Rapid: A Multimodal and Device-free Approach Using Noise Estimation for Robust Person Identification

YUANYING CHEN, WEI DONG, and YI GAO, Zhejiang University & AZFT
XUE LIU, McGill University
TAO GU, RMIT University

Device-free human sensing is a key technology to support many applications such as indoor navigation and activity recognition. By exploiting WiFi signals reflected by human body, there have been many WiFi-based device-free human sensing applications. Among these applications, person identification is a fundamental technology to enable user-specific services. In this paper, we present Rapid, a system that can perform *robust* person identification in a device-free and low-cost manner, using fine-grained channel information (i.e., CSI) of WiFi and acoustic information from footstep sound. In order to achieve high accuracy in real-life scenarios with both system and environment noise, we perform *noise estimation* and include two different *confidence values* to quantify the impact of noise to both CSI and acoustic measurements. Based on an accurate *gait analysis*, we then adaptively fuse CSI and acoustic measurements to achieve robust person identification. We implement low-cost *Rapid nodes* and evaluate our system using experiments at multiple locations with a total of 1800 gait instances from 20 volunteers, and the results show that Rapid identifies a subject with an average accuracy of 92% to 82% from a group of 2 to 6 subjects, respectively.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing**; • **Computing methodologies → Supervised learning by classification**; • **Computer systems organization →** *Embedded and cyber-physical systems*;

Additional Key Words and Phrases: Multimodal person identification, Channel State Information (CSI), audio sensing, noise estimation

ACM Reference format:
Yuanying Chen, Wei Dong, Yi Gao, Xue Liu, and Tao Gu. 2017. Rapid: A Multimodal and Device-free Approach Using Noise Estimation for Robust Person Identification . *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 41 (September 2017), 27 pages.
DOI: http://doi.org/10.1145/3130906

This work is supported by the National Basic Research Program of China (No. 2015CB352400), the National Science Foundation of China (Nos. 61472360, 61502417, 61272456, 61472312), Alibaba-Zhejiang University Joint Institute of Frontier Technologies, Zhejiang Provincial Key Research and Development Program (No. 2017C02044), and the Fundamental Research Funds for the Central Universities (No. 2017FZA5013).
Author's addresses: Yuanying Chen, Wei Dong and Yi Gao, College of Computer Science and Technology & Alibaba-Zhejiang University Joint Institute of Frontier Technologies (AZFT), Zhejiang University, Hangzhou, Zhejiang, China; E-mail: {chenyy,dongw,gaoy}@emnets.org; Xue Liu, school of Computer Science and Dept. of Electrical, McGill University; E-mail: xueliu@cs.mcgill.ca; Tao Gu, Computer Science and School of Science, RMIT University. E-mail: tao.gu@rmit.edu.au;
Wei Dong is the corresponding author.

2474-9567/2017/9-ART41 $15.00
DOI: http://doi.org/10.1145/3130906

1 INTRODUCTION

Device-free human sensing has recently attracted many interests for its applications in indoor navigation [29], activity recognition [27], and entertainment. Unlike traditional device-free technologies like 3D camera, WiFi-based sensing has gained increasing popularity. WiFi-based sensing basically exploits WiFi radio signals and captures signal change patterns due to human movements for different applications. Since WiFi sensing does not require any special hardware, and WiFi is widely available and low-cost, it has become an attractive solution for device-free human sensing.

One of the key challenges in WiFi-based human sensing is how to identify an individual uniquely, known as person identification. With person identification, it is possible to associate an activity or a location with a person, opening up tremendous opportunities for personalized applications. For example, if the system knows who is sitting in front of a smart TV, it can push personalized application such as recommendations of his/her favorite TV channels. Several recent work [26, 31, 33] have been focusing on using WiFi signals for person identification. These approaches obtain Channel State Information (CSI) from WiFi signals to analyze a person's walking gait, and use supervised learning for person identification. Their basic idea is to first train a classifier (i.e., training process), and then use the classifier to perform person identification (i.e., identifying process). To achieve the best result, supervised learning requires the same model appears in both training process and identifying process. For person identification using WiFi sensing, study in WiWho [31] shows that the walking path has to be in parallel to AP-client line-of-sight (LoS) link at a fixed distance, in both training process and identifying process.

However, this assumption does not always hold in practice, i.e., a user may not always walk in a predefined path. In order to quantify the impact of the above difference between the two processes to system performance, we have implemented WiWho [31] and conducted a measurement study. We follow the same experimental setup, and deploy a pair of WiFi transmitter and receiver on the same side in a 2 meter-wide corridor. In the training process, subjects are required to walk strictly along the path in the middle of the corridor, i.e., a path which is 1 m away from LoS. While in the identifying process, we control the distance between walking path and LoS path to be 0.2 m, 0.6 m, 1 m, 1.4 m, 1.8 m, respectively.

We repeat the experiments and find that when the distance in the identifying process is 1m (pre-trained path), the identification accuracy is as high as that is reported in WiWho. However, when the distance increases, the accuracy drops rapidly. The reason is that CSI features are very sensitive to location due to multipath effect. As pointed out in [26], when user walks in a different path with respect to WiFi transmitter/receiver (Tx/Rx), the perceived CSI features (such as Doppler frequency) change even if the user retains the same gait with the same walking speed. Therefore, because of incompleteness of training samples in supervised learning, the performance of WiWho decreases in this case. Obviously, an ideal solution is to train all the possible walking paths for each person to obtain a complete feature set for classification. However, this will result in a surge increase on both (1) training cost, and (2) computation cost due to growing features.

Aiming to address the limitations in WiFi-based human sensing, in this paper we first analyze and quantify incompleteness of CSI training samples in supervised learning, we then propose a multimodal system to bring in additional modality to complement the CSI-based supervised learning. Based on our analysis, the performance of a CSI-based human sensing system is mainly influenced by the type of noise as follows. The lack of complete feature set due to incomplete CSI training samples leads to performance degradation, and they can be regarded as *system noise*. As in WiWho, when the walking path in the identifying process is different from the pre-trained path, system noise increase and distort CSI features. To provide a better estimation of system noise, we propose a *CFR Power Variance-Distance* (CPV-Distance) model to quantify the distance between walking path and LoS path. The principle of our CPV-Distance model lies on the fact that the vibration of CSI signals tend to be stronger when a person walks near the LoS path. Therefore, we can derive CSI confidence value which is the difference between CPVs as a metric to indicate system noise quantitatively. In this way, the estimation of the distance

between the walking path and pre-trained path can be made more accurately. If high system noise is found, we propose to bring in additional sensing modality to complement performance degradation. Ideally, the additional modality should not be sensitive to location, and also does not have a conflict with the strength of CSI-based sensing. In this work, we select acoustic sensing from user's footstep as a complementary sensing modality for its low cost and wide availability. However, in acoustic sensing, *environment noise* can severely affect system performance. For example, noise from air-conditioning is a typical type of environment noise affecting acoustic sensing. To quantify environment noise, we use acoustic confidence value derived from *Arithmetic Segmental Signal-to-Noise Ratio* (SSNRA), and compute the ratio of environment noise to footstep sound.

Based on the quantitative measurements of noise level in both CSI sensing and acoustic sensing, we design a weighted fusion algorithm to perform person identification in a more robust manner. We implement these ideas in a prototype system, named Rapid - a multimodal system which performs *Robust* low-cost, device-free Person IDentification using both CSI and Acoustic sensing. We conduct comprehensive experiments to evaluate the impact of different noise in three different locations. Our results show that Rapid provides higher accuracy for person identification compared to the state-of-the-arts in most of the scenarios, achieving an overall accuracy improvement of 15% to 25% with a group size of 2 to 6 subjects in practical environments.

In summary, this paper makes the following contributions.

- We analyze existing CSI-based human sensing, and discover that system accuracy decreases rapidly when a subject does not walk in a pre-trained path.
- We propose a quantitative approach to estimate noise from both CSI and acoustic signals, and use the estimated confidence values to perform robust person identification.
- We design and implement a prototype using a COTS radio chipset and acoustic sensor, incorporating both radio-based sensing and acoustic sensing modalities.
- We conduct comprehensive experiments to examine the feasibility and performance of Rapid on a gait database containing over 1,800 gait instances collected from 20 human subjects. The results show that Rapid achieves an overall accuracy of 92% to 83% with a group size of 2 to 6 subjects.

2 RELATED WORK

Wireless Sensing: Recent years have witnessed the emergence and development of WiFi-based sensing, which brings various applications such as indoor localization [29], activity and gesture recognition [12, 21, 27] and other types of applications such as sleep monitoring [13] and LoS identification [28]. These applications use commercial off-the-shelf infrastructures which are low-cost and ubiquitous to perform sensing of human position and activity in a device-free manner. Person identification can be considered a prerequisite for these applications since without that, it is not possible to associate a sensed activity to a specific person and provide customized services. Our paper leverages both wireless and acoustic sensing, two device-free methods, to achieve a more robust person identification.

Gait-Based Person Identification: Gait has been recognized as a unique signature for human beings. Compared with other methods, gait-based person identification has no addition constraints other than a person walking as usual. Prior works mostly use information collected from sources such as cameras, floor sensors, wearable sensors and microphones. In [25], authors use video camera to record walking process and then analyze silhouette to extract gait information from video frames. In [16], authors build a large inertial sensors-based database including at most 744 subjects and further compare different sensor based gait identification approaches. In [17], authors leverage force measurement sensors to record the footstep force signal and build corresponding profiles, achieving an accuracy of 90% to 97.5% for identifying a single target person. In [8], authors use footstep sound to extract acoustic features such as Mel frequency cepstral coefficients (MFCC) and energy, their results show the accuracy of identification is highly related to the features selected.

Recently, WiFi-based human sensing [26, 31] has been used to perform person identification, they both leverage the features of CSI signal in both time domain or frequency domain to identify a person. However, their walking path is limited in a **predefined** line, since they only leverage CSI-based features which is sensitive to person's location. Our work expands the scenario and use acoustic features to complement CSI-based sensing.

Multimodal Person Identification: Compared with monomodal person identification, multimodal person identification tends to be more accurate. Previous approaches [4, 7, 14] fuse the audio-based classifier and video-based classifier to identify persons. The authors in [18] first recognize the face region of a user, then extract features from face, speech and the motion of muscles in face when speaking, finally use an autoassociative neural network for person authentication. Our system integrates two device-free and low-cost modules (CSI module and acoustic module) in an IoT device called Rapid node to perform multimodal person identification. Different from simply fusion of features, Rapid conducts noise estimation and gives adaptive weights to classifiers in decision phase, thus leads to a more robust result.

3 UNDERSTANDING CPV-DISTANCE MODEL

In this section, we will first review CSI and multipath effects in an indoor environment, and then introduce CFR power variance (CPV). Finally, we build a model between CPV and Line-of-Sight distance, which is used in system noise estimation.

