Wireless Systems that Extend Our Senses

Fadel Adib

Wireless Systems that Extend Our Senses

by

Fadel Adib

M.S. in Electrical Engineering and Computer Science, Massachusetts Institute of Technology (2013)
B.Eng. in Computer & Communications Engineering, American University of Beirut (2011)

Submitted to the
Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

February 2017

Author . . . .
Department of Electrical Engineering and Computer Science
November 1, 2016

Certified by . . . .
Dina Katabi
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by . . . .
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Theses

Wireless Systems that Extend Our Senses

by

Fadel Adib

Submitted to the Department of Electrical Engineering and Computer Science
on November 1, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Wireless signals, such as Wi-Fi, are traditionally used for communications. In this thesis, we show that these signals can also be used as sensing tools that enable us to learn about our environment without physically reaching out to the various objects in it. Specifically, as these signals travel in the medium, they traverse occlusions like walls and bounce off different objects and humans before arriving at a receiver; hence, they carry information about the environment. This thesis presents algorithms and software-hardware systems that extract this information to deliver a variety of new sensing capabilities.

We deliver four fundamental contributions: We present the first design that uses Wi-Fi signals to see through walls, enabling us to detect people behind walls by relying purely on the reflections of Wi-Fi signals off their bodies. Next, we demonstrate how we can use radio frequency (RF) reflections to track people's 3D locations and gestures in indoor environments without requiring them to wear or carry any devices. Beyond localizing people, we introduce the first system that can recover human silhouettes through walls; the captured silhouettes enable us to track the 3D positions of human limbs and body parts and to distinguish between different people behind a wall. Finally, we show how smart environments can monitor their inhabitants breathing and heart rates by relying purely on how the human body modulates reflected RF signals.

To deliver these contributions, we exploit physical properties of RF signals, work across software-hardware boundaries, and introduce new systems and new algorithms that require redesigning the entire computing stack, from the hardware to the applications. We implement and evaluate these systems demonstrating how they can enable many new real-world applications including baby monitoring, elderly fall detection, non-invasive vital sign tracking, gesture control, and human identification through walls.

Thesis Supervisor: Dina Katabi
Title: Professor of Electrical Engineering and Computer Science

To Salam, My Guardian Angel

Acknowledgments

When I sent an email to Professor Dina Katabi as an undergraduate student seeking a summer research internship under her supervision at MIT, I did not imagine that, 6 years later, I would be accompanying her to the White House to demo our research on live TV to the President of the United States. Part of our demo was to measure a person's heart rate from a distance without any body contact. Since I was the test subject for the demo, the President could see that my heart rate was over 110. What the President could not see, however, is that my entire PhD journey under Dina's supervision was as magical and as surreal as the moment my heart rate overshot 110 during that demo. Throughout this journey, Dina's role transcended that of an advisor. She believed in me and inspired me. She fought battles for me and by my side, all the while treating me as her friend. I shall always remain indebted to Dina, and I hope that I can fight by and for my students and inspire them as she has done for me.

I am likewise massively indebted to my amazing faculty collaborators without whom much of this research would not have been possible [22, 24, 26, 72]. Rob Miller invariably provided fascinating insights in directing my research to solve real-world problems. I vividly recall when I told Rob that I finally managed to localize people from wireless reflections; at the time, he asked me whether I could extend that capability to 3D so that we can detect elderly falls. Frédo Durand's enthusiasm for 'cool and crazy research' was absolutely contagious, and I thoroughly enjoyed our discussions on bridging RF with computer vision and graphics applications. I learned much from Piotr Indyk through working closely with him on QuickSync [72] and through his day-to-day insights about academia.

And, while I have not had the fortune to collaborate with Martin Rinard (yet), he has served as an insightful mentor at various points of my PhD journey.

I am profoundly grateful to my amazing thesis committee: Dina, Kevin Fu, Sanjay Sarma, John Guttag, and Victor Bahl. Beyond providing feedback on my thesis, they have been tremendously supportive throughout my entire PhD and faculty application process. They took calls from me at all hours of the day (and night!), patiently listened to my concerns, and shared with me invaluable advice.

Much credit for the research in this dissertation goes to my phenomenal student collaborators. Zach was my partner-in-crime for over two years, as we worked incessantly on building and debugging our systems. Hongzi stayed up with me many sleepless nights in the lab as we worked on Vital-Radio. And, Chen-Yu was a powerhouse throughout our collaboration on RF-Capture. I am very fortunate to have learnt from my other coauthors, who were fantastic collaborators, mentors, and mentees: Haitham Al-Hassanieh, Ross Knepper, Swarun Kumar, Nate Kushman, Jue Wang, and Mingmin Zhao.

I could not have wished for a more supportive and fun group of colleagues than my labmates at NETMIT and CSAIL: Omid, Ezz, Deepak, Jue, Lixin, Steph, Rahul, Haitham, Swarun, Nate, Mingmin, Jouya, Arthur, Diego, Peter, Abe, and Nabeel. I am indebted to them for unwaveringly volunteering in hundreds upon hundreds of experiments and for providing me with feedback on numerous iterations of my papers and talks. I owe many thanks to our group admin, Mary, for taking many burdens off our shoulders – with equipment, travel, food and any last-minute emergencies – and taking care of us before our deadlines.

◇

I'm blessed with an extraordinary group of friends who have been there for me over the past five years, cheering me up on my darkest days and celebrating with me every success. Haitham somehow managed to be a mentor, a friend, and an older brother all at once. Jad took flights and bus trips to visit me and support me as if international travel to Boston was a weekend trip home; I am blessed to have shared with him every step of my PhD journey even though he was in another country. Ozzy was a source of much joy and happiness for me during some of my most stressful times. I know I must have done one thing right when I see myself surrounded by my wonderful group of friends: Merhi, Bardan, Uzra, Assil, Deeni, Muneeza, Shadab, and Raafat.

Finally, yet most importantly, words cannot express my gratitude towards my parents, Salam and Fawaz, my sister and confidante, Hayat, and my brother(-in-law), Ahmad, for their perpetual support, guidance, sacrifices, and love. And in a final word, transcendent to gratitude, I turn to my mother, Salam, my torch of guidance, inspiration, and motivation: thank you for believing in me when I lost all belief in myself; this thesis is dedicated to you.

Previously Published Material

Chapter 3 revises a previous publication [25]: Fadel Adib and Dina Katabi. See Through Walls with Wi-Fi! ACM SIGCOMM 2013.

Chapter 4 revises a previous publication [24]: Fadel Adib, Zachary Kabelac, Dina Katabi, and Robert C. Miller. 3D Tracking via Body Radio Reflections. Usenix NSDI 2014.

Chapter 5 revises a previous publication [23]: Fadel Adib, Zachary Kabelac, and Dina Katabi. Multi-Person Localization via RF Body Reflections. Usenix NSDI 2015.

Chapter 6 revises a previous publication [22]: Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. Capturing the Human Figure Through a Wall. ACM SIGGRAPH Asia 2015.

Chapter 7 revises a previous publication [26]: Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C. Miller. Smart Homes That Monitor Breathing and Heart Rate. ACM CHI 2015.

Contents

List of Figures

List of Tables

Chapter 1
Introduction

This thesis explores the use of wireless signals to sense the human body and the environment. Our techniques all leverage the basic physical principles of radio frequency (RF) signals. As RF signals travel in indoor environments, they reflect off different objects – including the human body – before arriving at a receiver. Hence, they carry information about the environment. In this dissertation, I present techniques that extract information from the reflected RF signals to solve a variety of sensing problems and deliver new sensing capabilities.

Specifically, I design systems that send wireless signals, capture their reflections from the environment, and analyze these reflections. Based purely on how human motion interacts with these wireless signals, I show how to 1) accurately localize people in indoor environments, 2) track human gestures in mid-air and enable users to control smart environments through these gestures, 3) extract human breathing and heart pulses from a distance, and 4) achieve sufficient resolution (from wireless signals) to recover a human silhouette from behind a wall.

One of the key advantages of my sensing approach is that it does not require any physical contact with the human body. This enables us to address major societal challenges on multiple fronts. On the health front, noncontact sensing of vital signs – like breathing and heart rate – would enable us to monitor neonates and premature babies, whose sensitive skin gets damaged with traditional electrodes. It also allows us to monitor the health progress of infectious disease (e.g., Ebola) patients without requiring physical contact with these patients, which can be threatening to the caregiver's life. On the energy front, our ability to track people's locations indoors would enable us to tune heating and cooling in order to reduce energy consumption. More generally, non-contact sensing of the human body enables many applications in health-care, computer graphics, ubiquitous computing, surveillance, and user interaction. Finally, even though in this thesis I focus on the human body, my techniques can generalize to sense the entire environment using wireless signals.

■ 1.1 Wireless Signals as a Sensing Modality

In spirit, this research advances a long-standing desire to sense humans and the environment through various means including ultrasound, radar, terahertz, and vision techniques. The advantage of using RF signals, however, is that they traverse occlusions, enabling us to sense the human body through walls while leveraging low-cost massively-produced RF components. The key challenge in extracting semantics from these signals is that we live in a sea of RF waves which interact with each other and with the environment in complex ways. These signals bounce off all walls, furniture, and humans indoors, resulting in a plethora of reflections. These reflections are typically referred to as the "multi-path effect"; they are particularly challenging in indoor environments since reflectors are all close to each other. The problem of sensing human bodies using RF is exacerbated by two additional challenges: first, human body reflections are very weak, and second, our bodies are not rigid and their reflective surfaces change with motion.

Overcoming the above challenges requires designing new algorithms that exploit a model of human motion and the human body in indoor environments. For example, to prevent multi-path reflections from introducing errors, we design Hidden Markov Models that account for practical constraints on the velocity of human motion. Furthermore, to factor in the deformability of the human body with motion, we introduce a coarse skeletal model that recovers a human silhouette by regularizing measurements over time, exploiting the radar cross section of different human body parts.

Beyond algorithmic innovations, addressing the above challenges requires designing software-hardware systems that allow fine-grained control over the structure of the transmitted and received RF signals. For example, to sense breathing and heartbeats, I built radio front-ends that can isolate such minute movements from other sources of motion in the environment. Additionally, to scale my designs for tracking multiple people, I deployed and coordinated multiple radios at very fine time granularity; in order to achieve this coordination, I implemented new tight synchronization techniques at both the hardware and software layers.

The combination of these algorithms and hardware-software systems enables us to sense the human body using wireless signals. Our designs have three key properties that make them particularly powerful: First, they make no assumption about the environment they operate in and can effectively deal with multipath without any prior knowledge of that environment. Second, these systems are built using resources that are available to consumers, such as Wi-Fi and other low-power wireless technologies. Third, they are all implemented in practice; these implementations and are empirically evaluated to demonstrate their feasibility and high accuracy.

■ 1.2 Systems Developed

This dissertation chronicles the evolution of these technologies starting with the most basic capability we first developed: the ability to simply sense the presence of a person behind a wall using RF signals without placing a sensor on his or her body. As the thesis progresses, I describe how we were able to move from that basic capability to being able to capture a human figure - i.e., coarse skeleton - through occlusions like walls, and accurately track a person's breathing and heart rate without placing any sensor on his or her body. Below, I highlight the contributions of each of these systems in chronological order of their development:

1. **Detecting Humans through Walls with Wi-Fi:** I present Wi-Vi, the first system that uses Wi-Fi signals to detect people through walls without any device on their bodies and by relying purely on Wi-Fi reflections. The primary challenge in designing Wi-Vi comes from reflections from the wall itself, which are 10,000-100,000x stronger than the reflections from humans behind the wall. To overcome this challenge, our key insight is to treat the wall's reflection as interference in the context of sensing humans. This insight enables us to

(a) Person Moving in a Room

(b) WiTrack's Output

Figure 1-1: **Through-Wall Localization from RF Reflections.** WiTrack localizes people by relying purely on RF signals reflected off their bodies. (a) shows a person who is moving in a room different than the room WiTrack is placed in; the person does not wear or carry any wireless device. (b) shows the output of WiTrack; the red dot indicates where WiTrack localizes the person at the corresponding point in time. The spiral on the ground is only there to show the level of accuracy. Full video is available at: https://www.youtube.com/watch?v=sbFZPPC7REc

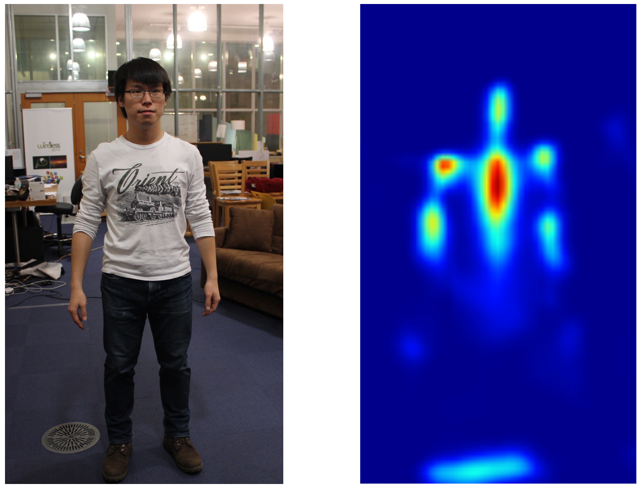
adapt techniques from MIMO (multiple-input multiple-out) communications for interference cancellation. Specifically, we encode our transmitted Wi-Fi signals across multiple antennas to cancel interference from the wall's reflection, as well as from all static objects in the environment. This enables us to sense minute reflections from a person's body behind the wall. In Chapter 3, I elaborate on this technique, and prove that our iterative algorithm for cancelling static reflectors (including the wall) converges exponentially fast, enabling us to detect people behind a wall in real time. I further present a prototype implementation of Wi-Vi and demonstrate that it can detect movements of people through different wall materials, including concrete, by relying purely on Wi-Fi reflections off their bodies.

