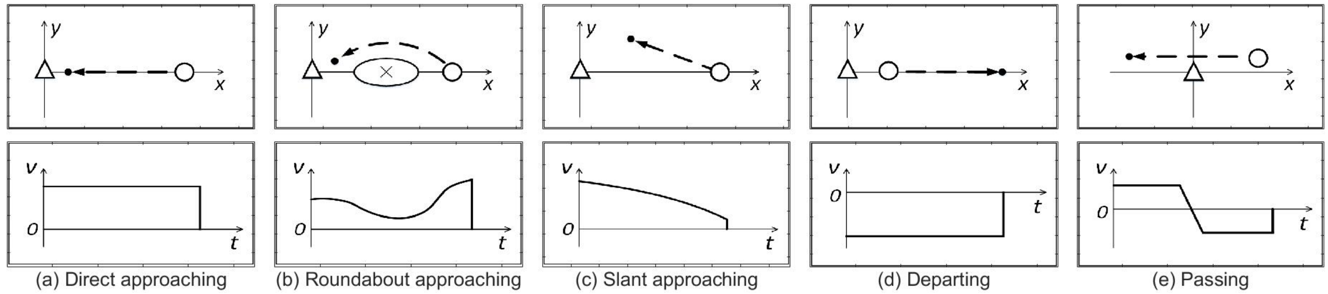


Fig. 3. Five categories of relative trajectories (the upper row) and their corresponding Doppler profiles (the lower row). One person (in triangle) is fixed in the axis origin. The relative movement of the other person (in circle) is depicted in a dashed line

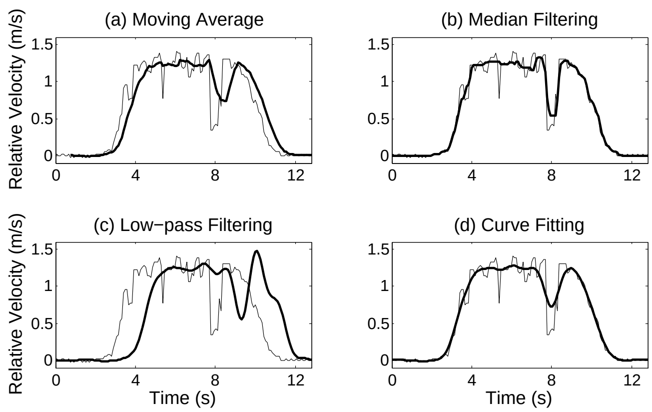


Fig. 4. Results of mainstream smoothing techniques. The valley at $8s$ cannot be removed

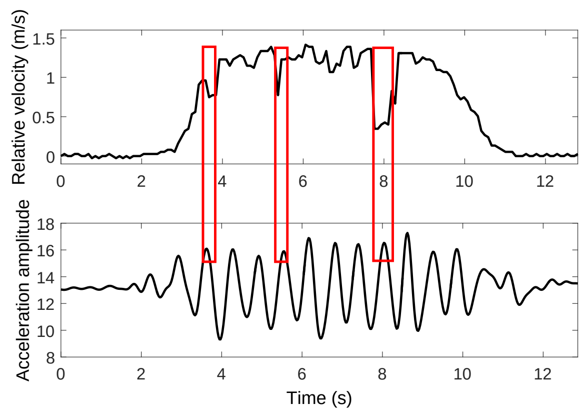


Fig. 5. Valleys in the relative velocity trace occur at $3.8s$, $5.4s$, and $8.0s$ when the person's foot strikes the ground.

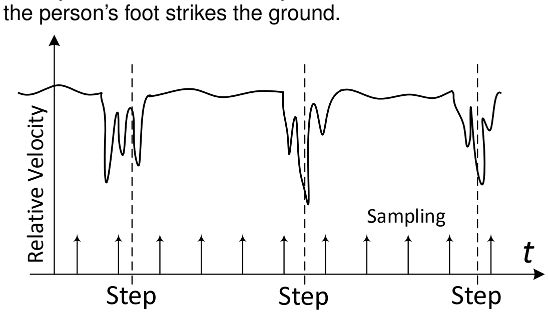


Fig. 6. Limited sampling rate contributes to occasional valleys. If the instantaneous body adjustment when human strikes the ground is sampled, then valley occurs.

in the Doppler profile indicates the effective trajectory when it remains *positive* and *non-decreasing* before the two persons stop walking.

2.3 Doppler Profiling in Practice

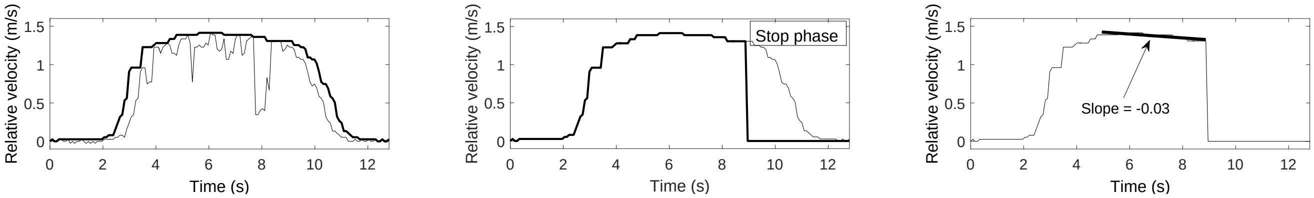
Trajectory classification illustrates the basic principle to interpret the Doppler profile and derives the types of approaching trajectories. However, the practical Doppler profile measured on smartphones differs from the ideal Doppler profile shown in Figure 3. A typical Doppler profile measured on smartphones as shown in Figure 2 is obtained when one person walks directly towards the other with normal gaits. While we expect their relative velocity to be constant, the measured trace in practice contains valleys of relative velocities. Without cautious data cleaning on those profiles, we may mistreat those valleys as the result of sudden shift of heading directions and thus wrongly classify such an effective trajectory into other ineffective trajectories. In our preliminary test, we measure 50 traces of Doppler profiles in practice, and similar valleys of relative velocities occur in more than 40 of them.

Another problem with the Doppler profile as demonstrated in Figure 2 is that the measured relative velocity gradually decreases in a period instead of immediately dropping to zero when the person stops walking (time period from $9s$ to $11s$ in Figure 2). The stop phase of human walking lasts for some time and may mislead to detection of decreasing velocity and thus misclassifying the trajectory into slant approach. From our preliminary 50 tests, all Doppler profiles contain non-negligible stop phases of human walkings. To remove the above-mentioned noises and errors contained in the Doppler profile in practice, DopEnc adopts the following three steps of data cleaning.

Step 1: smoothing the velocity trace. To filter out the valleys of velocity traces, we test mainstream smoothing techniques, including moving average, median filtering, low-pass filtering and curve fitting. Figure 4 shows the smoothing results. Although only low-pass filtering fails to flatten the velocity trace during the two persons' walking, all of these smoothing techniques cannot remove the deep valley at $8s$, which could lead to falsely classify effective trajectories into slant approaching.

We comparatively study the measured velocity traces and the accelerometer readings, and try to figure out the root cause of the valleys. Figure 5 depicts the relative velocities between two persons and the corresponding ac-

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.2017.2776915, IEEE Transactions on Mobile Computing



(a) Step 1: smoothing the velocity trace.

(b) Step 2: removing the stop phase of human walking.

(c) Step 3: linear fitting to derive the slope of velocity.

Fig. 7. Three steps of data cleaning to Doppler profile to remove noises and errors. Approaching trajectories are classified based on the slopes of the fitting lines.

celerometer readings measured by the walker's smartphone. The raw Three-Dimensional (3D) accelerometer readings are summed up and smoothed by low-pass filtering. A peak of acceleration occurs when the walker's foot strikes the ground [20]. We find that some of those peaks of acceleration perfectly correspond to the valleys observed in the velocity trace, because the instantaneous walking speed is disrupted by the body adjustment when the person's foot strikes the ground.

Figure 6 shows that limited sampling rate contributes to occasional valleys. Not all the foot strikes correspond to velocity valleys due to the limited sampling rate used in DopEnc, i.e., DopEnc is able to sample the human walking velocity at $11.7Hz$ and each human walking step takes approximately $0.5s$ [21], so only about six samples are measured during one walking step. As a result, a velocity valley appears only when the sample is taken close to the time of a foot strike.