3.1 CSI Preliminaries

WiFi NICs can continuously track variations in wireless channel and collect the Channel State Information (CSI) which characterizes the Carrier Frequency Response (CFR) of the wireless channel at the granularity of each subcarrier. We denote CFR of a time-varying channel as $H(f,t)$, where f is the frequency and t is the time. Let $X(f,t)$ and $Y(f,t)$ be the frequency domain representations of transmitted and received wireless signals, respectively. Then the following equation gives the relationship among these three quantities, $Y(f,t) = H(f,t) \cdot X(f,t)$. CSI measurements contain the CFR values in 30 selected OFDM subcarriers for a received 802.11 frame, i.e., CSI is samples of CFR values in 30 selected frequencies. Let N_{Tx} and N_{Rx} represent the number of transmitting and receiving antennas, respectively. Then every 802.11 frame brings a matrix of $30 \times N_{Tx} \times N_{Rx}$ of CSI values. Taking time into consideration, we can obtain a stream of $30 \times N_{Tx} \times N_{Tx} \times T$ (complex) CSI values overall for a duration with T packets.

3.2 Multipath Effect

Multipath effect is the reason why CSI vibrates when a subject walks near the LoS path of the WiFi signal. In cluttered indoor environments, wireless signals often propagate through multipath, since they are usually reflected by walls and other infrastructures. If a wireless signal arrives at the receiver through N different paths, then CFR ($H(f,t)$) can be represented as the following equation [22]:

$$H(f,t) = e^{-j2\pi\Delta f t} \sum_{k=1}^{N} a_k(f,t) e^{-j2\pi f \tau_k(t)}. \tag{1}$$

where $a_k(f,t)$ is the complex value representing the initial phase offset and attenuation of the k^{th} path, $e^{-j2\pi f \tau_k(t)}$ is the phase shift on the k^{th} path that has a propagation delay of $\tau_k(t)$, and $e^{-j2\pi\Delta f t}$ is phase shift caused by CFO (Carrier Frequency Offset) Δf between the sender and the receiver. The changes in the length of a path lead to the changes in the phase and attenuation of the WiFi signal on the corresponding path. In our gait identification scenario, when a subject walks near the LoS path of the WiFi signal, one or multiple paths could be affected, resulting in variations of CSI measurements.

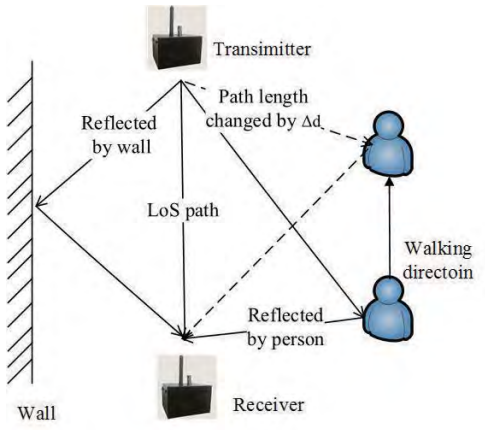


Fig. 1. Multipath effect and influence of walking on CFR.

3.3 CPV-Distance Model

In this subsection, we study how *CFR Power Variance* (CPV) indicates the distance between walking path and LoS path, and thus can be used as a metric measuring reliability of CSI features with a comparison to the CPV in trained path. Note that we have an assumption: the CSI features will be more unreliable if the walking path is further away from the pre-trained path. We will have a discussion about this assumption in Section 9. To show how *Variance* indicates the *Distance*, we first express CFR as a sum of *dynamic* CFR and *static* CFR. Dynamic CFR, represented by $H_d(f,t)$, is the sum of CFRs for paths whose lengths change with subject movement. For simplicity, we use a two-way model [32] which approximates all dynamic paths as one dominated path which is given by $H_d(f,t) = a(f,t)e^{-j2\pi d(t)/\lambda}$. Static CFR, represented by $H_s(f)$, is the sum of CFRs for static path. Thus, according to equation (1), the total CFR is given by the following equation.

$$H(f,t) = e^{-j2\pi\Delta f t}\left(H_s(f) + a(f,t)e^{-j\frac{2\pi d(t)}{\lambda}}\right). \tag{2}$$

We now consider how CFR power changes with a subject walking near the LoS path of the WiFi signal. We use a similar CFR power derivation method in [27] in which CFR power is leveraged to derive a CSI-speed model. Consider the scenario in Figure 1, a person moves at a constant speed such that the length of the reflected path changes at a constant speed v for a short time period. Let $d(t)$ be the length of this path at time t, i.e., $d(t) = d(0) + vt$. The instantaneous CFR power at time t can be derived as follows.

$$|H(f,t)|^2 = 2|H_s(f)a(f,t)|\cos\left(\frac{2\pi vt}{\lambda} + \frac{2\pi d(0)}{\lambda} + \phi\right) + |a(f,t)|^2 + |H_s(f)|^2, \tag{3}$$

where $\frac{2\pi d(0)}{\lambda} + \phi$ are constant values representing initial phase offsets, $|a(f,t)|^2$ and $|H_s(f)|^2$ are also constant CFR values. Then we observe the CFR power variance during a short time period of $[0,\tau]$, denote $\varphi = 2\pi v\tau/\lambda$ as the phase changed due to subject movement, now the CFR power variance can be expressed as the following equation (detailed derivations are illustrated in Appendix A).

$$Var(|H(f,t)|^2) = 2|H_s(f)a(f,t)|^2\left(sinc(2\varphi) - 2sinc^2(\varphi) + 1\right., \tag{4}$$

where $sinc(\varphi)$ is in its mathematic form which equals to $\sin(\varphi)/\varphi$, it can be easily proved that $sinc(2\varphi) - 2sinc^2(\varphi) + 1$ is positive. $H_s(f)$ is constant as mentioned before. We approximately treat attenuation of wireless signal as

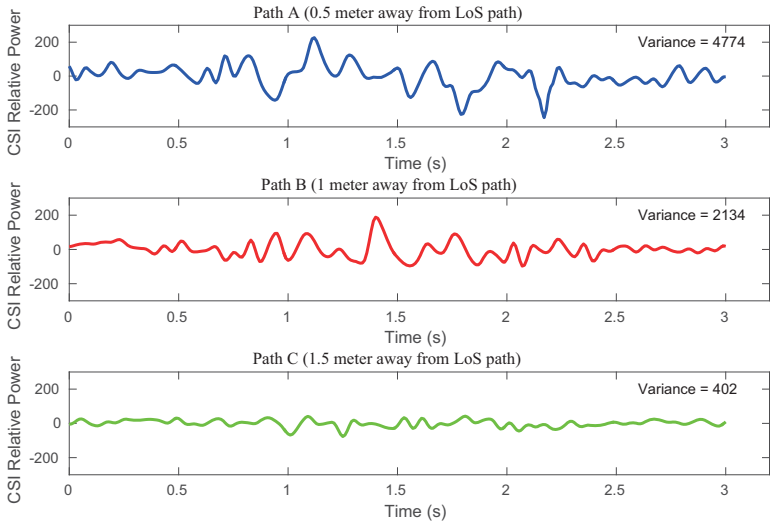


Fig. 2. Comparison between CSI power of different walking paths which are 0.5 meter, 1 meter, 1.5 meter away from LoS path

presented by Friis free space propagation model that the received power attenuates by a factor of square of the distance from the transmitter:

$$a(f,t)^2 \propto \frac{1}{d^2}. \tag{5}$$

Combined equation (4) with equation (5), we can deduce:

$$Var(|H(f,t)|^2) \propto \frac{1}{d^2}. \tag{6}$$

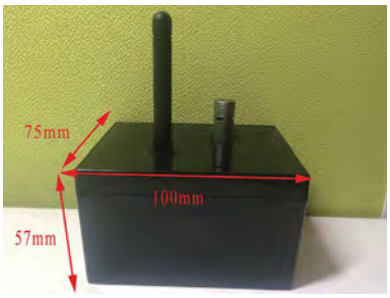
Equation (6) gives us a key insight: When a subject walks near the LoS path, the CFR power vibrates more strongly.

We further validate it in our experiments. We let a subject walk in different paths which are 0.5 meter, 1 meter, 1.5 meter away from LoS path of the WiFi signal, respectively. CSI samples obtained during this process are displayed in Figure 2. It is observed from Figure 2 that the CFR power variance (CPV) during the whole walk process is related to walker's distance from the LoS path. The CPV corresponding to three walking paths are 4774, 2134 and 402, respectively, i.e., an approximate ratio of 10:5:1. The experiment verifies our assumption above. This result motivates us to use the CPV during walking process as a metric to estimate the distance between a subject's walking path and LoS path. We leverage this model to measure the system noise of CSI in Section 7.

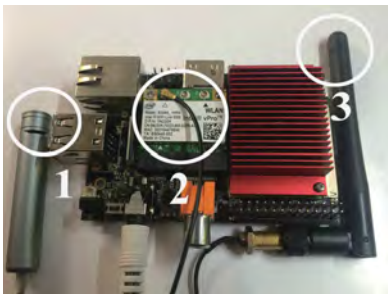
4 OVERVIEW OF RAPID

4.1 Rapid Node Prototype

We use an off-the-shelf laptop and a wireless router to record CSI and acoustic information at first. However, this implementation is somewhat obstructive especially in scenarios such as a narrow corridor. Therefore, we design and implement Rapid nodes which are much smaller and easier to deploy. The Rapid system consists of two small and easily deployed rapid nodes aiming to collect WiFi signal and acoustic signal. The Rapid node prototype (100 * 75 * 57 mm; 523 g) includes one HummingBoard Pro (HMB) [20] device powered by a 3000 mAH battery. The HMB is a low-cost ARM-based mini-computer that contains an on-board 1.2 GHz ARM Cortex-A9 processor and 1GB RAM. In order to enable CSI measurement, an Intel WiFi Link 5300 card together with an omnidirectional



(a) Appearance of Rapid node



(b) Components of Rapid node

Fig. 3. Our Rapid node consists of: 1) one unidirectional condenser microphone, 2) one Intel 5300 NIC with 3) one omnidirectional antenna attached to it. They are all controlled by one HummingBoard Pro (HMB) mini-computer.

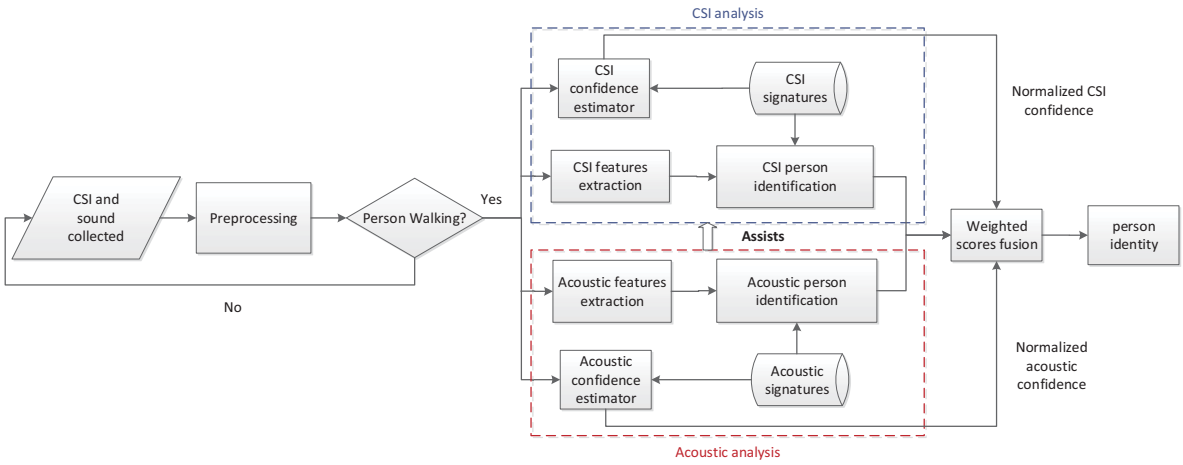


Fig. 4. System overview of Rapid

antenna is attached to HMB via the micro PCIe socket. We install the modified driver and firmware released by the CSI measurement tool [9] in the Cubox-i Debian Jessie OS running on the HummingBoard Pro. In our scenario, Rapid Node can be either a transmitter or a receiver, so we install hostapd [15] in each Rapid Node, enabling its NIC to work both in AP mode and client mode (but not simultaneously). In addition, a unidirectional condenser microphone is attached to HMB to collect acoustic signal. The appearance and components of Rapid node are illustrated in Figure 3.