2. **3D Localization from Human Body Reflections:** This dissertation also demonstrates how we can accurately localize the human body based purely on RF reflections, even if the person is behind a wall; a sample output of our system is shown in Fig. 1-1. RF-based localization is a classical research area in wireless networking, spanning over two decades of academic research [30, 44, 86, 119, 119, 130, 160, 170]. Traditional localization techniques require a user to hold or wear a wireless device. In contrast, I present WiTrack, the first system that enables 3D localization of the human body in indoor environments without requiring users to hold or wear any device. Moreover, WiTrack achieves the

same (centimeter-scale) accuracy as prior state-of-the-art localization systems which require users to wear or carry devices. WiTrack opens up RF localization to numerous new applications where it is either infeasible or inconvenient for a user to hold or wear a device including elderly fall detection, intrusion detection, smart homes, and search & rescue missions.

The fundamental challenge in designing WiTrack arises from the recursive nature of reflections in indoor environments, whereby each person's reflections bounce off all furniture and building structures – potentially multiple times – before arriving at a receiver. Unlike perfectly static reflections addressed in Chapter 3, these secondary and tertiary reflections are dynamic and move with the human body resulting in a "ghosting effect". To overcome this challenge, we observe that a person's direct reflection arrives earlier in time than indirect reflections; furthermore, indirect reflections vary significantly when measured from different vantage points indoors. Building on these observations, Chapter 4 presents our design of a software-controlled time-of-flight RF sensor that enables us to tease apart the various reflections spatially. Additionally, it elaborates on how we analyze these time-of-flight measurements to learn and identify human reflections and discard non-human reflections; our analysis incorporates Hidden Markov Models that account for practical constraints on the continuity of human motion. Chapter 5 describes how we can move from single-person to multi-user localization through an iterative stochastic algorithm that is inspired by successive interference cancellation, an algorithm widely used in wireless communications. It also presents a prototype implementation and evaluation which demonstrates that we can achieve centimeter-scale localization by relying purely on RF reflections off the human body.

3. **Capturing a Human Figure through Walls:** Beyond localizing the human as a single point in space, this thesis demonstrates how we can capture a human figure – i.e., a coarse skeleton – through occlusions from RF reflections, as shown in Fig. 1-2. Capturing the skeleton of a human body enables many applications in computer graphics, ubiquitous computing, surveillance, and user interaction [61, 62, 71, 116, 125, 127, 132, 153, 154]. Existing approaches for skeletal acquisition, like the Xbox Kinect, either require a user to stay within the devices' line-of-sight and cannot track him across rooms, or they require the subject to wear on-body sensors. In contrast, I present RF-Capture, the first system that can capture



(a) Human Subject **(b) Captured Figure**

Figure 1-2: **Capturing a Human Silhouette Through Walls.** We use RF signals in order to capture a human figure (i.e., coarse skeleton) through walls. (a) shows a person who is occluded by a wall from our sensors. (b) shows the output of our sensors; the output shows the person's captured skeleton in the form of a heatmap, where the background is in navy blue and the different body parts are in red, orange, and yellow. Full video is available at: https://www.youtube.com/watch?v=7LTr02cJkiA

a coarse human skeleton through walls, without requiring the subject to wear any sensor.

The key challenge in designing RF-Capture is that human body parts are specular when imaged with signals whose frequencies traverse walls – i.e., human body parts act as pure mirrors at these frequencies. As a result, due to the laws of reflections, RF sensors capture only a subset of human body parts and cannot recover a human silhouette. Our solution to this problem exploits the fact that due to human motion, consecutive RF snapshots tend to expose different body parts and diverse perspectives of the same body part. By exploiting the radar cross section of different body parts and incorporating a coarse model of the human body, we show in Chapter 6 how we can regularize these measurements over time to capture a coarse skeleton with the person's head, chest, arms, and feet. We further demonstrate how capturing such a skeleton enables us to distinguish between different subjects behind a wall and even trace a person's hand as he writes large letters in the air.

4. **Smart Homes that Monitor Breathing and Heart Rate:** Finally, this dissertation demonstrates how we can enable smart environments to accurately monitor a person's breathing and heart rate without any body contact. A sample output of our system – when it is used to monitor the vital signs of a sleeping baby – is shown in Fig. 1-3. To provide this capability, we exploit the fact that wireless signals are affected by motion in the environ-

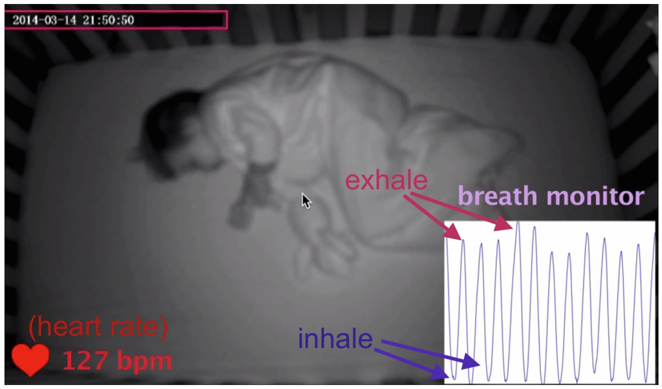


Figure 1-3: **Vital Sign Monitoring with Wireless Signals.** Our system captures the breathing and heart-rate of a baby in realtime without body contact. Full video is available at: https://www.youtube.com/watch?v=3Atky2Jt_-4

ment, including chest movements due to inhaling and exhaling and bodily vibrations due to heartbeats.

The key challenge in sensing these movements is that breathing and heart rate cause millimeter-scale variations, and hence can be easily overshadowed by any other source of movement in the environment. To overcome this challenge, we build on our earlier chapters on RF-localization and skeletal capture. Specifically, we re-formalize our localization technique as a filter – i.e., we use localization as a filter to isolate the reflected signals arriving from different locations in the environment, enabling us to boost the SINR (signal-to-interference-and-noise ratio) of each of the sources of motion. Chapter 7 elaborates on this reformulation and describes our algorithms to classify vital-sign-induced movements and discard other movements. It also presents our prototype implementation and demonstrates how we can accurately track users' breathing and heart rates based purely on RF reflections of their bodies, even when these users are 8 meters away from the device or in a different room.

■ 1.3 Beyond the Developed Systems

In this thesis, I focus on using wireless signals to sense the human body and enable new applications. The presented techniques generalize along multiple dimensions. First, my algorithmic contributions are not tied to a specific implementation technology. For example, the multi-antenna cancellation algorithm – which I use in Chapter 3 to cancel a wall's reflection – can be used to increase the dynamic range of a variety of (phase-based) imaging modalities. Second, my systems' abilities to sense extend beyond the human body to any moving object indoors. For example, I have successfully experimented with tracking an iRobot Create robot. Additionally, I have successfully used my vital-sign monitoring technique (of Chapter 7) to sense the breathing of a turkey embryo through its egg shell.

Fundamentally, this dissertation explores the use of wireless signals as a sensing modality to solve a variety of societal and technological challenges. Nonetheless, the synergies I identify between RF systems and a number of computer science communities (Vision, Graphics, HCI, etc.) extend beyond the problems I address in this thesis. This dissertation lays the foundation for an on-going exploration, and I believe that as our understanding of wireless signals in these contexts evolves, the sensing capabilities will expand and grow.

CHAPTER 2

Background & Related Work

The desire for using wireless signals to sense the environment dates back to the development of military radar and sonar systems in WWII [134]. These systems were originally designed for detecting and tracking large metallic objects in open spaces, like airplanes in the sky or tanks on the ground [36, 94].

Over the past two decades, there has been increasing interest in using military radar in urban warfare for detecting humans through walls [10, 55, 84, 135, 156, 171]. Multiple advances have been made for enabling these systems to detect humans through dense walls like concrete. However, the state-of-the-art system in this space, which was developed by the MIT Lincoln Lab [40, 121], consists of a large (8-foot long) antenna array that is mounted on a truck. The system transmits high-power signals in military spectrum to detect people behind a wall. It detects humans as blobs moving in a dim background (see the video at [11] for a reference).

The research presented in this dissertation builds on this past literature and is also motivated by a desire to sense humans using wireless signals. However, bringing these technologies to everyday life and delivering new consumer applications raises new challenges, which we address in this thesis:

1. **Indoor Multipath:** The first challenge arises from indoor multipath. Specifically, in indoor environments, wireless signals reflect off all objects in the environment, not just the human

body, including secondary and tertiary reflections off multiple objects. All of these reflections arrive at a receiver, making it hard to associate incoming reflections with a person in the environment. As a result, past work on see-through-wall radar has been mostly tested in an open field with an erected wall [40, 121]. Work tested on human subjects in indoor environments outputs a radar image with blobs in the environment [84, 171]; this includes commercial products like Xaver-100, Xaver-400, Xaver-800, and Range-R [79]. Academic research has shown that these systems suffer from a well known problem called the "ghosting effect" – i.e., the appearance of many ghost humans due to multipath [32, 121]. This effect does not prevent these systems from detecting humans, but precludes them from distinguishing between humans and multi-path reflections and hence curbs their ability to localize humans in real indoor environments. In contrast, in this dissertation, we introduce algorithms that enable us to deal with indoor multipath and accurately localize humans indoors. Our algorithms account for practical constraints on human motion, which are incorporated into Hidden Markov Models (e.g., Kalman filters), as we describe in Chapter 4. We also design stochastic techniques that allow us to track multiple people by learning how human reflections evolve in indoor environments over time, as we describe in Chapter 5.

2. **Human Body Deformation:** Another major challenge this research addresses arises from the deformation of a human body due to motion. Specifically, as humans move, their body parts move in a loosely coupled manner and their reflective surfaces change due to motion. Such behavior is different from standard radar targets, like airplanes and tanks, which have rigid bodies. Without addressing body deformation, one cannot reconstruct a human figure or track human limbs to enable gesture control. This is why past systems that aimed at reconstructing a human body demonstrated their results on a doll covered with foil and required an antenna array larger than the imaged object [181]. In contrast, our designs incorporate skeletal models that factor in the deformation of the human body; this enables us to capture the human figure and track human gestures. Furthermore, we miniaturize our design into a compact array about twice the size of a Kinect (in Chapter 6), as opposed to a large array that is of the size of the human body.

3. **Consumer Resources:** Finally, most past radar systems were designed for the military. These systems typically transmit high-power in military reserved spectrum [40, 171, 176].

In contrast, our designs are operate within FCC regulations for consumer electronics. This enables small, cheap devices that are available to ordinary users.

Furthermore, our designs extend beyond addressing these challenges to delivering full-fledged systems. Our systems can automatically discover the number of humans in the environment, identify intervals in time corresponding to the desired human movements (e.g., gestures/breathing/heartbeats), and discard the undesired movements, preventing them from introducing estimation errors.

Besides radar systems, past work on **thermal and infrared imaging** techniques also shares our desire in extending human vision beyond the visible electromagnetic range, allowing us to detect objects in the dark or in smoke. These techniques operate by capturing infrared or thermal energy reflected off the first obstacle in line-of-sight of their sensors. However, cameras based on these technologies cannot see through walls because they have very short wavelengths (few m to sub-mm) [145], unlike our designs which employ signals whose wavelengths are few cm and hence can traverse walls.[1]

Beyond prior literature which shares our vision in using wireless signals to sense humans, there is more focused related work that deals with individual applications and subcomponents of our different systems. We summarize that work below and describe it in details in the corresponding chapters:

- **Indoor Localization:** The topic of indoor localization has received much attention from the wireless networking community over the past two decades [30, 44, 86, 119, 119, 120, 130, 159, 160, 170]. Past work focused on localizing networked devices using their wireless signals. Thus, in order to localize users, these past systems required the users to wear or carry wireless devices. In contrast, this dissertation shows how we can localize users by relying purely on the reflections of RF signals off their bodies. Prior to our work, the networking community made few attempts to localize or detect people even when they don't carry a wireless device. These proposals either require deploying dozens to hundreds of sensors and detect users as they cross a link between these sensors affecting their signal strength [165], or require an extensive calibration phase during which a person stands in all

[1]The longer the wavelength of an electromagnetic wave is, the lower its attenuation is [140]. Infrared and thermal imaging devices employ signals whose wavelengths are very close to visible light; hence, they do not penetrate building materials such as wood or concrete.

possible locations in the area of interest to create a database of signal measurements [129]. Our designs are intrinsically different from these systems. While these past proposals rely on mapping between current measurements and a database, our designs directly extract the exact location from signal reflections. In §4.1, we elaborate on past literature on RF-localization in the networking community.

- **Capturing Human Skeleton:** Recent work in the computer graphics community has proposed techniques for seeing around corners. These proposals image hidden objects by using light that bounces off corner reflectors in a scene [74, 90, 150]. In contrast, our work uses RF signals that traverse occlusions rather than image around them. As such, our designs do not require the placement of corner reflectors in the scene. Furthermore, unlike these past systems, we do not require the hidden shape to be fully static during the acquisition time, and hence evaluate our systems on real human subjects. In §6.1, we describe these past proposals in details and and elaborate on our differences with their techniques.