Based on the above analysis, instead of using general smoothing techniques like low-pass filtering, curve fitting, median filtering or moving average, which cannot remove the deep valley of velocity trace (e.g., the valley at $8.0s$ in Figure 2), DopEnc uses a more suitable smoothing method, called *moving maximum*, to effectively filter the valleys in the relative velocity traces. Moving maximum exploits the fact that the human walking speed is steady most of the time except when the foot strikes the ground. Moving maximum replaces the relative velocity of each sample with the maximum velocity within a range of adjacent samples. Empirically, DopEnc sets the window of moving maximum to one human walking cycle of two steps. Because people take approximately $0.5s$ to make a step [21], the number of velocity samples for one walking cycle is 11.7 ($2\times 0.5s\times 11.7Hz$). Therefore, DopEnc replaces the relative velocity of sample i with

$$v_i = \max(v_{i-5}, v_{i-4}, ..., v_i, ..., v_{i+4}, v_{i+5}) \quad (2)$$

Figure 7(a) depicts the smoothing result of moving maximum on the velocity trace in Figure 5. Moving maximum successfully eliminates most valleys including the deep valley at $8s$ in the trace. Even when the mobile phone is not held static in hand, moving maximum is still able to effectively smooth the velocity trace because of the repetitive pattern of human walking.

Step 2: removing the stop phase of human walking. DopEnc removes the misleading stop phase of human walking in the velocity trace. The relative velocities before the stop phase of human walking are sufficient for trajectory classification. DopEnc removes the stop phase of human walking by the walking steps. According to our empirical experience and our preliminary 50 traces from five persons, the stop phase of human walking normally contains three walking steps. Each human step normally takes approximate $0.5s$ [21]. Conservatively, DopEnc removes last two seconds' data from the velocity trace. The end of the velocity trace is detected if the measured frequency offsets are zero. Figure 7(b) shows that the stop phase of human walking in the trajectory trace is successfully removed.

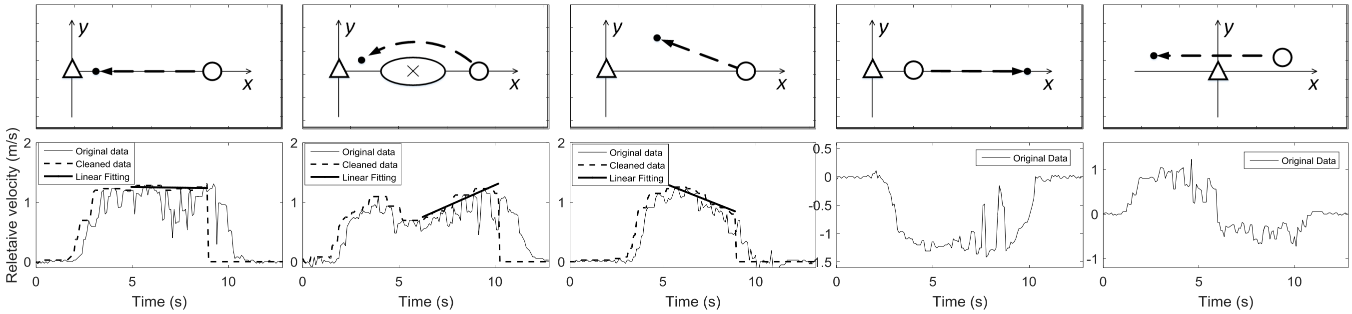
Step 3: linear fitting to derive the slope. After removing the valleys and the stop phase in the velocity traces, DopEnc applies linear fitting on the traces and uses the slope of the fitting line to classify the approaching trajectories. Figure 7(c) shows the fitting line and the slope. The fitting length in Figure 7(c) is set to $4s$. In this example, the slope of the fitting line is -0.03, very close to the ideal value of direct approach, i.e., zero. Figure 8 presents the measured relative velocity traces of different relative trajectory categories as well as the cleaned data. Based on the slopes of fitting lines, DopEnc is able to clearly distinguish direct approach, roundabout approach and slant approach. The fitting line in direct approach (Figure 8(a)) remains horizontal. The fitting lines in roundabout approach (Figure 8(b)) and slant approach (Figure 8(c)) are upwards and downwards respectively. The departing (Figure 8(d)) and passing by (Figure 8(e)) cases are easy to identify, as their velocity traces have a negative part before the persons stop walking.

After identifying the effective trajectories between two persons, DopEnc ascertains whether the two persons stop in proximity of each other. While there exists readily available proximity detection approaches based on time-of-arrival (TOA) [22], [23], signal strength [24] or interferometry [25], [26], DopEnc utilizes voice profiling to detect the user proximity. As the next section will detail, voice profiling is able to further confirm user engagement in conversations and thus firmly identifies the persons that one is interested to interact with.

3 Voice Profiling

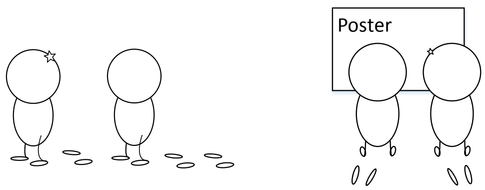
Following Doppler profiling, DopEnc is able to identify the effective trajectory that may lead to human interactions. The detection is based on the relative trajectory which however may misjudge the true intention of users in some cases. For example, (1) the relative trajectory may appear to be direct approach when two persons walk towards a same direction, but the latter one walks faster (Figure 9(a)); (2) the relative trajectory may appear to be roundabout approach

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.2017.2776915, IEEE Transactions on Mobile Computing



(a) Direct approach (b) Roundabout approach (c) Slant approach (d) Departing (e) Passing by

Fig. 8. Measured Doppler profiles of different categories of trajectories as well as the cleaned data.



(a) (b)

Fig. 9. Doppler profiling misidentifies "interactions" when (a) one person walks after the other, and (b) two persons walk towards a same target.

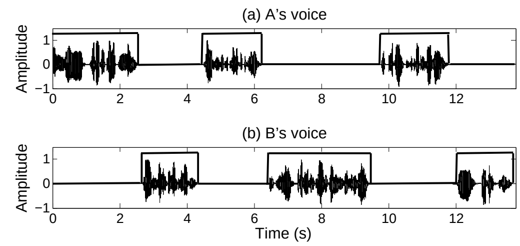


Fig. 10. Typical voice waveform during a conversation when two persons talk to each other. The conversation is of high alternativeness ratio and high duty ratio.

when two persons walk towards a same target, e.g., a poster or booth in the conference (Figure 9(b)). In both cases, the Doppler profiles indicate effective trajectories, whereas the two persons normally do not interact with each other. DopEnc further applies voice profiling to confirm that the two persons are engaged in an interactive conversation.

The work in [27] suggests that the conversation between two persons usually occurs within $0.5m-1.5m$. In DopEnc, voice profiling is performed to confirm the proximity and interest between two persons.

Existing voice processing systems, however, normally require pre-constructed voice feature databases of all speakers and complex machine learning based reasoning to recognize the voice segments [8], [9]. Some recent smartphone based systems reduce the processing overhead, but they still need voice feature database [28], or require voice pattern recognition across users [29], [30]. The computational overhead introduced by those approaches does not fit in the encounter based use cases in this application.

DopEnc leverages the result of Doppler profiling to trigger voice profiling. The voice profiling only needs to detect whether there is a conversation and does not purpose on recognizing speakers or their voices. Thus DopEnc employs a light-weight detection scheme that does not require any voice feature databases or complicated pattern recognition.

Figure 10 shows the voice waveform during a typical conversation between two persons. The conversation is of *alternativeness* [29], [30]. When two persons are talking to each other, they normally take turns to speak. Two persons thus alternate in speaking. DopEnc divides one voice trace into N time-slots. $v_i = 1$ if the person has voice in time-slot i; otherwise $v_i = 0$. Alternativeness ratio is defined as

$$\text{Alternativeness ratio} = 1 - \frac{\sum_{i=1}^{N}(v_{ai}\ \&\ v_{bi})}{N}, \qquad (3)$$

where $(v_{ai}\ \&\ v_{bi}) == 1\ iff.\ v_{ai} = 1$ and $v_{bi} = 1$. The conversation between person A and person B in Figure 10 is of high alternativeness ratio.

DopEnc adopts another parameter, *duty ratio*, which measures the time ratio of the combined voice segments from two speakers. The duty ratio is high when two persons are interested in each other, because they have topics to communicate. Duty ratio is defined as

$$\text{Duty ratio} = \frac{\sum_{i=1}^{N}(v_{ai} \mid v_{bi})}{N}, \qquad (4)$$

where $(v_{ai} \mid v_{bi}) == 0\ iff.\ v_{ai} = 0$ and $v_{bi} = 0$. The duty ratio is effective to characterize the conversation between two persons when they start talking.

