Master's Thesis

A Comparative Assessment on the Practicality of the Signal-based Human Motion Recognition Methods

Hayoung Jeong (정 하 영)

Department of Computer Science and Engineering

Pohang University of Science and Technology

2018


POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY
1986
포항공과대학교

신호 기반 모션 인식 기법의 실용성 비교 분석 및 평가

A Comparative Assessment on the Practicality of the Signal-based Human Motion Recognition Methods

A Comparative Assessment on the Practicality of the Signal-based Human Motion Recognition Methods

by

Hayoung Jeong
Department of Computer Science and Engineering
Pohang University of Science and Technology

A thesis submitted to the faculty of the Pohang University of Science and Technology in partial fulfillment of the requirements for the degree of Master of Science in the Computer Science and Engineering

Pohang, Korea
12. 26. 2017
Approved by
Jong Kim
Academic advisor

A Comparative Assessment on the Practicality of the Signal-based Human Motion Recognition Methods

Hayoung Jeong

The undersigned have examined this thesis and hereby certify that it is worthy of acceptance for a master's degree from POSTECH

12. 26. 2017

Committee Chair Jong Kim

Member Chanik Park

Member Hanjun Kim

MCSE 정 하 영. Hayoung Jeong
20162498
A Comparative Assessment on the Practicality of the Signal-based Human Motion Recognition Methods,
신호 기반 모션 인식 기법의 실용성 비교 분석 및 평가
Department of Computer Science and Engineering , 2018, 78p, Advisor : Jong Kim. Text in English.

ABSTRACT

Signal-based motion recognition schemes have received growing attention in the academia, but they failed to be adopted by the industry despite their flourish. In this thesis, we address the impractical requirements and conditions of signal-based human motion recognition methods. We propose seven practical facets – *granularity, stability, robustness, applicability, deployability, and efficiency* – that extensively evaluate the practicality of a scheme. We perform evaluation on the previous signal-based human motion recognition research based on the seven facets and unfold the findings we have earned through the analysis in addition to the emerging trends in the field and prospect on future research directions. We believe our work may serve as a standard for future signal-based human motion recognition research to assess the practicality of their schemes.


POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY
1986
포항공과대학교

Contents

List of Tables

List of Figures

I. Introduction

Motion recognition technology is becoming increasingly popular in pervasive computing. The ever-increasing number of smart devices and their technical advancements to sense various wireless signals have increased the density of accessible wireless signals in smart environments. These trends have provided opportunities for researchers to utilize wireless signals to develop techniques that are useful in smart environments. In particular, motion recognition approaches that leverage the readily available Wi-Fi and acoustic signals have emerged as a prominent research area in academia over the past few years.

There are several streams of research that aim to recognize human motions besides signal based methods. Some of the most widely used approaches are vision-based approach and motion sensor-based approach. Vision-based approach processes images collected from cameras to detect or recognize motion [1]. Motion sensor-based approach leverages the motion data (*e.g.,* accelerometer, gyrsocope, magnetometer) from the sensors that are attached to the target that takes motion [2]. While these streams of research have their own values and significance, signal-based approach has attracted the research society and received growing attention in recent years because of the following advantages they have over other approaches:

- Signal-based approach is insensitive to lighting conditions because it does not rely on vision inputs. Cameras, on the other hand, are sensitive to lighting conditions, so poor lighting condition in vision-based approach often leads to inaccurate results.

- The characteristic of signals to penetrate through objects enables signal-

based approach to work in non-line-of-sight. For instance, signal-based approach can recognize motions in through-the-wall scenarios or when a signal emitting device (*e.g.,* smartphone) is occluded (*e.g.,* inside the pocket). However, vision-based approach cannot capture motions when the target is in non-line-of-sight.

- Signal-based approach has less privacy risk than vision-based approach. Vision-based approach may lead to severe privacy breach when collected data are leaked. As a result, users may be reluctant to deploy cameras and feel uncomfortable with being continuously observed and monitored. However, the leak of signal data is not as fatal as leaked images, so signal-based approach has less potential in invading users' privacy.

- Signal-based approach is less obtrusive than motion sensor-based approach. The latter approach requires the data collecting device to be attached to the target that takes motion, but signal-based approach works even when the device is not in contact to the target. Therefore, we are convinced that the signal-based approach is more usable than the motion sensor-based approach.

For these reasons, signal-based approach has risen as a powerful alternative for motion recognition system and have facilitated active research in the academic community. Several works have proposed methods to recognize user's daily activities that occurs physically near the transmission channel [3–5]. Keystroke recognition has also been a popular topic in mainstream conferences [6–10]. Besides activity or keystroke recognition, vast research has leveraged signals to perform various applications including intruder detection [11], health monitoring [12, 13], authentication [14–16], lips reading [17], and fall detection [18].

Existing Surveys. By such flourish in the research community, recent

survey papers have investigated various signal-based activity recognition techniques [2,19,20]. These past surveys have categorized and examined the technical methods used in different approaches. However, no work have critically assessed the practicality of the proposed methods despite the fact that most of the considerable body of work that has been proposed have been unsuccessful in getting adopted by the industry.

The foremost reason why the increasing attention in the academic community did not lead to real-world adoption is because many signal-based motion recognition methods required infeasible requirements or impractical conditions. For instance, WiKey used 80 training samples for each of the 37 keys to obtain the classification accuracy of 97.5% [9]. The classification accuracy dropped to 82.87% when the training samples were reduced to 30. What's worse, whenever there was a significant change in the environment (*e.g.,* user's position changes), the system required users to provide 30 training samples for each key again. Users' participation in the training process, requirement of at least 30 training data for each of the keys, and the frequent retraining load were absurdly cumbersome to be adopted in real practical scenarios. Although the novelty of the proposed methods have significance and academic value, such impractical requirements have restricted wide deployment.

In order to go towards devising easily adoptable systems, it is imperative to set up a standard that measures the practicality of different methods. There is a clear tradeoff between accuracy and practicality. For instance, a scheme may achieve high accuracy by requiring exorbitantly many number of training samples; conversely, a scheme may require reasonable number of training samples and achieve lower accuracy. As such, there are many parameters that determine the accuracy of the method such as the number of training samples, distance from the transceivers to the target, sampling rate of the signal, and interferering

factors. In other words, accuracy is only a partial standard for comparing different works. With different parameters values used in different works, it is difficult to compare and to to fully assess the advancement of different methods. Therefore, a uniform standard that systematically integrates these parameters is necessary.

Our Approach. We propose a systematic evaluation standard that measures the practicality of different methods. Using this standard, we perform an in-depth comparative study that extensively evaluates prior state of the arts to point out the limitation and impracticality of the method. The focus of our work is not to enumerate and classify existing techniques, but to critically examine different schemes on different aspects exposing the barriers perceived by the industry that restrains real world adoption. We present remaining challenges and unresolved limitations to inspire future research the right way forward.

Our main contributions can be summarized as follows:

- We suggest a practicality measuring standard of signal-based motion recognition schemes.
- We address the impracticality of signal-based motion recognition schemes through performing a profound comparative assessment on the plethora of proposed solutions.
- We provide a comprehensive overview on the existing approaches and the status of latest research trends in signal-based motion recognition methods.
- We analyze the trends of current signal-based motion recognition research.
- We bring up remaining challenges and inspire future research directions towards constructing practical designs.

The key findings we have discovered through the assessment are as follows:

- State of the art signal-based motion recognitions have shown to be impractical. The average practicality score among the 25 targeted schemes was 38.9 over 70.

- Signal source selection's influence on recognizable motion types and robustness of a scheme.

- Richness of frequency features in a motion type determines the stability of machine learning based motion recognition schemes.

- Practicality parameters have different implications depending on the intent of the system of whether it is an attack or a service.

Roadmap. The remainder of this thesis is organized as follows. We present the basic background on signals in Chapter II. We provide an overview of the targeted signal-based motion recognition methods in Chapter III. We propose the seven evaluation facets in Chapter IV and perform an extensive assessment on the target schemes with the facets in Chapter V. In Chapter VI, we explain the findings we have earned through the investigation, followed by related works in Chapter VII. The limitation of our work and future research directions of signal-based motion recognition methods are discussed in Chapter VIII. We draw our concluding remarks in Chapter IX.

II. Background

In this chapter, we provide an overview of the background knowledge on wireless signals neccessary to understand the targeted methods in the survey.

2.1 Wireless Signals

2.1.1 Basics

A wireless signal is a wave that propagates through a distance. A wave can be characterized by its frequency, amplitude, and phase. The velocity of the signal is determined depending on the type of the signal (*i.e.,* mechanical waves and electromagnetic waves) and the medium that it traverses through. Acoustic signals travels at a speed of 343 m/s in air, and electromagnetic signal (*e.g.,* Wi-Fi signal) travels at a speed of light which is 299,792,458 m/s. For this reason, the type of the signal directly affects the interaction surface, interference boundary, and granularity of the signal-based motion recognition methods.

2.1.2 Multipaths

Signals travel from a transmitter to a receiver for a distance. When a signal encounters an object during its propagation, the signal experiences reflection, scattering, diffraction, or refraction, which causes the signal to propagate in multiple paths. Each multipath component takes different paths and are attenuated or delayed differently. In fact, the arriving signal that is received at the receiver is the superposition of the multipath components, and as a result, experiences destructive or constructive effect. In absense of movements, the multipaths are static - there are no changes in the paths of the multipaths. However, when an reflector object moves, the propagation paths changes which creates relevant

fluctuations in the received signal.

2.1.3 Channel State Information

Many research that utilizes signals for motion inference are built upon this principle; the changes in the received signal reflect the movement of the object. By assuming linearity and time invariance of the signal, it is possible to characterize each individual path by adopting Linear Time-Invariant (LTI) system.

Channel Impulse Response. Based on the LTI system model, a received signal is a transmitted signal convoluted by a temporal linear filter, the Channel Impulse Response (CIR):

$$y(t) = s(t) \circledast h(t), \tag{2.1}$$

where $s(t)$ is the transmitted signal, $h(t)$ is the CIR, and $y(t)$ is the received signal.

When there are a total of L paths, each individual path experiences different attenuation and delay. When we denote the amplitude, phase, and propagation delay of the i^{th} signal as a_i, θ_i, and τ_i respectively, the channel impulse response $h(t)$ is:

$$h(t) = \sum_{i=1}^{L} a_i e^{-j\theta_i} \delta(\tau - \tau_i), \tag{2.2}$$

which is the superposition of each individual paths. Each impulse represents each multipath component.

Channel Frequency Response. In the frequency domain, the spectrum experiences frequency-selective fading. Thus, the received spectrum is a trans-

mitted spectrum multiplied by the channel frequency response:

$$Y(f) = S(f) \times H(f), \tag{2.3}$$

where $S(f)$ is the transmitted spectrum, $H(f)$ is the CFR, and $Y(f)$ is the received spectrum. A CFR holds the amplitude-frequency and phase-frequency response of the channel.

$$H = |H| \, e^{jsin\{\angle H\}} \tag{2.4}$$

The amplitude response and phase response indicates the difference of amplitude and phase between the transmitter and the receiver.

2.1.4 Time Difference of Arrival

Time Difference of Arrival (TDoA) is the difference in two signal's Time of Arrival (ToA). TDoA is widely used for locating the target object that emits signal. When the target emits the signal to multiple receivers, the Time of Flight (ToF) of the signals are equal but the Time of Arrival (AoA) differs. The time difference can be translated into distance by multiplying the difference by the velocity of the signal.

2.1.5 Angle of Arrival

Angle of Arrival (AoA) is the angle of the a signal received on the receiver antenna. MUSIC algorithm [21] utilizes the phase of the CSI values to estimate AoA for each transmitter and receiver pairs. The output of the MUSIC algorithm is a pseudospectrum, where each peak is an estimated AoA. For linear antenna arrays, only 1-dimensional representation of AoA is obtainable.

AoA are often used in indoor localization methods [22]; given the location of the base stations and the AoA of the transmitted signal from the target object to each of the base station, the location of the target object can be determined using triangulation method.

2.1.6 Doppler-shift Effect

The shift in the carrier frequency that occurs when the target object that reflects the signal is moving is referred to as the Doppler-shift effect [23]. Doppler-shift can be measured by the wavelength and relative velocity between the transmitter and the target. When the target is moving towards the receiver, the Doppler-shift is negative; otherwise, the Doppler-shift is positive. For instance, the pitch of the siren in the police car rises as the car approaches, and falls as the car moves away. This happens because the frequency of the signal increases as the object that emits the signal (*i.e.,* police car) approaches the receiver (*i.e.,* the human listening to the siren) with negative Doppler-shift, and the frequency decreases as the object moves away with positive Doppler-shift.

2.2 Types of Wireless Signals

2.2.1 Wi-Fi Signal

Wi-Fi infrastructure has been primarily used to facilitate wireless communication between devices, until researchers began to utilize Wi-Fi side-channel information to achieve other purposes than communication.

From the intuition that the Received Signal Strength (RSS) from the receiver changes depending on the distance from the transmitter and interfering objects, several studies demonstrated utilizing the Wi-Fi RSSI data for localization or motion inference purposes [12, 24]. However, the RSS had too much heavy fluctuations and was too coarse-grained to extract precise information about diverse

motion because they only provided the information about the signal strength.

Some other work applied radar theories and techniques to Wi-Fi and used Wi-Fi signal to perform imaging or motion inference [25–28]. However, COTS Wi-Fi devices restricted access to physcial raw Wi-Fi data. Thus, these work had to simulate the Wi-Fi signal using special hardware like Universal Software Radio Peripheral (USRP).

In 2009, IEEE Computer Society released the IEEE 802.11n standard, which is an advanced Wi-Fi standard which supports Multiple Inputs Multiput Ouputs (MIMO) and Orthogonal Frequency Division Multiplexing (OFDM) technology that enables efficient communication [29]. In this standard, the modulation scheme changes adaptively depending on the physical state of the transmitting channel to provide reliable communication with high data rates. Accordingly, Wi-Fi NIC continuously monitors the state of the channel based on the Channel State Information (CSI), where the value reflects the difference in the amplitude and phase of the transmitted and received signal for each subcarrier (CFR). Such CSI data became accessible through the tools distributed by Halperin *et al.* [30] or Xie *et al.* [31]. The distribution of the CSI extraction tools fueled the motion inference research based on Wi-Fi CSI data [4, 5, 9, 10, 13, 14, 16, 32].

However, raw Wi-Fi phase information is rarely used due to Carrier Frequency Offsets (CFO) and Sampling Frequency Offsets (SFO). CFO is caused by the mismatch of the oscillators in the transmitters and receivers, which dynamically changing difference in carrier frequencies between a pair of Wi-Fi devices. SFO is caused at the receiver's side during the process of converting an analog signal to a digital signal. These offsets affect the phase value, burying the effects of motion in the signal [5].