- **Noncontact Vital Sign Monitoring:** Recently, the mounting interest in technologies for well-being has led researchers to investigate approaches for noncontact vital sign monitoring. However, past proposals either require placing a sensor (e.g., accelerometer, RF sensor, etc.) in very close proximity to the user's chest [57, 58, 59, 96] or have to ensure that there is no other source of motion in the environment (e.g., breathing of other humans, fans, etc.) [6, 28, 52, 57, 58, 59, 88, 113, 179]. Additionally, recent work proposed using vision-based techniques for noncontact vital sign monitoring; however, these techniques require the person to face a camera which analyzes video feeds to monitor subtle changes (e.g., in color) due to heartbeats [31, 169]. In contrast, our research enables us to monitor a person's breathing and heart rate even when the person is behind a wall or in a completely different room. In §7.1, we elaborate on the evolution of these noncontact vital sign monitoring systems.

Finally, the research presented in this dissertation has spurred significant follow-up work in the networking and mobile systems communities, and led to the exploration of various wireless systems for numerous tasks including gesture recognition [21, 85, 117], passive localization [164, 172], activity recognition [133, 162, 163], and breath monitoring [99, 157].

Chapter 3

Sensing Motion Through Walls with Wi-Fi

For many years humans have fantasized about X-ray vision and played with the concept in comic books and sci-fi movies. This chapter explores the potential of using Wi-Fi signals and recent advances in MIMO communications to build a device that can capture the motion of humans behind a wall and in closed rooms. Law enforcement personnel can use the device to avoid walking into an ambush, and minimize casualties in standoff and hostage situations. Emergency responders can use it to see through rubble and collapsed structures. Ordinary users can leverage the device for gaming, intrusion detection, privacy-enhanced monitoring of children and elderly, or personal security when stepping into dark alleys and unknown places.

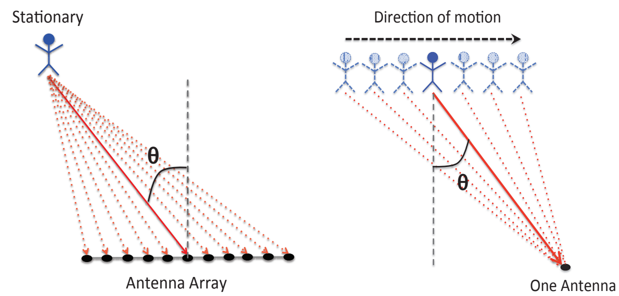
The concept underlying seeing through opaque obstacles is similar to radar and sonar imaging. Specifically, when faced with a non-metallic wall, a fraction of the RF signal would traverse the wall, reflect off objects and humans, and come back imprinted with a signature of what is inside a closed room. By capturing these reflections, we can image objects behind a wall. Building a device that can capture such reflections, however, is difficult because the signal power after traversing the wall twice (in and out of the room) is reduced by three to five orders of magnitude [41]. Even more challenging are the reflections from the wall itself, which are much stronger than the reflections from objects inside the room [41, 121]. Reflections off the wall overwhelm the receiver's analog to digital converter (ADC), preventing it from registering the minute variations due to reflections from objects behind the wall. This behavior is called the "Flash Effect" since it is analogous to how a mirror in front of a camera reflects the camera's flash and prevents it from capturing objects in the scene.

So how can one overcome these difficulties? The radar community has been investigating these issues, and has recently introduced a few ultra-wideband systems that can detect humans moving behind a wall, and show them as blobs moving in a dim background [121, 173] (see the video at [11] for a reference). Today's state-of-the-art system requires 2 GHz of bandwidth, a large power source, and an 8-foot long antenna array (2.4 meters) [40, 121]. Apart from the bulkiness of the device, blasting power in such a wide spectrum is infeasible for entities other than the military. The requirement for multi-GHz transmission is at the heart of how these systems work: they separate reflections off the wall from reflections from the objects behind the wall based on their arrival time, and hence need to identify sub-nanosecond delays (i.e., multi-GHz bandwidth) to filter the flash effect.[1] To address these limitations, an initial attempt was made in 2012 to use Wi-Fi to see through a wall [42]. However, to mitigate the flash effect, this past proposal needs to install an additional receiver behind the wall, and connect the receivers behind and in front of the wall to a joint clock via wires [42].

The objective of this chapter is to enable a see-through-wall technology that is low-bandwidth, low-power, compact, and accessible to non-military entities. To this end, the chapter introduces WiVi,[2] a see-through-wall device that employs Wi-Fi signals in the 2.4 GHz ISM band. WiVi limits itself to a 20 MHz-wide Wi-Fi channel, and avoids ultra-wideband solutions used today to address the flash effect. It also disposes of the large antenna array, typical in past systems, and uses instead a smaller 3-antenna MIMO radio.

So, how does WiVi eliminate the flash effect without using GHz of bandwidth? We observe that we can adapt recent advances in MIMO communications to through-wall imaging. In MIMO, multiple antenna systems can encode their transmissions so that the signal is nulled (i.e., sums up to zero) at a particular receive antenna. MIMO systems use this capability to eliminate interference to unwanted receivers. In contrast, we use nulling

[1]Filtering is done in the analog domain before the signal reaches the ADC.
[2]WiVi stands for Wi-Fi Vision.



(a) Antenna Array **(b) ISAR**

Figure 3-1: **A Moving Object as an Antenna Array.** In (a), an antenna array is able to locate an object by steering its beam spatially. In (b), the moving object itself emulates an antenna array; hence, it acts as an inverse synthetic aperture. WiVi leverages this principle in order to beamform the received signal in time (rather than in space) and locate the moving object.

to eliminate reflections from static objects, including the wall. Specifically, a WiVi device has two transmit antennas and a single receive antenna. WiVi operates in two stages. In the first stage, it measures the channels from each of its two transmit antennas to its receive antenna. In stage 2, the two transmit antennas use the channel measurements from stage 1 to null the signal at the receive antenna. Since wireless signals (including reflections) combine linearly over the medium, only reflections off objects that move between the two stages are captured in stage 2. Reflections off static objects, including the wall, are nulled in this stage. In §3.3, we refine this basic idea by introducing iterative nulling, which allows us to eliminate residual flash and the weaker reflections from static objects behind the wall.

Second, how does WiVi track moving objects without an antenna array? To address this challenge, we borrow a technique called inverse synthetic aperture radar (ISAR), which has been used for mapping the surfaces of the Earth and other planets. ISAR uses the movement of the target to emulate an antenna array. As shown in Fig. 3-1, a device using an antenna array would capture a target from spatially spaced antennas and process this information to identify the direction of the target with respect to the array (i.e., θ). In contrast, in ISAR, there is only one receive antenna; hence, at any point in time, we capture a single measurement. Nevertheless, since the target is moving, consecutive measurements in time emulate an inverse antenna array – i.e., it is as if the moving human is imaging the WiVi device. By processing such consecutive measurements using standard antenna array beam steering, WiVi can identify the spatial direction of the human. In §3.4.2, we extend this method to multiple moving targets.

Additionally, WiVi leverages its ability to track motion to enable a through-wall gesture-based communication channel. Specifically, a human can communicate messages to a WiVi receiver via gestures without carrying any wireless device. We have picked two simple body gestures to refer to "0" and "1" bits. A human behind a wall may use a short sequence of these gestures to send a message to WiVi. After applying a matched filter, the message signal looks similar to standard BPSK encoding (a positive signal for a "1" bit, and a negative signal for a "0" bit) and can be decoded by considering the sign of the signal. The system enables law enforcement personnel to communicate with their team across a wall, even if their communication devices are confiscated.

We built a prototype of WiVi using USRP N210 radios and evaluated it in two office buildings. Our results are as follows:

- WiVi can detect objects and humans moving behind opaque structural obstructions. This applies to $8''$ concrete walls, $6''$ hollow walls, and $1.75''$ solid wooden doors.
- A WiVi device pointed at a closed room with $6''$ hollow walls supported by steel frames can distinguish between 0, 1, 2, and 3 moving humans in the room. Computed over 80 trials with 8 human subjects, WiVi achieves an accuracy of 100%, 100%, 85%, and 90% respectively in each of these cases.
- In the same room, and given a single person sending gesture-based messages, WiVi correctly decodes all messages performed at distances equal to or smaller than 5 meters. The decoding accuracy decreases to 75% at distances of 8 meters, and the device stops detecting gestures beyond 9 meters. For 8 volunteers who participated in the experiment, on average, it took a person 8.8 seconds to send a message of 4 gestures.
- In comparison to the state-of-the-art ultra-wideband see-through-wall radar [121], WiVi is limited in two ways. First, replacing the antenna array by ISAR means that the angular resolution in WiVi depends on the amount of movement. To achieve a narrow beam the human needs to move by about 4 wavelengths (i.e., about 50 cm). Second, in contrast to [121], we cannot detect humans behind concrete walls thicker than $8''$. This is due to both the much lower transmit power from our USRPs and the residual flash power from imperfect nulling. On the other hand, nulling the flash removes the need for GHz bandwidth. It also removes clutter from all static reflectors, rather than just one wall. This includes other walls in the environments as well as furniture inside and outside the imaged room. To reduce clutter, the empirical results in past work are typically collected using a person-height standing wall, positioned either outdoors or in large empty indoor

spaces [121, 173]. In contrast, our experiments are in standard office buildings with the imaged humans inside closed fully-furnished rooms.

■ 3.1 Related Work

WiVi is related to past work in two major areas:

Through-wall radar. Interest in through-wall imaging has been surging for about a decade [10]. Earlier work in this domain focused on simulations [122, 156] and modeling [136, 137]. Recently, there have been some implementations tested with moving humans [42, 121, 173]. These past systems eliminate the flash effect by isolating the signal reflected off the wall from signals reflected off objects behind the wall. This isolation can be achieved in the time domain, by using very short pulses (less than 1ns) [10, 174] whereby the pulse reflected off the wall arrives earlier in time than that reflected off moving objects behind it. Alternatively, it may be achieved in the frequency domain by using a linear frequency chirp [41, 121]. In this case, reflections off objects at different distances arrive with different tones. By analog filtering the tone that corresponds to the wall, one may remove the flash effect. These techniques require ultra-wide bandwidths (UWB) of the order of 2 GHz [41, 174]. Similarly, through-wall imaging products developed by the industry [10, 12] hinge on the same radar principles, requiring multiple GHz of bandwidth and hence are targeted solely at the military.

As a through-wall imaging technology, WiVi differs from all the above systems in that it requires only few MHz of bandwidth and operates in the same range as Wi-Fi. It overcomes the need for UWB by leveraging MIMO nulling to remove the flash effect.

Researchers have recognized the limitations of UWB systems and explored the potential of using narrowband radars for through-wall technologies [123, 124]. These systems ignore the flash effect and try to operate in presence of high interference caused by reflections off the wall. They typically rely on detecting the Doppler shift caused by moving objects behind the wall. However, the flash effect limits their detection capabilities. Hence, most of these systems are demonstrated either in simulation [122], or in free space with no obstruction [89, 101]. The ones demonstrated with an obstruction use a low-attenuation standing wall, and do not work across higher attenuation materials such as solid wood or concrete [123, 124]. WiVi shares the objectives of these devices; however, it introduces a new approach for eliminating the flash effect without wideband transmission. This enables it to work with concrete walls and solid wood doors, as well as fully closed rooms.

The only attempt which we are aware of that uses Wi-Fi signals in order to see through walls was made in 2012 [42]. This system required both the transmitter and a reference receiver to be inside the imaged room. Furthermore, the reference receiver in the room has to be connected to the same clock as the receiver outside the room. In contrast, WiVi can perform through-wall imaging without access to any device on the other side of the wall.

Gesture-based interfaces. Today, commercial gesture-recognition systems – such as the Xbox Kinect [16], Nintendo Wii [9], etc. – can identify a wide variety of gestures. The academic community has also developed systems capable of identifying human gestures either by employing cameras [105] or by placing sensors on the human body [47, 87]. Prior work has also leveraged narrowband signals in the 2.4 GHz range to identify human activities in line-of-sight using micro-Doppler signatures [89]. WiVi, however, presents the first gesture-based interface that works in non-line-of-sight scenarios, and even through a wall, yet does not require the human to carry a wireless device or wear a set of sensors.

3.2 WiVi Overview

WiVi is a wireless device that captures moving objects behind a wall. It leverages the ubiquity of Wi-Fi chipsets to make through-wall imaging relatively low-power, low-cost, low-bandwidth, and accessible to average users. To this end, WiVi uses Wi-Fi OFDM signals in the ISM band (at 2.4 GHz) and typical Wi-Fi hardware.

WiVi is essentially a 3-antenna MIMO device: two of the antennas are used for transmitting and one is used for receiving. It also employs directional antennas to focus the energy toward the wall or room of interest.[3] Its design incorporates two main components: 1) the first component eliminates the flash reflected off the wall by performing MIMO nulling; 2) the second component tracks a moving object by treating the object itself as an antenna array using a technique called inverse SAR.

WiVi can be used in one of two modes, depending on the user's choice. In mode 1, it

[3]Directional antennas have a form factor on the order of the wavelength. At Wi-Fi frequencies, this corresponds to approximately 12 cm.

can be used to image moving objects behind a wall and track them. In mode 2, on the other hand, WiVi functions as a gesture-based interface from behind a wall that enables humans to compose messages and send them to the WiVi receiver.

In sections §3.3- §3.5, we describe WiVi's operation in detail.