To confirm a conversation in DopEnc, each smartphone first recognizes its owner's voice, and then the sender transmits the timestamps and durations of its owner's voice presences to the receiver. The conversation detection is performed at the receiver side by combining the trace of the sender's voice presence with the trace of the receiver's own voice presence.

Recognizing the phone owner's voice. Voice recognition has two phases: training and assessment. The owner trains her smartphone to recognize her voice features. During interactions with other people, the assessment phase verifies whether the recorded voice contains the owner's voice. DopEnc adopts similar techniques as in [28] to identify its owner's voice presence in the voice record. Since only the phone owner's voice needs to be recognized, DopEnc does not require any voice feature databases of other users or complicated reasoning and computation.

Exchanging the voice presence. The sender does not know the ID of the receiver, and vice versa. The acoustic signals transmitted by the sender for Doppler profiling cannot embed the sender's ID; otherwise, the Doppler offsets calcu-

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.2017.2776915, IEEE Transactions on Mobile Computing

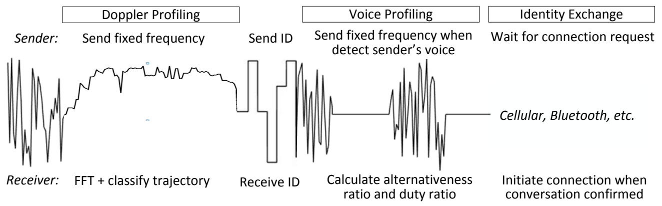


Fig. 11. Overall workflow of DopEnc and interactions between the sender and the receiver. The signals are depicted in frequency domain as detected by the receiver.

TABLE 1
Possible frequency bands supported when undersampling factor n is used. The inaudible frequency bands are underlined.

n	F_s^* (kHz)	Frequency band (kHz)	Δv (cm/s)
6	8.0	(8.0, 12.0), (16.0, 20.0)	3.6
7	6.9	(13.8, 17.25), (20.7, 24.15)	3.0
8	6.0	(12.0, 15.0), (18.0, 21.0)	2.7 *
9	5.3	(15.9, 18.6), (21.2, 23.9)	2.5

lated by the receiver will be incorrect. Due to the unknown ID of the opposite side, existing networking mechanisms, e.g., Bluetooth [31] or Wi-Fi [32], cannot be directly applied. DopEnc first transmits the sender ID in the same acoustic channel that was used for Doppler profiling. The channel is available since the sender has stopped walking and the Doppler profiling phase completes. Afterwards, when the sender's smartphone detects its owner's voice, it sends out an acoustic signal with the same duration of its voice presence. By measuring the start and end time points of such signals, the receiver can identify the timestamps and durations of the presence of the sender's voice.

Overall workflow of DopEnc. Figure 11 depicts the workflow of DopEnc, including three phases: Doppler profiling, voice profiling, and personal information exchanging.

DopEnc leverages the accelerometer on the smartphone to detect when a person starts walking. When the person is walking, her smartphone (the sender) transmits acoustic signals with a fixed frequency. When the sender stops walking, her ID is transmitted in the same acoustic channel using Frequency-Shift Keying (FSK) modulation. In our current implementation, we assign each smartphone with a unique two-digit number as the ID, whereas in practical usage, the sender and the receiver can consult with each other in the type of ID they use, such as phone number or Bluetooth ID. On the receiver's side, if no senders are present, the detected frequency of the recorded signals jitters dramatically as shown in Figure 11. Upon the detection of the sender's signals, the receiver extracts the frequency offsets and performs Doppler profiling. If the relative trajectory is detected effective, the receiver starts to receive the sender's ID. The security issue of binding devices is out scope of this paper.

Voice profiling is performed after the sender stops walking (which is inferred by the receiver by detecting zero frequency offset). If the conversation is confirmed, based on the choice of the users, the receiver may further transmit a connection request to the sender in order to initiate data exchange of related identity information. Both the connection request and the following data exchange can be performed with other higher data rate communication schemes, e.g., Bluetooth or cellular.

4 Coordinating Multiple Users

In previous sections, we explore the interaction between one pair of sender and receiver. This section presents the interactions among multiple users. The frequency band is shared by multiple users as in radio communication systems [33], [34]. In DopEnc, only smartphones of the moving persons transmit acoustic signals. DopEnc divides the acoustic frequency band into several channels and each walker's smartphone transmits on one channel. The smartphone is the sender as to the channel on which it transmits signals and meanwhile it is the receiver to other channels. The interactions among multiple users can be divided into the following categories:

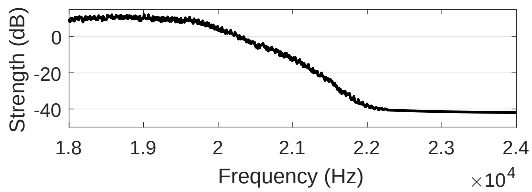
- *One sender to one receiver*. The interaction between one sender and one receiver is exhaustively discussed in the previous sections.
- *One sender to multiple receivers*. Since receivers do not transmit acoustic signals, the workflow of one sender to multiple receivers is the same as the workflow of one sender to one receiver.
- *Multiple senders to one receiver*. The signals from multiple senders will collide at the receiver side, which disrupts the receiver's Doppler profiling. Meanwhile, multiple moving persons whose smartphones transmit signals on the same channel cannot perform Doppler profiling with each other.
- *Multiple senders to multiple receivers*. The case of multiple senders to multiple receivers is the same as the case of multiple senders to one receiver.

Because receivers do not transmit acoustic signals, DopEnc only needs to coordinate senders to avoid collisions. DopEnc uses CSMA (carrier sense multiple access)-like multiple access to find a proper channel for each sender to transmit the acoustic signal. A walking person is a sender as to the channel that she uses; at the same time, she is a receiver on other channels. Collisions can be detected based on the root mean square (RMS) of the FFT signal on the channel. A collision occurs if the RMS is detected higher than a threshold. When a collision is detected, the sender switches to another channel after a random backoff interval. Since DopEnc performs FFT on the whole inaudible acoustic frequency band, the sender can obtain the RMSs for all channels with one time measurement. To reduce the probability of collisions in next channel access, DopEnc uses a special prioritization scheme for channel selection.

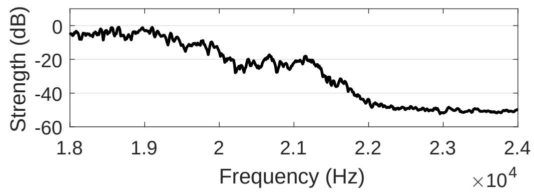
4.1 Channel Design

The number of channels that can be used in DopEnc is determined by the entire frequency bandwidth and the bandwidth of each channel. According to the undersampling theorem [11], the supported frequency band (f_L, f_H) follows the relationship:

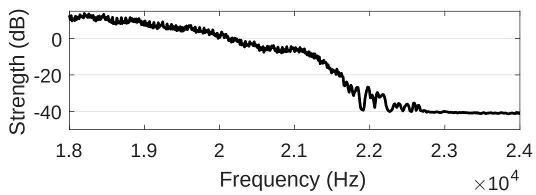
$$\frac{2 \cdot f_H}{n} \leq \frac{F_s}{n} \leq \frac{2 \cdot f_L}{n-1}, \forall n : 1 \leq n \leq \lfloor \frac{f_H}{f_H - f_L} \rfloor, \quad (5)$$



(a) Speaker of LG Nexus 5 to microphone of Samsung Galaxy Nexus i9250



(b) Speaker of Samsung Galaxy Nexus i9250 to microphone of LG Nexus 5

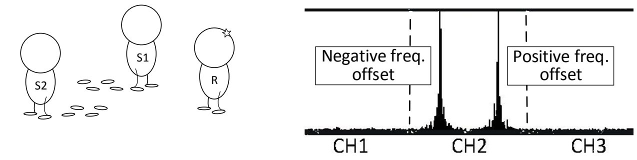


(c) Speaker of LG Nexus 5 to microphone of LG Nexus 5

Fig. 12. The signal responses between different transmitter-to-receiver pairs. Signal frequency of higher than $20kHz$ is not well supported on smartphones

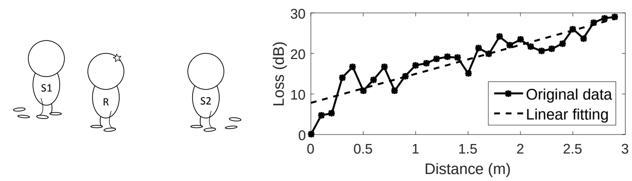
where n is the undersampling factor, F_s is the original sampling rate and $\lfloor \cdot \rfloor$ is the flooring operation. Table 1 summarizes the possible frequency bands supported in our system when different undersampling factors are used. DopEnc applies undersampling factor of 8 for the optimal setting of velocity resolution and temporal fidelity (Section 2.1), corresponding to an undersampling sampling rate F_s^* of $6kHz$. However, current smartphone speakers and microphones are tailored for low-frequency acoustic signals [35]. Figure 12 shows the amplitude response between different smartphone models. We can see that signal frequency of higher than $20kHz$ is not sufficiently supported. Therefore, DopEnc adopts a frequency band from $18kHz$ to $20kHz$, with a total bandwidth of $2kHz$.