2.2.2 Acoustic Signal

Although Wi-Fi CSI data can sensitively track fine-grained movements, commercial Wi-Fi devices restrict access to physical raw Wi-Fi data limiting opportunities for further analysis. Thus, acoustic signal is an attractive alternative because it is possible to access the physical raw acoustic signals easily by microphones. Moreover, RF signal travels much faster than sound signals. This may enlarge the interaction surface and the distance from the tranceivers, but at the same time, increase the influences of interferences. In addition, the smallest estimation in time may lead to substantial distance estimation error for RF signals.

Works that utilize acoustic signals can be divided into two categories: emission source target tracking and around-device target tracking. Emission source target tracking receives acoustic signals and localizes the target that have emitted the sound. These methods either construct a template of different location and find the best match for localization or obtain TDoA of two receivers (*i.e.,* microphones) to estimate the distance between the target and the object. Around-device target tracking transmits and receives acoustic signals and analyzes the changes in the multipath components caused by the target object that is near transmission channel. These work leverages the Doppler shift effect, echo profile, or phase information that are affected by the target's movement.

However, when a method that performs around-device target tracking transmits audible signals, such sounds may be unpleasant for users to hear. Thus, schemes that took around-device target tracking approach transmitted signals in an inaudible frequency band above 18 kHz. According to the Nyquist-Shannon sampling theorem, a sufficient sampling rate to capture B Hz frequency should be more than 2 B samples/second [33]. Thus, to transmit and analyze inaudible acoustic signal, the hardware must support at least 36 kHz sampling rate. Nev-

ertheless, this is not a problem for today's COTS devices; the common sampling rate of COTS smartphones is 44 kHz; in general, COTS smartphones support up to 48 kHz. Any sampling rate below 48 kHz do not harm the deployability of a scheme.

III. Signal-based Motion Recognition

In this chapter, we give an overview of the signal-based motion recognition methods we target to assess. We introduce the targeted schemes of this survey according to their signal source and briefly explain the techniques each method have leveraged to achieve motion recogntion. Finally, we explain the application domains of the methods.

3.1 Targeted Schemes

3.1.1 Wi-Fi

Many work has availed themselves of the ubiquity of Wi-Fi signals. Different work has exploited different data from the Wi-Fi signals to achieve their goals. The followings are the types of most widely used Wi-Fi data form: CSI magnitude, CSI phase, RSSI, and AoA.

RSSI. Several works have utilized Wi-Fi RSSI data. WiGest [34] performed gesture recognition using RSSI data. They applied DWT on raw RSSI signal and extracted the local maximum of the 5th level detailed wavelet. They distinguished three primitives from the wavelet: rising edges, falling edges, and pauses. They marked the start of the gesture by requesting users to perform a predefined preamble action and marked the end of the gesture when the silence period exceeds the predefined threshold period. When preamble and silence period was detected, they compared the sequence of primitives with gesture templates to find the best match.

UbiBreathe [12] estimated the breathing pattern. Their aim was not to

classify the breathing pattern, but to extract breathing signal and to detect signs of apnea. Thus, they performed various de-noising techniques to obtain the breathing signal from the RSSI signal that is unstable and has heavy fluctuations. UbiBreathe performed both frequency-domain and time-domain techniques to remove noises. In frequency-domain, UbiBreathe applied band-pass filter of cut-off frequencies from 0.1 to 0.5 Hz to eliminate data outside the human breathing frequency range (6 to 30 bpm). They further removed frequencies that have amplitude below threshold. In time-domain, they applied within-window local mean removal to handle sudden changes, fused overlapping consecutive windows, and applied α-trimmed mean filter to handle both impulse and gaussian noise. For apnea detection, UbiBreathe used DWT as a de-noising mechanism; they applied dynamic thresholding to the wavelet detailed coefficients and performed inverse wavelet transform on the detailed and approximated signal to reconstruct final de-noised signal. When the minimum and maximum value within the 10 seconds moving window is less than the threshold, UbiBreathe regarded it as a sign of apnea and alarmed the user.

CSI Amplitude. Many works utilized the Wi-Fi CSI data for motion recognition because CSI provides fine-grained and detailed information of each subcarriers compared to RSSI which provides averaged power strength over the whole channel bandwidth. In particular, most works that perform motion recognition using Wi-Fi signals from the commercial devices utilized the amplitude of CSI because phase information is often inaccurate due to CFOs and SFOs as mentioned in Chapter 2.1.3. In fact, many works shared similar processing techniques. Thus, instead of enumerating the methods by repeatedly stating their techniques, we group the frequently used techniques and describe how they were used in the methods.

1. De-noising: Prior to any analysis, they performed a de-noising preprocessing

step on the raw signal because CSI amplitudes contains high density noises and other information irrelevant to the motions the methods aim to recognize. Many works applied band-pass or low-pass filter as a de-noising method. WiHear [17] was interested in mouth movements that are within 2 to 5-Hz frequency range, so they applied band-pass filter to eliminate all the other frequency range outside their interest. Similarly, WiKey and WindTalker [9, 10] was interested in hand and finger typing movements that falls below 80 Hz, so they applied low-pass filter with cut-off frequency of 80 Hz to eliminate high frequency noise. E-eyes [4], which aimed to recognize human activities, adopted dynamic exponential smoothing filter (DESF) as for their low-pass filter. WifiU [14] applied a low-pass filter just for the smoothing purposes. WiWho [15] applied low-pass filter with cut-off frequency down to 2 Hz for step analysis, and FreeSense [16] with cut-off frequency of 10 Hz. Another widely used method was to apply Principal Component Analysis (PCA). Applying PCA on CSI data was suggested by the authors of CARM [5] from the idea that CSI streams of different subcarriers show high correlation when there is a movement. PCA component enabled extraction of the data that is highly relevant to the motion while reducing dimension of the multiple subcarriers data. [9, 10, 14, 16] have all applied PCA to eliminate noise and reduce dimension.

2. Domain transformation: After the analysis, many works transformed the signal into time-frequency or frequency domain signal. This step was necessary because even the slightest changes in the movement (*e.g.,* a bob of the head) alters the propagation paths which results in different CSI amplitude wave shape. In other words, same motion would result in different wave shape; making a wave signature or extracting features from such signal would yield inaccurate result. Thus, most of the works did not use raw time-series CSI am-

plitude signal, but instead applied Discrete Wavelet Transform was commonly used to capture both the time and frequency To capture both the time and frequency domain information from the signal, the following works have applied Discrete Wavelet Transform (DWT): [9,10,16,17]. DWT also reduces the computation cost because the size of the segment is reduced into half its original size for each order. An alternative method to capture the time and frequency domain information is to use Short-time Fourier transform (STFT). [14] applied STFT to draw the spectrogram with three dimensions: time, frequency, and FFT amplitude. Meanwhile, [13] transformed the time-domain signal into frequency-domain by applying Power Spectral Density (PSD), and [4] extracted the histogram (*i.e.,* the amplitude distribution) from the signal.

3. Identification: There were two approaches in classifying the motion: wavelet signature-based and feature vector-based. Wavelet signature-based methods trained the classifier by constructing signature templates of a particular motion directly from the wavelet. During testing time, these methods compared the input wavelet with the template wavelets of different motions. The template with the smallest distance was selected as the final choice of motion. To compare and measure the similarity of two waveforms, most works used the same algorithm: Dynamic Time Warping (DTW)—a method that calculates the minimum distance alignment between two temporal sequences [4, 9, 10, 16, 17]. However, each work used different types of classifier according to the purpose. WiHear used Spectral Regression Discriminant Analysis (SRDA), an advanced machine learning scheme that mitigates computational redundancy. WiKey and FreeSense used kNN classification scheme, in which the classifier searches for the majority class label among k nearest neighbors of the corresponding input. WindTalker and ARM built a classifier that selects the key with the minimum average distance between the training waveforms and

the input waveform. For more complex activities, ARM used Hidden Markov Model (HMM), which observes the sequence of data to infer hidden states. E-eyes took a similar approach to WindTalker and ARM, but used multiple-dimensional DTW for multiple subcarrier groups to identify the walking trajectory.

Another approach was feature-based approach. Instead of directly comparing the wavelet with the template, these works extracted features from the wavelets for classification. WifiU used LibSVM with Radial basis Function kernel as their classifier with the following features: time between two consecutive gait instances, torso movement speed, leg speed, and spectrogram signature. The spectrogram signature was constructed by dividing each half-gait cycles by four stages and calculating the energy of those stages. The shape of energy change in the frequency domain served as the spectrogram signature. These features were fed to the classifier to recognize user's gait. WiWho also proposed a gait-based identification method, but used a decision tree-based classifier with different features. For every 30-subcarrier CSI input, WiWho appened 8 features (min, max, median, standard deviation, skewness and kurtosis) to form a vector of 38 elements. From each of the 38 streams, WiWho extracted 16 time-domain features from activity band (0.3 to 2 Hz) for step features. For walk features, WiWho extracted either 16 time-domain features, or four frequency-domain features, or both from low-energy band (0 to 0.7 Hz), activity band (0.3 to 2 Hz), and high energy band (0.7 to 10 Hz) in addition to the number of steps per second. Wi-Fall used an SVM classifier with the following seven features to detect a falling activity: normalized standard deviation, offset of signal strength, motion period, median absolute deviation, interquartile range, signal entropy, and signal velocity change.

CSI phase. ARM [35] utilized CSI phase instead of amplitude to perform gesture recognition. As previously mentioned in Chapter 2.1.3, CSI phase information in current commercial Wi-Fi devices are inaccurate due to CFOs and SFOs. To resolve this problem, ARM performed both hardware and software synchronization. For hardware synchronization, ARM used the same GPS clock source for both the receiver and the transmitter. For software synchronization, ARM adopted maximum likelihood estimate algorithm by repeatedly transmitting the same data symbol and comparing the phases of each carriers. However, this requires access to the physical raw Wi-Fi signal, which is not possible for commercial Wi-Fi devices. Thus, instead of using commercial COTS devices, used USRP hardware, a software radio platform, to simulate Wi-Fi signal. Then ARM applied DWT to extract features and matched the constructed wavelet using DTW as the distance measure, selecting the gesture with the minimum DTW distance. For complex activities, ARM used HMM to observe CSI phase variations and recognize the activities.

AoA. WiDraw [36] utilized CSI to estimate AoAs of each tranmistter-receiver pair by employing the MUSIC algorithm [21] to track 3D hand trajectory. When a hand blocks a signal coming from a certain direction, the signal strength of the corresponding AoA experiences a sharp drop. Based on this idea, WiDraw tracked the signal drops of multiple AoAs to track the hand movement assuming an environment with a single receiver and multiple transmitter (*i.e.,* multiple incoming signals and AoAs). Once the AoA that experienced a sharp drop is determined, WiDraw performed a calibration process of rotating the antenna array over the z-axis to extract azimuth and elevation values from 2D AoA values in linear antenna array. From the obtained azimuth and elevation, WiDraw computed the horizontal and vertical coordinates of the hand. WiDraw further estimated

the depth of the hand's location by tracking the changes in the average RSSI values. They used RSSI to measure the depths because as the hand approaches the receiver, the signal strength will decrease since more incoming signals from different transmitters are blocked. To eliminate environmental noise, they applied low-pass filtering and eliminated signal strength drop that are smaller than 3 dB. With the horizontal and vertical coordinates and depth of the hand, WiDraw implemented an in-air drawing application.

3.1.2 Acoustic

Recent work utilized acoustic signals to infer physical motions. As previously mentioned in Chapter 3.1.2, it is easy to obtain raw physical acoustic signal through the microphone compared to Wi-Fi signals that restrict access to raw physical Wi-Fi signal. Thus, the use of acoustic signals enable fine-grained motion recognition, even up to mm-level hand or finger tracking.

Works that utilize acoustic signals can be divided into two categories: around-device target tracking and emission source target tracking. Around-device target tracking transmits and receives acoustic signals and analyzes the changes in the multipath components caused by the target object that is near transmission channel. These works leverages the Doppler shift effect, echo profile, or phase information that are affected by the target's movement. Emission source target tracking receives acoustic signals and localizes the target that have emitted the sound. These methods either construct a template of different location and find the best match for localization or obtain TDoA of two receivers (*i.e.,* microphones) to estimate the distance between the target and the object.

Doppler shift. AudioGest [37] leveraged the Doppler shift effect of acoustic signals that occurs when the hand (*e.g.,* reflector) is in motion. To remove noise,

AudioGest normalized the magnitudes of frequency bins for each audio frame, squared the difference of subsequent frames, and applied Gaussian smoothing. From the preprocessed signal, AudioGest derived an equation for computing the hand radial velocity. From the estimated velocity curve, they computed the direction of the hand based on the sign of the curve, and hand in-air duration based on the time interval of non-zero velocity. Furthermore, they computed the relative speed and waving range of the hand movement. This enabled AudioGest to recognize four linear movements (up, down, left, right) and two circular movements (clockwise and anticlockwise).

Echo profile. FingerIO [38] proposed a solution to track finger movements through analyzing the changes in the multipath components. FingerIO transmitted inaudible acoustic signals in OFDM format with silence intervals in between the symbols. To distinguish each individual multipath components, they constructed an echo profile of a symbol, which is the correlation of the original transmitted signal with the received signal as a function of the time delay. The echo profile had peaks (*i.e.,* high correlation) when a multipath component arrived at the receiver because the multipath component and the transmitted signal had strong resemblance. Through the echo profile, FingerIO obtained the amplitude and time of arrival for each multipath component. When a finger moves, the corresponding multipath undergoes change. Thus, FingerIO compared two subsequent echo profiles, tracked the changes in the multipaths, and find the arrival time of the multipath that showed a significant difference. From it, they estimated the length of the multipath (*i.e.,* the distance traveled via finger to the microphone) by multiplying the delay of the multipath with the speed of sound (343 m/s). Given the distance of the finger to the microphone, the finger can lie on any point on an ellipse where the estimated distance is equal to the sum of the

distance from the microphone and the speaker. Thus, given two microphones, FingerIO locates the finger by the intersection of two ellipses.

Phase. LLAP and Strata [39, 40] both used the change in the phase of the acoustic signal caused by the hand movement to track hand trajectory. LLAP and Strata both transmits and receives continuous wave acoustic signal. Since microphones can only receive real values, LLAP and Strata perform up-conversion and down-conversion to obtain complex values. With the obtained complex values that consist of real and imaginary parts, LLAP performed Local Extreme Value Detection (LEVD) to estimate the static vector which corresponds to the multipath component that travels through LOS path or are reflected by static movements. With a static vector removed, they obtained the dynamic vector that corresponds to the reflection caused by the moving hand. Strata, on the other hand, computed the CIR of the received data and subtracted two consecutive channel taps to remove static paths and obtain the phase change. Either way, multiplying the phase change with the wavelength of sound divided by 2π gave the path length change. These two works are very similar, but their differences are well stated in Strata. While LLAP measured the phase change caused by all surrounding objects, Strata separately measured the phase change with different delays, which resulted in better accuracy.