4.2 System Overview

In this section, we give an overview of the Rapid system. We assume there are two stationary Rapid nodes in a room where Rapid is deployed. One Rapid node acts as a radio transmitter that sends packets periodically and the

other node acts as the radio receiver that continuously receives packets, CSI will be recorded in receiver during this process. At the same time, both nodes records acoustic signal using a unidirectional condenser microphone.

The Rapid system operates in the following two processes:

(1) Training process: In the training process, we build a database containing both CSI and acoustic features induced by footstep during walking. Note that in this stage, subjects are required to walk in a fixed path in a quiet environment for obtaining reliable features.
(2) Identification process: Identification process is conducted in real-life situations, i.e., any walking path and may be in a noisy environment. In this process, we estimate both CSI confidence and acoustic confidence, and derive two parameters to depict the level of noise.

In the identification process, Rapid leverages two modules to perform signal processing:

(1) CSI signal processing module: After collecting WiFi signals, this module goes through a pipeline of signal processing operations that consists of environment noise removal, walking detection, CSI confidence estimation and feature extraction.
(2) Acoustic signal processing module: Similar to the CSI processing module, this module deduces acoustic features from footstep sound for identification. In addition, this module helps the CSI processing module to *segment* CSI signal with the help of gait cycle.

The reason why we segment CSI signal is that Rapid uses segmented CSI to obtain *step* features which are more fine-grained in gait analysis. Concretely, according to signal length, there are two types of features for identification.

(1) Step feature: We use step cycle to segment collected signals, and then extract various features in each detected step.
(2) Walk feature: Walk feature is extracted from the entire walk process (multiple steps). It depicts the overall walking behavior of a subject.

We will discuss how fine-grained step features influence identification accuracy in Section 9.8. Obtained features will be then compared to pre-trained subject walking signatures using multiple classifiers. Then we combine the result of classifiers by a weight fusion algorithm and finally predict a subject's identity. The overview of Rapid system is shown in Figure 4.

5 SIGNAL PREPROCESSING

We now begin to introduce the whole pipeline of signal processing in Rapid. The first module is signal preprocessing. The original CSI stream extracted from commodity WiFi NICs is inherently noisy due to imperfect hardware and other unwanted multipath signals, while acoustic signal is degraded mostly by environment noise such as air-conditioning noise. Therefore, Rapid estimates and preprocesses these two types of noise sources using different methods at first.

5.1 CSI Preprocessing

CSI preprocessing consists of three de-noising stages: (1) Distant Multipath Removal, (2) Outlier Removal, (3) Low-pass Filtering.

5.1.1 Distant Multipath Removal. CSI data is the frequency form of CIR from which we can deduce a rough distant multipath [30]. In Rapid, the reflection from a distant object or person will introduce distant multipath part in CSI that distracts our experiment. We first convert CSI to CIR through IFFT (Inverse Fast Fourier Transform) and then remove the part which is more than 0.1 microseconds (i.e., paths longer than 30m) as displayed in Figure 5. We then transform CIR back to CSI, thus remove this kind of CSI noise introduced by long paths.

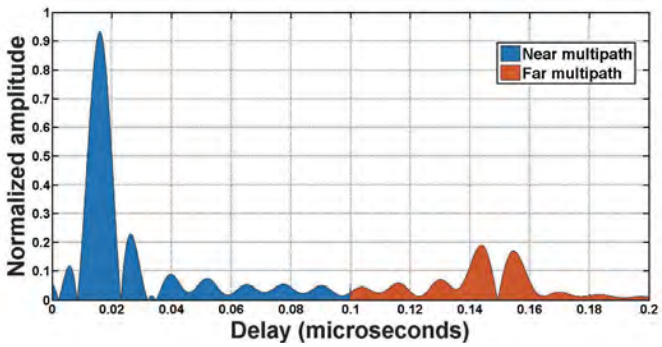
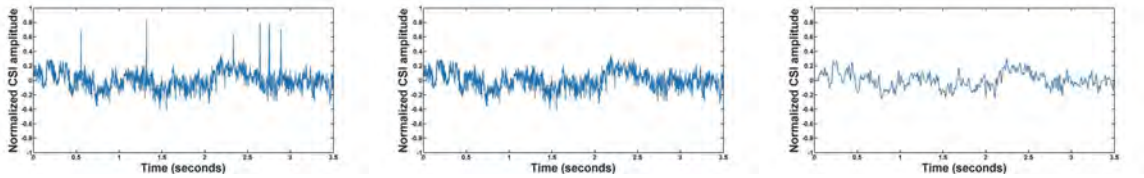


Fig. 5. CIR denoising: channel impulse response showing distant multipath.

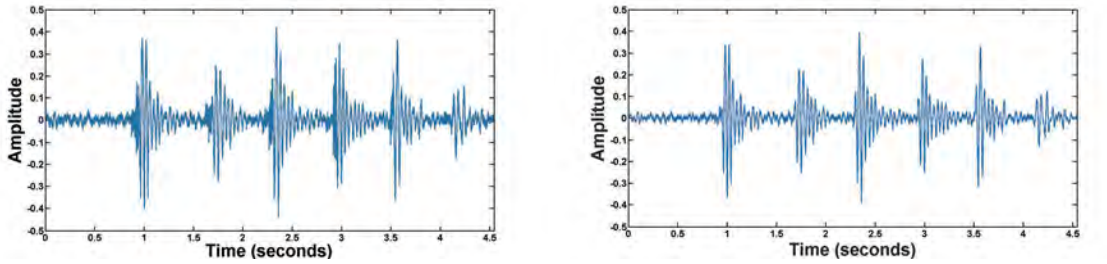


(a) Original CSI signal (b) CSI signal after outlier removal (c) CSI signal after low-pass filtering

Fig. 6. Preprocessing of CSI signal.

5.1.2 Outlier Removal. Internal state transition such as transmission rate adaptation and transmission power changes introduces burst noise in the CSI streams. Figure 6a shows the CSI signal of one walking process. It can be observed that there are some abrupt fluctuations in the signal. However, these outliers are not induced by body movements, and these burst noise undermine extracting walking features for identification. Rapid utilizes Hampel identifier [6] to remove these outliers. It treats all points out of the interval $[\mu - \gamma \times \sigma, \mu + \gamma \times \sigma]$ as outliers, where μ and σ are the median and median absolute deviation of CSI stream in the current window. γ is a parameter to control the sensitivity of detecting outliers and the most widely used value is 3. Figure 6b plots the CSI signal after removing outliers.

5.1.3 Low-pass Filtering. The second kind of CSI noise is high-frequency noise induced by hardware limitation, e.g., carrier frequency offset has a relatively high frequency. Since CSI power is the sum of a constant offset and a set of sinusoids [27], we use a sinusoid based bandpass filter to remove high-frequency noise. In Rapid, the walking speed of human is around $0.5 \sim 2$ m/s, hence the frequency of the variations in the CSI stream due to walking is often in a range of 10Hz to 80Hz (the speed of arms or legs in walking is about twice as much as the speed of body). We use Butterworth low-pass filter to cut off the out-of-scope frequency when applying it to the CSI signal in Figure 6b, with sampling rate $F_s = 1000$ *packets/s*. The filtered CSI signal is shown in Figure 6c. We can observed Butterworth low-pass filter successfully removes high frequency noise from CSI signal. Now the processed CSI stream can be utilized to detect walking and extract features for identification.



(a) Original acoustic signal (b) Denoised acoustic signal

Fig. 7. Preprocessing of acoustic signal using spectral substraction method.

5.2 Acoustic Signal Preprocessing

Acoustic signal collected by microphone is usually noisy. Figure 7a shows the acoustic signal of one walking process. In real-life environments, footstep signal is mostly degraded by additive noise that is uncorrelated with footstep signal like white Gaussian noise (WGN), colored noise. Therefore, noisy signal can be modeled as a sum of the clean footstep signal and the noise signal as

$$y(n) = s(n) + d(n), n = 0, 1, 2, ..., (N - 1) \tag{7}$$

where n is the discrete-time index, and N is the number of samples in the signal. Also, $y(n)$, $s(n)$, and $d(n)$, are the n^{th} sample of the discrete-time signal of mixed signal, clean footstep and random noise, respectively. The difference between CSI signal and footstep signal is that the latter is non-stationary in nature, therefore we use short-time Fourier transform (STFT) to divide footstep signal in small frames for further processing. Now representing the STFT of the time windowed signal by $Y_W(\Omega)$, $D_W(\Omega)$, and $S_W(\Omega)$ can be written as,

$$Y_W(\Omega) = S_W(\Omega) + N_W(\Omega) \tag{8}$$

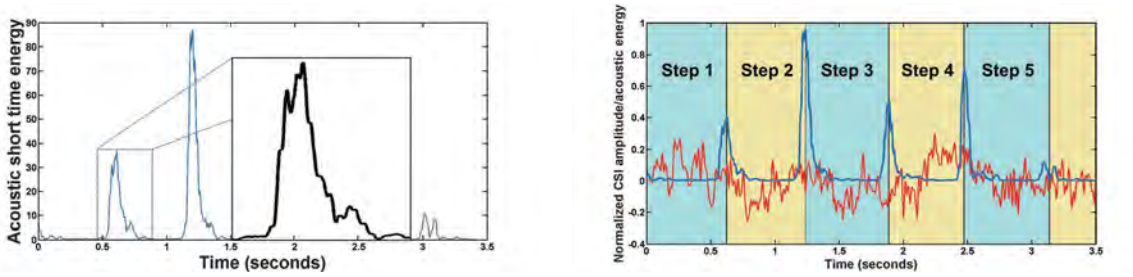
where Ω is the discrete-frequency index of the frame and W is the hamming window. In Rapid, we estimate the noise and apply an efficient audio denoising method - spectral subtraction [23][10] to denoise the footstep signal. Concretely, we estimate the noise during footstep pauses (i.e., the silent interval between each step), then subtract estimated noise spectrum $N_W(\Omega)$ from the noisy footstep spectrum $Y_W(\Omega)$ to denoise the acoustic signal induced from Equation 8 . The processed acoustic signal is shown in Figure 7b.

6 WALKING AND STEP DETECTION

The main effect of human walking on the received CSI stream is the vibration of signal. The vibration pattern is critical for detecting walking process and the uniqueness of different vibration patterns are exploited to classify different subjects. For acoustic signal, subject movement generates continuous step sound. We fuse two signals to detect the starting and ending point of each step, then detect more fine-grained information of steps.