The bandwidth of each channel is determined by the maximum frequency offset caused by human walking. People normally walk at $1.4m/s$ [36], corresponding to a frequency offset of $80Hz$. If two persons both walk, the relative velocity can be two times of each person's walking speed and thus the bandwidth is doubled to $160Hz$. Considering the positive Doppler offset (two persons walk closer) and the negative Doppler offset (two persons walk away), the bandwidth used of each channel should be further doubled to $320Hz$. DopEnc finally sets each channel with bandwidth of $400Hz$ to accommodate practical abnormality (e.g., extra frequency offsets introduced in some instant movements) and to be dividable of $2kHz$ frequency band. DopEnc evenly divides the entire frequency band into 5 ($2kHz/400Hz$) channels.



(a) (b)

Fig. 13. For receiver R, the frequency offset for the approaching sender S1 is positive and the frequency offset for the departing sender S2 is negative.



(a) (b)

Fig. 14. For receiver R, the far sender's signals are "overwhelmed" by near sender's signals. The attenuation rate of acoustic signals is as high as $6.6dB/m$.

4.2 Channel Prioritization

When many DopEnc users densely coexist, the limited number of channels may result in frequent collisions of Doppler profiling. Once a collision is detected, the collided senders switch to other "clean" channels. Since the Doppler profiling of DopEnc only needs to measure the relative velocities of a few seconds before the users stop walking, sufficient data can still be obtained after the channel switch. Random switch to another channel, however, causes high probability of new collisions. DopEnc minimizes the probability of making new collisions with a channel prioritization scheme. DopEnc classifies all available channels into the following three priority categories and switch the collided sender to the highest available one.

Free channels: A channel is free if the FFT RMS over the channel is detected lower than a threshold. A free channel has the highest priority.

Channels of empty positive sector: Each channel is composed of two sectors, a positive sector and a negative sector, corresponding to positive frequency offsets and negative frequency offsets in Doppler profiling. Figure 13 illustrates the measured signals by a stationary receiver (R) from two senders (S1 and S2). S1 and S2 are transmitting on a same channel. For the receiver R, the frequency offset for the approaching sender S1 is positive and the frequency offset for the departing sender S2 is negative.

In practical usage, DopEnc only concerns positive frequency offsets as they provide the information on how the approaching persons walk closer, i.e., direct approach, roundabout approach or slant approach. Therefore, the noises on the negative sector will not affect the detection of effective trajectories for Doppler profiling. Therefore, in DopEnc, a sender can switch to a busy channel as long as the positive sector contains no interferences.

Channels of low interference: Even if multiple senders walk towards the same receiver and use the same channel as shown in Figure 14(a), it is still possible to perform Doppler profiling for one of them. The attenuation of acoustic signals

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.2017.2776915, IEEE Transactions on Mobile Computing

Algorithm 1: Channel Coordination

Data:
$currentCh$ - Current channel
$switchCh$ - Channel to switch
t_{timer} - Timer for delay switching
$Var(\cdot)$ - Calculate variance
$Interf(\cdot)$ - Calculate the interference on channel's positive area
$GetBestCh(\cdot)$ - Get one highest-priority channel
$SetTimer(\cdot)$ - Set timer to a random value
T_{acce} - Threshold for human walking
T_{interf} - Threshold for collided channels

Result: Switch to a highest-priority channel when collision occurs

1 **while** $Var(acceleration) > T_{acce}$ **do**
2 // Sender is walking;
3 **if** $Interf(currentCh) > T_{interf}$ **then**
4 // Collision occurs
5 **if** $GetBestCh(\&switchCh)$
6 && $Interf(switchCh) < T_{interf}$ **then**
7 $SetTimer(\&t_{timer})$;
8 // Avoid simultaneous switching
9 **while** $t_{timer} > 0$ **do**
10 **if** $Interf(switchCh) > T_{interf}$ **then**
11 // Channel is seized by others
12 Cancel the switching task;
13 Break;
14 **end**
15 $t_{timer} \leftarrow t_{timer} - 1$;
16 **if** $t_{timer} == 0$ **then**
17 Switch to $switchCh$ channel;
18 Return success;
19 **end**
20 **end**
21 // Cannot switch to a non-collided channel
22 // Repeat the algorithm immediately
23 Continue;
24 **end**
25 **end**
26 **end**

is approximate $6.6dB/m$ as measured in Figure 14(b) (consistent with the estimation provided in [37]). As a result, the received signal strengths of two senders merely $2m$ apart may differ $20\times$ at the receiver side. Since DopEnc adopts FFT-based detection to locate the frequency offset with maximum amplitude, the signal of the closest sender may easily "overwhelm" the signals of far senders. Therefore, in DopEnc, a sender at last can switch to a busy channel as long as the FFT RMS over the positive sector of that channel is lower than a threshold.

4.3 Channel Coordination

Algorithm 1 describes the procedure of channel coordination in DopEnc. A collided sender switches to a channel of the highest priority (line 3-6). The sender waits for a random backoff time before switching to the new channel to avoid repeated collisions caused by the simultaneous switching of multiple senders (line 7-20). If there is no available channel detected or during the backoff interval the target channel has been occupied by other senders, the sender cancels the pending switching and re-executes the algorithm immediately (line 21-23).

5 Evaluation

We implement and test DopEnc with different use cases and on different smartphone models including Motorola Nexus 6, LG Nexus 5, LG Nexus 4, Huawei P7, LG G3, HTC Verizon, etc. In this section, we first present the operating range of DopEnc, and then the performance of the three key components in DopEnc separately, i.e., Doppler profiling, voice profiling and multiple access. We further examine the end-to-end performance of DopEnc in controlled experiments as well as a real world use case. We also measure the power profile of DopEnc in different use conditions. DopEnc bases on the change of relative walking speeds to perform Doppler profiling, and thus the absolute walking speeds of two users or whether one of users is static do not affect the system performance. Different stopping distances between two users only affect the voice detection accuracy, which is an active research topic of voice processing.

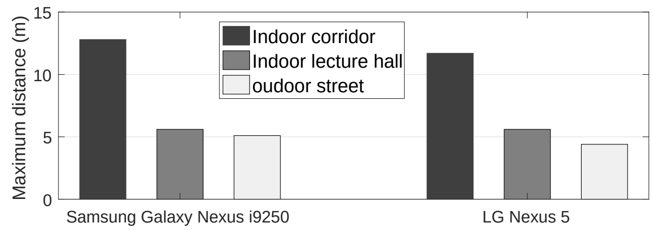


Fig. 15. Maximum transmission distance vs. different receivers and locations. We use another LG Nexus 5 as the transmitter

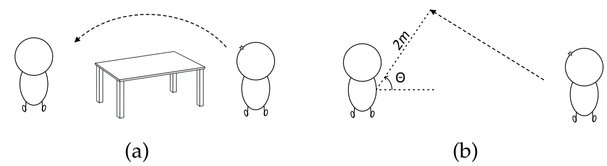


Fig. 16. Testing scenarios for roundabout approaching and slant approaching. (a) Roundabout approaching; (b) Slant approaching.

5.1 Operating Range

We measure the maximum propagation distance that the received signal can be detected after FFT. We consider different receiver models and places of types. We use a LG Nexus 5 as the transmitter with the volume set to the maximum. Figure 15 presents the results. As we can see that different locations dramatically affect the maximum distance. Indoor corridors provide most favorable transmission environment, and support acoustic signal transmission of about 12 meters, whereas outdoor open-space streets can only support about 5 meters. Smartphone models also affect the transmission distance but in a slighter degree. From our experiments, Samsung Galaxy Nexus i9250 (production year: 2011) is more sensitive to acoustic signals than the newer LG Nexus 5 (production year: 2013).