Spectrum density. UbiK [6] exploited the consistent fading patterns of acoustic signals caused by constructive and destructive multipaths to enable key entry with only a printed keyboard layout. They characterized the distinct fading patterns by their amplitude spectrum density (ASD). Using ASD as the feature vector, they found the best match of the input vector with the templates using nearest neighbors algorithm with Euclidean distance as the measure for dis-

tance. For optimization, they adopted the optimal separating hyperplane problem, which is used in statistical learning, to find the critical frequency range to lessen computational load.

TDoA. There were two keystroke inference attacks that leveraged the difference of arrival time in two different microphones to localize keystrokes. They both utilized the fact that most COTS smartphones are equipped with at least two microphones. They observed that the emitted acoustic signal from the key press has different paths for two microphones, which would result in different arrival time.

Tong Zhu *et al.* [7] also utilized the two microphones of the COTS smartphones and estimated the TDoA by finding the offset when the two signals from the two microphones yield maximum Generalized cross-corelation. Then they translated the time difference into distance by multiplying the TDoA with the speed of sound. The estimated difference of distance, Δd, is equal to the difference of the distance of the key to the first microphone and the second microphone. Thus, they drew a hyperbola by connecting all the points that has a constant value Δd with the two microphones as the focal points, where any point on the other half hyperbola would become a candidate location of the pressed key. After eliminating half of the parabola that is farther away from the pressed key, they applied an upper and lower bound to Δd to compensate for estimation error, which resulted in a hyperbola band. To locate the key, they used multiple smartphones and found the overlapping regions of the half parabolas.

Jian Liu *et al.* [8] also performed a keystroke inference attack, but they only used a single phone to achieve the same goal as [7]. Prior to localizing keystrokes, they first precomputed the expected TDoA values for each key given known keyboard layout and phone placement. Then they grouped the 26 expected TDoA

values into 13 key groups. After precomputation, they used cross-correlation to estimate the TDoA of the pressed key. Then they categorized the TDoA according to the key groups. Within key groups, they extracted MFCC features from the audio samples and clustered the keystrokes uisng k-means clustering. After clustering the keystroke, they calculated the mean TDoA of the cluster, which would be very close to the precomputed theoretical TDoA. Finally, they labeled the keystrokes in the clusters by selecting the theoretical TDoA that has minimum distance with the mean TDoA of the cluster.

IV. Evaluation Factors

In this chapter, we propose the seven practicality facets to evaluate the practicality of the signal-based motion recognition methods. ISO/IEC 9126-1 standard [41] classified the quality of a software into six categories: functionality, reliability, usability, efficiency, maintainability, and portability. We base our facets on the ISO/IEC 9126-1 standard, but eliminate the components that are less relevant or inapplicable to the signal-based motion recognition and expand the components that must be considered. We remove maintainability from the six categories because maintainability is related to how the solution is implemented, rather than the techniques or the solution itself. Functionality, according to the definition of the ISO/IEC standard, is a set of attributes related to a set of functions and their specified properties that satisfy the stated or implied needs of a solution. Based on its definition, we expand the functionality component into granularity, efficiency, robustness, and applicability, of which are the assets relevant to the functionality of signal-based motion recognition scheme. We expand reliability to robustness and stability based on the definition of reliability – a set of attributes related to the capability of a solution to maintain its level of performance under different conditions. We replace portability to deployability – a set of attributes related to the ability of a solution to be installed and deployed in a stated environment. The details of each attributes, of why they account to the quality of a scheme in terms of practicality, is described in the following sections. To sum up, we come up with the following seven practicality facets: *granularity, robustness, applicability, efficiency, usability, stability,* and *deployability.*

4.1 Granularity

A scheme's granularity determines how much motion variations a scheme can capture. We categorize the granularity of the schemes into four levels: detection, counting, classification, and tracking.

1. **Detection-level** has the lowest granularity in which it gives binary results of whether certain motion has occurred. Because the schemes that belong to the detection-level only give binary results, they cannot identify which specific type of motion has occurred. In general, detection-level schemes examines if certain data extracted from the signal exceeds a predefined threshold.

2. **Counting-level** estimates the number of occurrence of a specific motion or the number of the objects within certain boundary. Their granularity is higher than simply detecting the presence of motion. However, counting-level schemes also cannot identify which specific type of motion has occurred.

3. **Classification-level** classifies motion into categories. An example of classification would be classifying user's activities into running, walking, or jumping categories. In general, classification-level schemes builds profiles of specific motions and compares the data collected during run-time with the profiles to identify the motion. The more classes a scheme can distinguish, the more types of motions a scheme can capture. For instance, if scheme A can recognize three types of hand gestures while scheme B can recognize ten types of hand gestures, scheme A captures more types of motion than scheme B. Thus, we regard scheme A to be more fine-grained than scheme B.

4. **Tracking-level** has the highest granularity in which traces a target moving object in a form of localization. The schemes that belong to the tracking-level tracks the trajectory of a target object. Thus, the output is a 2-D drawing that projects the 3-D trajectory. The smaller the error distance between the projected trajectory from the original trajectory, the more fine-grained a method can trace the object.

4.2 Robustness

A scheme must be robust against interferences. Unless a scheme is targeting a motion that occurs in a chamber with shielding that blocks RF or acoustic signals, interferences are present in real-world settings. Thus, a scheme must be designed in consideration of such interferences. We consider two types of interferences: motion and same channel signal.

Interfering motion. A scheme must function properly in presence of interfering motions. Consider a scheme that aims to recognize the motion of an object, which we refer to as the target object. The scheme analyzes the reflected signal from the target object to recognize the motion. However, when other moving objects are nearby the target objects, reflected signals from the moving objects may interfere with the reflections of the target object altering the propagation paths. Such interfering objects' motion may result in degradation of the accuracy. Thus, a scheme must consider such interfering motions in their design.

To assess the robustness against interfering motions, we examine the interfering boundary of a scheme. We define an interfering boundary as a limit of distance that does not significantly affect the scheme's accuracy. As the distance of the object and the receiver increases, the signal-to-noise (SNR) decreases. Thus, any interfering motions outside the interfering boundary do not harm the accuracy. We regard a scheme as *strongly robust* (✓) against interfering motion

when the size of the motion and the size of the interfering boundary are similar. However, some schemes may have large difference of size between the motion and interfering boundary. In such a case, it is inevitable to include interfering objects within the boundary. We regard a scheme as *moderately robust* (△) when a scheme handles the influences of the interfering motions within the interfering boundary. The more interfering motions a scheme can handle, the stronger the robustness. When a scheme only operates in a setting without any interferences, we regard such schemes as *sensitive* (✗).

Interfering signal. A scheme must function properly in presence of interfering signals. A scheme that utilizes acoustic signal as its signal source must be robust against other sounds that may interfere with the reflected acoustic signal. Similarly, a scheme that utilizes Wi-Fi signal as their signal source must be robust against other RF-signals in the channel. We regard a scheme that tolerates such interfering signals as *strongly robust* (✓). The heavier the interferences, the stronger the robustness. When a scheme only operates in a setting without interfering sounds or RF-signals, we regard such schemes as *sensitive* (✗).

4.3 Applicability

A scheme must be applicable in various circumstances. We consider whether a scheme is applicable in the following two cases: a multiuser case and a non-line-of-sight case.

Multiuser case. We regard a scheme that has capability to recognize motions of multiple users simultaneously to have considered a multiuser case. Consider a scheme that aim to recognize activities in a smart home environment. A scheme that only handles a single resident smart home has restricted applications than a scheme that handles multiple users; all the more so because multiple person households are the common typical cases [42]. Thus, we regard a scheme as

highly applicable (✓) when it works for multiple users. Otherwise, we regard a scheme as *less applicable* (✗) when it only works for single users. The more users a scheme can handle, the higher applicability.

Non-line-of-sight case. A scheme that operates in NLOS has wider view than a scheme that only operates in LOS. When a scheme supports through-the-wall cases, a single transmitter-receiver pairs to capture motions from outside the building or from a different room. The NLOS capability is particularly more useful for attacks because attackers can launch the attack more covertly by hiding the tranceivers. Thus, we regard such schemes as *highly applicable* (✓). Otherwise, we regard them as *less applicable* (✗).

4.4 Efficiency

A scheme must require minimum computation cost. We consider how much power, time, and memory consumption a scheme consumes along with the sampling rate and the device type of the receiver.

Power, Time, and Memory (PTM) A scheme that consumes too much time, power, and memory is unlikely to be deployed in real-world settings. Thus, we examine the scheme's evaluation on time, power, and memory. We regard schemes that proved the efficiency of their model through evaluation on PTM as *highly efficient* (✓). We regard schemes that did not evaluate the PTM as *inefficient* (✗).

Sampling rate. Higher sampling rates indicates higher computing cost – the transmitter and the receiver haves to continuously transmit and receive the signal at a higher rate, consuming energy. We regard a scheme that requires higher sampling rates to incur higher overhead.

Device type. We also consider the device type in the assessment. Depending on the type of the device, the criticality of the overhead to the device

type a scheme uses for data collection and processing may vary because different devices have different computing power and memory space. A scheme that consumes much energy and memory may be more critical for devices that have limited computing power and memory space. Thus, we regard schemes that uses desktop terminals or laptops to be less critical (✓) to the overhead than schemes that uses mobile devices such as smartphones and tablets (✗).

4.5 Usability

A scheme must be convenient for users to use because users may avoid adopting a scheme with low usability. We consider two factors to assess the usability.

Training samples. A scheme must require minimum training samples to achieve high usability. This factor only applies to the schemes that used machine learning technique for recognition. A training sample in a user's perspective is performing and labeling each motion types repeatedly up to the number of required training samples. The lesser the training samples a scheme requires, the higher the usability. Note that the number of training samples increase with the number of classes; thus, increase in the number of classes may enhance the granularity (Chapter 4.1) but increases the training cost.

Training samples and the four factors of stability in Chapter 4.6 are also tightly related. When a scheme satisfies all four stability factors, one time training may be applicable to all users; thus, users may not have to individually train the classifier. However, if any of the stability factors (time, location, user, and the position of the transmitter/receiver) are unsatisfactory, user involvement is inevitable. For example, if a scheme is unstable over different locations, only the data collected in the user's location can be used as the training samples for the classifier; thus, a user has to train the classifier manually at his location. Furthermore, when a scheme is unstable over one of the four factors, and such

factor experiences significant changes significant changes (*e.g.,* the position of the device changes), a user may have to retrain the classifier and repeat the process all over again. Therefore, the four stability factors must be examined together with the number of required training samples.

Manual preadjustments. Requiring users to manually take part in the preadjustment process (*e.g.,* calibration) greatly harms the usability. We regard a scheme as *highly usable* (✓) when the scheme does not require any manual preadjustment process. When preadjustment process is necessary, we regard a scheme that requests simpler process as more usable.

4.6 Stability

A scheme must be stable over variable conditions. Real world environments change dynamically (*e.g.,* change in the location of the furnitures). Thus, a scheme that only works in fixed conditions and fails to adapt to variances is impractical for real world circumstances that changes dynamically. To assess the stability of a scheme, we examine the following four factors: time stability, place stability, user stability, and device stability.

Time stability. A scheme remains stable with respect to time. For machine-learning based approaches, a scheme has high time stability when the model functions properly with the testing data that is collected certain time difference before or after the collection of the training data. The longer the time difference, the higher the stability. When training and testing data are not collected separately (*i.e.,* cross-validation), the longer the data collection period, the higher the stability.

Place stability. A scheme remains stable with respect to target object's location. For machine-learning based approaches, a scheme achieves high location stability when the model functions properly with the testing data collected from

an untrained location. The farther the location, the higher the stability.

User stability. A scheme remains stable regardless of the user. For machine-learning based approaches, a method is user agnostic when it operates on untrained user's data.

Device stability. A scheme remains stable with respect to the position or orientation of the signal transmitting or receiving device. The more variation of the device position and orientation a scheme allows, the higher the device stability.

4.7 Deployability

A scheme must be easily deployable. We consider two factors that may restrict the realization of a scheme in real-world.

Specialized hardware requirement. Schemes that require specialized hardware or modification of hardware are less likely to be deployed by users than Commercial Off-the-Shelf (COTS) devices. We categorize the hardware requirement into the following levels:

1. **Modified hardware:** A scheme requires modification of a hardware to operate. This level has the lowest deployability because user has to modify the hardware to adopt the scheme, which requires high technical ability. The more complex the modification process, the lower the deployability.

2. **Simulation of COTS hardware:** A scheme uses a prototyping tool to simulate the signals of the COTS devices. For example, many studies used Universal Software Radio Peripheral (USRP) to simulate the and obtain the raw physical Wi-Fi signal. However, COTS Wi-Fi devices do not allow access to the physical raw signal; thus, such schemes cannot be adopted unless the COTS Wi-Fi device corporate adopts the scheme.

3. **COTS hardware:** A scheme only uses COTS devices. This level achieves the highest deployability because individuals can easily install and deploy the scheme on readily available COTS devices such as smartphones or COTS Wi-Fi devices.

Multiple hardware requirement. A scheme must require minimum number of devices to operate to achieve high deployability. For example, a scheme that requires three smartphones to recognize user's keystroke is less practical than a scheme that only requires a single phone to achieve the same purpose. The more the number of required hardware, the lower the deployability.

4.8 Accuracy

Accuracy determines the correctness of a scheme. If a scheme that satisfies all of the seven facets above but achieves low accuracy, such scheme are likely to be neglected. However, as mentioned in Chapter I, accuracy is determined by various parameters such as the number of training samples, number of devices, or the distance in between the target object and the transceivers. For this reason, we do not include accuracy in the seven facets, but instead, show how the accuracy changes with the different parameters in each practicality assessment facet. A practical method must achieve high accuracy while satisfying the seven practicality facets.

4.9 Scoring Method

Depending on the values on the table, a scheme is given a score for each of the seven practicality facets.

1. **Quantify.** A facet consists of multiple factors. We quantify the status of each factors. We can easily quantify the factors that have numeric values

Table 4.1: Conversion of a factor value to a numerical value

Factor value	Numeric value
N	N
Level N	N
✗ or NC	0
δ	0.5
✓ or NA	1

such as the number of training samples or devices, but factors with symbols must be converted to a numeric value. Thus, we assign values to the different markings. Symbols ✗, △, and ✓are given a value of 0, 0.5, and 1, respectively. An unconsidered factor (NC) or not specified factors (–), are given a value of 0, while a inapplicable factor (NA) is given a value of 1. A level is given a value from 1 to the number of levels.