6.1 Double-checked Walking Detection

For robust detection, we use two methods to analyze signal separately. The CSI-based method such as [31] leverages *motion energy* as an indicator to detect walking. The acoustic-based method inspects the continuous step sound to detect walking. To achieve robust identification, we extract the walking features only when two methods *both* detect the walking process. The two methods are illustrated as follows.



(a) Multiple peaks in one footstep

(b) CSI signal (red) segmented by acoustic signal (blue) of five steps

Fig. 8. Step detection

6.1.1 CSI-based Method. This method uses *energy* as a metric to detect walking. We first remove the Direct Current (DC) component of CSI stream by subtracting the constant offsets. The constant offsets can be calculated through a long-term averaging. We then convert the signal into frequency domain using FFT and obtain the coefficients of the FFT results. The motion energy can be calculated as $Energy = \sum_{i=1}^{windowlength/2} magnitude^2$, where magnitude values are the normalized FFT coefficients calculated over the time window. Compared to other activities such as sitting and standing, energy observed during walking activity is much higher, so the CSI-based method leverages energy as a way to detect if a subject is walking or not.

6.1.2 Acoustic-based Method. The acoustic-based method detects the start of walking based on the fact that moving humans generate continuous impact sound. When there are no human walking, the variances in acoustic measurement are mostly caused by noise. Since the footstep introduces large variance in acoustic measurement, we simply use a thresholding algorithm to detect walking. Rapid first calculates the variance *var(noise)* for a segment in 5 seconds at the beginning of an experiment. Then we set the detection threshold as 3 times the noise *var(noise)* so that a step sound is detected when the variance of the measurement substantially deviates from the average noise level. Our system detects walking when this abrupt change (pulse) is detected.

Moreover, we conduct an experiment in Section 9.9 to investigate on the relation between accuracy of identification and the numbers of required step segments. This experiment answers the question how the end of the walk interval is defined. The result shows that only $4 \sim 5$ steps are practically enough to build the classifier as illustrated in Figure 17. Therefore, the end of the walk interval in trace is defined as the timestamp of 6th acoustic pulse induced by footstep (i.e., selected trace contains 5 steps). In the next subsection, we will present how to detect steps for walking traces, a walking trace containing less than 5 steps will be regarded as an invalid trace and then be discarded.

6.2 Step Detection

Prior work [16] has shown that the shape of steps varies noticeably for different people. Therefore, we can segment the whole walking process into steps to extract more fine-grained features, e.g., the CSI signal shape of the first detected step. In Rapid, the acoustic signal collected consists of step sound (peaks in waveform) which marks the start timestamp or end timestamp of each step naturally. However, to detect more accurate steps, we first process the original acoustic waveform with short-time energy estimation, we use a typical short-time interval of 20 ms. Figure 8b displays synchronized CSI waveform and acoustic short-time energy waveform of one

walking process. The peaks in acoustic waveform divide the whole CSI waveform into different areas. However, because of the variety of gait, one footstep may have two peaks as shown in Figure 8a. From the figure, We notice another increase, though not as prominent, before one footstep sound terminates. The second burst happens depending on the walking style and the footwear of the person, when the metatarsophalangeal joint structure makes contact with the floor or lifts off from the surface [1]. We need to identify whether or not the peaks belong to one footstep sound in the acoustic signal. Therefore, here we use a robust step detection algorithm with the help of minimum and maximum interval between two footsteps. This detection method involves the following five processes(The first three roughly segment the signal, the latter two verify and store the step segments):

(1) Scan the input acoustic waveform, find the first sampling point whose amplitude exceeds the threshold that is given in walking detection, denote its timestamp as Ts_1;
(2) Scan the remaining sequence, find the first sampling point whose amplitude below the threshold and denote its timestamp as Te_1. Let $Bun_1 = [Ts_1, Te_1]$ be a timestamp bundle, i.e., a potential step, and Tm_1 be the median timestamp in Bun_1.
(3) Repeat the above two processes, find continuous $Bun_1, Bun_2, ..., Bun_N$ and corresponding $Tm_1, Tm_2, ..., Tm_N$, where N is the number of timestamp bundles.
(4) Calculate the distance between each adjacent bundle in time domain, then aggregate the bundles with distance smaller than δ_{tmin}, clean up the bundles with distance bigger than δ_{tmax}, push the bundles between them into output vector $\boldsymbol{St}$, where δ_{tmin} is the minimum interval between human steps, δ_{tmax} is the maximum interval between human steps.
(5) Repeat the 4th process until the the size of output vector $\boldsymbol{St}$ reaches to 5, which means we obtain vector containing 5 detected step segments.

The details are illustrated in Algorithm 1. Since we only scan input acoustic signal and pulses in it for one time, the time complexity of this algorithm is $O(N)$, N is the length of input signal.

The algorithm set two thresholds δ_{tmin} and δ_{tmax} to verify a step segment and then push it into output vector until there are 5 step segments in vector. The minimum interval δ_{tmin} (0.2s in our system) is used to detect whether the event that one footstep induces two pulses (peaks) as displayed in Figure 8a. If we find the distance between two pulses in time domain is less than δ_{tmin}, we aggregate two pulses into one pulse of a footstep. While maximum interval δ_{tmax}(2s in our system) is to cope with the situation that user walks, stops and then starts walking again. If we find the interval between two consecutive steps is larger than δ_{tmax}, we do not consider them belonging to the same walking process, i.e., event that the subject pauses while walking has happened. The two thresholds are obtained according to the ground truth of our experiments, i.e., the normal gait cycle time of subjects will not be less than 0.2s or more than 2s. A walking trace containing less than 5 steps will be regard as a invalid trace and then be discarded (I.e., we clean up the output vector St as illustrated in Algorithm 1) . After step detection, the whole walking trace is segmented as illustrated by Figure 8b Note in the whole process of step detection, we get features including interval length of each step and gait cycle time (average of the former), which we will use to construct a classifier in Section 8.1.

7 NOISE ESTIMATION

A straightforward way to identify a subject is to build a classifier that uses all the CSI and acoustic features equally. However, the performance of this method deteriorates due to the influence of system noise (deviation of walking path) or environment noise (e.g., the noise from air-condition). The reason is that the two types of noise undermine the CSI and acoustic features separately [26][23] and make the classifier unreliable. Therefore, we need to estimate the level of system noise and environment noise separately to adaptively give weights to features. Noise estimation answers the following question: *In what situation, we should use more acoustic features*

Algorithm 1 Step Detection Algorithm

Input: S: amplitude vector of acoustic stream; T: timestamp vector of acoustic stream; δ_s: threshold denoting minimum amplitude of footstep sound; $[\delta_{tmin}, \delta_{tmax}]$: interval range between human steps; N_{req}: number of required step segments (5 in our system);
Output: St: vector containing timestamps of N_{req} detected step segments;

```
// Part 1: segment the whole walking process based on δ_s;
initialize N = 0;
for i = 1 to |S| do
    // search start of a footstep pulse;
    for j = i to |S| do
        if S_j > δ_s then
            i = j; Ts_N = T_j; break;
    // search end of a footstep pulse;
    for j = i to |S| do
        if S_j < δ_s then
            i = j; Te_N = T_j; N = N + 1; break;
    Tm_N = (Ts_N + Te_N)/2;
    Bun_N = [Ts_N, Te_N];
// Part 2: verify and push step segments based on [δ_tmin, δ_tmax].
initialize numOfSteps = N − 2;
initialize St empty;
for i = 1 to numOfSteps do
    if |St| == N_req then break;
    if |Tm_{i+1} − Tm_i| < δ_tmin then
        // two peaks for one footstep
        Tm_{i+1} = (Tm_i + Tm_{i+1})/2; Continue;
    if |Tm_{i+1} − Tm_i| > δ_tmax then
        // user stops and then starts
        set St empty; Continue;
    push [Tm_i, Tm_{i+1}] to St;
```

than CSI features and vice versa? We first propose two *Confidence Values* to quantitatively estimate system noise and environment noise.

7.1 CSI Confidence

We consider system noise, i.e., deviation of walking path, as the main factor that distorts the CSI features. In Section 3, we have proved that *CFR power variance* (CPV) can be used as a metric to estimate the distance between actual walking path and the LoS path. Note that our assumption is that CSI features are more unreliable if the walking path is further away from pre-trained path, which we will validate in Section 9. We can now use CPV to calculate CSI confidence value, i.e., quantify this derivation. In the training phase, we let different subjects walk in a fixed path for multiple times and collect Q CSI traces overall. We calculate the CFR power variance in each CSI trace and denote the average as $\overline{CPV_{train}}$. In the identifying process, we denote the calculated CFR

power variance as CPV_{test}. The CSI confidence value ξ_{csi} is defined as

$$\xi_{csi} = \frac{1}{|\overline{CPV_{train}} - CPV_{test}| + 1} \tag{9}$$

This heuristic equation normalizes the confidence value in a range from 0 to 1. We use this value to indicate the reliability of CSI features.

7.2 Acoustic Confidence

Previous work [19][8] have used features extracted from footstep sound to perform person identification. However, environment noise distorts original footstep signal just like it influences speech signal. Many existing works focus on suppression of noise to enhance acoustic signal in the field of signal processing over the past few decades. A classic and effective denoising method is spectral subtraction [3] as we use in Section 5. However, this method introduces another annoying perceptible tonal characteristic, known as remnant musical noise [2] which is not easy to reduce. Therefore, we need to estimate the noise level to calculate the acoustic confidence, which is another important factor to weight acoustic feature. Our approach towards estimating noise in footstep signal stems from speech noise estimation since these two kinds of signals are both non-stationary. The acoustic confidence value $\xi_{acoustic}$ is defined as

$$\xi_{acoustic} = \frac{SSNRA_{test}}{\overline{SSNRA_{train}}} \tag{10}$$

where $SSNRA_{test}$ is the Arithmetic Segmental Signal-to-Noise Ratio (SSNRA) [24] in the identification process, and $\overline{SSNRA_{train}}$ is the average SSNRA value in the training process. SSNRA is used to depict Signal-to-Noise in non-stationary signal. The prerequisite of SSNRA calculation is the segmentation of signal, i.e., dividing the whole signal into purely noise segment and footstep with noise segment. Figure 9b plots this segmentation. Instead of previous segmentation method that we have used in Section 6.2, here we employ a noise threshold to find footstep sound. This process is similar to the VAD (Voice Activity Detection) process in speech recognition. However, here the signal is footstep instead of voice. Note in Algorithm 1, we detect the start and end of all footstep pulses, and denote them as **Ts** and **Te** separately, then $[Ts_N, Te_N]$ can be treated as part of target segment. SSNRA is then calculated as

$$SSNRA = 10 \log\left(\frac{1}{K}\sum_{i=1}^{K}\frac{\sum_{j=1}^{M_i} s_i^2[j]}{\sum_{j=1}^{M_i} n_i^2[j]}\right) \tag{11}$$

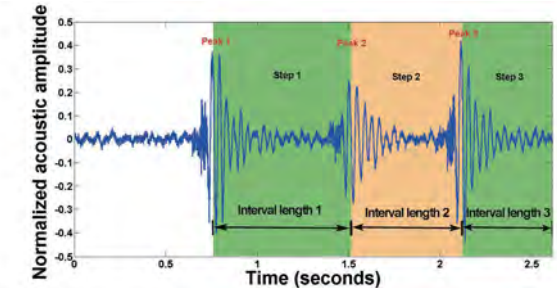
where $s_i[j]$ and $n_i[j]$ denote j-th footstep sound and noise sample in the i-th segment of acoustic signal respectively, M_i denotes the number of samples in the i-th segment and K is the number of segment.