Larger maximum transmission distances provide us more flexibility since the transmission distance can be easily tuned by adjusting the speaker volume. How to select the optimal transmission distance remains further exploitation.

5.2 Doppler Profiling

Since DopEnc can easily identify the departing and passing-by trajectories by detecting their negative relative velocities, our experiments mainly focus on the three different types of approaching trajectories for evaluating Doppler profiling. We collect 100 traces for each category of the approaching trajectories (i.e., direct approach, roundabout approach and

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.2017.2776915, IEEE Transactions on Mobile Computing

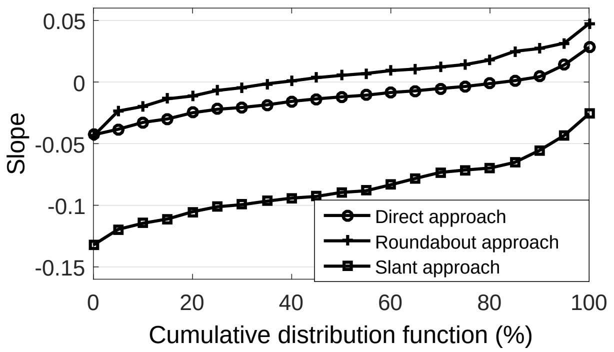


Fig. 17. CDF of the slopes for the Doppler profiling traces of the three different categories of approaching trajectories

slant approach, whereby the previous two constitute effective trajectories and the last one does not). The volunteers stand $6m$ away and walk towards each other with their normal gaits. For direct approach, the volunteers walk directly towards each other. Figure 16 illustrates the testing scenarios for roundabout approaching and slant approaching. For roundabout approach, three scenarios are tested where the paths between the volunteers are obstructed by objects with sizes of $1m\times1m$, $2m\times2m$ and $3m\times3m$. For slant approach, the volunteers' walking directions shift from each other with an angle of θ, which varies from $30°$ to $60°$ in our experiments. Since the angles of $10°$ and $20°$ are hard to be classified into direct approach or slant approach, we skip the experiments of these angles.

The classification of approaching trajectories is based on the slope of the fitted Doppler profiling. Figure 17 shows the Cumulative Distribution Function (CDF) of the slopes of Doppler profiling traces of the three types of trajectories obtained in our experiments. The slope difference between the slant approach trajectories and the other two types of effective approaching trajectories is obvious.

Table 2 further breaks down the measured slopes of slant approach trajectories according to different direction shift θ. The table summarizes the mean, max, and min slopes for slant approach trajectories of different θ ($\theta = 30° - 60°$) and compares with those of direct approach (i.e., $\theta = 0°$). As θ increases, the mean, max, and min slopes of slant approach trajectories dramatically decrease. The difference between the mean slopes of the direct approach trajectories (i.e., -0.0128) and the slant approach trajectories (i.e., -0.0613 for $\theta = 30°$, -0.0803 for $\theta = 40°$, -0.1037 for $\theta = 50°$, and -0.1010 for $\theta = 60°$) is large, which enables easy thresholding to distinguish these two approach cases. A threshold set to the smallest slope of direct approach trajectories, i.e., -0.0428, results in as low as 5% rate of misclassifying slant approach trajectories into effective ones. Since voice profiling is performed after the Doppler profiling, the effect of the low misclassification rate can be further mitigated.

In DopEnc implementation, another parameter affects the accuracy of trajectory classification, namely the length of the linear fitting to derive the slope. Figure 18 depicts the rate of misclassification when different fitting lengths are applied to the Doppler profiles. As DopEnc applies longer fitting length, the misclassification rate decreases because

TABLE 2
Mean, max, and min slopes for slant approach trajectories ($\theta = 30° - 60°$) in comparison with direct approach trajectories ($\theta = 0°$).

θ	$0°$	$30°$	$40°$	$50°$	$60°$
Mean	-0.0128	-0.0613	-0.0803	-0.1037	-0.1010
Max	0.0282	-0.0257	-0.0382	-0.0727	-0.0736
Min	-0.0428	-0.0967	-0.1110	-0.1276	-0.1322

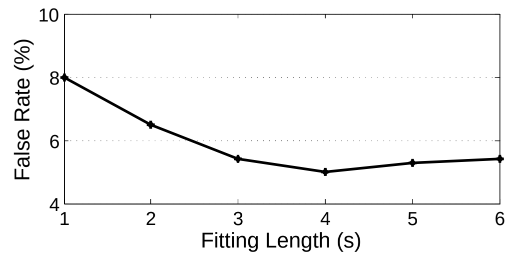


Fig. 18. Rate of misclassifying slant approach trajectories to effective trajectories versus different linear fitting lengths.

TABLE 3
Performance of one-on-one conversation confirmation with and without duty ratio.

Location		hall	corridor	street	mall
alternative	false negative	8.8%	7.0%	5.7%	9.2%
	false positive	15.4%	13.8%	18.3%	14.6%
alternative + duty ratio	false negative	9.4%	8.1%	6.7%	10.7%
	false positive	5.2%	4.7%	9.6%	8.7%

longer time of samples provide more stable observations on the relative velocity. Having too long fitting lengths, however, does not help in accuracy but reduces the flexibility of Doppler profiling since more data are required. According to our experimental observation we set the fitting length to $4s$, which brings approximately 5% misclassification rate.

5.3 Voice Profiling

We evaluate the performance of voice profiling in DopEnc with three volunteers. They stand within a square of $1m\times1m$. Two volunteers have one-on-one conversation with each other, while the third volunteer speaks to his phone as an interferer. Each volunteer's mobile phone records the time-stamps and durations of its owner's voice presences as the ground truth. We perform the experiment at four types of places, i.e., conference hall, corridor, street and shopping mall. The corridor is the quietest place and the shopping mall has the strongest background noise. Separate test traces combined to a total length of one hour are collected for each place. The voice processing parameters of each smartphone to identify its owner's voice is similar to [28]. The accuracy of voice profiling is evaluated off-line.

Table 3 summarizes the accuracy of voice profiling in two different versions of DopEnc. In the first version, DopEnc only uses alternativeness ratio to confirm conversation. In the other version, both alternativeness ratio and duty ratio are used. False negative measures the ratio that no conversation is indicated when two volunteers actually talk with each other. False positive measures the ratio that a

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.2017.2776915, IEEE Transactions on Mobile Computing

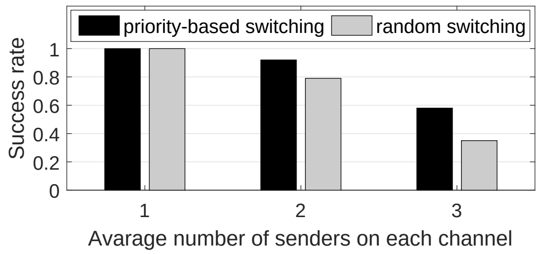


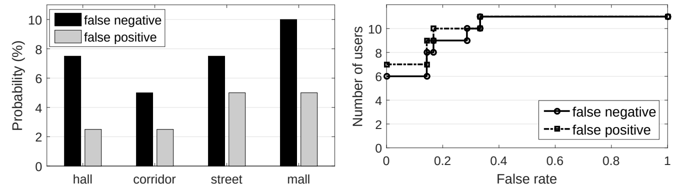
Fig. 19. Success rate of channel access vs. average number of senders per channel.

conversation is confirmed when two volunteers are not talking with each other. From the experiment results in Table 3, we see that only using alternativeness ratio, the false positives are high (the average false positive of the four places is 15.5%). When both alternativeness ratio and duty ratio are applied, the average false positive is reduced by approximate 8.5% and the increase of false negative is negligible (i.e., 1.0%). We want to emphasize that DopEnc proposes a new framework of conversation confirmation without voice database and can be implemented on COTS smartphones. Other related works, such as SeapkerSense [28] and Sociophone [29], presents detailed performance analysis of voice recognition under different levels of noises. Compared with purely voice processing based conversation confirmation works, DopEnc provides better performance, since DopEnc leverages Doppler profiling to filter persons with ineffective trajectories, resulting in smaller number of candidates for conversation confirmation.