2. **Normalize.** We normalize and multiply the value of each factor by 10 to derive the score of a factor. When a larger value has higher qualities (*e.g.,* number of classes), we normalize the value of a scheme by the maximum value. When a smaller value has higher qualities (*e.g.,* required training samples), we normalize the value of a scheme by the reciprocal of the minimum value. We normalize the numeric factors (*i.e.,* number of class, training samples, number of devices, and sampling rate) with the minimum or maximum value within their groups. We normalize the levels and the non-numeric factors by the obtainable minimum or maximum value. However, we make an exception to the sampling rate because it strongly depends on the type of a signal. Thus, we normalize the sampling rate of a scheme by the minimum value of its group and its type.

3. **Average.** The overall score of a facet is the average score of the factors except for granularity. A granularity facet has three different scores: gran-

ularity level score, classification score, and a tracking error score. We treat granularity as an exceptional case because the the scores are not of equal levels; classification score and tracking error score are under the granularity level score. Thus, we put more weight on the granularity level score and compute the weighted average to obtain the overall granularity score.

4. **Sum.** After averaging, each factor in the seven facets have a score that range from 0 to 10. We sum up the scores of the seven facets and obtain the practicality score of a scheme. Thus, the maximum practicality score of a scheme sums up to 70. We give equal weights to the seven facets, but they can be given different weights.

   Nigel Bevan categorized the primary concerns of the usability standards into four perspectives: 1) the use of the product 2) the user interface and interaction 3) the process used to develop the product 4) the capability of an organization to apply user-centered design [43]. In view of the first perspective, granularity, efficiency, robustness, and applicability facets can be given heavier weights than others because these factors examine how much a scheme effectively, efficiently, and satisfactorily achieves the context of use. In view of the second perspective, the usability and stability of a scheme are the facets can be given heavier weights because they examine how much user involvement a scheme requires. In view of the third perspective, deployability can be given heavier weights because it is related to the development process. (No factors among the seven facets correspond to the fourth perspective because the the organizational capability to develop a easily manageable life cycle of a product is less relevant to the technique itself.) Depending on the different perspectives of a user or an application, the seven factors can be given different weights.

Table 4.2: Gait Recognition schemes

		Granularity		**Stability**				**Robustness**		**Applicability**		**Usability**		**Deployability**		**Efficiency**		
		Inference lev.	Class or Error dist.	Time	Place	User	Device	Motion interference	Signal interference	Multi-user	NLOS	Training req.	Pre-adjustment	Hardware req. lev.	Multiple devices	PTM	Samples	Device
Gait	FIMD [44]	1	NA	✓	✓	✓	✓	NC	NC	NA	✗	–	✓	3	1TX 1RX	NC	–	✓
	Banerjee *et al.* [45]	1	NA	✓	✓	✓	△	NC	NC	NA	✓	NA	✓	3	1TX 2RX	NC	10/12 Hz	✓
	Freesense [16]	3	2, 9	NC	✗	✓	✗	NC	NC	✗	✗	20 per user	✗	3	1TX 1RX	NC	1000 Hz	✓
	WiWho [15]	3	2, 7	△	✗	✓	✗	✗	NC	✗	✗	–	✗	3	1TX 1RX	NC	100 Hz	✓
	WifiU [14]	3	7, 50	✓	✗	✓	△	✗	△	✗	✗	40 per user	✗	3	1TX 1RX	✓	2500 Hz	✓
Activity	WiFall [18]	1	NA	✓	✗	NC	NC	NC	NC	✗	✗	2×4	✓	3	1TX 1RX	NC	100 Hz	✓
	ARM [35]	3	6	NC	NC	NC	NC	NC	NC	✗	✗	–	✓	2	1TX 1RX	NC	–	✓
	CARM [5]	3	10	✓	✓	✓	✓	△	✓	✗	✗	(6-30)×10	✓	3	1TX 1RX	✓	2500 Hz	✓
	E-eyes [4]	3	17	✓	NC	NC	✗	NC	NC	✗	✗	–	✓	3	1TX 3RX	NC	20 Hz	✓
	WiSee [28]	3	9	✓	✓	✓	✓	✓	NC	△	✓	NA	✗	2	2TX 1RX	NC	–	✓
Gesture	AudioGest [37]	3	6	✓	✓	✓	✗	✓	✓	✗	✗	NA	✓	3	1 TXRX	NC	44.1 kHz	✗
	WiGest [34]	3	4	✓	✓	✓	✓	△	NC	✗	✓	NA	✗	3	1TX 2.5RX	NC	50 Hz	✓
	Melgarejo *et al.* [46]	3	14, 25	✗	✗	✗	✗	NC	NC	✗	✗	40×25	✓	2	1TX 1RX	✓	–	✓
	WiDraw [36]	4	5.4 cm	✓	✓	✓	✓	△	NC	✗	✓	NA	✗	3	30TX 1RX	NC	25 Hz	✓
	LLAP [39]	4	4.6 mm	✓	✓	✓	✓	✓	✓	✗	✓	NA	✓	3	1 TXRX	✓	48 kHz	✗
	FingerIO [38]	4	8 mm	✓	✓	✓	✓	✓	NC	✗	✓	NA	✓	3	1 TXRX	✓	48 kHz	✗
	Strata [40]	4	10.1 mm	✓	✓	✓	✓	✓	✓	✗	✗	NA	✗	3	1 TXRX	✓	48 kHz	✗
Keystroke	Liu *et al.* [8]	3	26	✓	✓	✓	△	✓	NC	✗	✗	NA	✓	3	1 TXRX	NC	48 kHz	✗
	Zhu *et al.* [7]	3	28	✓	✓	✓	△	✓	✓	✗	✗	NA	△	3	3 TXRX	NC	44.1 kHz	✗
	WiKey [9]	3	37	✗	✗	✗	NC	NC	NC	✗	✗	30×37	✓	3	1TX 1RX	NC	2500 Hz	✓
	UbiK [6]	3	56	NC	✗	✗	△	NC	✓	✗	✗	3×56	✗	3	1TX 1RX	✓	48 kHz	✗
	WindTalker [10]	3	10	✗	✗	✗	NC	NC	NC	✗	✓	10×10	✓	3	1TX 1RX	NC	800 Hz	✓
Others	UbiBreathe [12]	2	NA	✓	✓	✓	△	NC	NC	✓	✓	NA	✓	3	1TX 1RX	✓	10 Hz	△
	Wang *et al.* [13]	2	NA	✓	✓	✓	△	NC	NC	✓	✗	NA	✓	3	1TX 1RX	NC	20 Hz	✓
	WiHear [17]	3	33	NC	NC	NC	NC	✓	✓	✓	✗	50×33	✓	3	1TX 1RX	NC	100 Hz	✓

V. Evaluation

5.1 Overview

We apply the seven facets we described in Chapter IV to assess the schemes. For better comparison, we group the schemes by the following motion types: gait, activity, hand/finger gesture, keystroke, and others. Table 4.2 shows the result of the investigation.

5.2 Gait Motion

We assessed five studies that used the reflected signals caused by the gait motion to either detect or recognize users. Figure 5.1 is the chart that represents the scores of the seven facets of the gait based schemes.

5.2.1 Granularity

Two studies achieved detection level granularity. The goal of FIMD was to perform motion detection and monitor the position changes of entities. We included their target motion in the gait motion category because they had subjects walk back and forth when collecting the data with motion in their experiments. A. Banerjee *et al.* detected a line-crossing of the links between the legitimate transmitter and the maliciously deployed receiver outside a concrete wall. In addition to the line-crossing detection, they also estimated the direction of movement (left or right).

Three studies achieved classification level granularity. They all performed gait-based human identification using Wi-Fi signal reflections. The number of classes for these schemes indicated the number of users the schemes can identify.

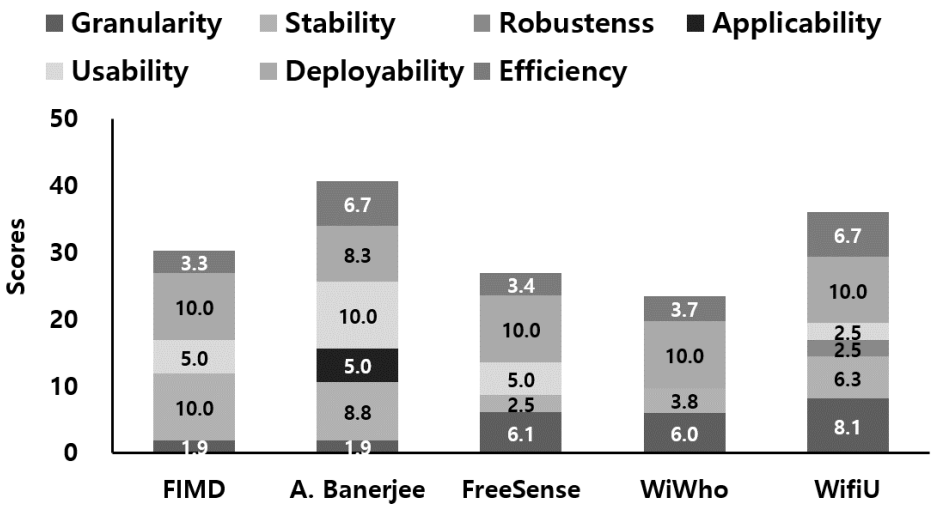

Figure 5.1: Assessment of gait-based motion recognition schemes

The accuracy was greatly affected by the number of users; it decreased with the increase in the number of users. FreeSense achieved an identification accuracy of 94.5% for two users, but it drastically decreased to 75.5% for nine users. Similarly, WiWho achieved an identification accuracy of 92% for two users, but it decreased down to 75% for 7 users. WifiU classified more users with higher accuracy than the other two schemes; it achieved 92.31% and 79.28% for 10 and 50 users, respectively.

The core difference between WifiU and other gait identification schemes was the machine learning features. FreeSense extracted the following time-domain features in their classifier: maximum, minimum, mean, median, standard deviation, skewness and kurtosis. WiWho used the time-frequency wavelet derived from the Discrete Wavelet Transform as their feature. However, WifiU extracted gait cycle length, estimated footstep length, the maximum, minimum, average, and variance for torso and leg speeds during the gait cycle along with the time-frequency spectrogram signature. Through the evaluation results, WifiU's features have shown to be more effective gait features for recognizing users.

5.2.2 Stability

Detection-level schemes [44, 45] were highly stable over time, place, users, and device position. In contrast, classification level schemes were less stable over the stability factors.

Time stability. FreeSense did not specify the data collecting period of time nor performed experiments on time stability. WiWho performed cross-validation in their evaluation; their data collection period was four to seven hours. This guarantees the stability of the classifier up to seven hours; thus, a user may have to re-label the gait every seven hours. WifiU also performed cross-validation in their evaluation but collected their data over four months of period. Therefore, compared to the other two classification level schemes, WifiU achieved the highest time stability.

Place stability. All three schemes required users to walk only the predefined paths that the users walked during the training period. Therefore, all three schemes had low place stability.

User stability. The goal of the schemes were to identify users. Thus, we labeled the user stability degree of the schemes as 'Not Applicable' (NA) in the table.

Position stability. All three schemes were sensitive to the changes of the transmitter/receiver positions. They all required fixed deployment of the devices. Although authors of WifiU mentioned that there were slight changes in the location of the router and the laptop during the four months data collection period, they conducted no experiments regarding the stability of the device position nor the orientation. Thus, we marked them as '△' in the table.

5.2.3 Robustness

All five gait motion detection or recognition schemes did not deeply investigate on the robustness of the scheme regarding interfering motions and signals. We marked the schemes that did not consider the effect of interferences as 'Not Considered' (NC) in the table. WiWho and WiFall explicitly mentioned that only one person should be walking while the scheme runs. Thus, we marked them as ✗.

5.2.4 Applicability

FIMD's goal was to detect motion and [45]'s goal was to detect line-crossings. Thus, consideration on multiuser cases were unnecessary for these work. FreeSense, WiWho, and WifiU also did not supported multiuser cases. Among five studies, only the model proposed by A. Banerjee *et al.* considered a NLOS case. They deployed receivers outside a 1-ft thick concrete wall to infer users' line-crossing within the building.

5.2.5 Usability

All five schemes did not require any form of preadjustment process. Thus, they all achieve high usability in terms of preadjustment process. However, three gait identification schemes required training instances of user walking a predefined path because they took a machine learning approach. FreeSense used 20 training samples each for the nine test subjects. They also included an evaluation that investigated the effect of training size, which showed that as the training samples changed from 10 to 30, the accuracy increased from 75.0% to 91.7%, but they did not specify the number of users. WiWho collected 20 samples for each users but they did not specify how much among the 20 samples are used for training. WifiU required the most number of training samples to build a classifier, which

was 40 training instances.

5.2.6 Deployability

All schemes except the model of A. Banerjee *et al.* used only a single COTS transmitter and receiver pair. Thus, they achieved high deployability. They also used COTS Wi-Fi devices but their accuracy was affected by the number of receivers and the antenna spacing that they manually customized. At best conditions with two receivers and large antenna spacing, they achieved 0% FPR and FNR on line-crossing detection with 100% accuracy in direction detection. However, when they used a single Wi-Fi receiver with normal antenna spacing, the FNR increased to 1.92% and the accuracy dropped down to 59.62%.

5.2.7 Efficiency

Except for WifiU, no work evaluated the time, power, or energy overhead of their system. WifiU measured the CPU processing time required for processing and training; their scheme took 0.275 seconds to process one second of data and 100 ms to train a gait model with less than 800 training samples. Their evaluation results proved the efficiency of their scheme. The scheme with the lowest sampling rate was the model of A. Banerjee *et al.*, which transmitted and received 10 packets per second. Particularly, they had to minimize the sampling rate because they devised an attack that utilizes transmitted packets from the devices of the legitimate users who are unaware of their system. The scheme that required the most amount of sampling rate was WifiU which continuously transmitted and received 2500 packets per second. Nevertheless, FIMD, FreeSense, WiWho, and WifiU used stationary devices such as a laptop that has relative high computing power and memory.