8 MULTIMODAL PERSON IDENTIFICATION

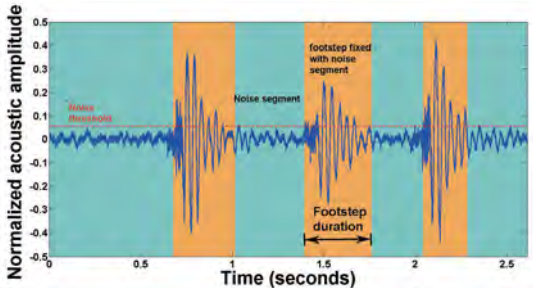
In this section, we first present the gait features we extracted and then use machine learning techniques to build a multimodal person identification system.

8.1 Constructing Features

Before constructing features, we first present two segmentation methods of acoustic signal to give a clear definition to some features. Figure 9a plots the segmentation that is similar to the step detection process in Section 6.2, where step segments are denoted by peaks in acoustic signal. From the figure, we can calculate the *length interval of each step*. The average of all the length interval of steps is *gait cycle time*. Figure 9b plots the segmentation that we use in Section 7.2, it divides the whole acoustic signal into purely noise segments and footstep mixed with noise segment. Noise threshold is important in this process since it denotes the start and end



(a) Step segments in acoustic signal using peaks



(b) Noise segments vs. footstep fixed with noise segments in acoustic signal using noise threshold

Fig. 9. Comparison between two different segmentation methods for extracting interval lengths of each step, footstep duration

of footstep signal. We derive the *footstep duration* in this segmentation. We now divide the whole features into three kinds based on the sensitivity to two kinds of noise:

- **Confidence Independent**: This kind of features are *almost* resistant to noise in our scenarios. Gait cycle time measures the time duration between two consecutive events that the right heel touches the ground, which is calculated in walking and step detection phase using Algorithm 1. Gait cycle time and length intervals of each step belong to this kind. This is because no matter how big the noise is, as long as the noise is white gaussian noise and the walking process is detected, their value will not change. As shown in Figure 7, the gait cycle time does not change before and after denoising.
- **CSI Confidence Sensitive**: The reliability of these features are sensitive to system noise from CSI signal but resistant to environment noise, e.g., *shape of CSI signal* for the whole walking process and each step.
- **Acoustic Confidence Sensitive**: On the contrary, acoustic features such as *MFCCs* are typical features that are sensitive to environment noise. Besides, *footstep duration* is also related to the environment noise level because we detect the start time and end time of each step sound based on the noise threshold as shown in Figure 9b.

Table 1 describes the general feature set and sources of each feature. Table 2 describes features which describe *the shape of CSI signal* for the whole walking process and each step in Table 1. Our features for CSI classifier in the table is similar to features used by Zeng et al. in [31], who also evaluate the impact of these features on person identification. Since scenarios in our system for CSI classifier is similar to [31], we leverage these features which have been proved to be effective. We conduct experiments to investigate how the feature – *shape of segmented CSI signal* in the step level affect the accuracy of identification in Section 9.8.

8.2 Constructing Classifiers

After designing the feature space that capture subject's walking and step characteristics, we start to construct three classifiers for each kind of features. We use the LibSVM tool [5] with the Radial Basis Function (RBF) kernel in the training and identification process. In the training process, we build a database containing both CSI and acoustic features, and then we divide the features into three unrelated types as illustrated in Table 1. Then three classifiers with optimal values for parameters are constructed separately according to three feature spaces. Note that in order to quantify the result of classification, we produce *probabilistic classifiers*, which are able to predict a fitness probability distribution over a set of classes rather than output the most likely class.

Table 1. General feature set in Rapid

Feature type	Feature	Source
Confidence independent	Gait cycle time	Acoustic
	Interval lengths of each step	Acoustic
CSI confidence sensitive	*Shape* of CSI signal for walking process	CSI
	Shape of CSI signal for each step	CSI/Acoustic
Acoustic confidence sensitive	MFCCs	Acoustic
	Energy	Acoustic
	Footstep duration	Acoustic

Table 2. Features extracted from shape of CSI signal for the entire walking process and each step.

Min	Max	Mean
Standard Deviation	Skewness	Kurtosis
Spectral Entropy	The first quartile	Median
The third quartile	Mean Crossing Rate	

In the identifying process, we collect gait instances of the target human subject. We also construct three feature space from the instance, and use corresponding classifier to identify them. Each classifier outputs a fitness vector, representing the fitness probability of instance in the given set of candidates. We denote three fitness vectors as $\boldsymbol{F_{independent}}$,$\boldsymbol{F_{csi}}$,$\boldsymbol{F_{ac}}$ separately, where the length of each vector is the number of candidates in the set.

8.3 Fusing Classifiers

In a multimodal system to perform person identification, fusion is possible at three levels: feature extraction level, matching score level or decision level. Fusion at the feature extraction level combines different features in the recognition process, while decision level fusion performs logical operations upon the monomodal system decisions to reach a final resolution. Score level fusion matches the individual scores of different recognition systems to obtain a multimodal score, which is usually preferred by most of the multimodal systems [11].

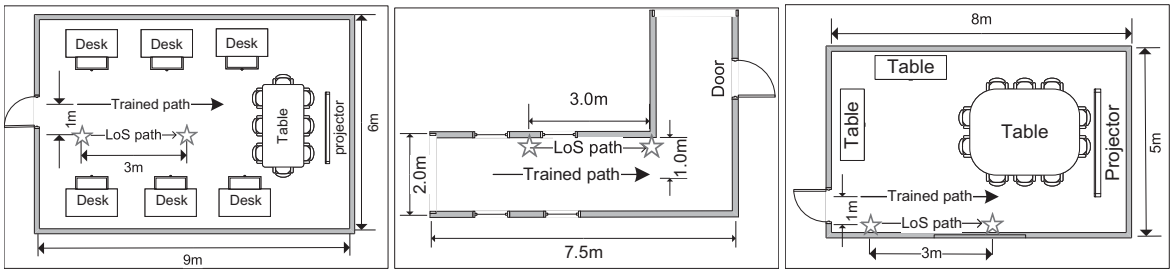
In Rapid, we use a three-step process to perform a score-based fusion based on the fitness vectors from three classifiers.

(1) Score Normalization: Since monomodal scores are usually non-homogeneous, the normalization process transforms the different scores of each monomodal system into a comparable range of values. From constructed classifiers, we obtain three fitness vectors $\boldsymbol{F_{independent}}$, $\boldsymbol{F_{csi}}$, $\boldsymbol{F_{ac}}$. In Rapid, we utilize *z-score* to transform them into a distribution with zero mean and unitary variance, take $\boldsymbol{F_{csi}}$ for example, for every f_j in the vector, we calculate:

$$Norm(f_j) = \frac{f_j - mean(\boldsymbol{F_{csi}})}{std(\boldsymbol{F_{csi}})} \quad j = 1, 2, ..., J \tag{12}$$

$Norm(\boldsymbol{F_{csi}})$ is the normalized score vector of CSI, J is the number of classes in the candidate set. Similarly, we obtain other normalized score vectors $F_{independent}$ and F_{ac}.

(2) Modality Weighting: From noise estimation, we derive normalized ξ_{csi} and ξ_{ac} to estimate the reliability of CSI and acoustic features for each instance. Rapid uses these two confidence values to adaptively perform modality weighting. Weighted score is calculated as $\xi_{csi}\boldsymbol{F_{csi}}$ and $\xi_{ac}\boldsymbol{F_{ac}}$.



(a) laboratory (b) corridor (c) meeting room

Fig. 10. Layouts of three locations and corresponding LoS path and trained path and the stars denote Rapid nodes.

(3) Classifier Combination: Since $\boldsymbol{F_{independent}}$ corresponds to the classifier whose features are confirmed in walking detection, which is reliable regardless of two confidences, we combine the modalities using a classifier combination method Sum rule[11] as follows (Note that ξ_{csi} and ξ_{ac} have been normalized into a value between 0 to 1):

$$\boldsymbol{ES} = \xi_{csi}\boldsymbol{F_{csi}} + \xi_{ac}\boldsymbol{F_{ac}} + \boldsymbol{F_{independent}} \tag{13}$$

where $\boldsymbol{ES}$ represents the eventual score vector for each candidate and we select the candidate corresponding to the maximum value es_{max} as the result of identification.

9 EVALUATION

9.1 Experimental Methodology

To examine the feasibility of Rapid and evaluate its performance, we conduct extensive experiments. We divide the evaluation into eight experiments. The objective of the first two experiments is to quantify limitations of the existing systems, which examine the impact of system noise on performance of radio sensing modality and CSI confidence value, respectively. Then we conduct the third and fourth experiments to examine the feasibility of the Rapid system, as well as to make a quantitative comparison of performance with purely radio- and audio-based identification approaches. To achieve this objective, we compare Rapid with a radio-based (CSI-based) approach [31] and an audio-based identification system [8] since they all use gait information for personal identification. To fully examine the performance in different conditions, the third experiment is conducted in a clean environment, and the fourth is in a noisy (including system noise) environment. At the same time, we also evaluate the performance of different modules in Rapid, i.e., only using acoustic module or CSI module in above two experiments. The fifth experiment examines the performance of walking detection which is the trigger of gait-based person identification systems. In the sixth experiment, we examine the impact of fine-grained step-based features on the performance of Rapid. The seventh experiment solves the question how many steps Rapid system needs to identify a person. It also answers the question how the end of the walk interval is defined in our system. The eighth experiment evaluates the performance of stranger identification.

9.2 Experimental Scenarios and Data Collection

9.2.1 Scenarios: Since Rapid targets for smart home or smart office scenarios, our experiments are carried out in three different locations as follows:

- **Corridor** First, we perform experiments in a corridor of a university building with two Rapid nodes placed next to the wall. The width of corridor is 2 m.

- **Laboratory** Second, we deploy Rapid in a research laboratory that covers an area of $9m \times 6m$. It is surrounded by multiple office facilities such as desks and chairs, and therefore is subject to multipath effects.
- **Meeting room** Third, we carry out experiments in a meeting room with a size of $8m \times 5m$ where a table covers a large area.

The layouts of three locations and trained walking path are displayed in Figure 10. The stars denote Rapid nodes which are deployed on the wall in the corridor and the meeting room, while deployed on the floor in the laboratory. The distance between Tx and Rx is 3 meters, because the experiment in Section 9.9 shows that only 4 $\sim$ 5 steps (around 2 meters to 3 meters) are practically enough to build the classifier as displayed in Figure 17. We set 1 meter as the distance between pre-defined training path and LoS path. Note that our trained path is the most likely path that user may walk along, e.g., the middle path in the corridor is 1 meter away.