3.3 Eliminating the Flash

In any through-wall system, the signal reflected off the wall, i.e., the flash, is much stronger than any signal reflected from objects behind the wall. This is due to the significant attenuation which electromagnetic signals suffer when penetrating dense obstacles. Table 3-1 shows a few examples of the one-way attenuation experienced by Wi-Fi signals in common construction materials (based on [3]). For example, a one-way traversal of a standard hollow wall or a concrete wall can reduce Wi-Fi signal power by 9 dB and 18 dB respectively. Since through-wall systems require traversing the obstacle twice, the one-way attenuation doubles, leading to an 18-36 dB flash effect in typical indoor scenarios.

Building Materials	2.4 GHz
Glass	3 dB
Solid Wood Door 1.75 inches	6 dB
Interior Hollow Wall 6 inches	9 dB
Concrete Wall 18 inches	18 dB
Reinforced Concrete	40 dB

Table 3-1: One-Way RF Attenuation in Common Building Materials at 2.4 GHz [3].

This problem is exacerbated by two other parameters: First, the actual reflected signal is significantly weaker since it depends both on the reflection coefficient as well as the cross-section of the object. The wall is typically much larger than the objects of interest, and has a higher reflection coefficient [41]. Second, in addition to the direct flash caused by reflections off the wall, through-wall systems have to eliminate the direct signal from the transmit to the receive antenna, which is significantly larger than the reflections of interest. WiVi uses interference nulling to cancel both the wall reflections and the direct signal from the transmit to the receive antenna, hence increasing its sensitivity to the reflections of interest.

3.3.1 Nulling to Remove the Flash

Recent advances show that MIMO systems can pre-code their transmissions such that the signal received at a particular antenna is cancelled [66, 142]. Past work on MIMO has used this property to enable concurrent transmissions and null interference [95, 118]. We observe that the same technique can be tailored to eliminate the flash effect as well as the direct signal from the transmit to the receive antenna, thereby enabling WiVi to capture the reflections from objects of interest with minimal interference.

At a high level, WiVi's nulling procedure can be divided into three phases: initial nulling, power boosting, and iterative nulling, as shown in Alg. 1.

Initial Nulling. In this phase, WiVi performs standard MIMO nulling. Recall that WiVi has two transmit antennas and one receive antenna. First, the device transmits a known preamble x only on its first transmit antenna. This preamble is received at the receive antenna as $y = h_1 x$, where h_1 is the channel between the first transmit antenna and the receive antenna. The receiver uses this signal in order to compute an estimate of the channel $\hat{h_1}$. Second, the device transmits the same preamble x, this time only on its second antenna, and uses the received signal to estimate channel $\hat{h_2}$ between the second transmit antenna and the receive antenna. Third, WiVi uses these channel estimates to compute the ratio $p = -\hat{h_1}/\hat{h_2}$. Finally, the two transmit antennas transmit concurrently, where the first antenna transmits x and the second transmits px. Therefore, the perceived channel at the receiver is:

$$h_{res} = h_1 + h_2 \left(-\frac{\hat{h_1}}{\hat{h_2}} \right) \approx 0 \tag{3.1}$$

In the ideal case, where the estimates $\hat{h_1}$ and $\hat{h_2}$ are perfect, the received signal h_{res} would be equal to zero.

Hence, by the end of this phase WiVi has eliminated the signals reflected off all static objects as well as the direct signal from the transmit antennas to the receive antenna. If no object moves, the channel will continue being nulled. However, since RF reflections combine linearly over the medium, if some object moves, its reflections will start showing up in the channel value.

Power Boosting. Simply nulling static reflections, however, is not enough because the signals due to moving objects behind the wall are too weak. Say, for example, the flash

1 Pseudocode for WiVi's Nulling

INITIAL NULLING:
▷ Channel Estimation
Tx ant. 1 sends x; Rx receives y; $\hat{h_1} \leftarrow y/x$
Tx ant. 2 sends x; Rx receives y; $\hat{h_2} \leftarrow y/x$
▷ Pre-coding: $p \leftarrow -\hat{h_1}/\hat{h_2}$
POWER BOOSTING:
Tx antennas boost power
Tx ant. 1 transmits x, Tx ant. 2 transmits px concurrently
ITERATIVE NULLING:
$i \leftarrow 0$
repeat
  Rx receives y; $h_{res} \leftarrow y/x$
  if i even **then**
    $\hat{h_1} \leftarrow h_{res} + \hat{h_1}$
  else
    $\hat{h_2} \leftarrow \left(1 - \frac{h_{res}}{\hat{h_1}}\right) \hat{h_2}$
  $p \leftarrow -\hat{h_1}/\hat{h_2}$
  Tx antennas transmit concurrently
  $i \leftarrow i + 1$
until Converges

effect was 30 to 40 dB above the power of reflections off moving objects. Even though we removed the flash effect, we can hardly discern the signal due to moving objects since it will be immersed in the receiver's hardware noise. Thus, we next boost the transmitted signal power.[4] Note that because the channel has already been nulled, i.e., $h_{res} \approx 0$, this increase in power does not saturate the receiver's ADC. However, it increases the overall power that traverses the wall, and, hence, improves the SNR of the signal due to the objects behind the wall.

Iterative Nulling. After boosting the transmit power, residual reflections which were below the ADC quantization level become measurable. Such reflections from static objects can create significant clutter in the tracking process if not removed. To address this issue, WiVi performs a procedure called iterative nulling. At a high level, the objective is simple: we need to null the signal again after boosting the power to eliminate the residual reflections from static objects. The challenge, however, is that at this stage, we cannot separately estimate the channels from each of the two transmit antennas since, after nulling, we only receive a combined channel. We also cannot remove the nulling and re-estimate the channels, because after boosting the power, without nulling, the ADC would saturate.

[4] In our USRP implementation, we boost the power by 12 dB. This value is limited by the need to stay within the linear range of the USRP transmitter. After nulling, we can also boost the receive gain without saturating the receiver's ADC. On average, we null 42 dB of the signal, which allows a large boost in the receive gain.

However, WiVi can leverage the fact that errors in the channel estimates are much smaller than the channel estimates themselves, and use this observation to refine its estimates. Specifically, by assuming that the estimate for h_2 is accurate (i.e., $\hat{h_2} = h_2$), Eq. 3.1 is left with only one unknown variable h_1. By solving for this unknown variable, we obtain a better estimate of h_1. In particular, the new estimate $\hat{h_1}'$ is:

$$\hat{h}_1' = h_1 = h_{res} + \hat{h_1} \tag{3.2}$$

Similarly, by assuming that the estimate for h_1 is accurate (i.e., $\hat{h_1} = h_1$), we can solve Eq. 3.1 for a finer estimate for h_2:

$$\hat{h}_2' = h_2 = \left(1 - \frac{h_{res}}{\hat{h_1}}\right)\hat{h_2} \tag{3.3}$$

Therefore, WiVi iterates between these two steps to obtain finer estimates for both h_1 and h_2, until the two estimates $\hat{h_1}$ and $\hat{h_2}$ converge. This iterative nulling algorithm converges exponentially fast. In particular, in the appendix, we prove the following lemma:

Lemma 3.3.1. *Assume that* $|\frac{\hat{h_2}-h_2}{h_2}| < 1$, *then, after* i *iterations*, $|h_{res}^{(i)}| = |h_{res}^{(0)}||\frac{\hat{h_2}-h_2}{h_2}|^i$

A few points are worth noting about WiVi's procedure to eliminate the flash effect:

- Besides removing the wall's reflection, it also removes reflections received from other stationary objects both in front of and behind the wall, such as the table on which the radio is mounted, the floor, the radio case itself, etc. In addition, it removes the direct signal from the transmitting antennas to our receive antenna. Note that the direct channels between WiVi's transmit antennas and its receive antenna are significantly attenuated because WiVi uses directional transmit and receive antennas focused towards the wall (and away from the direct path).
- WiVi's nulling algorithm provides a 42 dB mean reduction in signal power, as shown in §3.6.6. This reduction is sufficient to remove the flash effect from a wide range of wall structures including solid wood doors, 6" hollow walls, and most indoor concrete walls. Further, since WiVi uses directional antennas focused on the imaged wall, the direct signal from the transmit antennas to WiVi's receive antenna is weaker than in typical MIMO systems, and becomes negligible after nulling.
- Nulling can be performed in the presence of objects moving behind the wall; it can also

be performed in the presence of objects moving in front of the wall as long as they are outside the field of view of WiVi's directional antennas. Because nulling is mathematically equivalent to subtraction, the presence of such moving objects leads to a small additive constant at the output of WiVi after nulling. Such additive constants do not prevent later tracking of moving objects.

3.4 Identifying and Tracking Humans

Now that we have eliminated the impact of static objects in the environment, we can focus on tracking moving objects. We will refer to moving objects as humans since they are the primary subjects of interest for our application; however, our system is general, and can capture other moving bodies.[5] Below, we first explain how WiVi tracks the motion of a single human. We then show how to extend our approach to track multiple moving humans.

3.4.1 Tracking a Single Human

Most prior through-wall systems track human motion using an antenna array. They steer the array's beam to determine the direction of maximum energy. This direction corresponds to the signal's spatial angle of arrival. By tracking that angle in time, they infer how the object moves in space.

WiVi, however, avoids using an antenna array for two reasons: First, in order to obtain a narrow beam and hence achieve a good resolution, one needs a large antenna array with many antenna elements. This would result in a bulky and expensive device. Second, since WiVi eliminates the flash effect using MIMO nulling, adding multiple receive antennas would require nulling the signal at each of them. This would require adding more transmit antennas, thus making the device even bulkier and more expensive.

To capture the benefits of an antenna array while avoiding its drawbacks, WiVi leverages a technique called inverse synthetic aperture radar (ISAR). ISAR exploits the movement of the target to emulate an antenna array. Existing systems which use antenna arrays capture the signal reflected off a target from spatially spaced antennas and processes this

[5] For example, we have successfully experimented with tracking an iRobot Create robot.

information to identify the direction of the target with respect to the array. In contrast, in ISAR, there is only one receive antenna; hence, at any point in time, the receiver captures a single measurement. However, as the target moves, he/she samples the received signal at successive locations in space, as if we had a receive antenna at each of these points. Furthermore, because of channel reciprocity, successive time samples received by WiVi correspond to successive spatial locations of the moving target. Hence, WiVi effectively receives in time what an antenna array would receive in space. By treating consecutive time samples as spatial samples, WiVi can emulate an antenna array and use it to track motion behind the wall.

In what follows, we formalize the above discussion. Let $y[n]$ be the signal sample received by WiVi at a discrete time point n. Define the **spatial angle** θ as the angle between the line connecting the human to WiVi and the normal to the motion, as shown in Fig. 3-1(b). Note that the sign of θ is positive when the vector from the human to WiVi and the vector of the motion are in the same direction, and negative when these two vectors are in opposite directions.

We are interested in computing $A[\theta, n]$, a function that measures the signal along the spatial direction θ at time n. To compute this value, WiVi first processes the received samples to remove the effect of the transmitted signal, and obtain the channel as a function of time, i.e., $h[n] = y[n]/x[n]$. To emulate an antenna array of size w, WiVi considers w consecutive channel measurements $h[n] \ldots h[n+w]$, as shown in Fig. 3-2. WiVi then computes $A[\theta, n]$ by applying standard antenna array equations [139] as follows:

$$A[\theta, n] = \sum_{i=1}^{w} h[n+i] e^{j \frac{2\pi}{\lambda} i \Delta \sin\theta}, \tag{3.4}$$

where λ is the wavelength, and Δ is the spatial separation between successive antennas in the array.[6] At any point in time n, the value of θ that produces the highest value in $A[\theta, n]$ will correspond to the direction along which the object is moving.

To compute $A[\theta, n]$ from the above equation, we need to estimate Δ, the antenna spacing in the emulated array. Since human motion emulates the antennas in the array, $\Delta = vT$, where T is WiVi's sampling period, and v is the velocity of the motion. Of course, WiVi

[6] Δ is twice the one-way separation to account for the round-trip time.

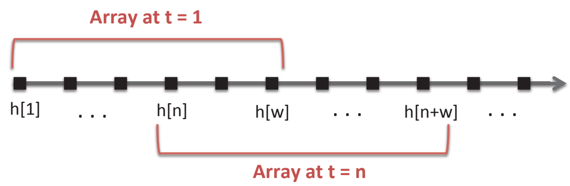
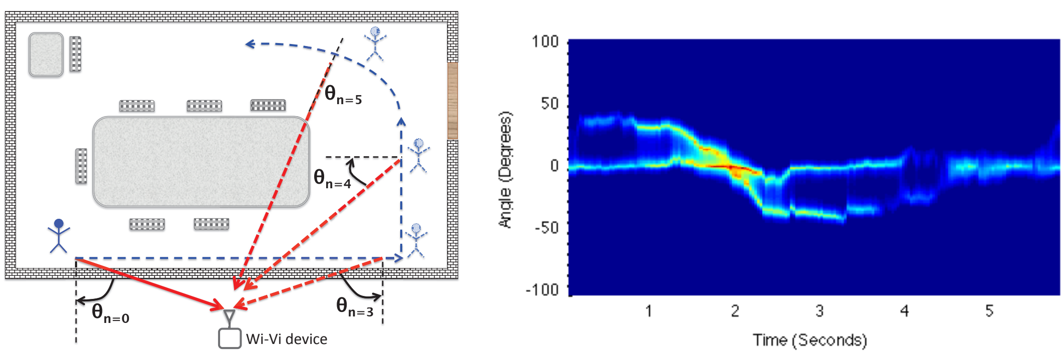


Figure 3-2: **Time samples as Antenna Arrays.** WiVi groups consecutive time samples into overlapping windows of size w, then treats each window $h[n]\ldots h[n+w]$ as an antenna array. This allows it to track the direction of a moving object with respect to the receiver.