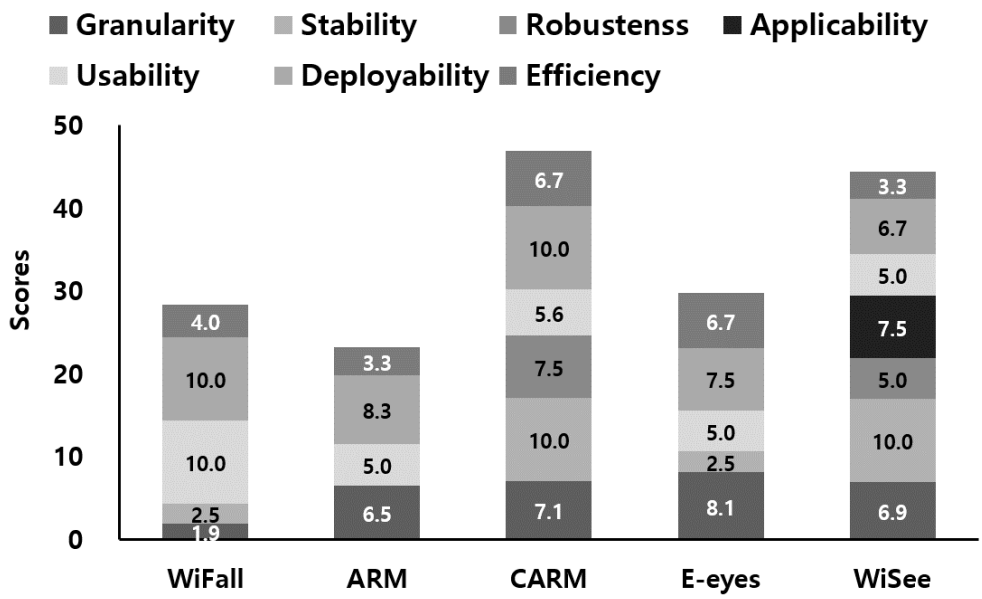

Figure 5.2: Assessment of activity recognition schemes

5.3 Activity

We assessed five schemes that performed activity recognition. Activity recognition can be used to monitor user's daily life pattern to provide services such as healthcare, eldercare, or fitness tracking. Figure 5.2 summarizes the assessment on the gait based schemes.

5.3.1 Granularity

Although WiFall's classifier classified motions into three to four activities (sit,stand up, walk, and fall) to differentiate falling action from other motions, they achieved detection level granularity because the accuracy evaluation was based on how much falling actions it can detect instead of how much motions they can classify. ARM identified six in-place activities.WiSee utilized the positive and negative Doppler shifts present in the Wi-Fi signal to recognize the motion. Thus, they selected nine gestures that can be represented by sequence of positive and negative Doppler Shifts.CARM identified eight in-place activities and two moving

activities such as walking and running. E-eyes identified nine in-place activities and eight walking trajectories.

5.3.2 Stability

Two studies achieved high stability over the four stability factors. WiSee modeled the gestures by constructing a profile of positive or negative Doppler shift sequence. The Doppler shift profile was stable over time or location or device position compared to CSI signal streams, although they required users to perform the action towards the receiver. Their model also worked with different users performing the same gesture with different speed. Authors of CARM considered the stability factors in their work. They tested their model on untrained location, user, and device position and achieved 80% classification accuracy. CARM's experiment on untrained location also implied time stability because of the separately collected testing and training data. Thus, CARM did not require on-site training data collection process.

Three studies had less consideration on the stability factors. WiFall was stable over time because it collected the training data from their testbed for one week and collected the testing data separately afterward. However, their model achieved low place stability because they only tested their data on particular trained spots. They also did not consider their model on untrained users or altered device position. E-eyes collected the training data for one day and the testing data over different days. To adapt to changes over time, they also added a semi-supervised mechanism that clusters daily activities and requests users to label newly clustered activity set. However, they also only tested their data on the trained spots within the room with the trained people. They also required the Wi-Fi signal transmitting and receiving devices to be stationary, failing to meet the device position stability factor. ARM did not perform any evaluation

on all stability factors. Thus, they did not satisfy any of the stability factors.

5.3.3 Robustness

Only two studies investigated the scheme's robustness to interfering motions and signals. CARM experimented on the effect of same channel interferences. The test results showed their model's tolerance to such interferences. WiSee evaluated the detection and classification accuracy of their scheme in presence of other interfering humans. They had 12 subjects within the room for 24 hours and counted the false alarms. When they repeated the preamble motion three times, average false alarms per hour was only 0.13. They also tested the classification accuracy in the presence of interfering users. They used a 5-antenna receiver in their experiment. When there were three interfering users, their model successfully classified the activities with 90% accuracy. However, with four interfering users, the accuracy dropped down to 60%. CARM did not conduct an additional experiment on interfering motions like WiSee, but they collected their data while other people were sitting or using computers in the same room. This showed some level of robustness against interfering motions, so we marked them as '△'.

5.3.4 Applicability

Only WiSee considered multiuser and NLOS cases in their design. WiSee treated each users as transmitters reflecting Wi-Fi signals. They adopted the MIMO technology, in which they leveraged the repetitive motion of a user as a preamble to compute the channel that maximizes the energy of the user's reflections. Although recognizing activities of multiple users simultaneously should work by principle, they did not conduct an experiment to verify the idea. They only experimented on the target user's classification accuracy in the presence of other users. Thus, we marked them as '△'. WiSee also considered various NLOS cases. They experimented the feasibility of their model by measuring the Doppler

SNR in different transmitter and receiver deployments including through-the-wall, through-the-corridor, and through-the-room cases. In the whole-home scenario, they achieved 94% classification accuracy on data that in included data collected in NLOS and through-the-wall spots.

5.3.5 Usability

WiSee did not require any training samples because it did not use machine learning for classification. WiFall required the least amount of training samples next to WiSee. They asked subjects to perform four activities three to five times. They did not specify how much of them were used as training samples in the chamber and laboratory testbed, but in the dorm testbed, they used two sets for training and two sets for testing. CARM collected different number of samples for different activities and provided the values on a table. The collected samples of the ten activities varied from 60 to 300 samples and used 10% of them as the training data. In the table, we put the average number of training samples among the activities. E-eyes collected 50 known activities, 100 unknown activities, 20 known trajectories, and 20 aimless trajectories in their data collection, but they did not specify how much among the collected data was used for the training data. ARM also omitted details of the data collection.

5.3.6 Deployability

[4,5,18] used only the COTS devices to recognize activities. However, ARM and WiSee simulated Wi-Fi signal using USRP to obtain physical Wi-Fi signal. ARM used the raw signal to compute CSI phase difference, and WiSee used the raw signal to obtain the Doppler shift of a signal.

E-eyes and WiSee required multiple devices. E-eyes used 1 AP as a receiver and used three Wi-Fi devices as the transmitters and achieved 97% accuracy. With a single receiver, the accuracy dropped to 90%. WiSee deployed 4-antenna

receiver and two 1-antenna transmitters in the room in the whole-home scenario. When they only used 1 transmitter, they could not obtain high accuracy in three of the nine locations that were the blind spots.

5.3.7 Efficiency

Only CARM conducted an evaluation on the time cost. CARM took 100.55 seconds to train the 9 activities (1400 samples), but the overhead was reasonable because it was a one-time cost. However, they also did not evaluate on the real-time time, power, or memory cost. Among the five activity recognition schemes, their model required the most sampling rate, which was 2500 packets per second. WiFall transmitted 100 packets per second. WiSee required the least sampling rate, which was only 20 packets per second.

Three schemes, WiFall, CARM, and E-eyes, were designed to run on relatively high computing devices like laptops. ARM and WiSee ran their models on the daughterboards because they used USRP to simulate the Wi-Fi signal, but we still considered their overhead as less critical because their models would run on high computing devices when they are commercialized.

5.4 Hand/finger Gesture

Gesture recognition technology has emerged as a means for future human-computer interaction, all the more so due to the decreasing interface sizes of portable devices such as smartphones and smartwatches. Reflecting their popularity and trend, many studies [34, 36–40, 46] proposed gesture recognition or tracking solutions that uses signal propagation characteristics.

5.4.1 Granularity

Three studies achieved classification-level granularity. WiGest distinguished four primitive hand gestures: up-down, left-right, infinity shape, and fist open-

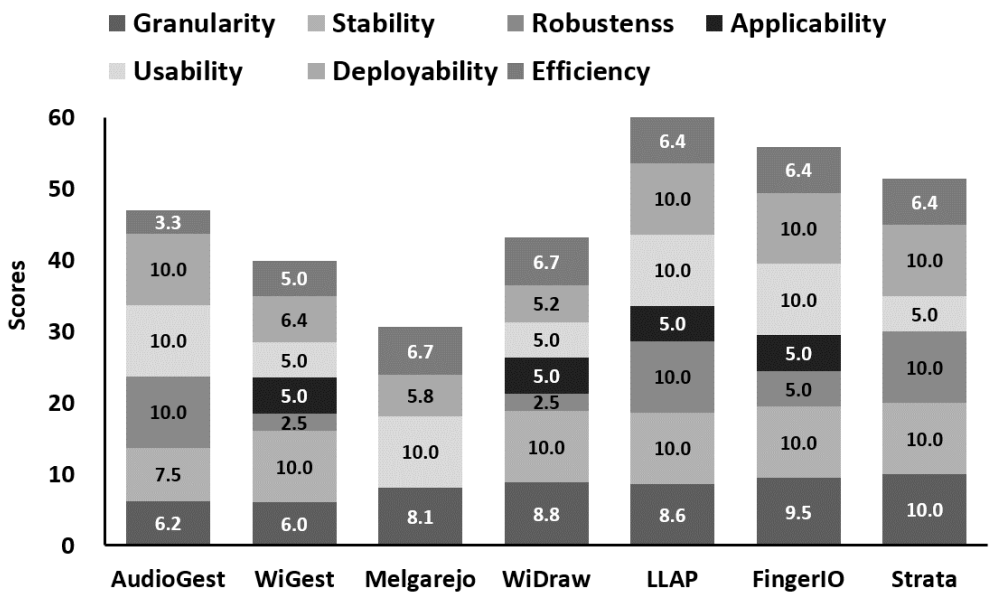

Figure 5.3: Assessment of hand or finger gesture recognition schemes

close motion. However, this model could not distinguish the direction of the movement—up from down, and left from right. AudioGest resolved this problem by using acoustic signal to compute the sequence of hand-microphone angle and its corresponding duration. This approach enabled distinction between the direction of the hand movement, and thus distinguished six primitive hand gestures: up, down, left, right, clockwise circular, and anti-clockwise circular motion. P. Melgarejo *et al.* modeled American Sign Language (ASL) words by their Received Signal Strength (RSS) and phase difference information and classified them with the nearest neighbors algorithm; they estimated the distance by the cross correlation coefficient or Dynamic Time Warping (DTW). They selected 25 ASL words for the wheelchair scenario. For the in-vehicle scenario that is more challenging due to denser multipath propagations, they selected 14 words from the 25 word set but did not specify which words they selected with what standard.

Four studies achieved tracking-level granularity. WiDraw used the AoA links between the user's receiver and multiple transmitters to track user's hand tra-

jectory. They achieved 5.4 cm median tracking error with 30 neighboring transmitters. The interaction surface plane supported by WiDraw was approximately 30 cm by 100 cm wide. FingerIO, LLAP and Strata transmitted and received inaudible acoustic signals to track user's finger movement. FingerIO and LLAP achieved 8 mm and 4.6 mm tracking error granularity. However, Strata implemented their model along with the models of FingerIO and LLAP for comparison. In their experimental results with the initial distance of 20 cm, the 2-D tracking errors of Strata, LLAP and FingerIO were 10.1 mm, 19 mm, and 34.7 mm. The drawing surface place for the finger tracking schemes were approximately 5 cm by 5 cm to 10 cm by 10 cm wide.

5.4.2 Stability

P. Melgarejo *et al.*'s model was unstable over the stability factors. They performed cross-validation to evaluate the classification accuracy of their scheme, but did not specify the data collection period nor include experiments on the time stability. Their model was also sensitive to changes in the antenna placement and differences of gesture across users. When the orientation of the device shifted by 5°, the accuracy dropped from 92% to 67%. Besides P. Melgarejo *et al.*, all other schemes did not employ machine learning methods. AudioGest and WiGest took rule-based approach. Thus, they were tolerant to time, location, and users. However, AudioGest was sensitive to the orientation angle of the smartphone (*i.e.,* transceiver) and could only tolerate up to orientation angle below $\frac{\pi}{4}$. The three finger tracking schemes took localization-based approach. Thus, they showed high stability regarding time, location, users, and position.

5.4.3 Robustness

Gesture recognition schemes that use acoustic signals as their signal source have limited interference boundary. Their operation boundary is below 50 cm.

Motion interferences outside this boundary do not significantly affect the scheme's accuracy. Other interfering motions are less likely to occur within 50 cm boundary, so we regard these schemes to be robust against interfering motions. WiGest and WiDraw have much larger interference boundary because they use Wi-Fi signal as their signal source. However, WiGest collected their data in presence of up to seven interfering users in the room with the target user. WiDraw also tested their model in a a busy cafeteria. The average tracking error was approximately 5 to 6 cm, slightly higher than controlled environment.

Three studies that uses acoustic signal (AudioGest, LLAP, and Strata) investigated the effect of interfering signals, but they all concluded that such interferences do not have critical impact on the accuracy of the schemes. They tested their model in environments with loud voices and music but all confirmed that their model was nearly unaffected the noises.

5.4.4 Applicability

No gesture recognition schemes investigated multiuser cases. However, four out of six schemes considered non-line-of-sight cases. WiGest considered various through-the-wall scenarios (up to two walls) in their evaluation. In the whole-home gesture recognition case study, they achieved 96% accuracy for classifying seven gestures across all locations that includes line-of-sight, through-one-wall, and through-two-walls cases. WiDraw, LLAP, and FingerIO placed the receiver inside a bag or a pocket and evaluated the NLOS cases. WiDraw achieved 1-2 cm higher tracking error in the NLOS compared to LOS case. LLAP's measurement error slightly increased by 1.4 mm on average when when the device is inside a bag. FingerIO's median tracking error across all the participants was 1 cm compared to 8 mm when the smartphone was not occluded.

5.4.5 Usability

P. Melgarejo *et al.*'s model was the only scheme that used machine learning and required training samples. In their experiment, they collected 200 samples for each of the 25 ASL gestures and performed 5-fold cross validation. They also conducted an additional experiment on the impact of training samples and confirmed that the accuracy only slightly increased when the training samples were increased up to four times.

Three schemes had a preadjustment process that required user involvement. WiGest detected the beginning of the gesture with the preamble motion (two up-down motions) to maximize energy efficiency. WiDraw required a calibration process that requested users to rotate the laptop (receiver) over the z-axis. This process was necessary to to estimate the azimuth and elevation of an AoA because the MUSIC algorithm can only compute a 2-D AoA with linear antenna array. Strata required users to perform a few trials before performing a gesture to obtain the appropriate parameter values such as the frame detection threshold.

5.4.6 Deployability

All gesture recognition schemes except [46] used COTS Wi-Fi devices or smartphones. P. Melgarejo *et al.* used WARP v3 boards, a RF signal prototyping device to obtain CSI phase information. They also used directional antennas to perform beamforming.