5.4 Multiple Access

We evaluate the performance of the priority-based channel coordination scheme of DopEnc at our office ($15m{\times}10m$). We compare DopEnc's priority-based switching scheme with random switch and backoff scheme which is a natural generalization of the existing CSMA based method. During the experiment, we evaluate whether two volunteers can successfully avoid collisions and perform Doppler profiling. We put other smartphones on site and configure them to imitate DopEnc senders and transmit acoustic signals on different channels. We test the success rate of channel access when the number of DopEnc senders varies. Since channel access is the premise to the following operations in DopEnc (e.g., Doppler profiling and voice profiling), its success rate must be high (e.g., higher than 90%).

Figure 19 presents the experiment results for different numbers of senders per channel. When there is only one sender on each channel, the success rates of both channel switch schemes are 100%. When the average number of senders per channel becomes two, the success rate of random channel switch drops to 79%. DopEnc's priority-based channel switch still achieves a success rate of 92%, since it selects the best channels (i.e., no or low interference in the positive Doppler sector). For the case of three senders per channel, the performance of both schemes drops to below 60%. However, DopEnc is still better than the random channel switch as DopEnc considers the low interference in the positive Doppler sector.



(a) Controlled lab experiment (b) Real world use case

Fig. 20. End-to-end performance of DopEnc.

5.5 End-to-End Performance

Controlled lab experiment: We first evaluate the end-to-end performance of DopEnc with a pair of users at four different types of places, including conference hall, corridor, street and shopping mall. 80 traces are collected for each place. In the collected 80 traces, 40 traces have interactions (i.e., need to exchange identity information) and the other 40 traces do not. Among the 80 traces, 50 traces are effective trajectories and the other 30 traces are not. For the traces with effective trajectories, 40 traces have conversation and the other 10 traces do not. Figure 20(a) summarizes the accuracy of DopEnc, where the false negative and false positive are shown. The false positive corresponds to the cases where two users do not have interactions, but DopEnc detects an encountering. The false negative corresponds to the missed identification of encountering. On average, DopEnc achieves false negative of 7.5%, and false positive as low as of 3.8%. False positive is small because Doppler profiling has filtered a majority of non-interested persons with ineffective relative trajectories. The results suggest that DopEnc has high confidence on the identified persons (96.2% of them are accurate) but may miss some events (7.5% are missed).

Real world use case: We further test DopEnc in a real event with 11 participants. We host the school mentor-meets-mentee session for the lab students (including senior and junior postgraduates as well as final year undergraduates). The session takes place at the university bistro corner and lasts for approximate 1 hour. The participants include 2 female students/staff and 9 male students/staff. To facilitate the experiment, the participants record and submit a list of the people they have interacted with in sequence. Their records are used as groundtruth to verify the DopEnc results.

Figure 20(b) depicts the CDF of the number of users versus the false negative rate and the false positive rate during the event. From the results in Figure 20(b), we see that the majority of users (6 out of 11) have no false positive or false negative, and most of the rest users (3 out of 5) achieve low false positive ($< 15.0\%$) or false negative ($< 17.0\%$). On average, the false negative is only 9.7% and false positive is only 6.9%, which are consistent with what we observed in the above controlled experiments with a pair of users.

5.6 Power Profile

The power profiles of DopEnc on different smartphone models are similar, and we show the measurement result on HTC Verizon using Monsoon power monitor [38]. We put two smartphones near to each and adjust the speaker volumes so that smartphones can clearly “hear” each other.

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.2017.2776915, IEEE Transactions on Mobile Computing

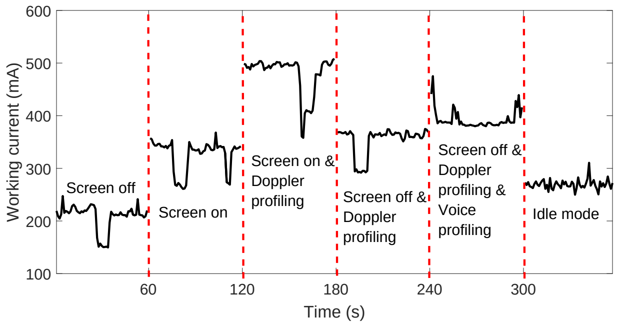


Fig. 21. Power profile of DopEnc.

Figure 21 presents the smartphone's working current (mA). The working voltage is $3.7V$. Multiplying the current with the voltage, we obtain the power of consumption. The total energy consumption can be got by multiplying the power with the considered time duration. The measurement result in Figure 21 is smoothed in one second. When the smartphone is initially turned on and the screen is off, the working current is fluctuating around $210mA$. When the screen is turned on at the $60s$, the working current rises to $340mA$. Doppler profiling of DopEnc is performed from $120s$ to $180s$ with the screen on and from 180s to 240s with the screen off. The working current maintains at around $500mA$ and $360mA$ respectively. At the $240s$, voice profiling is enabled and the working current becomes around $390mA$. Compared with the original working current of $210mA$, only $180mA$ ($390mA$–$210mA$) additional current is incurred by DopEnc.

The power consumption of DopEnc can be further reduced. In DopEnc, FFT computation drains the largest portion of energy, because 11.7 times 4096-FFT per second are performed to track relative trajectories. A simple optimization would be when no senders are present, DopEnc goes to idle mode, i.e., only to execute 1.5 times FFT per second by adjusting the overlapping ratio to zero (Section 2.1). As shown in Figure 21, in such a way the working current of idle mode ($300s$ onward) is reduced to $50mA$ ($260mA - 210mA$), corresponding to a working lifetime of 29 hours on the smartphone with $1450mAh$ battery capacity. DopEnc reduces FFT frequency only when there are no frequency offsets detected, i.e., no other users moving nearby. If reliable frequency-shifts are detected, DopEnc returns to its normal operation, and thus does not affect the accuracy.

6 DISCUSSION

Accuracy. DopEnc is not 100% accurate. Small false negative and false positive exist, which however is tolerable, since the encounter profile from DopEnc is usually used for non-critical applications, e.g., life logging, name card exchange, etc. The configurable settings in DopEnc help to adapt the system according to the user preference, making the system more conservative or aggressive. Three parameters including the slope threshold of Doppler profiling, alternativeness ratio and duty ratio of voice profiling can be adjusted to achieve different trade-offs in false positives and false negatives.

Scalability. Although our real world case study only test with 11 persons (mainly due to the limited number of available smartphones), we believe that DopEnc is scalable to support many more in big events. Firstly, DopEnc can support any number of stationary persons since only the smartphones of moving persons need to transmit acoustic signals. Secondly, the acoustic signal attenuates very fast in the air. Our measurement study shows that the maximum propagation distance is proximate $11m$ before the signal can make any detectable energy after FFT (tested with Nexus series and Huawei P7 with the largest sound volume setting). Such phenomenon means that the same acoustic channel can be comfortably reused every $11m$ away. The traditional hidden terminal problem in CSMA will not be an issue in DopEnc, as the Doppler profiling process focuses more on the signal trace when the sender moves close to the receiver where the strong near-far effect of acoustics makes it unlikely interfered by other unaware senders.

Phone positions. Doppler profiling is robust against phone positions (e.g., in a bag, in a pocket), because it bases on the tendency of relative velocity trace to classify the walking trajectory. The performance of voice processing is affected by different phone positions, making conversation confirmation systems based totally on voice processing frequently fail to work in different and harsh settings. Instead, DopEnc leverages the context information of human interactions, i.e., the relative trajectories of people walking, to filter majority of non-interest speakers, and thus conversation confirmation in DopEnc is more robust than the purely voice processing based systems.

Alternatives. There are a few human sensing techniques which might be adapted to address the encounter profiling, but with their own limits. High5 [4] is able to detect human hand-to-hand touch which might be extended to detect handshake as an indicator for identity exchange. The adapted solution, however, requires everyone wears the special device on the wrist. It will also be difficult for the devices of the two handshaking users mutually discover and pair up with each other. Human tracking and localization approaches can be applied to detect the encounters of people [5], [22], [39]. Existing approaches, however, mostly rely on infrastructure support. Human proximity detection based schemes like BLE or NFC [6], [7] give another indicator. The detection of encounters or proximity without contextual understanding, however, results in high false positives.

7 RELATED WORKS

Many intelligent mobile systems have been proposed to sense and monitor contexts and human activities in order to facilitate human daily lives [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51]. Unlike previous works, DopEnc targets at a new application of encounter profiling. This section summarizes existing works related to the three key components of DopEnc, i.e., Doppler profiling, voice profiling, and audio networking.