(a) **Experimental Setup** (b) **WiVi's output**

Figure 3-3: **WiVi tracks a single person's motion.** (a) shows the experimental setup of a trial which consisted of a single person moving around in a conference room. (b) shows how WiVi is able to track the motion of the person by computing the variation of the inverse angle of arrival with time, i.e. $A'[\theta, n]$ for θ in $[-90°, 90°]$.

does not know the exact speed at which the human is moving. However, the range of speeds that humans have in a confined room is fairly narrow. Hence, we can substitute a value for v that matches comfortable walking (our default is $v = 1$m/s [37]). Note that errors in the value of v translate to an underestimation or an overestimation of the exact direction of the human.[7] Errors in velocity, however, do not prevent WiVi from tracking that the human is moving closer (i.e., angle is positive) or moving away from the WiVi device (angle is negative). In other words, because we do not know the exact v, we cannot pinpoint the location of the human, but we can track her/his relative movements.

Fig. 3-3 shows results from one of our experiments. In particular, 3-3(a) shows a diagram of the movement, and 3-3(b) plots the magnitude of $A[\theta, n]$ (in dB) as a heat map. There are two lines in Fig. 3-3(b): the first one is a zero line, which represents the DC (i.e., the average energy from static elements).[8] This line is present regardless of the number

[7]For example, in one of our experiments, WiVi estimated the human's direction of motion at 30° when the actual direction was 40° but she was moving at a speed around 1.2m/s

[8]Recall that nulling mitigates these reflections so that they do not saturate the receiver's ADC, enabling

of moving objects. Second, there is a curved line with a changing angle. This line tracks the human motion. Around $n = 0$ seconds, the person starts moving towards the WiVi device. As a result, the spatial angle θ is positive and decreasing. (It is positive because the vector of motion and the line from the human to WiVi are in the same direction, and it is decreasing because the absolute angle between the normal on the motion and the line from the human to WiVi is getting smaller.) Around $n = 1.8$s, the person crosses in front of the WiVi device, at which time his angle becomes zero. From $n = 1.8$s to $n = 3$s, the person is moving away from WiVi, and hence, his angle is negative. But the absolute value of the angle is decreasing. At $n = 3$, the person turns and starts moving inward, causing the angle to go back toward zero, but the signal becomes weaker as he is now relatively far from the WiVi receiver.[9]

■ 3.4.2 Tracking Multiple Humans

In this section, we show how WiVi extends its tracking procedure to multiple humans. Our previous discussion about using human motion to emulate an antenna array still holds. However, each human will emulate a separate antenna array. Since WiVi has a single antenna, the received signal will be a superposition of the antenna arrays of the moving humans. In particular, instead of having one curved line as in Fig. 3-3(b), at any time, there will be as many curved lines as moving humans at that point in time.

However, with multiple humans, the noise increases significantly. On one hand, each human is not just one object because of different body parts moving in a loosely coupled way. On the other hand, the signal reflected off all of these humans is correlated in time, since they all reflect the transmitted signal. The lack of independence between the reflected signals is important. For example, the reflections of two humans may combine systematically to dim each other over some period of time.

WiVi to register the minute channel variations due to moving objects behind the wall. However, minuscule errors in channel estimates during the nulling phase would still be registered as a residual DC by WiVi.

[9]Interestingly, even when the direction of motion is perpendicular to the line connecting the person to the device, WiVi registers this motion (note how the DC line is much wider at $n = 5$ than at $n = 0$). This is because Eq. 3.4 approximates WiVi as a monostatic radar, i.e., it simplifies the model by assuming all antennas are co-located. A more detailed model that accounts for the fact that the antennas are not completely co-located shows that for a trajectory to be invisible (i.e., coincide with the DC line) two conditions have to hold: (1) the person moves on an ellipse whose foci are the first transmit antenna and the receive antenna, (2) she moves on an ellipse whose foci are the second transmit antenna and the receive antenna. However, the locus of such motion is discontinuous.

The problem of disentangling correlated super-imposed signals is well studied in signal processing. The basic approach for processing such signals relies on the smoothed MUSIC algorithm [131, 170]. Similar to the standard antenna array processing in Eq. 3.4, smoothed MUSIC computes the power received along a particular direction, which we call $A'[\theta, n]$ because it estimates the same function in Eq. 3.4 but in manner more resilient to noise and correlated signals [139].

For a given antenna array $\mathbf{h} = (h[n], \ldots, h[n+w])$ of size w, MUSIC first computes the $w \times w$ correlation matrix $R[n]$:

$$R[n] = E[\mathbf{h}\mathbf{h}^{\mathbf{H}}], \tag{3.5}$$

where H refers to the hermitian (conjugate transpose) of the vector. It then performs an eigen decomposition of $R[n]$ to remove the noise and keep the strongest eigenvectors, which in our case correspond to the few moving humans, as well as the DC value. For example, in the presence of only one human, MUSIC would produce one main eigenvector (in addition to the DC eigenvector). On the other hand, if 2 or 3 humans were present, it would discover 2 or 3 eigenvectors with large eigenvalues (in addition to the DC eigenvector). MUSIC partitions the eigenvector matrix $U[n]$ into 2 subspaces: the signal space $U_S[n]$ and the noise space $U_N[n]$, where the signal space is the span of the signal eigenvectors, and the noise space is the span of the noise eigenvectors. MUSIC then projects all directions θ on the null space, then takes the inverse. This causes the θ's corresponding to the real signals (i.e., moving humans) to spike. More formally, MUSIC computes the power density along each angles θ as:

$$A'[\theta, n] = \frac{1}{\sum_{k=1}^{K} || \sum_{i=1}^{w} e^{-j\frac{2\pi}{\lambda} i \Delta \sin\theta} U_N[n](i,k)||^2}. \tag{3.6}$$

where K is the total number of noise eigenvectors.

In comparison to the conventional MUSIC algorithm described above, smoothed MUSIC performs an additional step before it computes the correlation matrix. It partitions each array $\mathbf{h}$ of size w into overlapping sub-arrays of size $w' < w$. It then computes the correlation matrices for each of these sub-arrays. Finally, it combines the different correlation matrices by summing them up before performing the eigen decomposition. The additional step performed by smoothed MUSIC is intended to de-correlate signals arriving from spatially different entities. Specifically, by taking different shifts for the same antenna array, reflections from different bodies get shifted by different amounts depending on the distance and orientation of the reflector, which helps de-correlating them [131].

Fig. 3-4 shows the result of applying smoothed MUSIC on the signal captured from two moving humans. Similar to Fig. 3-3(b), the y-axis corresponds to the angle, and the x-axis corresponds to time. As before, the zero line corresponds to DC. At any point in time, we see significant energy at two angles (besides the DC). For example, at time $n = 0.5$s, both humans have negative angles and, hence, are moving away from WiVi. Between $n = 1$s and $n = 2$s, only one angle is present. This may be because the other human is not moving or he/she is too far inside the room. Again, from $n = 2$s to $n = 3$s, we see both humans, one moving towards the device and the other moving away (since one has a positive angle while the other has a negative angle).

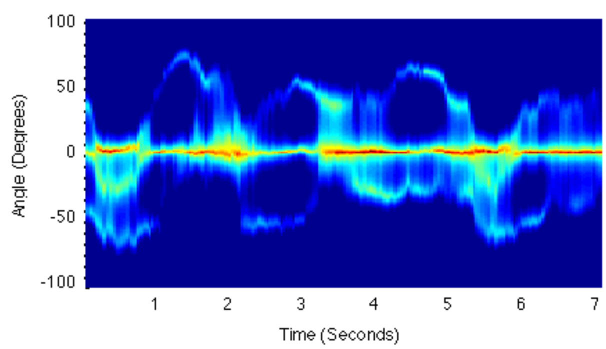


Figure 3-4: **WiVi tracks the motion of two humans.** The figure shows how the presence of two humans translates into two curved lines whose angles vary in time, and one straight line which corresponds to the DC.

One point is worth emphasizing: the smoothed MUSIC algorithm is conceptually similar to the standard antenna array beamforming discussed in §3.4.1; both approaches aim at identifying the spatial angle of the signal. However, by projecting on the null space and taking the inverse norm (as described in Eq. 3.6), MUSIC achieves sharper peaks, and hence is often termed a super-resolution technique [139]. Because smoothed MUSIC is similar to antenna array beamforming, it can be used even to detect a single moving object, i.e., the presence of a single person. In fact, Fig. 3-3(b) was generated by the smoothed MUSIC algorithm.[10]

[10]Plotting the magnitude of $A[\theta, n]$ as opposed to $A'[\theta, n]$ gives the same figure but with more noise. This is because, unlike standard beamforming, the MUSIC algorithm does not incur significant side lobes which would otherwise mask part of signal reflected from different objects.

Finally, to enable WiVi to automatically detect the number of humans in a closed room, one option is to train a machine learning classifier using images like those in Fig. 3-3(b) and Fig. 3-4. We discovered, however, that a simple heuristic based on spatial variance works well in practice. As explained earlier, moving humans appear as curved lines in the 2-D function $A'[\theta, n]$. Any human can be only at one location at any point in time. Thus, at any point in time, the larger the number of humans, the higher the spatial variance. The spatial variance is computed as follows. First, WiVi computes the spatial centroid as a function of time:

$$C[n] = \sum_{\theta=-90}^{90} \theta \cdot 20 \log_{10} A'[\theta, n], \tag{3.7}$$

where $A'[\theta, n]$ is given by Eq. 3.6. It then computes the spatial variance as:

$$VAR[n] = \sum_{\theta=-90}^{90} \theta^2 \cdot 20 \log_{10} A'[\theta, n] - C[n]^2 \tag{3.8}$$

This variance is then averaged over the duration of the experiment to return one number that describes the spatial variance in the room for the duration of the measurement. WiVi uses a training set and a testing set to learn the thresholds that separate the spatial variances corresponding to 0, 1, 2, or 3 humans. The testing and training experiments are conducted in different rooms. In §3.6.4, we evaluate this scheme and measure its ability at automatically capture the number of moving humans.

■ 3.5 Through-Wall Gesture-Based Communication

For a human to transmit a message to a computer wirelessly, she typically has to carry a wireless device. In contrast, WiVi can enable a human who does not carry any wireless device to communicate commands or short messages to a receiver using simple gestures. WiVi designates a pair of gestures as a '0' bit and a '1' bit. A human can compose these gestures to create messages that have different interpretations. Additionally, WiVi can evolve by borrowing other existing principles and practices from today's communication systems, such as adding a simple code to ensure reliability, or reserving a certain pattern of '0's and '1's for packet preambles. At this stage, WiVi's interface is still very basic, yet we believe that future advances in through-wall technology can render this interface more expressive.

Below, we describe the gesture-based communication channel that we implemented with WiVi.

3.5.1 Gesture Encoding

At the transmitter side, the '0' and '1' bits must be encoded using some modulation scheme. WiVi implements this encoding using gestures. One can envision a wide variety of gestures to represent these bits. However, in choosing our encoding we have imposed three conditions: 1) the gestures must be composable – i.e. at the end of each bit, whether '0' or '1', the human should be back in the same initial state as the start of the gesture. This enables the person to compose multiple such gestures to send a longer message. 2) The gestures must be simple so that a human finds it easy to perform them and compose them. 3) The gestures should be easy to detect and decode without requiring sophisticated decoders, such as machine learning classifiers.

Given the above constraints, we have selected the following gestures to modulate the bits: a '0' bit is a step forward followed by a step backward; a '1' bit is a step backward followed by a step forward. This modulation is similar to Manchester encoding, where a '0' bit is represented by a falling edge of the clock, (i.e., an increase in the signal value followed by a decrease,) and a '1' bit is represented by a rising edge of the clock, (i.e., a reduction in signal value followed by an increase) [7]. These gestures are simple, composable and easy to decode as we show in §3.5.2.

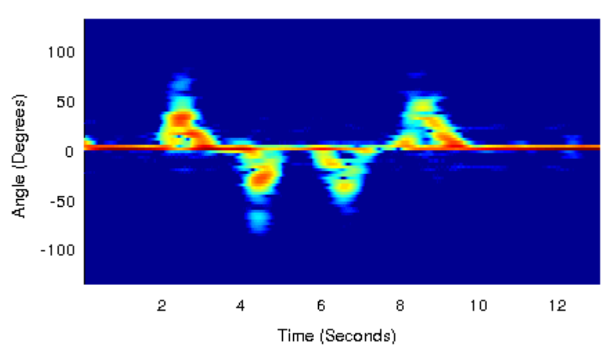


Figure 3-5: **Gestures as detected by WiVi.** The figure shows a sequence of four steps: step forward, step backward, step backward, step forward. Forward steps appear as triangles above the zero line; backward steps appear as inverted triangles below the zero line. Each pair of steps represents a gesture/bit: the first two represent bit '0', the second two represent bit '1'.