WiGest deployed two APs in the apartment testbed and three APs in the engineering building testbed. With a single AP, they achieved 87.5% accuracy in edge detection. The accuracy increased to 96% when they used three APs. WiDraw required the most number of transmitters among all target schemes. In the experiment, they used 30 neighboring devices as transmitters because higher signal density implies higher granularity. They required at least dozen transmit-

ters to achieve moderate accuracy (median error of 6.4 cm). Four schemes [37–40] required a speaker and microphones to transmit and receive inaudible acoustic signal. However, users did not have to deploy multiple microphones or speakers but instead deploy a single smartphone that is equipped with the microphones and speakers. Thus, we marked such schemes that deploy smartphones or other smart devices as '1 TXRX', which means that the schemes only require deployment of a single device.

5.4.7 Efficiency

Evaluation on the overhead was essenstial for schemes acoustic signal based schemes because smartphones or smart devices that the schemes used to transmit and receive acoustic signals had limited computing power and memory. The measured processing time of LLAP at each interval was 4.3 ms. They also fully charged an iPhone 6s and tested how long their model continuously runs. The results showed that LLAP can continuously run for 10.57 hours, incurring less than 3% CPU time on iOS platform. FingerIO also conducted a similar test on Samsung Galaxy S4; their model lasted four hours. Strata measured the energy consumption of a phone during one hour period. A phone that is not actively running apps consumed 14% battery in an hour. With FingerIO running, the smartphone spent 8% more battery. The average processing time was at each interval was 2.5 ms. AudioGest also used mobile devices (laptop, tablet, and smartphone) as tranceivers, but did not conduct evaluation on the overhead. P. Melgarejo *et al.*'s model ran on a daughterboard, but they still emphasized the light weight complexity of their algorithm (O(25×11) for NN algorithm and O(25) for DTW). They also measured the required memory space for their classification algorithm. Their model required 640 KB of memory space per gesture.

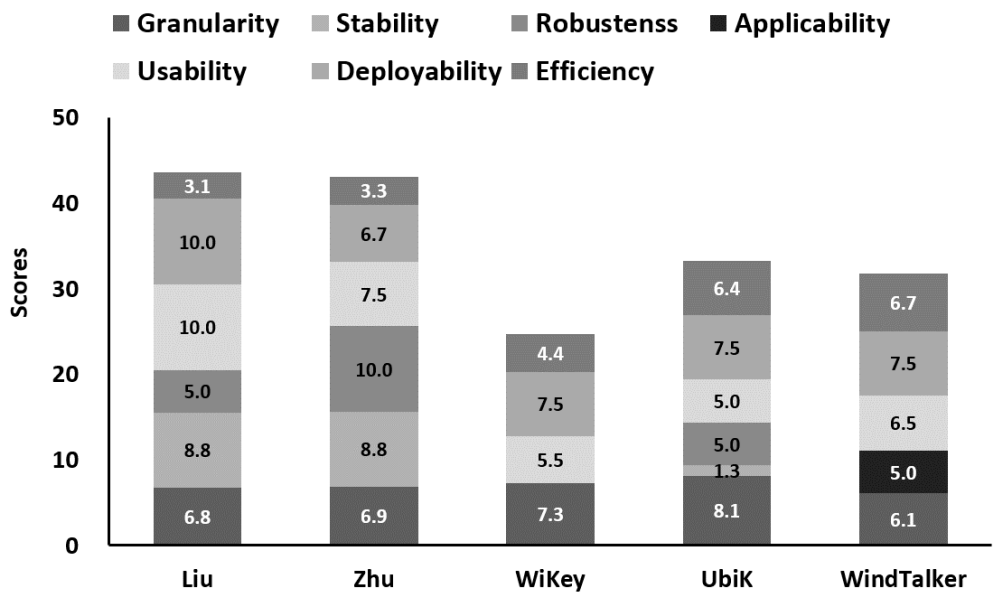

Figure 5.4: Assessment of keystroke recognition schemes

5.5 Keystroke

By localizing the key by the key's acoustic emanations, or distinguishing a unique signal wave generated by the key press, several works have performed keystroke recognitions. Two studies proposed a keystroke recognition method that can replace physical keyboard hardware [6, 9], and three studies proposed a keystroke snooping attacks that infer victim user's keystrokes [7, 8, 10]. Figure 5.4 shows the practicality scores of each scheme.

5.5.1 Granularity

All five keystroke recognition schemes achieved classification level granularity. Four schemes [6–9] recognized computer keyboard keystrokes while [10] extended the idea to the smartphone keystrokes. The number of distinguishable keys were different for all schemes (Table 5.1). The classification accuracy of these schemes were lower than other motion types because the keys are located close to each other.

Table 5.1: Granularity of keystroke recognition schemes

Scheme	Accuracy	Class
Liu *et al.* [8]	85.50%	26 alphabets (a-z)
Zhu *et al.* [7]	72.20%	25 alphabets (a-y) and three special characters (space, enter, and delete)
WiKey [9]	82.87%	26 alphabets (a-z) and 10 digits (0-9)
UbiK [6]	82.9%	all keys except the PC-specific functional keys
WindTalker [10]	73%	10 digits (0-9)

5.5.2 Stability

The two attacks proposed by Liu *et al.* and Zhu *et al.* exhibited high stability because they took a localization method. They used the time difference of the keystroke acoustic emanation arriving at the two microphones to estimate the distance of the key from the microphone. Thus, these two schemes were independent of time, location, and user. However, the orientation of the device could affect the coverage area of the keyboard. Thus, we marked them as '△' in the table. The three remaining schemes [6, 9, 10] all used machine learning approach. WiKey and WindTalker trained the CSI's DWT wavelet. UbiK trained the acoustic sound leveraging multipath fading effect. Thus, they achieved low stability. UbiK mentioned in their work that minor displacement is only tolerable up to the key's edge size. Thus, we marked them as '△'.

5.5.3 Robustness

Liu *et al.* and Zhu *et al.* utilized the arrival times of the signals; nearby motions are not likely to alter the arrival time of the direct path. Thus, we considered them to be robust against interfering motions. On the other hand, UbiK also used an acoustic signal but it utilized the multipath effect, which can be altered by the nearby motion. Thus, we marked them as 'NC'.

Zhu *et al.* and Wang *et al.* [6] conducted experiments in noisy environments

with interfering acoustic signals. Zhu *et al.*'s model achieved at least 64% detection accuracy in a noisy office and meeting room. UbiK's accuracy in a noisy airplane (76.5 dB) was 92.4%.

5.5.4 Applicability

Only WindTalker considered the NLOS case. In their real-world simulation attack, they put the receiver behind the counter to hide it from the victim. Among ten 6-digit passwords, they recovered 2 passwords with top 5 candidate accuracy.

5.5.5 Usability

WiKye required the most number of training samples. They obtained the baseline classification accuracy with 30 training samples per key. WindTalker used much less training samples than WiKey, but WindTalker was an attack – a much more challenging case to obtain user's training samples. Thus, they also performed evaluations on the impact of the number of training samples. However, the results were only to show the shortcomings of their scheme because their recovery rate decreased to 68.3% when one training sample was used.

Zhu *et al.*'s model required a keyboard pre-entering process to reconstruct the keyboard layout. However, we marked them as △ because user did not have to label the keys. Liu *et al.*'s model did not require such a process because they assumed that the keyboard shape and phone placement is known to the attacker. UbiK executed a run-time calibration process. It displayed 5 candidate keys and requested users to click the correct key that the user has pressed. However, they disabled the calibration function when measuring the classification accuracy.

5.5.6 Deployability

The keystroke recognition schemes all used COTS smartphones or Wi-Fi devices. Except for the model of Zhu *et al.*, they all used single transmitter and receiver or a single smartphone. However, Zhu *et al.*'s keystroke snooping attack required an attacker to place three smartphones (at least two) adjacent to user's keyboard in order to conduct the attack.

5.5.7 Efficiency

Three schemes [6–8] used smartphones as the data processing device, but only UbiK have thoroughly investigated the power consumption of their model. They have used the Monsoon power monitor to profile the power cost of their model; the results showed that their model incurs an additional power consumption of 194.7 mW on top of the base power consumption (18.5% of power cost).

5.6 Others

Besides gait, activity, gesture, and keystroke recognitions, there were schemes that analyzed RF-signal to monitor user's breathing pattern or read user's lip shape. The scores of the schemes are shown in Figure 5.5.

5.6.1 Granularity

We categorized the two breathing estimator schemes [12,13] in the counting-level because they counted the number of breaths per minute. UbiBreathe estimated different breathing rates with less than 1 breaths per minute error. H. Wang *et al.*'s contribution came from applying the concept of Fresnel zone in the Wi-Fi CSI model. Instead of measuring the accuracy of correctly detected breaths, they performed a case study provided a figure that illustrates the ground truth respiration rate and the estimated respiration rate.

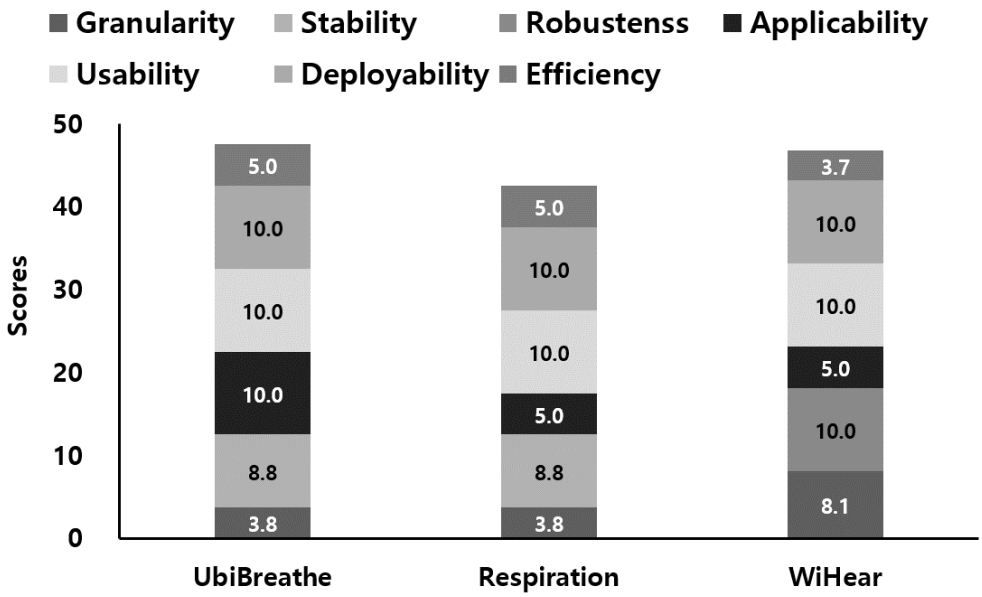

Figure 5.5: Assessment of other recognition schemes

WiHear was designed to identified subtle lip motions to hear people's talks. They selected 33 simple words (*e.g.,* are, you, good) to evaluate the classification accuracy. However, by age of 5, children have an expressive vocabulary of 2,100 to 2,200 [47]. 33 words are far less than the common words humans use to communicate daily.

5.6.2 Stability

UbiBreathe and the respiration detection model of H. Wang *et al.* were stable over time, place, or user. However, their performance was affected by the orientation or position of the device or user. WiHear achieved low stability. They assumed that people do not move when they, and did not investigate on the untrained location and user, nor the stability of their model against time.

5.6.3 Robustness

WiHear explored the influence of other ISM-band interferences and irrelevant human movements. Their evaluation results confirmed that their model is

resistant to ISM band interferences and irrelevant to human motions that are 3-m away from the receiver.

5.6.4 Applicability

All three work considered multiuser cases. However, UbiBreathe and Wang's model could only detect multiple users' breathing rates only if the users breathe with different breathing rates. WiHear supported multiple target's lip reading when they use multiple transceivers. When they deployed three pairs of transceivers and performed lip readings of three users speaking simultaneously, they achieved 74% accuracy with less than 3 spoken words. With more than 6 words, they achieved less than 60% accuracy.

5.6.5 Usability

Only WiHear required a training sample. Their model required a minimum of 50 training samples per word. All three schemes did not require any preadjustment phase.

5.6.6 Deployability

WiHear used a directional antenna that is dedicated for the lip reading purpose. They also required multiple transceivers to perform multiple users' lip reading.

5.6.7 Efficiency

WiHear's lip motion was more fine-grained and required higher sampling rate to capture the detailed lip motion. Thus, compared to UbiBreathe and Wang's model, they used higher sampling rates. UbiBreathe used smartphone and laptop as receivers, which are both the lightweight and relatively strong devices. Thus, we marked their efficiency level as '△.

VI. Findings

In this chapter, we report the findings we earned through the assessment in various perspectives.

6.1 Wi-Fi vs. Acoustic

The signal type is strongly related to the size of the motion (Figure 6.1). This is because of the propagation speed difference of the two signals. Wi-Fi signal travels much faster than acoustic signals. Thus, Wi-Fi signal based schemes could recognize wide spectrum of motion size because the received signal contains information of motions that occur both far and near the receiver. However, acoustic signals were mostly used to recognize hand gestures or keystrokes because they could only capture motions within a meter boundary. For the same reason, acoustic signals were more robust to nearby interferences because their interference boundary was small enough to limit influences of nearby motions achieving higher robustness compared to Wi-Fi signals (Figure 6.2).

All schemes that used acoustic signal as their signal source used portable devices (*e.g.,* smartphones) for their transceivers. Consequently, they had to minimize the performance overhead because of the limited computation power and memory of the devices. Thus, they performed evaluation on the computation power and time overhead more extensively than Wi-Fi signal based schemes. Among five acoustic signal-based schemes that proposed a means for Human-Computer Interaction (HCI) [6,37–40], four of them included evaluation on power and time overhead. However, the overall efficiency scores of the Wi-Fi based schemes and acoustic signal-based schemes did not show much difference because

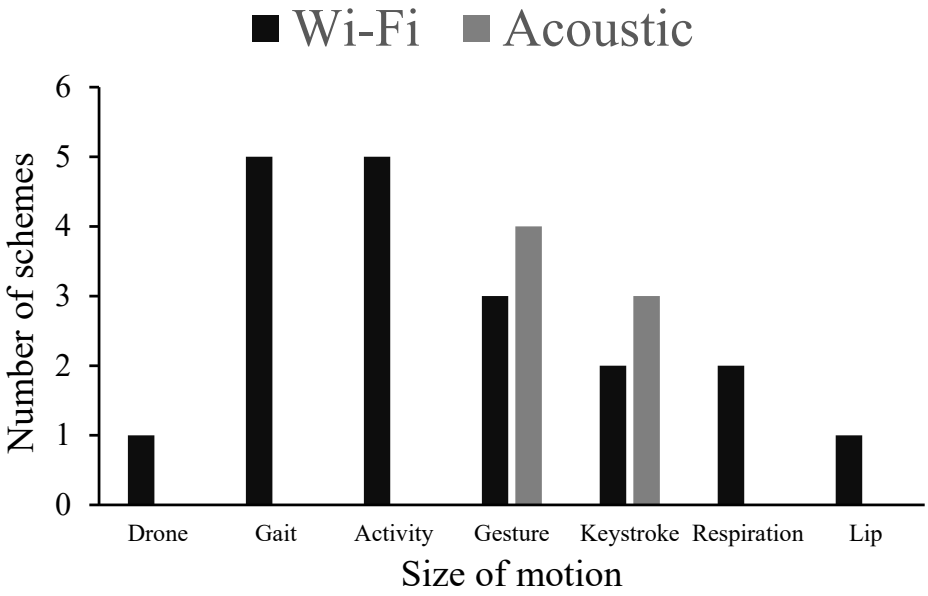

Figure 6.1: The number of proposed schemes for each motion type of different size

acoustic signal-based keystroke snooping attacks did not consider the power, time, or memory overhead although they used smartphones as their data processing device. Further analysis on the difference between services and attacks are explained in Section 6.2.