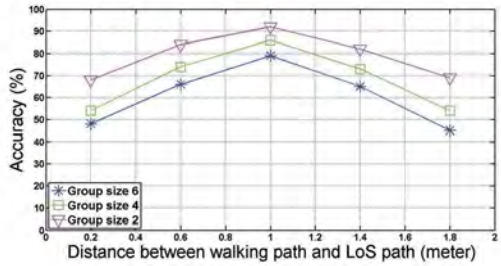
9.2.2 Data Sets: We recruit 20 volunteers who are university students and researchers with a mix of males and females. They have an age ranging from 20 to 33 years (μ = 26.9; σ = 4.7), body height ranging from 161 cm to 185 cm (μ = 173.2 cm; σ = 8.2 cm) and body weight ranging from 53 kg to 76 kg (μ = 68.2 kg; σ = 5.4 kg). For our experiments at each location, we use a camera to record subject's identity for ground truth. Two Rapid nodes constantly communicate with a ping rate of 1K samples/second, while the acoustic signals are recorded in PCM format at a 44-kHz sampling rate, with 16 bits per sample. We average the multi-channel acoustic signal to obtain a mono input for processing. Since Rapid node integrates both radio and audio sensing modalities, we conduct all the experiments (including the replicated works) in them. For the same reason, collected CSI signal and acoustic signal are time aligned. The Rapid node is connected to a laptop using cable for transferring data in real time. We process the raw data and perform person identification on the laptop.

The detailed data collection process is as follows. First, we explain two kinds of noise that may be brought in:

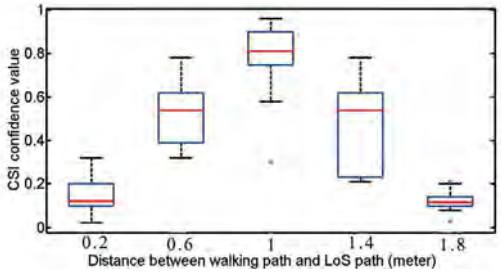
(1) **system noise**: noise that influences CSI modality. When we require subjects to walk in the pre-defined path(1 meter away from LoS path and parallel to it), this kind of noise is non-existent. When subjects are required to walk *freely*, we just give them a direction (e.g., we ask them just to pass through the corridor in the corridor scenario), and this kind of noise may be introduced. However, in any case, subjects are asked to walk in their *natural* way without intentional speed up or slow down, so strange behaviors like turning around on half way and then turning around again are not allowed.
(2) **environment noise**: noise that influences acoustic modality. To introduce a noisy environment for audio data collection, we generate audio noise using a computer speaker in the background such that the noise levels we measure at our Rapid nodes are -18dB and -13dB (We use the Audacity software to perform noise level measurements). The noise level at -18 dB is an average measurement in the acoustic environment where students kept talking to each other in a whisper-like voice. The noise level at -13 dB is an average measurement in the acoustic environment where students talk to each other in a natural volume. We also conduct experiments in quiet environment, where this kind of noise is negligible.

There are two data sets collected in our experiments as follows:

(1) **clean data set**: We collect the first data set in all three locations. In each location, we collect 20 walking instances for each subject *without system noise and environment noise*, thus we obtain $3 * 20 * 20 = 1200$ walking instances. Note that to remove environment noise, the whole experiment is kept quiet except footstep sound. To remove system noise, the subjects are required strictly in the pre-defined path(1 m away from LoS and parallel to it).
(2) **noisy data set**: We also collect the second data set in all three locations. However, this time system noise and environment noise will be brought in. In laboratory, environment noise is set as -18 dB. While in



(a) Impact of system noise on radio-based identification.



(b) Impact of system noise on CSI confidence value. The middle path is the trained path.

Fig. 11. Impact of system noise on radio-based module and the corresponding change of CSI confidence value. The middle path (1 m from LoS path) is the trained path.

corridor and meeting room, it is set as -13 dB. And subjects are required to walk freely. We collect 10 walking instances for each subject, thus we obtain $3 * 10 * 20 = 600$ walking instances.

In total, we collect $1200 + 600 = 1800$ walking instances in the form of audio and radio.

9.3 Impact of System Noise on Radio-based Identification

9.3.1 Objective: In this experiment, we examine the impact of the system noise on the radio-based identification performance.

9.3.2 Experimental Setup: This experiment has been conducted in all the three locations. We only use radio sensing modality to identify a subject. The training data comes from *clean data set*, i.e., each subject is required to walk along a predefined path. The testing data is obtained as follows. Each subject is required to walk along a path which is 0.2 m, 0.6 m, 1 m (trained path), 1.4 m, 1.8m away from LoS path, respectively. We examine the identification performance with different group sizes of 2, 4, 6.

9.3.3 Results and Implication: Because of the similar result in different locations, Figure 11a illustrates the overall accuracy of three locations with different walking paths and different group sizes. As shown, the identification performance decreases when walking paths are different from pre-trained path. And when the walking path is further away from the pre-trained path, the performance decreases more, e.g., drops from an average of 79% (trained path) to 62% (0.4 m away) and 45% (0.8 m away) of group size 6. The result indicates system noise has a large impact on the performance of radio-based identification.

9.4 Impact of System Noise on CSI Confidence

9.4.1 Objective: In this experiment, we examine the impact of system noise on the CSI confidence value.

9.4.2 Experimental Setup: This experiment has the similar setup as Section 9.3. The training data come from *clean data set*, i.e., each subject is required to walk along a predefined path. The testing data are obtained as follows. Each subject is required to walk along a path which is 0.2 m, 0.6 m, 1 m (trained path), 1.4 m, 1.8m away from LoS path, respectively. We examine the relationship between CSI confidence values and walking paths.

9.4.3 Results and Implication: Figure 11b illustrates the distribution of CSI confidence values with different walking paths. As shown, the confidence decreases when walking paths are different from pre-trained path in

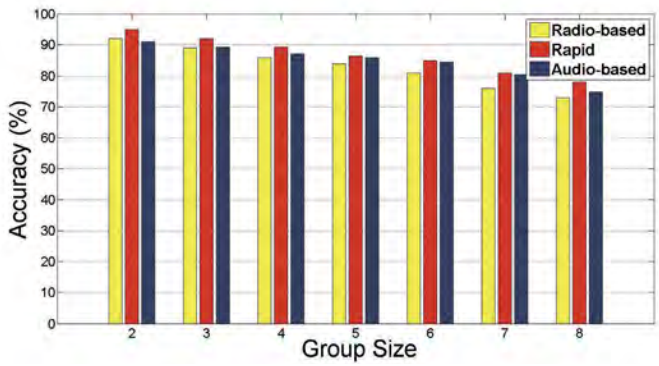


Fig. 12. Overall identification accuracy comparison among three systems(modules) in a clean environment with different group sizes.

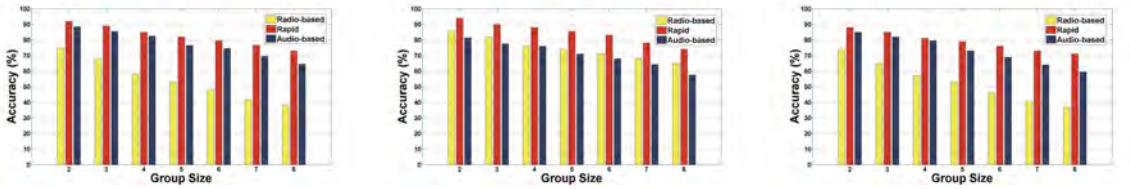
general, although rarely walking in a different path corresponds to a high confidence (this may be induced by heterogeneity of physical characteristic such as height). And there is a tendency that walking in path further away from trained path generates a smaller confidence value. This experiment together with the fourth experiment verify our previous assumption: CSI features are more unreliable if the walking path is further away from pre-trained path.

9.5 Performance Comparison between Monomodal- and Multimodal-based Identification in Clean Environments

9.5.1 Objective: In this experiment, we evaluate the identification performance of Rapid with different group sizes and compare it with purely radio-based and purely audio-based identification systems in a clean environment.

9.5.2 Experimental Setup: This experiment has been conducted in all the three locations. The training data set and testing data set all come from *clean data set*. We implement a radio-based (CSI-based) approach [31] and an audio-based identification system [8] for comparison. We also using the data to evaluate the performance of radio module and audio module in Rapid separately. Since our target scenario is smart home or smart office where we aim to identify a person from a small group, we randomly select from 2 to 8 subjects from the group and use 10-fold cross-validation separately. For each group size, we repeat this process 10 times and calculate its overall accuracy.

9.5.3 Results and Implication: As the similar results in different locations, we only display the overall accuracy of three locations. Our experiment shows purely radio-based and purely audio-based identification system achieve the similar performance compared with radio module and audio module in our system respectively. This is expected because we use similar feature set, therefore, we only display one in the result. Figure 12 illustrates the overall accuracy of Rapid and two monomodal systems (modules) in a clean environment with different group sizes. We observe that the accuracy decreases for all the systems as the group size increases. This is expected because identifying more subjects implies a larger possibility that there are similar gait features. The overall accuracy of Rapid is as high as 96% in a group size of 2, and 80% in a group size of 8, a little higher than two monomodal systems (modules). This result demonstrates that monomodal systems achieve good results without the interference of system noise or environment noise. However, Rapid achieves more competitive identification performance because of feature fusion.



(a) Laboratory(-18 dB noise), less likely to walk in the trained path (b) Corridor(-13 dB noise), more likely to walk the trained path (c) Meeting room(-13 dB noise), less likely to walk in the trained path

Fig. 13. Identification accuracy comparison among three systems(modules) in three noisy locations with different group sizes.

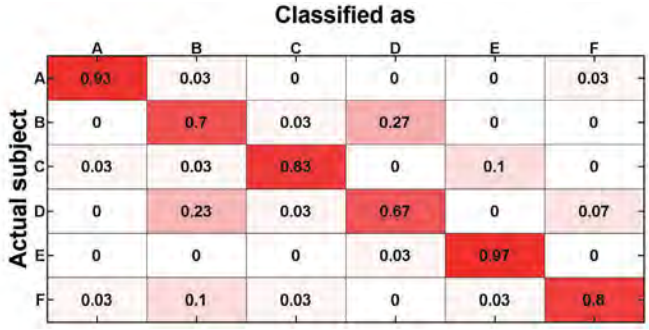


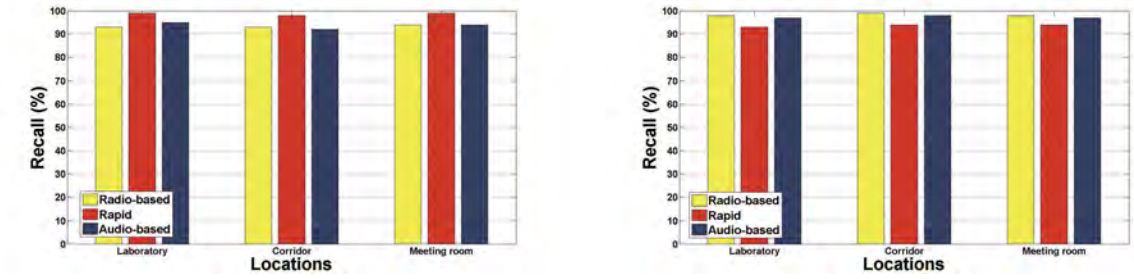
Fig. 14. The confusion matrix of person identification with 6 people in three locations

9.6 Performance Comparison between Monomodal- and Multimodal-based Identification in Noisy Environments

9.6.1 Objective: In this experiment, we evaluate the identification performance of Rapid with different group sizes and compare it with a purely radio-based and a purely audio-based identification systems in a noisy environment.