Fig. 3-5 shows the signal captured by WiVi, at the output of the smoothed MUSIC algorithm for each of these two gestures. Taking a step forward towards the WiVi device produces a positive angle, whereas taking a step backward produces a negative angle. The exact values of the produced angles depend on whether the human is exactly oriented towards the device. Recall that the angle is between the vector orthogonal to the motion and the line connecting the human to the WiVi device, and its sign is positive when the human is moving toward WiVi and negative when the human moves away from WiVi. As shown in Fig. 3-6, if the human is directly oriented towards the device, the two angles are +90° and -90°. If the human does not know the exact location of the WiVi device and simply steps in its general direction, the absolute value of the angle is smaller, but the shape of the bit is maintained.

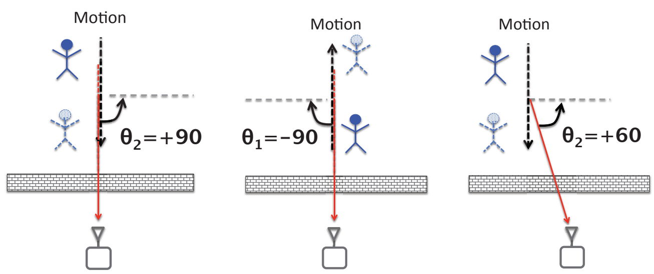
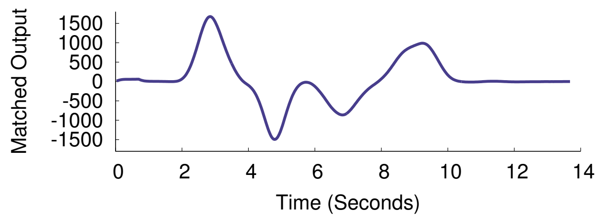


Figure 3-6: **Gestures as Angles.** Recall θ's magnitude and sign as defined in §3.4.1. In (a), the subject takes one step forward; the emulated antenna array's normal forms an angle of 90° with the line from the human to WiVi. Because the vector of the motion and the vector from the human to WiVi are in same direction, θ is positive; hence, it is +90°. In (b), the subject takes a step backward, and $\theta = -90$ degrees. In (c), the subject does not exactly know where the WiVi device is, so he performs the steps towards the wall, without orienting himself directly toward WiVi. Note that the vector of motion and the vector from the human to WiVi are in the same direction; hence, θ is positive. However, due to the slanted orientation, it is now +60° (rather than +90°).

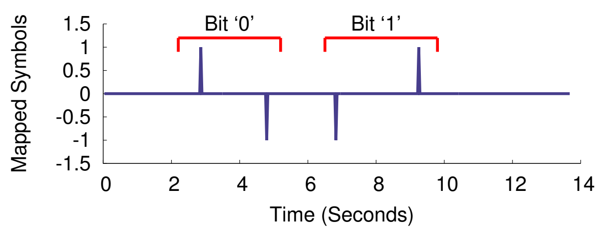
■ 3.5.2 Gesture Decoding

Decoding the above gestures is fairly simple and follows standard communication techniques. Specifically, WiVi's decoder takes as input $A'[\theta, n]$. Similar to a standard decoder [50], WiVi applies a matched filter on this signal. However, since each bit is a combination of two steps, forward and backward, WiVi applies two matched filters: one for the step forward and one for the step backward. Because of the structure of the signal shown in Fig. 3-5, the two matched filters are simply a triangle above the zero line, and an inverted triangle below the zero line. WiVi applies these filters separately on the received signal, then adds up their output.

Fig. 3-7 shows the results of applying the matched filters on the received signal in Fig. 3-5. Note that the signal after applying the matched filters looks fairly similar to a BPSK signal, where a peak above the zero line represents a '1' bit and a trough below the zero line represents a '0' bit. (Though, in WiVi, our encoding is such that a peak or a trough alone only represents half a bit.) Next, WiVi uses a standard peak detector to detect the peaks/troughs and match them to the corresponding bits. Fig. 3-7 shows the identified peaks and the detected bits for the two-bit message in Fig. 3-5.



(a) Output of matched filter.



(b) Decoded bits.

Figure 3-7: **Gesture Decoding in WiVi.** The figure shows how WiVi decodes the gestures of Fig. 3-5. (a) shows the output of the matched filter step. (b) shows the output of the peak detector. The sequence $(1, -1)$ represents bit '0', whereas the sequence $(-1, 1)$ represents bit '1'.

3.6 Implementation and Evaluation

In this section, we describe our implementation and the results of our experimental evaluation.

3.6.1 Implementation

We built WiVi using USRP N210 software radios [13] with SBX daughter boards. The system uses LP0965 directional antennas [15], which provide a gain of 6 dBi. The system consists of three USRPs connected to an external clock so that they act as one MIMO system. Two of the USRPs are used for transmitting, and one for receiving. MIMO nulling is implemented directly into the UHD driver, so that it is performed in real-time. Post-processing using the smoothed MUSIC algorithm is performed on the obtained traces offline in Matlab R2012a under Ubuntu 11.10 on a 64-bit machine with Intel i7 processor. Matlab already has a built-in and highly optimized smoothed MUSIC implementation. Processing traces of 25-second length took on average 1.0564s per trace, with a standard deviation of 0.2561s.

We implement standard Wi-Fi OFDM modulation in the UHD code; each OFDM symbol consists of 64 subcarriers including the DC. The nulling procedure in §3.3 is performed on a subcarrier basis. The channel measurements across the different subcarriers are combined to improve the SNR. Since USRPs cannot process signals in real-time at 20 MHz, we reduced the transmitted signal bandwidth to 5 MHz so that our nulling can still run in real time.

Finally, the emulated antenna array was taken over 0.32 seconds. The collected samples during this duration were averaged into an antenna array of size $w = 100$, which was provided as an input to the smoothed MUSIC algorithm.

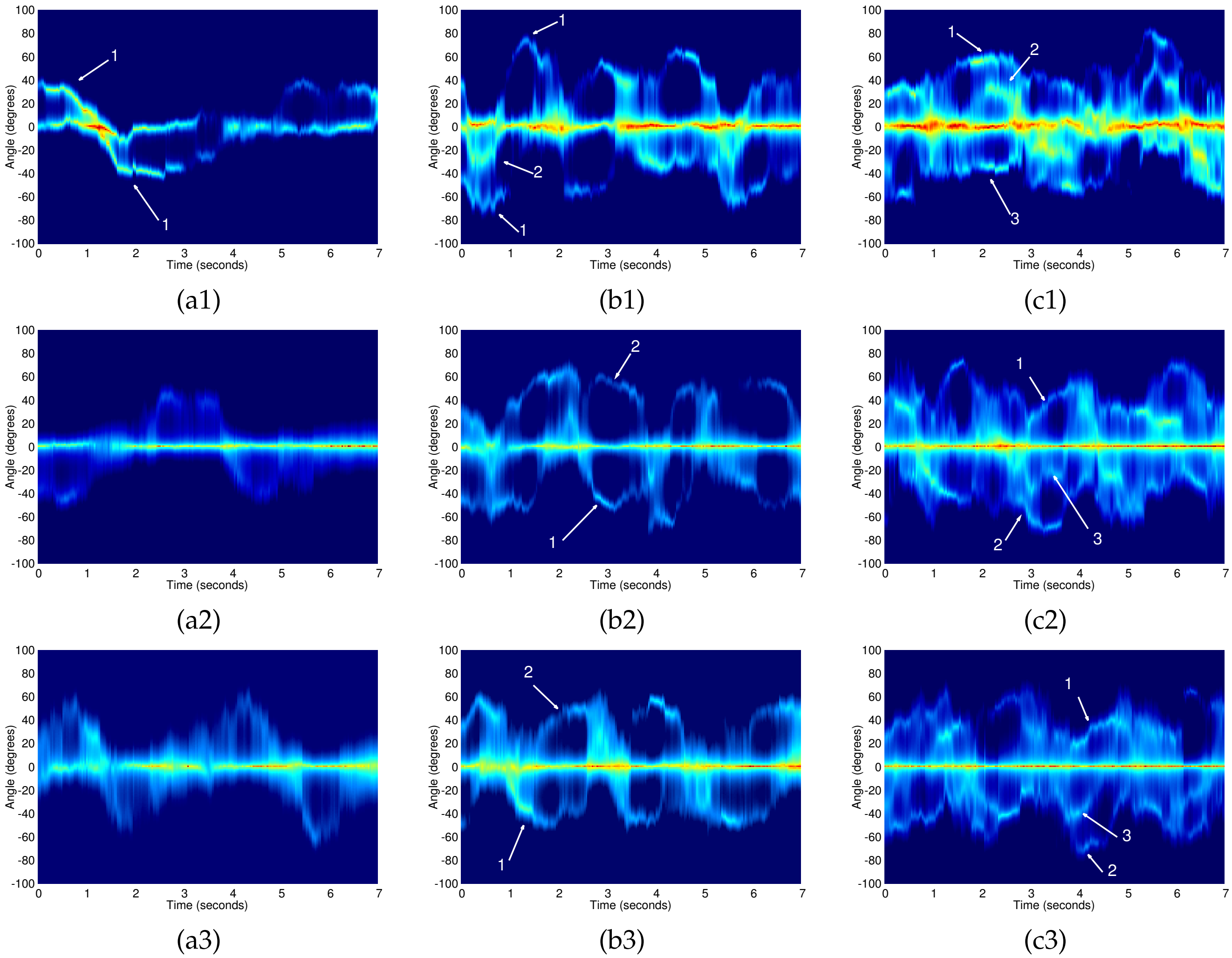
■ 3.6.2 Experimental Setup

Most of our experiments were run in one office building using two different conference rooms. The rooms have standard furniture: tables, chairs, boards, etc. The interior walls of the building are 6-inch hollow walls supported by steel frames with sheet rock on top. The first conference room is 7×4 meters; the second is 11×7 meters. We also conducted some experiments in a second building on our campus which has 8-inch concrete walls.

The experiments were conducted with eight human subjects, three women and five men, of different heights and builds. For the tracking experiments, we asked the subjects to enter a room, close the door, and move at will. The through-wall gesture experiments were performed with four subjects (one woman and three men). The persons were shown the gestures in advance and tried them a few times. Then, each of them entered the room separately and performed the gestures. The experiments are repeated in different locations in different rooms, and in different locations in each room.

■ 3.6.3 Micro Benchmarks

First, we would like to get a better understanding of the information captured by WiVi, and how it relates to the moving objects. We run experiments in two conference rooms



(a) One Human **(b) Two Humans** **(c) Three Humans**

Figure 3-8: **Tracking human motion with WiVi.** The figures show output traces with a different number of humans after processing with the smoothed MUSIC algorithm. They plot $A'[\theta, n]$ where θ is the angle in $[-90, 90]$ is plotted on the y-axis and time is on the x-axis. (a) shows traces for one human; (b) for two humans; and (c) for three humans moving behind the wall of a closed room.

in our building. Both conference rooms have $6''$ hollow walls supported by steel frames with sheet rock on top. In all of these experiments, we position WiVi one meter away from a wall that has neither a door nor a window. For each of our experiments, we ask a number of humans between 1 and 3 to enter the room, close the door, and move at will. WiVi performs nulling in real time and collects a trace of the signals. We perform each experiment with a different subset of our subjects. We process the collected traces using the smoothed MUSIC algorithm as described in §3.4.2.

Fig. 3-8 shows the output of WiVi in the presence of one, two, or three humans moving in a closed room. Consider the plots with one human in Figs. 3-8(a). Besides the DC, the graphs show one fuzzy curved line. The line tracks the spatial angle of the moving human. Compare these figures with the set of figures in 3-8(b), which capture two moving humans. In 3-8(b), we can discern two curved lines that track the angular motion of these humans with respect to WiVi. If we take a vertical line at any time, in any of the two-human figures, we see at most two bright lines, besides the DC. This is because, in these figures, at any point in time, there are at most two moving bodies in the room. Let us zoom in on the interval $[1s, 2s]$ in 3-8(b1). During this interval, we see only one curved line. This has two possible interpretations: either one of the two people stopped moving or he/she was too deep inside the room that we could not capture his/her signal. As we move to 3-8(c), the figures get fuzzier since we have more people moving in the same area. However the general observations carry to these figures. Specifically, we can identify the presence of three humans from observing multiple intervals in which we can discern three curved lines. For example, consider the interval $[1.8s, 2.5s]$ in 3-8(c1); it shows two lines with positive angles and one with a negative angle. These lines indicate that two people are moving towards WiVi, while one person is moving away.

One can also make multiple observations based on the shape of the lines. First, a positive angle means the human is moving toward WiVi, while a negative angle means that he is moving away. The value of that angle depends on the orientation of the human and the direction of motion. Each line looks like a wave because, given a confined space, a person that moves towards WiVi will eventually have to move away or stop. Second, the brightness of the line typically indicates distance. Note that for the same spatial angle, one may be close or far from WiVi. Hence, some large angles appear bright or dim depending on the part of the trace we look at.

A third observation is that as the number of humans increases, it becomes harder to separate them. The problem is that the curved lines are fuzzy both due to residual noise and the fact that a human can move his body parts differently as he moves. For example, waving while moving makes the lines significantly fuzzier as in 3-8(a3).