Acoustic signals also achieved higher granularity, stability, and usability because many schemes took a rule-based approach instead of machine learning (Figure 6.2). Thus, many acoustic signal-based schemes achieved tracking level granularity with high stability while not requiring training samples. They also earned higher scores in deployability because a single device (*i.e.,* smartphone) contained a transmitter and two receivers by itself eliminating the deployment cost.

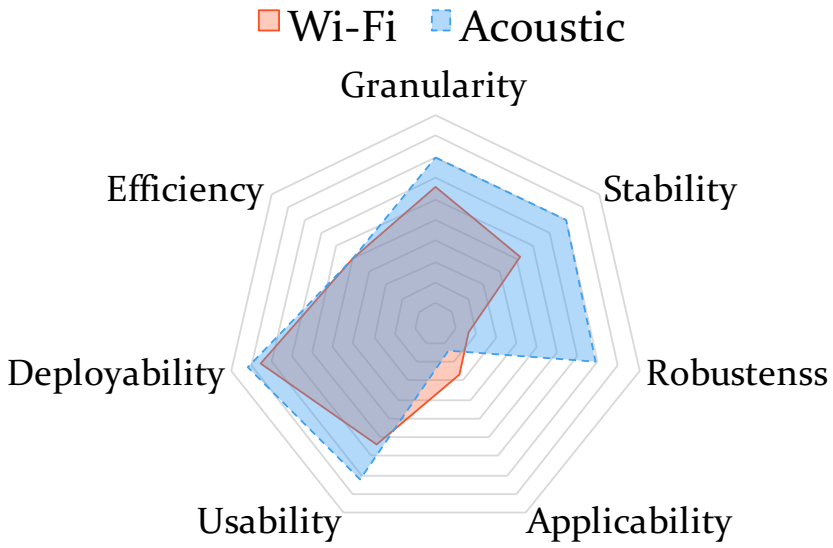

Figure 6.2: The practicality scores of Wi-fi based schemes and acoustic signal-based schemes

6.2 Service vs. Attack

Motion recognition is a double-edged sword. Motion recognition can be used to provide useful services such as a HCI application or activity monitoring tool. On the other hand, an attacker can abuse such techniques to devise an attack and invade other's privacy. Depending on the intent of the scheme of whether the technique is used as a service or an attack, the notion of the seven facets is different.

Efficiency. A scheme that used motion recognition to provide services to users performed more evaluation on the PTM consumption than motion inference attacks. Among seven schemes that used a portable light devices to process their data, five schemes were services and two schemes were attacks. Four out of five services included a section that evaluated the PTM overhead of the model. In contrast, none of the two attacks evaluated the PTM. Perhaps it is because users are likely to neglect a services that imposes too much overhead. In contrast,

attackers are likely to bear the overhead of the system only if the attack can be successful.

NLOS. A scheme that supports NLOS case could be more useful for attackers to covertly conduct the attack because attackers can hide the presence of data collecting devices. Out of four attacks [7, 8, 10, 45], two attacks leveraged the NLOS property. Banerjee *et al.* deployed malicious receivers outside the concrete wall building to infer the line-crossing of the transmitter-receiver link. WindTalker also hid the receiver behind the counter in their experiment so that the victim remains as unnoticed. However, the other two attacks [7, 8] did not consider NLOS cases because the transceivers had to be placed right next to the victim's keyboard.

There were schemes that supported NLOS cases even when the intent was to provide services. In such cases, they were to support whole-home recognition [12, 28, 34], enable inside-pocket recognition [38, 39], or enhance the flexibility in the position of the device [36].

Sampling rate. In a service, a user can control the sampling rate of the transmitter and the receiver. Thus, sampling rate is only a factor that incurs overhead. However, from an attack perspective, an attacker may not be able to control the sampling rate if a legitimate user's device is involved as a transmitter. In such a case, designing a system that requires high sampling rate could not be feasible. Therefore, Banerjee *et al.* had to minimize the sampling down to 10 Hz per second. Yet, WindTalker used a sampling rate of 800 Hz even though it was an attack. This is because both the transmitter and the receiver were attacker's devices that are deployed near the user. Thus, the attacker had full control over the conditions.

Training. Supervised learning is more critical for attack cases because attackers must collect labeled training samples of the victim. Training may be a

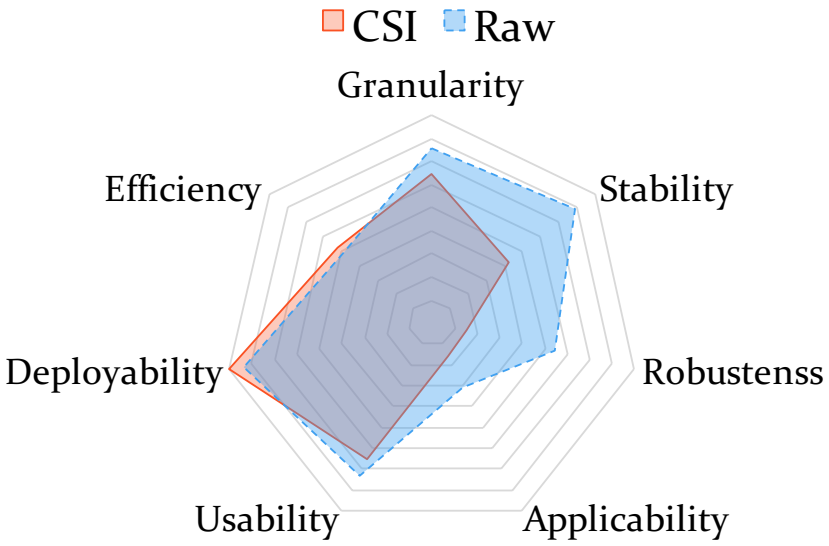

Figure 6.3: The practicality scores of schemes that used Wi-Fi CSI data vs. raw acoustic or Wi-Fi data

cumbersome process even for a motion recognition service, but users may be willing to provide training samples if they can get the service in return. However, not many victims would be willingly provide the training samples just for their privacy to be violated. Thus, three out of four attacks did not use machine learning in their approach. They took a rule-based approach that are not user specific to infer victim's motion. WindTalker used a machine learning approach and used 10 training samples to evaluate the classification accuracy of the keystrokes, but to claim the feasibility of the attack, they also performed evaluations on the impact of the number of training samples. However, the results were only to show the shortcomings of their scheme because their recovery rate decreased to 68.3% when one training sample was used.

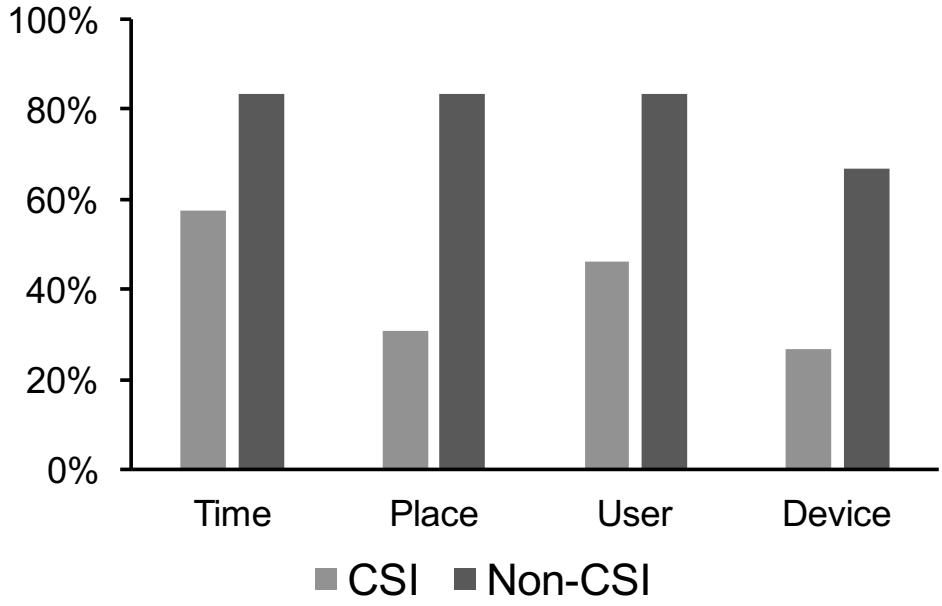

Figure 6.4: The number of proposed schemes for each motion type of different size

6.3 Raw signal vs. CSI

Some schemes used the raw acoustic signal or simulated USRP RF-signal, while the other schemes used the CSI signal from the COTS Wi-Fi devices. Attackers could eliminate the training process by using the physical raw signal because information such as Doppler-shift or phase difference could be extracted from the raw signal. WiSee and AudioGest used raw RF-signal and acoustic signal to obtain the Doppler-shift information. Doppler-shift could distinguish the direction of the movement and did not require machine learning to model different motions. [8, 38–40] and [7] used the raw acoustic signal to obtain the phase difference or apply cross-correlation. They translated the phase difference or the TDoA derived through the cross-correlation into the difference in distance and localized the moving object. Through the localization-based approach, these schemes also could eliminate training and achieve fine-grained granularity. UbiK

Table 6.1: Motion type's influence on the stability of CSI-based schemes

	Scheme	Motion type	Stability score
Rich frequency feature	Wang *et al.* [13]	Respiration	8.75
	FIMD [44]	Gait	10
	Banerjee *et al.* [45]	Gait	8.75
	FreeSense [16]	Gait	2.5
	WiWho [15]	Gait	3.75
	WifiU [14]	Gait	6.25
	WiFall [18]	Falling action	2.5
	CARM [5]	Activity	10
	E-eyes [4]	Activity	2.5
Poor frequency feature	Melgarejo *et al.* [46]	Gesture	0
	WiKey [9]	Keystroke	0
	WindTalker [10]	Keystroke	0
	WiHear [17]	Lip motion	0

was the only work that used an acoustic signal used training to classify different keys. However, researchers could only obtain the physical raw RF-signal through simulation because COTS Wi-Fi devices prohibits access.

Time-series CSI signal was vulnerable to the changes in the environment. Slight changes in the environment changed the propagation paths of the signals, making inconsistent pattern. Figure 6.4 shows the percentage of schemes that achieved the four stability factors. As the figure shows, the schemes that utilized CSI signal exhibited low stability than schemes that utilized the raw signal or RSSI value.

In particular, the schemes that used CSI to classify motions that have less frequency characteristics were all unstable against the stability factors. That is because CSI signal could capture the frequency of the movement but could not capture the direction of the movement. Thus, CSI waveform of motions without much frequency features (mostly quick and short motions) such as keystrokes or lip movements maintained its form only for a short while and easily changed its form even with a slight environmental changes. However, CSI waveform of motions that contains rich frequency features were relatively stable against the

changes (Table 6.1).

On the other hand, WiDraw which used CSI as their signal source could break from machine-learning based classification-level granularity because they used multiple AoAs derived from CSI instead of raw CSI value; their approach achieved tracking level granularity with high stability.

VII. Related Work

Recent surveys have investigated sensor-based or Wi-Fi based motion recognition system. [19] have explored the schemes that perform contactless activity recognition through Wi-Fi signal. They have reviewed the historical overview of the field as well as the theoretical models that serve as the base for the schemes. They also explained the key techniques used in the schemes – the base signal, the preprocessing techniques, and the features used.

[20] also have compared the techniques of the Wi-Fi CSI based behavior recognition schemes. More specifically, they categorized Wi-Fi CSi behavior recognition schemes into two categories: histogram-based schemes, DWT-based schemes, and Deep Learning-based schemes. They explained the processing steps of the schemes and have suggested using Deep Learning for activity recognition. To verify the idea, they've collected CSI data of users performing six activities and applied three different models: random forest, hidden markov model, and long short term memory. They compared the classification accuracy of the three models and confirmed that Deep Learning model outperforms the other two models.

Although the existing surveys provide detailed analysis of the techniques used to achieve activity recognition, they fail to address the shortcomings of existing schemes, the infeasibility and impracticality aspect of the signal based motion recognition methods that hinders adoption in the industry. Furthermore, these surveys have not reflected the progress of recent motion recognition studies that have eliminated machine learning to achieve more fine-grained, stable, and practical system. Our work has investigated on the factors that degrades the practicality of motion recognition schemes. We have proposed seven practicality

facets that can be used to extensively evaluate a scheme's practicality. Moreover, we have assessed more than 20 schemes with the standard we have set, and have drawn meaningful findings we have obtained through the assessments.

VIII. Discussion

8.1 Limitation of our work

8.1.1 Evaluation on the Correctness of the Facets

8.1.2 Not Included Impractical Conditions

Although we have suggested an evaluation standard to assess the practicality of signal-based motion recognition schemes, not all infeasible requirements could be assessed through the standard we have proposed. It was because the requirements were specific to a few schemes, and thus could not be generalized to signal-based motion recognition schemes in general.