9.6.2 Experimental Setup: This experiment has been conducted at all the three locations. The training data set comes from *clean data set*. However, the testing data set comes from *noisy data set*. The noise level in each scenario is not the same. We generate noise of -18 dB in the laboratory, while -13 dB in the corridor and meeting room. We implement the purely radio-based approach and the purely audio-based identification system for comparison. Besides, we evaluate the performance of radio module and audio module of Rapid separately. We randomly select from 2 to 8 subjects from the group and for each group size, train a classifier respectively from training data. We repeat this process 10 times and calculate its overall accuracy.

9.6.3 Results and Implication: Since the experiment shows purely radio-based and purely audio-based identification system achieve the similar performance compared with CSI module and acoustic module of Rapid system respectively, we combine the similar results for simplicity. The performance of three methods and two modules in Rapid in three locations is displayed in Figure 13. We observe that the average accuracy of person



(a) Precision of walk detection

(b) Recall of walk detection

Fig. 15. Precision and Recall of Rapid and other two monomodal systems.

identification of Rapid is as high as 92% in a group size of 2 and 72% in a group size of 8. Rapid achieves nearly 78% of accuracy for a group size of 6 or lower for all the three locations. We observe that the accuracy of radio-based system(module) decreases sharply when a subject walks along a different path. For example, the accuracy in meeting room decreases from 79% to 47% in a group size of 6. However, the accuracy is relatively high in corridor as shown in Figure 13b. This is because the corridor in our scenario is narrow as illustrated in Figure 10b, thus a subject is more likely to walk along the pre-trained path. The performance of audio-based system drops from an average of 85% (clean) to 69% (-13 dB) and 76% (-18dB) in a group size of 6, respectively. We note the audio-based system (module) performed more badly in corridor compared with meeting room although their noise level is both at -13 dB. We think it may because there are other noise far away except man-made noise (meeting room is closed thus has relatively high sound insulation to resist far-away noise). However, with noise estimation, we fuse the classifiers adaptively and thus obtain a higher accuracy compared with monomodal identification. Figure 14 shows the confusion matrix for the case of 6 subjects at three locations. Note that the result in this figure could be seen as a subset of those reported in Figure 13.

9.7 Performance of Walking Detection

9.7.1 Objective: In this experiment, we measure the walking detection performance of Rapid and compare it with purely radio-based and purely audio-based methods in a noisy environment.

9.7.2 Experimental Setup: The experimental setup is similar to the experiment that evaluates the identification performance, since walking detection is the beginning of the whole identification process. The experiment is conducted at all the three locations with both system noise and environment noise. We obtain the data set from the original *noisy data set*. We use precision and recall as metrics to evaluate the performance of walking detection.

9.7.3 Results and Implication: Figure 15 illustrates the precision and recall of three systems at the three locations. As shown, all three systems achieve a relatively satisfying result (all the metrics are higher than 90%). However, due to double-checked detection method used, Rapid achieves a higher precision and a lower recall compared with two monomodal identification systems. There is a tradeoff that monomodal system is more sensitive but more prone to be affected by noise. In contrast, to get fused features such as CSI steps segmented by acoustic signal, Rapid verifies walking only when radio modality and audio modality both detect it.

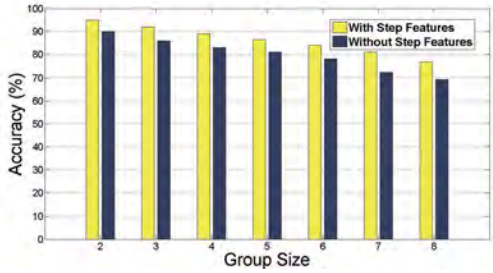


Fig. 16. Impact of fine-grained step-based features on performance of Rapid.

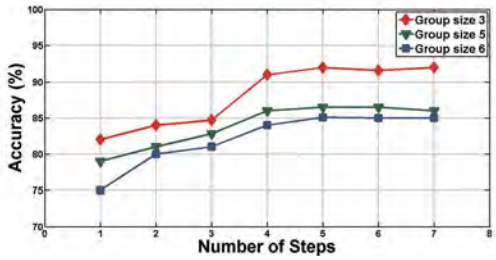


Fig. 17. Identification accuracy comparison for different group sizes with different numbers of steps.

9.8 Impact of Step-based Features

9.8.1 Objective: Our step-based features fuse both radio-based information and audio-based information. Compared with features extracted from the entire walking process, step-based features are more fine-grained. Therefore, in this experiment, we focus on examining how step-based features influence the identification performance of Rapid.

9.8.2 Experimental Setup: Since we focus on the influence of step-based features, we conduct this experiment without system noise and environment noise in three locations. The training and testing data all come from *clean data set* We use 10-fold cross-validation to measure the identification performance with step-based features and without step-based features. We repeat the above process for 10 times. For each group size from 2 to 8, we calculate its overall accuracy.

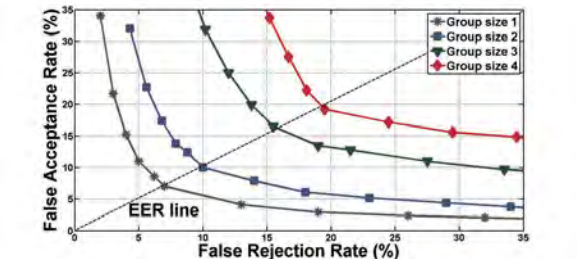
9.8.3 Results and Implication: Figure 16 illustrates the overall accuracy with step-based features and without step-based features in different group sizes. As shown, the overall accuracy increases (e.g., from 78% to 84% in a group size of 6) when step-based features are introduced. This can be attributed to the fact that acoustic signal *accurately* segment the CSI signal to get more fine-grained features in each step, i.e., the shape of each step. This result indicates our system leverages more fine-grained features with the fusion of radio-based information and audio-based information, leading to better identification performance.

9.9 Robustness with the Numbers of Steps

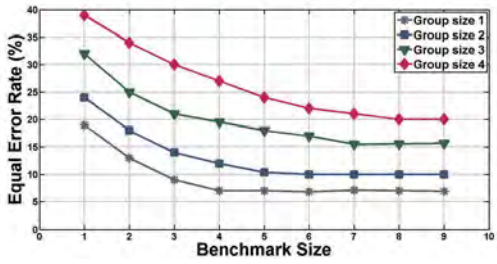
9.9.1 Objective: Another important issue for the application of Rapid is how many steps the system needs to identify a person. We conduct an experiment here to investigate on the relation between accuracy of identification and the numbers of required step segments. This experiment also answers the question how the end of the walk interval is defined in our system.

9.9.2 Experimental Setup: Since we focus on the influence of numbers of steps, We conduct this experiment without system noise and environment noise in three locations. We use the training and testing data from *clean data set*. We use 10-fold cross-validation for 5 times to measure the identification performance with group sizes of 3, 5 and 6, respectively. For each group size, we construct features set from different numbers of steps and calculate their overall accuracy in three locations.

9.9.3 Results and Implication: Figure 17 plots the identification accuracy with different numbers of steps and different group sizes. As shown, the accuracy either increases or remains similar with increase in the number of



(a) Tradeoff between FAR and FRR with different group sizes



(b) EER with different sizes of benchmark sets under different group sizes

Fig. 18. Performance of stranger identification

steps. The figure shows that Rapid can identify person with high accuracy even with data trace containing only 4 ~ 5 steps(approximately 2.5 meters), increasing its applicability in space-constrained indoor environments.

9.10 Performance of Stranger Identification

9.10.1 Objective: The objective of stranger identification is to detect whether or not a subject is within the known group. It has many applications in security areas such as intruder detection.

9.10.2 Experimental Setup: This function can be triggered in Rapid system before person identification. However, in our system, this task is not easy because the influence of two kinds of noise, which will distort the features and make it difficult to judge if a "stranger" is a true stranger or just noise distort the features of a subject in database. Therefore, we leverage two confidence values to measure the current noise level and the process of stranger identification will start only if the level of system noise and environment noise are both low, otherwise our system will output "unknown". To avoid the "unknown" output, we leverage the dataset in all three locations without the interference of two kinds of noise (from *clean data set*) for both training and testing, i.e., the dataset is collected in a quiet environment and the subjects walk in the same path (1 meter away and parallel to the LoS path).

In the training process, we use the gait instances from groups with sizes from 1 to 4 as known group, respectively, and use gait instances from other 8 subjects as benchmark for negative class. Note the benchmark gait instances are employed to determine the decision boundary for the known group. The remaining of subjects are regarded as strangers which do not occur in the training process. We build probabilistic classifiers for two classes, which can calculate the fitness probability that an unknown gait instance belongs to the known group. We regard gait instances with fitness probability higher than a given threshold as instances that belong to the known group. The classifiers can also identify a stranger which does not participate in the training process since the gait instance of him/her will also have low fitness probability.

We evaluate the identification accuracy in terms of False Acceptance Rate (FAR) and False Rejection Rate (FRR). The FAR is defined as the rate that a stranger is wrongly classified as the subject in the known group and the FRR is the rate that the known subject is wrongly classified as a stranger. Since we can tradeoff between the FAR and FRR by changing the fitness probability threshold for identification, we define the Equal Error Rate (EER) point as the point that FAR and FRR are equal. We calculate FAR and FRR of identification by using all the available training data of selected subjects (i.e., 20 for each subject) and using 10-fold cross validation. After getting the

FAR and FRR for one group, we repeat the whole process with different subjects as known group and benchmark set. The final results are averaged over 20 randomly selected known groups and benchmark sets.

9.10.3 Results and Implication: Figure 18a plots FAR vs FRR under different sizes of known group. This figure shows that Rapid achieves an average EER of 19.5% when the size of known group is set to 4. When the group size decreases, the EER also decreases, e.g., EER is 16% when the size is 3 and 10% when the size is 2. We believe Rapid is effective for scenarios such as smart home (typical size of members is 3). Figure 18b plots EER with different benchmark sizes under different group sizes. This figure shows that using the benchmark set size of 8 is effective. We observe that using more benchmark subjects gives lower EER, but the result for having 8 or 9 benchmark subjects are almost the same for all group sizes. Therefore, we choose to use 8 benchmark subjects in Rapid.

10 DISCUSSION AND LIMITATIONS

Rapid is a multimodal system establishing the feasibility of using combined acoustic radio (CSI of WiFi) and acoustic information to identify persons through their gait patterns. In order to achieve high accuracy in real-life scenarios with system noise and environment noise, Rapid includes noise estimation to quantify the impact of noise to both CSI and acoustic measurements. Then based on an gait analysis, together with confidence estimations, Rapid adaptively fuses the CSI and acoustic measurements. However, since Rapid uses features from CSI and acoustic measurements, some factors that distort these features will degrade the performance of Rapid. *First*, the change of person's footwear (E.g, wearing a flat-shoe vs. a high-heal) will have a nonnegligible impact on acoustic module, the accuracy of identification will deteriorate more or less. This is because the change of footwear will bring in different footstep sound and then acoustic-based features such as MFCCs will be distorted. The degree to which the performance of our system is impacted depend upon the comparison between CSI confidence value and acoustic confidence value. If CSI confidence value is much larger, i.e., CSI based is in the dominate positon during the identification process, the influence of footwear change will be small. *Second*, when there are multiple users walking nearby at the same time, the gait patterns captured by Rapid are complex mixtures of multiple activities of the users. For CSI measurement, other users will introduce multiple interference paths. For acoustic measurement, other users will introduce other footstep sound. It is difficult to separate the signals from other walking users. Since our Rapid nodes are light-weight and easy to deploy, we plan to explore the potential in using multiple nodes and building multiple wireless links to separate signals from multiple users.