Doppler effect. Doppler effect has been explored in many HCI systems, such as connecting multiple devices [52], inferring user gestures [53], and imitating remote controllers [54]. Doppler effect has also been leveraged by some tracking systems to locate or navigate users [55], [56]. None of existing works study the construction or analysis of human walking trajectories. Although most of the above works [52], [53], [54], [55] work on acoustic signals to measure Doppler effect, they all focus on single-link measurements and do not address the multiple access problem.

Voice processing. Most existing voice processing systems, like [8], [9], aim at detecting conversations or identifying speakers in the conversation. SpeakerSense [28] can automatically recognize the person the user is talking with based on the processing of collected voice data and training data on smartphones. The previous works normally require a voice feature database of all potential speakers. SocioPhone [29] and SocialWeaver [30] group people based on conversation sessions and do not require voice databases. However, they involve complicated pattern recognition and comparison algorithms of high computational overhead and cannot be directly used for our application due to long processing time and high energy consumption. None of these works leverage the context information of human behavior, e.g., the trajectory of people walking as in DopEnc, to facilitate conversation confirmation.

Audio networking. Some works have been done to enable acoustic communications between devices that are equipped with microphones and speakers. U-Wear [57] enables data dissemination between ultrasonic wearable devices. Jiang et al. [58] utilize audible sound for near filed communications between smartphones, using OFDM and FSK modulations. All these works focus on high data rate and throughput for communication. Conventional CSMA based multiple access methods of Radio Frequency (RF) communications are used for coordinating multiple users on a single channel. No existing works consider coordinating multiple access for Doppler effect measurement as DopEnc does.

Multipath. Multipath effect could result in falsely classifying trajectories in Doppler profiling. For example, frequency shifts can be positive even if two users are moving apart because of the signal reflection of the wall. However, thanks to the high attenuation of acoustic signals and thus small transmission distances, we did not observe many such corner cases during our experiments.

8 CONCLUSION

DopEnc is the first smartphone system that facilitates automatic identification of persons that users interacted with. The encounter profiling in DopEnc incorporates Doppler effect of acoustic signals to identify the effective trajectories when people approach each other and voice profiling to identify their interactive conversations. Our experiment and user study demonstrate that with above techniques DopEnc can effectively to identify the persons in encounter-based human interactions.

ACKNOWLEDGMENTS

This work is supported by Singapore MOE AcRF Tier 1 MOE2013-T1-002-005 and NTU NAP grant M4080738.020.

REFERENCES

[1] Huanle Zhang, Wan Du, Pengfei Zhou, Mo Li, and Prasant Mohapatra. Dopenc: Acoustic-based encounter profiling using smartphones. In *Proc. ACM MobiCom*, 2016.

[2] LifeLog. Available at Google Play, visited Jun. 2016.

[3] Keep. Available at Google Play, visited Jun. 2016.

[4] Yuhwan Kim, Seungchul Lee, Inseok Hwang, Hyunho Ro, Youngki Lee, Miri Moon, and Junehwa Song. High5: Promoting interpersonal hand-to-hand touch for vibrant workplace with electrodermal sensor watches. In *Proc. ACM UbiComp*, 2014.

[5] Rijurekha Sen, Yongki Lee, Kasthuri Jayarajah, Archan Misra, and Rajesh Krishna Balan. Grumon: Fast and accurate group monitoring for heterogeneous urban spaces. In *Proc. ACM SenSys*, 2014.

[6] Carles Gomez, Joaquim Oller, and Josep Paradells. Overview and evaluation of bluetooth low energy: An emerging low-power wireless technology. *Sensors*, 12(9):11734–11753, 2012.

[7] Jianwei Niu, Wenfang Song, Chuang Liu, Lei Shu, and Canfeng Chen. Necas: Near field communication system for smartphone based on visible light. In *Proc. IEEE WCNC*, 2014.

[8] Douglas A. Reynolds, Thomas F. Quatieri, and Robert B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital Signal Processing*, 10(1-3):19–41, 2000.

[9] Frederic Bimbot, Jean Francois Bonastre, Corinne Frdouille, Guillaume Gravier, Ivan Magrin Chagnolleau, Sylvain Meignier, Teva Merlin, Javier Ortega Garcia, Dijana Petrovska Delacretaz, and Douglas A. Reynolds. A tutorial on text-independent speaker verification. *EURASIP Journal on Applied Signal Processing*, 2004(1):430–451, 2004.

[10] Joe Rosen and Lisa Quinn Gothard. *Encyclopedia of Physical Science*. Facts on File, 2009.

[11] Zheng Sun, Aveek Purohit, Raja Bose, and Pei Zhang. Spartacus: Spatially-aware interaction for mobile devices through energy-efficient audio sensing. In *Proc. ACM MobiSys*, 2013.

[12] Tektronix. Understanding fft overlap processing fundamentals. Technical report, Tetronix Inc., 2011.

[13] Wenguang Mao, Jian He, and Lili Qiu. Cat: High-precision acoustic motion tracking. In *Proc. ACM MobiCom*, 2016.

[14] Wei Wang, Alex X. Liu, and Ke Sun. Device-free gesture tracking using acoustic signals. In *Proc. ACM MobiCom*, 2016.

[15] Gerald Goertzel. An algorithm for the evaluation of finite trigonometric series. *The American Mathematical Monthly*, 65(1):34–35, 1958.

[16] Anshul Rai, Krishna Kant Chintalapudi, Venkata N. Padmanabhan, and Rijurekha Sen. Zee: Zero-effort crowdsourcing for indoor localization. In *Proc. ACM MobiCom*, 2012.

[17] Jue Wang, Deepak Vasisht, and Dina Katabi. Rf-idraw: Virtual touch screen in the air using rf signals. In *Proc. ACM SIGCOMM*, 2014.

[18] Richard W. Bohannon. Comfortable and maximum walking speed of adults aged 20-79 years: Reference values and determinants. *Age and Ageing*, 26(1):15–19, 1997.

[19] David C. Brogan and Nicholas L. Johnson. Realistic human walking paths. In *Proc. CASA*, 2003.

[20] Nirupam Roy, He Wang, and Romit Roy Choudhury. I am a smartphone and I can tell my user's walking direction. In *Proc. ACM MobiSys*, 2014.

[21] Mingjing Yang, Huiru Zheng, Haiying Wang, Sally McClean, and Nigel Harris. Assessing the utility of smart mobile phones in gait pattern analysis. *Health Technol.*, 2(1):81–88, 2012.

[22] Chunyi Peng, Guobin Shen, and Yongguang Zhang. Beepbeep: A high accuracy acoustic ranging system using cots mobile devices. In *Proc. ACM SenSys*, 2007.

[23] Moustafa Youssef, Adel youssef, Chuck Rieger, Udaya Shankar, and Ashok Agrawala. Pinpoint: An asynchronous time-based location determination system. In *Proc. ACM MobiSys*, 2006.

[24] Chengwen Luo, Long Cheng, Mun Choon Chan, Yu Gu, Jianqiang Li, and Zhong Ming. Pallas: Self-bootstrapping fine-grained passive indoor localization using wifi monitors. *IEEE Transactions on Mobile Computing*, 16(2):466–481, 2017.

[25] Branislav Kusy, Vladimir Protopopescu, and Morey Parang. Radio interferometric tracking of mobile wireless nodes. In *Proc. ACM MobiSys*, 2007.

[26] Miklos Maroti, Peter Volgyesi, Sebestyen Dora, Branislav Kusy, Andras Nadas, Akos Ledeczi, Gyorgy Balogh, and Karoly Molnar. Radio interferometric geolocation. In *Proc. ACM SenSys*, 2005.

[27] Edward T. Hall. *The Hidden Dimension*. Anchor, 1990.

[28] Hong Lu, A.J. Bernheim Brush, Bodhi Priyantha, Amy K. Karlson, and Jie Liu. Speakersense: Energy efficient unobstrusive speaker identification on mobile phones. In *Proc. Pervasive*, 2011.