We have listed some of the strong and impractical assumptions the target schemes required or missing evaluations that should have been conducted. A. Banerjee *et al.* [45] required legitimate transmitters to transmit packets regularly and frequently. It also required the target user to walk with constant speed of 0.5 m/s and assumed that the attacker knows the location of the packet transmitting devices. E-eyes [4] made an assumption that in-place activities (*e.g.,* cooking) occur at dedicated locations (*e.g.,* kitchen). Thus, they could only distinguish activities that are linked to a location. P. Melgarejo *et al.* [46] performed gesture recognition on a wheelchair or inside a car based on the RSS and signal phase difference. A wheelchair and a car are mobile vehicles; a user using a wheelchair or a car is likely to be mobile as well. However, they did not perform any evaluation on a moving wheelchair on a car. Liu *et al.* [8] and Zhu *et al.* [7] devised a keystroke snooping attack that uses TDoA of keyboard acoustic emanation on the two microphones of the smartphone. However, Liu [8] assumed the attacker knows the keyboard layout and the phone placement; Zhu [7] assumed

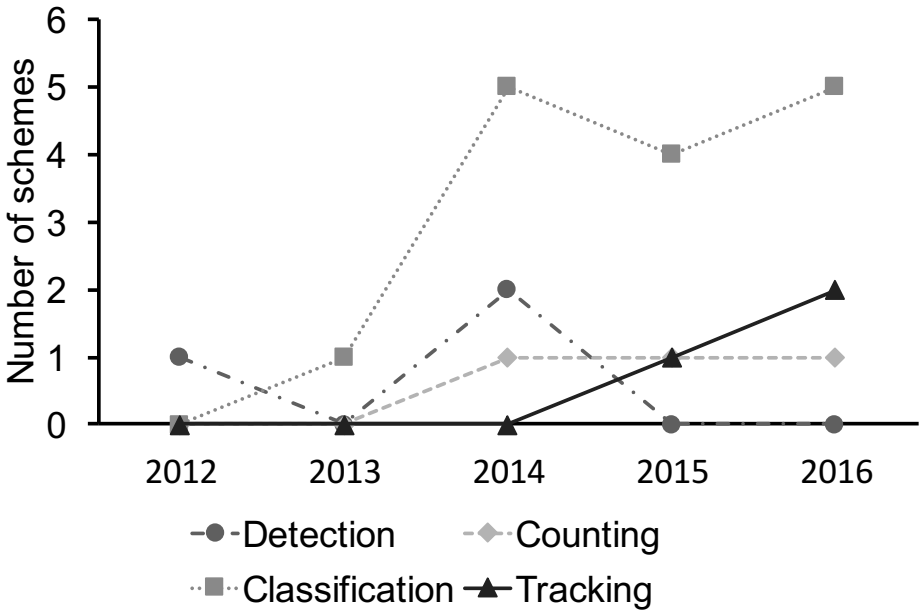

Figure 8.1: Granularity trend of signal-based motion recognition schemes

the attacker knows the phone placement.

8.2 Emerging Trends

Early studies have only considered demonstrating that recognizing certain motions are possible using signal propagation characteristics. Schemes with rather impractical requirements or assumptions have been accepted in top-tier conferences because the idea was novel and unessayed. However, recent work began go beyond demonstrating a possibility and has progressed towards devising a more practical and feasible design. Researchers have moved from using simulated RF-signals to COTS CSI signals and machine-learning to rule-based. Particularly, the granularity of the schemes have progressed from detection-level to classification-level and tracking-level (Figure 8.1). Therefore, proposing a coarser-grained method that has low practicality in terms of the seven practicality facet would be a regression.

8.3 Future Research Directions

Machine learning. Despite the disadvantages of machine learning, recent studies, especially Wi-Fi signal-based studies, still rely on machine learning. Except [36], all tracking-level schemes used acoustic signals as their signal source. Wi-Fi is an attractive source signal to utilize because of their rich presence in today's general environment. Wi-Fi signal will be even more denser in the future due to emerging trends such as IoT that are based on Internet connections. Therefore, a practical Wi-Fi based method that does not rely on training must be devised.

Authentication. Wi-Fi signal based authentication has been only proposed recently. They are at the beginning stage in which they all require users to walk predefined paths, have low stability, robustness, applicability, usability, and efficiency. A practical authentication scheme that satisfies the practicality requirements is imperative.

Modeling. Many previous work has empirically observed a signal pattern to recognize motion. However, future research should go beyond empirical analysis, and go towards accurately modeling the motion. For instance, instead of classifying the wavelet pattern of certain motions, [13] has considered the size of the chest movement (4.2 to 4.3 mm) and modeled how much the CSI value must change according to the change in the wavelength. They also modeled the distance that their receiver can capture the chest movement using the Fresnel Zone. Future research on motion recognition should follow the footsteps of [13] and construct an accurate model of the motion and the signal.

IX. Conclusion

In this thesis, we addressed the impractical requirements and conditions of signal-based human motion recognition methods that are perceived as barriers to the industry. We proposed seven practical facets – *granularity, stability, robustness, applicability, deployability, and efficiency* – that extensively evaluate the practicality of a scheme. We performed evaluation on the previous signal-based human motion recognition research based on the seven facets and unfolded the findings we have earned through the analysis in addition to the emerging trends in the field and prospect on future research directions. We believe our work may serve as a standard for future signal-based human motion recognition research to assess the practicality of their schemes.

요 약 문

신호 기반 모션 인식 기법은 최근 학계에서 많은 주목을 받았음에도 불구하고 시장으로 뻗어나가는 데에는 실패하였다. 우리는 신호 기반 모션 인식 기법들이 가진 비현실적인 요구사항의 한계를 지적하며 일곱 개의 실용성 기준 – 입상도(granularity), 안정성(stability), 강건성(robustness), 응용성(applicability), 배치성(deployability), 효율성(efficiency) – 을 제안하였다. 이에 기초하여 기존에 발표된 시그널 기반 모션 기법들을 비교 분석 및 평가하였고 이를 통해 얻은 발견과 최신 동향 및 장래의 연구 방향을 제시하였다. 실용성을 평가하는 기준을 통해 미래 시그널 기반 모션 인식 연구 방향이 연구를 위한 연구를 넘어서 실질적으로 보급되는 방향으로 나아갈 것을 기대한다.

References

[1] Thomas B. Moeslund and Erik Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81(3):231 – 268, 2001.

[2] Liming Chem, Jesse Hoey, Chris D. Nugent, Diane J. Cook, and Zhiwen Yu. Sensor-based activity recognition. volume 42, pages 790–808, November 2012.

[3] Youngwook Kim and Hao Ling. Human activity classification based on micro-doppler signatures using a support vector machine. volume 47, pages 1328–1337, May 2009.

[4] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. E-eyes: Device-free location-oriented activity identification using fine-grained wifi signatures. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, pages 617–628, 2014.

[5] Wei Wang, Alex X. Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 65–76, 2015.

[6] Junjue Wang, Kaichen Zhao, Xinyu Zhang, and Chunyi Peng. Ubiquitous keyboard for small mobile devices: harnessing multipath fading for fine-grained keystroke localization. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*, pages 14–27, 2014.

[7] Tong Zhu, Qiang Ma, Shanfeng Zhang, and Yunhao Liu. Context-free attacks using keyboard acoustic emanations. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 453–464, 2014.

[8] Jian Liu, Yan Wang, Gorkem Kar, Yingying Chen, Jie Yang, and Marco Gruteser. Snooping keystrokes with mm-level audio ranging on a single phone. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 142–154, 2015.

[9] Kamran Ali, Alex X. Liu, Wei Wang, and Muhammad Shahzad. Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 90–102, 2015.

[10] Mengyuan Li, Yan Meng, Junyi Liu, Haojin Zhu, Yao Liu, and Na Ruan. When csi meets public wifi: Inferring your mobile phone passwords via wifi signals. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 1068–1079, 2016.

[11] Simon Birnbach, Richard Baker, and Ivan Martinovic. Wi-fly?: Detecting privacy invasion attacks by consumer drones. In *Proceedings of the Network and Distributed System Security Symposium*, 2017.

[12] Heba Abdelnasser, Khaled A. Harras, and Moustafa Youssef. Ubibreathe: A ubiquitous non-invasive wifi-based breathing estimator. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 277–286, 2015.

[13] Wei Wang, Alex X. Liu, and Muhammad Shahzad. Gait recognition using wifi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 25–36, 2016.

[14] Hao Wang, Daqing Zhang, Junyi Ma, yasha Wang, Yuxiang Wang, Dan Wu, Tao Gu, and Bing Xie. Human respiration detection with commodity wifi devices: Do user location and body orientation matter? In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 363–373, 2016.

[15] Wiwho: Wifi-based person identification in smart spaces. In *Proceedings of the 15th International Conference on Information Processing in Sensor Networks*, 2016.

[16] Tong Xin, Bin Guo, Zhu Wang, Mingyang Li, Zhiwen Yu, and Xingshe Zhou. Freesense: Indoor human identification with wi-fi signals. In *Proceedings of the 2016 IEEE Global Communications Conference*, 2016.

[17] Guanhua Wang, Yongpan Zhou, Zimu Zhou, Kaishun Wu, and Lionel M. Ni. We can hear you with wi-fi! volume 15, pages 2097–2920, November 2014.

[18] Yuxi Wang, Kaishun Wu, and Lionel M. Ni. Wifall: Device-free fall detection by wireless networks. volume 16, pages 581–594, Feb 2017.

[19] Junyi Ma, Hao Wang, Daqing Zhang, Yasha Wang, and Yuxiang Wang. A survey on wi-fi based contactless activity recognition. In *Proceedings of the 2016 International IEEE Conference on Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*, pages 1086–1091, 2016.

[20] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee. A survey on behavior recognition using wifi channel state information. *IEEE Communications Magazine*, 55(10):98–104, OCTOBER 2017.

[21] Ralph O. Schmidt. Multiple emitter location and signal parameter estimation. In *IEEE Transactions on Antennas and Propagation*, volume 34, pages 276–280, 1989.

[22] H. Liu, H. Darabi, P. Banerjee, and J. Liu. Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(6):1067–1080, Nov 2007.

[23] Victor C. Chen, fanyin Li, Shen-shyang Ho, and Harry Wechsler. Micro-doppler effect in radar: Phenomenon, model, and simulation study. volume 42, pages 2–21, January 2006.

[24] P. Bahl and V. N. Padmanabhan. Radar: an in-building rf-based user location and tracking system. In *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*, volume 2, pages 775–784 vol.2, 2000.

[25] Fadel Adib and Dina Katabi. See through walls with wifi! *SIGCOMM Comput. Commun. Rev.*, 43(4):75–86, August 2013.

[26] Donny Huang, Rajalakshmi Nandakumar, and Shyamnath Gollakota. Feasibility and limits of wi-fi imaging. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*, SenSys '14, pages 266–279, New York, NY, USA, 2014. ACM.

[27] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Fredo Durand. Capturing the human figure through the wall. volume 34, November 2015.

[28] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th annual international conference on Mobile computing and networking*, pages 27–38, 2013.

[29] Ieee standard for information technology– local and metropolitan area networks– specific requirements– part 11: Wireless lan medium access control (mac)and physical layer (phy) specifications amendment 5: Enhancements for higher throughput. *IEEE Std 802.11n-2009 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, and IEEE Std 802.11w-2009)*, pages 1–565, Oct 2009.

[30] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. Tool release: Gathering 802.11n traces with channel state information. *SIGCOMM Comput. Commun. Rev.*, 41(1):53–53, January 2011.

[31] Yaxiong Xie, Zhenjiang Li, and Mo Li. Precise power delay profiling with commodity wifi. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, MobiCom '15, pages 53–64, New York, NY, USA, 2015. ACM.

[32] Wei Xi, Jizhong Zhao, Xiang-Yang Li, Kun Zhao, Shaojie Tang, Xue Liu, and Zhiping Jiang. Electronic frog eye: Counting crowd using wifi. In *Proceedings of the 2014 IEEE International Conference on Computer Communications*, pages 361–369, 2014.

[33] C. E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, Jan 1949.

[34] Heba Abdelnasser, Moustafa Youssef, and Khaled A. Harras. Wigest: A ubiquitous wifi-based gesture recognition system. In *Proceedings of the 2015 IEEE Conference on Computer Communications (INFOCOM)*, 2015.

[35] Wei Xi, Dong Huang, Kun Zhao, Yubo Yan, Yuanhang Cai, Rong Ma, and Deng Chen. Device-free human activity recognition using csi. In *Proceedings of the 1st Workshop on Context Sensing and Activity Recognition*, pages 31–36, 2015.

[36] Li Sun, Souvik Sen, Dimitrios Koutsonikolas, and Kyu-Han Kim. Widraw: Enabling hands-free drawing in the air on commodity wifi devices. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 77–89, 2015.

[37] Wenjie Ruan, Quan Z. Sheng, Lei Yang, Tao Gu, Peipei Xu, and Longfei Shangguan. Audiogest: Enabling fine-grained hand gesture detection by decoding echo signal. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 474–485, 2016.

[38] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. Fingerio: Using active sonar for fine-grained finger tracking. In *Proceedings of the 34th Annual ACM Conference on Human Factors in Computing Systems*, pages 1515–1525, 2016.

[39] Wei Wang, Alex X. Liu, and Ke Sun. Device-free gesture tracking using acoustic signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 82–94, 201.

[40] Sangki Yun, Yi-Chao Chen, Huihuang Zheng, Lili Qiu, and Wenguang Mao. Strata: Fine-grained acoustic-based device-free tracking. In *Proceedings of*

the 15th Annual International Conference on Mobile Systems, Applications, and Services, pages 15–28, 2017.

[41] ISO/IEC. *ISO/IEC 9126. Software engineering – Product quality.* ISO/IEC, 2001.

[42] United States Cencus Bureau. Quickfacts united states, 2016.

[43] NIGEL BEVAN. International standards for hci and usability. *Int. J. Hum.-Comput. Stud.*, 55(4):533–552, October 2001.

[44] Jiang Xiao, Kaishun Wu, Youwen Yi, Lu Wang, and L. M. Ni. Fimd: Fine-grained device-free motion detection. In *2012 IEEE 18th International Conference on Parallel and Distributed Systems*, pages 229–235, 2012.

[45] Arijit Banerjee, Dustin Maas, Maurizio Bocca, Neal Patwari, and Sneha Kasera. Violating privacy through walls by passive monitoring of radio windows. In *Proceedings of the 2014 ACM Conference on Security and Privacy in Wireless & Mobile Networks*, WiSec '14, pages 69–80, New York, NY, USA, 2014. ACM.

[46] Pedro Melgarejo, Xinyu Zhang, Parameswaran Ramanathan, and David Chu. Leveraging directional antenna capabilities for fine-grained gesture recognition. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 541–551, 2014.

[47] Susie Loraine. Vocabulary development, 2008.

Acknowledgements

석사 기간동안 많이 부족한 저를 지켜봐주시고 열정으로 이끌어주신 김종 교수님께 큰 감사드립니다. 교수님의 지도를 통해 성장할 수 있었습니다. 훌륭한 교수님 밑에서 연구할 수 있음을 감사하게 생각합니다.

HPC 연구실 동료들에게도 논문을 빌어 감사의 마음을 전합니다. 함께 먹은 수많은 맛있는 음식과 통나무집, 배드민턴, 생일파티, 일본 사가 워크숍은 절대로 잊지 못할 소중한 추억입니다. 많이 그리울 것입니다.

또 저의 소중한 인연들에게도 감사합니다. 그런 인연들이 있어서 웃을 일이 없는 날에도 웃을 수 있었습니다. 마지막으로 항상 응원해주시고 힘이 되어주시는 가족에게도 감사의 마음을 전합니다.

Curriculum Vitae

Name : Hayoung Jeong

Education

2011. 3. – 2016. 2. Department of Computer Science and Electrical Engineering, Handong University (B.S.)

2016. 3. – 2018. 2. Department of Computer Science and Engineering, Pohang University of Science and Technology (M.S.)


POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY
1986
포항공과대학교