Finally, our experiments are conducted in multipath-rich indoor environments. Thus, the results in Fig. 3-8 show that WiVi works in the presence of multipath effects. This is because the direct path from a moving human to WiVi is much stronger than indirect paths which bounce off the internal walls of the room. A moving human acts like a large antenna. In order to block the direct path, the human body must be obstructed by a pillar or a large piece of furniture, and stay obstructed for the duration of WiVi's measurements.[11]

■ 3.6.4 Automatic Detection of Moving Humans

We are interested in evaluating whether WiVi can use the spatial variance described in §3.4.2 to automate the detection of moving humans. As in the previous section, we run our experiments in the same conference rooms described in §3.6.3. Again, we position WiVi such that it faces a wall that has neither a door nor a window. For each of our experiments, we ask a number of humans between 0 and 3 from our volunteers to enter the room and move at will. Each experiment lasts for 25 seconds excluding the time required for iterative nulling. We perform each experiment with a different subset of subjects, and conduct a total of 80 experiments, with equal number of experiments spanning the cases of 0, 1, 2, and 3 moving humans. We process the collected traces offline and compute the spatial variance as described in §3.4.2.

Fig. 3-9 shows the CDFs (cumulative distribution functions) of the spatial variance for the experiments run with each number of moving humans: 0, 1, 2, and 3. We observe the following:

- The spatial variance provides a good metric for distinguishing the number of moving hu-

[11]We note that the experiments in this chapter were performed in scenarios where the separator is homogeneous wall (e.g., concrete, wooden, glass, etc.). There might be scenarios in which the separator is non-homogeneous (e.g., the field of view of WiVi's directional antenna captures a side of a wall and a glass window), which may cause some indirect paths to be stronger than the direct path. In this case, WiVi will still detect a moving object but may have errors in tracking the angle of the movement or predicting the number of moving humans.

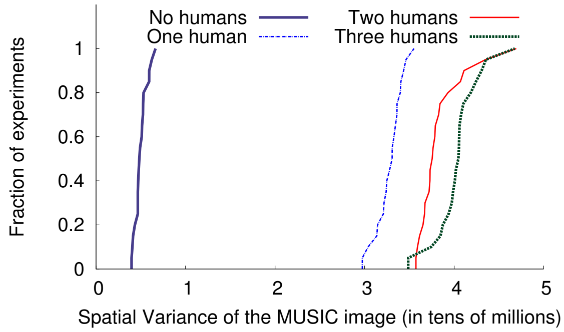


Figure 3-9: **CDF of spatial variance for a different number of moving humans.** As the number of humans increases, the spatial variance increases.

mans. In particular, the variance increases as the number of humans involved in each experiment increases. This is also evident from the figures in 3-8, where one can visually see that the spatial variance is higher with more moving bodies in the room.

- Interestingly, the separation between successive CDFs decreases as the number of humans increases. In particular, the separation is larger between the CDFs of no humans and one human, than between the CDFs of one human and two humans. The separation is the least between the CDFs of 2 humans and 3. To understand this behavior, recall that because the room has a confined space, as the number of people increases, the freedom of movement decreases. Hence, adding a human to a congested space is expected to add less spatial variance than adding her to a less congested space where she has more freedom to move.

Next, we would like to automate the thresholds for distinguishing 0, 1, 2, and 3 moving humans. To do so, we divide the data into a training set and a testing set. To ensure that WiVi can generalize across environments, we ensure that the training examples are all conducted in one conference room, while the testing examples are conducted in another conference room (Recall that the two rooms have different sizes). We use the training set to learn the thresholds to separate the spatial variances corresponding to 0, 1, 2, and 3 humans. We then use these thresholds to classify the experiments in the testing set. Finally, we perform cross-validation, i.e., we repeat the same procedure after switching the training and testing sets.

Table 3-2 shows the result of the classification. It shows that WiVi can identify whether there is 0 or 1 person in a room with 100% accuracy; this is expected based on the CDFs in Fig. 3-9. Also, row 3 shows that two humans are never confused with 0 or 1. How-

Actual \ Detected	0	1	2	3
0	100%	0%	0%	0%
1	0%	100%	0%	0%
2	0%	0%	85%	15%
3	0%	0%	10%	90%

Table 3-2: **Accuracy of Automatic Detection of Humans.** The table shows the accuracy of detecting the number of moving humans based on the spatial variance.

ever, WiVi confused 2 humans with 3 humans in 15% of the trials, whereas it accurately identified their number in 85% of the cases.

3.6.5 Gesture Decoding

Next, we evaluate WiVi's ability to decode the bits associated with the gestures in §3.5. In each experiment, a human is asked to stand at a particular distance from the wall that separates the room from our device, and perform the two gestures corresponding to bit '0' and bit '1'. Each human took steps at a length they found comfortable. Typical step sizes were 2-3 feet. The experiments are repeated at various distances in the range [1m, 9m]. All experiments are conducted in the same conference rooms described above and under the same experimental conditions. One of our conference rooms is only 7m wide, whereas the other is 11m wide. Hence, the experiments with distances larger than 6 meters are conducted in the larger conference room, whereas for all distances less than or equal 6 meters, our experiments included trials from both rooms. The obtained traces are processed using the matched filter and decoding algorithm described in §3.5.2.

Fig. 3-10 plots the fraction of time the gestures were decoded correctly as a function of the distance from the wall separating WiVi from the closed room. We note the following observations:

- WiVi correctly decoded the performed gestures at all distances less than or equal to 5m. It identified 93.75% of the gestures performed at distances between 6m and 7m. At 8m, the performance started degrading, leading to correct identification of only 75% of the gestures. Finally, WiVi could not identify any of the gestures when the person was standing 9m away from the wall.
- It is important to note that, in our experiments, WiVi never mistook a '0' bit for a '1' bit

or the inverse. When it failed to decode a bit, it was because it could not register enough energy to detect the gesture from the noise. This means that WiVi 's errors are erasure errors as opposed to standard bit errors.

- We measured the time it took the different subjects to perform a one bit gesture. Averaged over all traces, our subjects took 2.2s to perform a gesture, with a standard deviation of 0.4s.

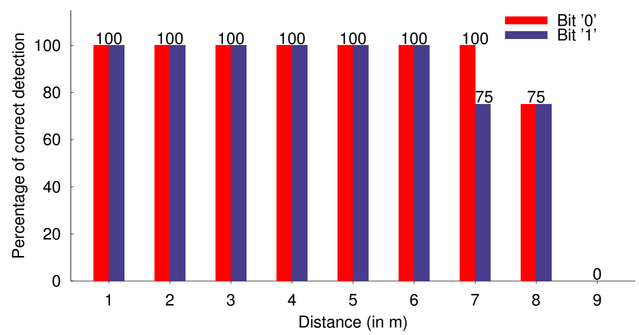


Figure 3-10: **Accuracy of Gesture Decoding as a Function of Distance.** The figure shows the fraction of experiments in which WiVi correctly decoded the bit associated with the performed gesture at different distances separating the subject from the wall. Note that WiVi decodes a gesture only when its SNR is greater than 3dB; this explains the sharp cutoff between 8 and 9 meters.

To gain further insight into WiVi's gesture decoding, Fig. 3-11 plots the CDFs of the SNRs of the '0' gesture and the '1' gesture, across all the experiments. Interestingly, the gesture associated with a '0' bit has a higher SNR than the gesture associated with a '1' bit. This is due to two reasons: First, the '0' gesture involves a step forward followed by a step backward, whereas the '1' gesture requires the human to first step backward then forward. Hence, for the same starting point, the human is on average closer to WiVi while performing the '0' gesture, which results in an increase in the received power. Second, taking a step backward is naturally harder for humans; hence, they tend to take smaller steps in the '1' gesture. This observation is visually evident in Fig. 3-5 where a '0' gesture has a higher power (red) than the '1' gesture.

We note that the main factor limiting gesture decodability with increased distance is the low transmit power of USRPs. The linear transmit power range for USRPs is around 20 mW (i.e., beyond this power the signal starts being clipped), whereas Wi-Fi's power limit is 100mW. Hence, one would expect that with better hardware, WiVi can have a higher decoding range.

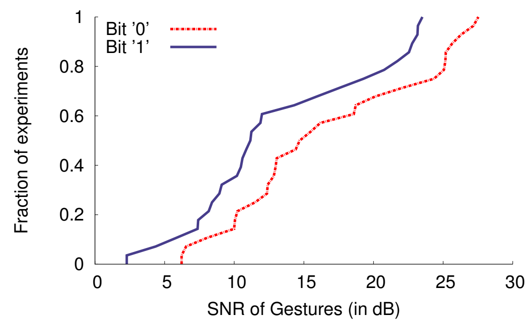


Figure 3-11: **CDF of the gesture SNRs.** The figure shows the CDFs of the SNR after applying the matched filter taken over different distances from WiVi.

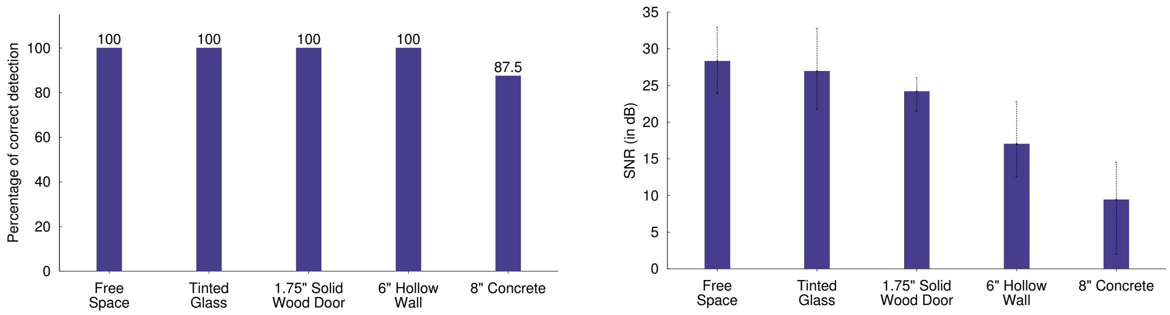
3.6.6 The Effect of Building Material

Finally, we evaluate WiVi's performance with different building materials. Thus, in addition to the two conference rooms described before, we also test WiVi in a second building in our university campus, where the walls are different. In particular, we experiment with 4 types of building materials: $8''$ concrete wall, $6''$ hollow wall supported by steel frames with sheet rock on top, $1.75''$ solid wood door, and tinted glass. In addition, we perform experiments in free space with no obstruction between WiVi and the subject.

In each experiment, the subject is asked to stand 3 meters away from the wall (or WiVi itself in the case of no obstruction) and perform the '0' bit gesture described above. For each type of building material, we perform 8 experiments.

Fig. 3-12 shows WiVi's performance across different building materials. Specifically, Fig. 3-12(a) shows the detection rate as the fraction of experiments in which WiVi correctly decoded the gesture, whereas Fig. 3-12(b) shows the average SNRs of the gestures. The figures show that WiVi can detect humans and identify their gestures across various indoor building materials: tinted glass, solid wood doors, $6''$ hollow walls, and to a large extent $8''$ concrete walls. As expected, the thicker and denser the obstructing material, the harder it is for WiVi to capture reflections from behind it.

Detecting humans behind different materials depends on WiVi's power as well as its ability to eliminate the flash effect. Fig. 3-13 plots the CDF of the amount of nulling (i.e., reduction in SNRs) that WiVi achieves in various experiments. The plot shows WiVi's nulling reduces the signal from static objects by a median of 40 dB. This number indicates that WiVi can eliminate the flash reflected off common building material such as glass,



(a) Detection Accuracy **(b) SNR**

Figure 3-12: **Gesture detection in different building structures.** (a) plots the detection accuracy of WiVi for different types of obstructions. (b) shows the average SNR of the experiments done through these different materials, with the error bars showing the minimum and maximum achieved SNRs across the trials.

solid wood doors, interior walls, and concrete walls with a limited thickness [3]. However, it would not be able to see through denser material like re-enforced concrete. To improve the nulling, one may use a circulator at the analog front end [78] or leverage recent advances in full-duplex radio [45], which were reported to produce 80 dB reduction in interference power [83].

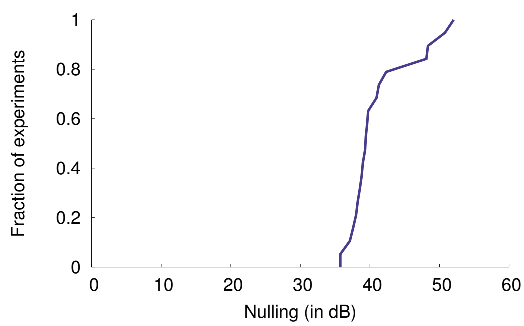


Figure 3-13: **CDF of achieved nulling.** The figure plots the CDF which shows the ability of nulling to reduce the power received along static paths.

■ 3.7 Summary & Discussion

In this chapter, we have presented WiVi, a wireless technology that uses Wi-Fi signals to detect moving humans behind walls and in closed rooms. In contrast to previous systems, which are targeted for the military, WiVi enables small cheap see-through-wall devices that operate in the ISM band, rendering them feasible to the general public. WiVi also establishes a communication channel between itself and a human behind a wall, allowing him/her to communicate directly with WiVi without carrying any transmitting device.

We believe that WiVi is an instance of a broader set of functionality that future wireless networks will provide. Future Wi-Fi networks will likely expand beyond communications and deliver services such as indoor localization, sensing, and control. WiVi demonstrates an advanced form of Wi-Fi-based sensing and localization by using Wi-Fi to track humans behind wall, even when they do not carry a wireless device. It also raises issues of importance to the networking community pertinent to user privacy and regulations concerning the use of Wi-Fi signals.

Finally, WiVi bridges state-of-the-art networking techniques with human-computer interaction. It motivates a new form of user interfaces which rely solely on using the reflections of a transmitted RF signal to identify human gestures. However, in its current form, WiVi's gesture-based interface remains quite coarse; in the next few chapters, we design new algorithms and build more sensitive hardware to capture higher quality images which enable the gesture-based interface to become more expressive.