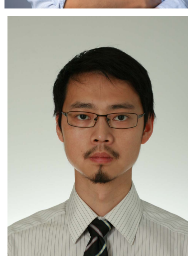
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2017.2776915, IEEE Transactions on Mobile Computing

[29] Youngki Lee, Chulhong Min, Chanyou Hwang, Jaeung Lee, Inseok Hwang, Younghyun Ju, Chungkuk Yoo, Miri Moon, Uichin Lee, and Junehwa Song. Sociophone: Everyday face-to-face interaction monitoring platform using multi-phone sensor fusion. In *Proc. ACM MobiSys*, 2013.
[30] Chengwen Luo and Mun Choon Chan. Socialweaver: Collaborative inference of human conversation networks using smartphones. In *Proc. ACM SenSys*, 2013.
[31] Goutam Chakraborty, Kshirasagar Naik, Debasish Chakraborty, Norio Shiratori, and David Wei. Analysis of the bluetooth device discovery protocol. *Wireless Networks*, 16(2):421–436, 2010.
[32] Daniel Camps-mur, Andres Garcia-saavedra, and Pablo Serrano. Device-to-device communications with wi-fi direct: Overview and experimentation. *IEEE Wireless Communications*, 20(13):96–104, 2013.
[33] Erfan Meskar, Terence D. Todd, Dongmei Zhao, and George Karakostas. Energy aware offloading for competing users on a shared communication channel. *IEEE Transactions on Mobile Computing*, 16(1):87–96, 2017.
[34] Yuan Wu, Jiachao Chen, Li Ping Qian, Jianwei Huang, and Xuemin Sherman Shen. Energy-aware cooperative traffic offloading via device-to-device cooperations: An analytical approach. *IEEE Transactions on Mobile Computing*, 16(1):97–114, 2017.
[35] Yu-Chih Tung and Kang G. Shin. Echotag: Accurate infrastructure-free indoor location tagging with smartphones. In *Proc. ACM MobiCom*, 2015.
[36] Raymond C. Browning, Emily A. Baker, Jessica A. Herron, and Rodger Kram. Effects of obesity and sex on the energetic cost and preferred speed of walking. *Journal of Applied Physiology*, 100(2):390–398, 2005.
[37] Lawrence E. Kinsler, Austin R. Frey, Alan B. Coppens, and James V. Sanders. *Fundamentals of Acoustics*. Wiley, 1999.
[38] Moonsoon Solutions Inc., http://www.msoon.com/, visited Mar. 2015.
[39] Jie Xiong, Karthikeyan Sundaresan, and Kyle Jamieson. Tonetrack: Leveraging frequency-agile radios for time-based indoor wireless localization. In *Proc. ACM MobiCom*, 2015.
[40] Shubham Jain, Carlo Borgiattino, Yanzhi Ren, Marco Gruteser, Yingying Chen, and Carla-Fabiana Chiasserini. Lookup: Enabling pedestrian safety services via shoe sensing. In *Proc. ACM MobiSys*, 2015.
[41] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. Contactless sleep apnea detection on smartphones. In *Proc. ACM MobiSys*, 2015.
[42] Cagdas Karatas and Marco Gruteser. Printing multi-key touch interfaces. In *Proc. ACM UbiComp*, 2015.
[43] Yuanchao Shu, Kang G. Shin, Tian He, and Jiming Chen. Last-mile navigation using smartphones. In *Proc. ACM MobiCom*, 2015.
[44] Shuyu Shi, Lin Chen, Wenjun Hu, and Marco Gruteser. Reading between lines: High-rate, non-intrusive visual codes within regular videos via implicitcode. In *Proc. ACM UbiComp*, 2015.
[45] Li Sun, Souvik Sen, Dimitrios Koutsonikolas, and Kyu-Han Kim. Widraw: Enabling hands-free drawing in the air on commodity wifi devices. In *Proc. ACM MobiCom*, 2015.
[46] Swarun Kumar, Stephanie Gil, Dina Katabi, and Daniela Rus. Accurate indoor localization with zero start-up cost. In *Proc. ACM MobiCom*, 2014.
[47] Hui-Yu Lee, Hao-Min Lin, Yu-Lin Wei, Hsin l Wu, Hsin-Mu Tsai, and Kate Ching-Ju Lin. Rollinglight: Enabling line-of-sight light-to-camera communications. In *Proc. ACM MobiSys*, 2015.
[48] Bo Hand, Feng Qian, and Moo-Ryong Ra. Human assisted positioning using textual signs. In *Proc. ACM HotMobile*, 2015.
[49] Kamran Ali, Alex X. Liu, Wei Wang, and Muhammad Shahzad. Keystroke recognition using wifi signals. In *Proc. ACM MobiCom*, 2015.
[50] Teng Wei and Xinyu Zhang. mtrack: High-precision passive tracking using millimeter wave radios. In *Proc. ACM MobiCom*, 2015.
[51] Tianxing Li, Chuankai An, Zhao Tian, Andrew T. Campbell, and Xia Zhou. Human sensing using visible light communication. In *Proc. ACM MobiCom*, 2015.
[52] Md Tanvir Islam Aumi, Sidhant Gupta, Mayank Goel, Eric Larson, and Shwetak Patel. Doplink: Using the doppler effect for multi-device interaction. In *Proc. ACM UbiComp*, 2013.
[53] Sidhant Gupta, Dan Morris, Shwetak N Patel, and Desney Tan. Soundwave: Using the doppler effect to sense gestures. In *Proc. ACM CHI*, 2012.
[54] Sangki Yun, Yi-Chao Chen, and Lili Qiu. Turning a mobile device into a mouse in the air. In *Proc. ACM MobiSys*, 2015.
[55] Wenchao Huang, Yan Xiong, Xiang-Yang Li, Hao Lin, Xufei Mao, Panlong Yang, and Yunhao Liu. Shake and walk: Acoustic direction finding and fine-grained indoor localization using smartphones. In *Proc. IEEE INFOCOM*, 2014.
[56] Lan Zhang, Xiang-Yang Li, Wenchao Huang, Kebin Liu, Shuwei Zong, Xuesi Jian, Puchun Feng, Taeho Jung, and Yunhao Liu. It starts with igraze: Visual attention driven networking with smart glasses. In *Proc. ACM MobiCom*, 2014.
[57] G. Enrico Santagati and Tommaso Melodia. U-wear: Software-defined ultrasonic networking for wearable devices. In *Proc. ACM MobiSys*, 2015.
[58] Rajalakshmi Nandakumar, Krishna Kant Chintalapudi, Venkata N. Padmanabhan, and Ramarathnam Venkatesan. Dhwani: Secure peer-to-peer acoustic nfc. In *Proc. ACM SIGCOMM*, 2013.

Huanle Zhang received the B.S. degree in Communication Engineering from Hangzhou Dianzi University, China, in 2011, and the M.S. degree in Communication and Information System from University of Electronic Science and Technology of China, in 2014. He was a Project Officer with the School of Computer Science and Engineering, Nanyang Technological University, Singapore, from 2014-2016. He is currently a Ph.D. student with the Department of Computer Science, University of California, Davis. His research interests include mobile systems and applications, cellular communication systems, and Internet of Things. He is a member of the IEEE.

Wan Du is currently an Assistant Professor in Computer Science and Engineering at the University of California, Merced. Before moving to Merced, Dr. Du had worked as a Research Fellow in the School of Computer Science and Engineering, Nanyang Technological University, Singapore, from 2011 to 2017. He received the B.E. and M.S. degrees in Electrical Engineering from Beihang University, China, in 2005 and 2008, respectively, and the Ph.D. degree in Electronics from the University of Lyon (cole centrale de Lyon), France, in 2011. His research interests include the Internet of Things, cyber-physical system, distributed networking systems, and mobile systems.

Pengfei Zhou received the B.E. degree from the Automation Department at Tsinghua University, Beijing, China, in 2009. He received the Ph.D. degree from the School of Computer Engineering, Nanyang Technological University, Singapore, in 2016. His current research interests include mobile computing and systems, localization, and cellular network communications. He is a member of the IEEE and ACM.



Mo Li received the BS degree in computer science and technology from Tsinghua University, Beijing, China, in 2004, and the PhD degree in computer science and engineering from Hong Kong University of Science and Technology, in 2009. He is a Nanyang assistant professor with the Computer Science Division, School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include distributed systems, wireless sensor networks, pervasive computing and RFID, and wireless and mobile systems. He is a member of the IEEE and ACM.

Prasant Mohapatra is a Professor in the Department of Computer Science and is serving as the Dean and Vice-Provost of Graduate Studies at University of California, Davis. He was the Editor-in-Chief of the IEEE Transactions on Mobile Computing. He has served on the editorial board of the IEEE Transactions on Computers, IEEE Transactions on Mobile Computing, IEEE Transaction on Parallel and Distributed Systems, ACM WINET, and Ad Hoc Networks. He is a Fellow of the IEEE and a Fellow of AAAS.