11 CONCLUSION

In this paper, we present the design, implementation and evaluation of Rapid, a system that can perform robust person identification in a device-free, effortless and low-cost manner, using radio (CSI of WiFi) and acoustic information. In order to achieve high accuracy in real-life scenarios with system noise and environment noise, Rapid includes noise estimation to quantify the impact of noise to both CSI and acoustic measurements. Then based on an accurate gait analysis, together with confidence estimations, Rapid adaptively fuses the CSI and acoustic measurements, achieving robust person identification. We evaluate Rapid using experiments at multiple locations with a total of 20 volunteers and 1800 gait instances, and our results show that Rapid identifies a subject with an average accuracy of 92% to 82% from a group of 2 to 6 people, respectively.

Currently, we are expanding Rapid with multiple (more than two) Rapid nodes to create multiple Tx/Rx links. This will enable a large-scale person identification system to improve the performance further. We also plan to investigate the potential of identifying a subject when there are other subjects nearby.

A DERIVATION OF CPV

We now consider the CFR power variance during a short time period of $[0, \tau]$. From equation (3), CFR power is expressed as:

$$|H(f,t)|^2 = 2|H_s(f)a(f,t)| \cos\left(\frac{2\pi vt}{\lambda} + \frac{2\pi d(0)}{\lambda} + \phi\right) + |a(f,t)|^2 + |H_s(f)|^2. \tag{14}$$

Note $\frac{2\pi d(0)}{\lambda} + \phi$ are constant values representing initial phase offsets, $|H_s(f)|^2$ are constant static CFR values, since $|a(f,t)|^2$ is also constant when the length of path reflected by body remains almost the same during a short time period. Therefore, CFR power variance can be expressed as:

$$Var(|H(f,t)|^2) = Var\left(2|H_s(f)a(f,t)| \cos\left(\frac{2\pi vt}{\lambda}\right)\right). \tag{15}$$

We know variance can be calculated by "mean of square minus square of mean", i.e., for a variable X,

$$Var(X) = E\left[X^2\right] - (E[X])^2. \tag{16}$$

Therefore, we first calculate the part of "mean of square", given the time period of $[0, \tau]$,

$$E\left[X^2\right] = \frac{\int_0^\tau \left(2|H_s(f)a(f,t)|cos\left(\frac{2\pi vt}{\lambda}\right)\right)^2 dt}{\tau}, \tag{17}$$

where X represents $2|H_s(f)a(f,t)| \cos\left(\frac{2\pi vt}{\lambda}\right)$. Similarly, "square of mean" can be calculated as:

$$(E[X])^2 = \left(\frac{2|H_s(f)a(f,t)| \int_0^\tau cos\left(\frac{2\pi vt}{\lambda}\right) dt}{\tau}\right)^2 \tag{18}$$

Subtracting equation (18) from equation (17), we obtain CPV as follows:

$$Var(|H(f,t)|^2) = 2|H_s(f)a(f,t)|^2 \left(\frac{sin(4\pi v\tau/\lambda)}{2\pi v\tau/\lambda} - \frac{2sin^2(2\pi v\tau/\lambda)}{(2\pi v\tau/\lambda)^2} + 1\right). \tag{19}$$

Denote φ as $2\pi v\tau/\lambda$, and cardinal sine function $sinc(\varphi)$ equals to $\sin(\varphi)/\varphi$, we finally obtain:

$$Var(|H(f,t)|^2) = 2|H_s(f)a(f,t)|^2 \left(sinc(2\varphi) - 2sinc^2(\varphi) + 1\right), \tag{20}$$

which is the equation (4) presented in Section 3.3.

REFERENCES

[1] M Umair Bin Altaf, Taras Butko, and Biing-Hwang Fred Juang. 2015. Acoustic Gaits: Gait Analysis With Footstep Sounds. *IEEE Transactions on Biomedical Engineering* 62, 8 (2015), 2001–2011.

[2] Michael Berouti, Richard Schwartz, and John Makhoul. 1979. Enhancement of speech corrupted by acoustic noise. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'79.*, Vol. 4. IEEE, 208–211.

[3] Steven Boll. 1979. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on acoustics, speech, and signal processing* 27, 2 (1979), 113–120.

[4] Roberto Brunelli and Daniele Falavigna. 1995. Person identification using multiple cues. *IEEE transactions on pattern analysis and machine intelligence* 17, 10 (1995), 955–966.

[5] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2, 3 (2011), 27.

[6] Laurie Davies and Ursula Gather. 1993. The identification of multiple outliers. *J. Amer. Statist. Assoc.* 88, 423 (1993), 782–792.

[7] Benoît Duc, Elizabeth Saers Bigün, Josef Bigün, Gilbert Maître, and Stefan Fischer. 1997. Fusion of audio and video information for multi modal person authentication. *Pattern Recognition Letters* 18, 9 (1997), 835–843.

[8] Jürgen T Geiger, Martin Hofmann, Björn Schuller, and Gerhard Rigoll. 2013. Gait-based person identification by spectral, cepstral and energy-related audio features. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 458–462.

[9] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. 2011. Tool release: gathering 802.11 n traces with channel state information. *ACM SIGCOMM Computer Communication Review* 41, 1 (2011), 53–53.
[10] Marc Karam, Hasan F Khazaal, Heshmat Aglan, and Cliston Cole. 2014. Noise removal in speech processing using spectral subtraction. *Journal of Signal and Information Processing* 2014 (2014).
[11] Josef Kittler, Mohamad Hatef, Robert PW Duin, and Jiri Matas. 1998. On combining classifiers. *IEEE transactions on pattern analysis and machine intelligence* 20, 3 (1998), 226–239.
[12] Hong Li, Wei Yang, Jianxin Wang, Yang Xu, and Liusheng Huang. 2016. WiFinger: talk to your smart devices with finger-grained gesture. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 250–261.
[13] Xuefeng Liu, Jiannong Cao, Shaojie Tang, and Jiaqi Wen. 2014. Wi-Sleep: Contactless sleep monitoring via WiFi signals. In *Real-Time Systems Symposium (RTSS), 2014 IEEE*. IEEE, 346–355.
[14] Jordi Luque, Ramon Morros, Ainara Garde, Jan Anguita, Mireia Farrus, Dušan Macho, Ferran Marqués, Claudi Martínez, Verónica Vilaplana, and Javier Hernando. 2006. Audio, video and multimodal person identification in a smart room. In *International Evaluation Workshop on Classification of Events, Activities and Relationships*. Springer, 258–269.
[15] Jouni Malinen and others. 2014. *hostapd: Ieee 802.11 ap, ieee 802.1 x.* Technical Report. WPA/WPA2/EAP/RADIUS Authenticator. online: http://hostap. epitest. fi/hostapd.
[16] Thanh Trung Ngo, Yasushi Makihara, Hajime Nagahara, Yasuhiro Mukaigawa, and Yasushi Yagi. 2014. The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication. *Pattern Recognition* 47, 1 (2014), 228–237.
[17] Robert J Orr and Gregory D Abowd. 2000. The smart floor: A mechanism for natural user identification and tracking. In *CHI'00 extended abstracts on Human factors in computing systems*. ACM, 275–276.
[18] S Palanivel and B Yegnanarayana. 2008. Multimodal person authentication using speech, face and visual speech. *Computer Vision and Image Understanding* 109, 1 (2008), 44–55.
[19] Yasuhiro Shoji, Takashi Takasuka, and Hiroshi Yasukawa. 2004. Personal identification using footstep detection. In *Intelligent Signal Processing and Communication Systems, 2004. ISPACS 2004. Proceedings of 2004 International Symposium on*. IEEE, 43–47.
[20] SolidRun. 2014. HummingBoard Pro. (2014). http://wiki.solid-run.com/doku.php?id=products:imx6:hummingboard
[21] Li Sun, Souvik Sen, Dimitrios Koutsonikolas, and Kyu-Han Kim. 2015. WiDraw: Enabling Hands-free Drawing in the Air on Commodity WiFi Devices. In *Proceedings of ACM Annual International Conference on Mobile Computing and Networking (MobiCom)*. 77–89.
[22] David Tse and Pramod Viswanath. 2005. *Fundamentals of wireless communication*. Cambridge university press.
[23] Navneet Upadhyay and Abhijit Karmakar. 2013. Spectral Subtractive-Type Algorithms for Enhancement of Noisy Speech: An Integrative Review. *International Journal of Image, Graphics and Signal Processing* 5, 11 (2013), 13.
[24] Martin Vondrasek and Petr Pollak. 2005. Methods for speech SNR estimation: Evaluation tool and analysis of VAD dependency. *Radioengineering* (2005).
[25] Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. 2003. Silhouette analysis-based gait recognition for human identification. *IEEE transactions on pattern analysis and machine intelligence* 25, 12 (2003), 1505–1518.
[26] Wei Wang, Alex X Liu, and Muhammad Shahzad. 2016. Gait recognition using wifi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 363–373.
[27] Wei Wang, Alex X. Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2015. Understanding and Modeling of WiFi Signal Based Human Activity Recognition. In *Proceedings of ACM Annual International Conference on Mobile Computing and Networking (MobiCom)*. 65–76.
[28] Chenshu Wu, Zheng Yang, Zimu Zhou, Kun Qian, Yunhao Liu, and Mingyan Liu. 2015. PhaseU: Real-time LOS identification with WiFi. In *2015 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2038–2046.
[29] Jiang Xiao, Kaishun Wu, Youwen Yi, Lu Wang, and L. M. Ni. 2013. Pilot: Passive Device-Free Indoor Localization Using Channel State Information. In *Proceedings of IEEE International Conference on Distributed Computing Systems (ICDCS)*. 236–245.
[30] Zheng Yang, Zimu Zhou, and Yunhao Liu. 2013. From RSSI to CSI: Indoor Localization via Channel Response. *Comput. Surveys* 46, 2, Article 25 (2013), 32 pages.
[31] Yunze Zeng, Parth H. Pathak, and Prasant Mohapatra. 2016. WiWho: WiFi-based Person Identification in Smart Spaces. In *Proceedings of International Conference on Information Processing in Sensor Networks (IPSN)*.
[32] Dian Zhang, Jian Ma, Quanbin Chen, and Lionel M Ni. 2007. An RF-based system for tracking transceiver-free objects. In *Fifth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom'07)*. IEEE, 135–144.
[33] J. Zhang, B. Wei, W. Hu, and S. S. Kanhere. 2016. WiFi-ID: Human Identification using WiFi signal. In *Proceedings of International Conference on Distributed Computing in Sensor Systems (DCOSS)*.

Received February 2017; revised May 2017; accepted July 2017.