CHAPTER 4

3D Tracking via Body Radio Reflections

Recent years have witnessed a surge in motion tracking and localization systems. Multiple advances have been made both in terms of accuracy and robustness. In particular, RF localization using WiFi and other communication devices has reached sub-meter accuracy and demonstrated its ability to deal with occlusions and non-line of sight scenarios [86, 170]. Yet these systems require the user to carry a wireless device in order to be localized. In contrast, systems like Kinect and depth imaging have revolutionized the field of human-computer interaction by enabling 3D motion tracking without instrumenting the body of the user. However, Kinect and imaging systems require a user to stay within the device's line-of-sight and cannot track her across rooms. We envision that if an RF system can perform 3D motion tracking without requiring the user to wear a radio, it will motivate the integration of such a technology in systems like Kinect to expand their reach beyond direct line of sight and enable through-wall human-computer interaction.

Motivated by this vision, this chapter introduces WiTrack, a system that tracks the 3D motion of a user using radio reflections that bounce off her body. It works through walls and occlusions, but does not require the user to carry any wireless device. WiTrack can also provide coarse tracking of a body part. In particular, the user may lift her hand and point at objects in the environment; the device detects the direction of the hand motion, enabling the user to identify objects of interest. In comparison to WiVi (the device described in Chapter 3), WiTrack is much more powerful. It can track the exact location of a person, whereas WiVi can only detect coarse (forward-backward) movements and gestures. In addition, WiTrack provides a much more expressive gesture-based interface than WiVi. However, to achieve these capabilities, WiTrack does not rely on standard WiFi (like in Chapter 3); nonetheless, it still limits itself to FCC regulations for consumer electronics.

WiTrack has one antenna for transmission and three antennas for receiving. At a high level, WiTrack's motion tracking works as follows. The device transmits a radio signal and uses its reflections to estimate the time it takes the signal to travel from the transmitting antenna to the reflecting object and back to each of the receiving antennas. WiTrack then uses its knowledge of the position of the antennas to create a geometric reference model, which maps the round trip delays observed by the receive antennas to a 3D position of the reflecting body.

Transforming this high-level idea into a practical system, however, requires addressing multiple challenges. First, measuring the time of flight is difficult since RF signals travel very fast – at the speed of light. To distinguish between two locations that are closer than one foot apart, one needs to measure differences in reflection time on the order of hundreds of picoseconds, which is quite challenging. To address this problem, we leverage a technique called FMCW (frequency modulated carrier wave) which maps differences in time to shifts in the carrier frequency; such frequency shifts are easy to measure in radio systems by looking at the spectrum of the received signal.

A second challenge stems from multipath effects, which create errors in mapping the delay of a reflection to the distance from the target. WiTrack has to deal with two types of multipath effects. Some multipath effects are due to the transmitted signal being reflected off walls and furniture. Others are caused by the signal first reflecting off the human body then reflecting off other objects. This is further complicated by the fact that in non-line-of-sight settings, the strongest signal is not the one directly bouncing off the human body. Rather it is the signal that avoids the occluding object by bouncing off some side walls. WiTrack eliminates reflections from walls and furniture by noting that their distance (and time of flight) does not change over time. Hence, they can be eliminated by subtracting consecutive frames of the signals. Reflections that involve a combination of a human and some static object are more complex and are addressed through filters that account for practical constraints on the continuity of human motion and its speed in indoor settings.

We have built a prototype of WiTrack and evaluated it empirically. Since off-the-shelf radios do not perform FMCW, we built an analog FMCW radio frontend, which operates as a daughterboard for the USRP software radio. In our evaluation, we use the VICON motion capture system [14] to report the ground truth location. VICON can achieve sub-centimeter accuracy but requires instrumenting the human body with infrared markers and positioning an array of infrared cameras on the ceiling. Since VICON cannot operate in non-line-of-sight, the human moves in the VICON room while our device is placed outside the room and tracks the motion across the wall. Our evaluation considers three applications, each of them uses the developed 3D tracking primitive in a different way.

In the first application, we consider 3D tracking of human motion through a wall. The objective of such an application is to augment virtual reality and gaming systems to work in non-line-of-sight and across rooms. We compute the tracking error as the difference between the location reported by our device and the actual location of the body center as reported by VICON. Our results show that WiTrack localizes the center of the human body to within 10 to 13 cm in the x and y dimensions, and 21 cm in the z dimension. This high accuracy stems from WiTrack's ability to eliminate errors due to multipath and the combined performance of FMCW and our geometric mapping algorithm. The results also show that even the 90^{th} percentile of the measurements stays within one foot along the x/y-axis and two feet along the z-axis.

In the second application, we consider elderly fall detection. Current solutions to this problem include inertial sensors which old people tend to forget to wear [51], or cameras which infringe on privacy, particularly in bedrooms and bathrooms [109]. In contrast, WiTrack does not require the user to wear any device and protects her privacy much better than a camera. However, simply looking at the change in elevation cannot allow us to distinguish a fall from sitting on the floor. Thus, WiTrack identifies a fall as a *fast* change in the elevation that reaches the ground level. In a population of 11 users and over 133 experiments, WiTrack distinguishes a fall from standing, walking, sitting on a chair and sitting on the floor with an accuracy of 96.9% (the F-measure is 94.34%).

In the third application, we consider a user who desires to control appliances by pointing at them (e.g., the user can turn her monitor on or turn the lights off by simply pointing at these objects.) We consider a gesture in which the user lifts her arm, points at an appliance, and drops her arm. By comparing the position of the arm over time, WiTrack can identify the pointing direction. Our prototype estimates the pointing direction with a median of 11.2 degrees and a 90^{th} percentile of 37.9 degrees.

Our results also show that the prototype operates in realtime, and outputs the 3D location within 75 ms from the time the antennas receive the signal. Further, it operates at a fairly low-power, transmitting sub-milliWatt power. However, the prototype described in this chapter can track a single person, and requires the person to move to obtain an initial estimate of his location. Chapter 5 describes how we can move from single-person to multi-person localization and how we can localize static users.

■ 4.1 Related Work

WiTrack is related to prior art in the following areas:

Indoor wireless localization: WiTrack builds on recent advances in RF-based localization [30, 86, 160, 170]. These systems localize a wireless device using RSSI [30, 119], fine-grained-OFDM channel information [130], antenna arrays [86, 170], or RFID backscatter [159, 160]. In contrast, WiTrack localizes a human using body radio reflections.

Some past works in radio tomography use a network of tens or hundred sensors to track a person even if she does not carry any wireless device [165, 166]. These works measure the RSSI for each of the resulting n^2 links between their sensors, and attribute the variation of RSSI on a link to a human crossing that link. Other works on device-free localization rely on RSSI fingerprints [129, 178], which are generated in a training phase by asking a person to stand in different locations throughout the area of interest. In the testing phase, they localize a person by mapping the resulting RSSI to the closest fingerprint. While WiTrack shares the objective of tracking a person's motion without instrumenting her body, it differs in both technology and accuracy. Specifically, WiTrack does not require prior training and uses a few antennas that generate FMCW signals and measure the time-of-flight of the signal reflections to infer location of a human. Its technique extends to 3D, and its 2D accuracy is more than $5\times$ higher than the state of the art RSSI-based systems [129, 180].

See through-wall & gesture recognition using WiFi: WiTrack is motivated by our earlier work on WiVi (described in Chapter 3) and other contemporary research that used WiFi signals to detect users through walls and identify some of their gestures [42, 117]. Similar to these systems, WiTrack captures and interprets radio reflections off a human body. WiTrack, however, differs from these systems both in capability and technology. Specifically, these systems rely on the Doppler shift of WiFi signals. Hence, they can distinguish only between getting closer or getting further away, but cannot identify the location of the person.[1] In contrast, WiTrack measures the time of flight and, hence, can identify the exact location of a person. Among these past systems, WiVi focuses on tracking through dense walls such as concrete by leveraging interference nulling to eliminate the wall's reflection. In contrast, WiTrack focuses on accurate 3D motion tracking that operates through *interior walls* (which are less dense than concrete)[2], pinpointing the exact location of a user at any point in time.

FMCW Radar: WiTrack builds on past work on FMCW radar, including work that used FMCW for see-through-wall that is targeted for the military [41, 121]. WiTrack however differs along multiple dimensions, as we explained in Chapter 2. Specifically, recall that FMCW radios in past work were high-power and heavy (needed to be mounted on a truck). Their tracking capabilities hinge on using large antenna arrays that can achieve a narrow beam, which enables tracking a moving target. In contrast, we present a light weight, low-power FMCW radio that complies with the FCC regulations for consumer devices. We are able to perform accurate tracking with a low-power, relatively cheap FMCW prototype because of two innovations: first, a geometric localization algorithm that combines multiple measurements from different antenna locations and fits them within a geometric reference to pinpoint an accurate 3D location, and second, novel techniques that enable rejecting errors that are due to both static and dynamic multi-path in indoor environments. Further, WiTrack extends its techniques to tracking the motion of body parts, e.g., tracking a hand as it points in a particular direction.

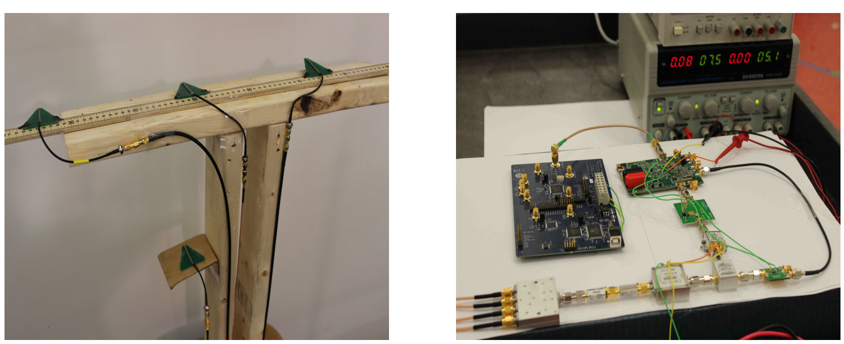
[1]The gestures recognized by WiVi and WiSee are sequences of getting closer or getting further away, which translate into positive and negative Doppler shifts. The work in [42] provides a distance estimate with an accuracy of about 30 meters.

[2]To enable WiTrack to track through thicker walls such as concrete (as in WiVi), one may add a filter to remove the wall's reflection.

Motion tracking in user interfaces: Finally, WiTrack is related to an emerging body of motion-tracking user interfaces. These include devices that the person needs to hold (such as the Nintendo Wii [9]) or wear (e.g., on-body sensors such as wristbands [2, 47, 69]). They also include vision and infrared-based systems, like Xbox Kinect [16] and Leap Motion [8], which can track a person's movement without requiring her to hold or wear any transmitter or receiver but require the user to maintain a line-of-sight path to their sensors. Similar to these systems, WiTrack enables more natural human-computer interaction. However, in comparison to these systems, WiTrack does not require the user to hold/wear any device or to maintain a line-of-sight path to its sensors; it can track a user and her gestures in non-line-of-sight and across different rooms.

4.2 WiTrack Overview

WiTrack is a wireless system that performs 3D motion tracking in both line-of-sight and through wall scenarios. It can also provide coarse tracking of body parts, like an arm movement. WiTrack uses multiple directional antennas: one antenna is used for transmitting, and three antennas for receiving. In its default setup, the antennas are arranged in a "T" shape, as shown in Fig. 4-1(a). In its current version WiTrack tracks one moving body at any time. Other people may be around but should be either behind the antenna beam or they should be approximately static.[3]



(a) **Antenna "T" Setup** (b) **FMCW Signal Generation**

Figure 4-1: **WiTrack's Setup and Signal Generation.** (a) shows WiTrack's directional antennas (dimension of each antenna: $5cm \times 5cm$) arranged in a "T": the transmit antenna is placed at the crossing point of the T, whereas the receive antennas are on the edges. (b) shows the hardware we built to generate FMCW signals.

[3]Small moving objects which do not have significant reflections (e.g., a plastic fan) create some noise but do not prevent WiTrack's 3D tracking.

WiTrack operates by transmitting an RF signal and capturing its reflections off a human body. It tracks the motion by processing the signals from its received antennas using the following three steps:

1. *Time-of-Flight (TOF) Estimation:* WiTrack first measures the time it takes for its signal to travel from its transmit antenna to the reflecting body, and then back to each of its receive antennas. We call this time the TOF (time-of-flight). WiTrack obtains an initial measurement of the TOF using FMCW transmission technique; it then cleans this estimate to eliminate multipath effects and abrupt jumps due to noise.
2. *3D Localization:* Once it obtains the TOF as perceived from each of its receiving antennas, WiTrack leverages the geometric placement of its antennas to localize the moving body in 3D.
3. *Fall Detection and Pointing:* WiTrack builds on the 3D localization primitive to enable new functionalities. Specifically, WiTrack can detect a fall by monitoring fast changes in the elevation of a human and the final elevation after the change. WiTrack can also differentiate an arm motion from a whole body motion; it can track the motion of raising one's arm, localize the initial and final position of the arm, and determine the direction in which the arm is pointing.

4.3 Time-of-Flight Estimation

The first step for WiTrack is to measure the TOF from its transmit antenna to each of its receive antennas and clean this estimate from the effect of multi-path.

4.3.1 Obtaining Time-of-Flight Estimates

A straightforward approach for estimating the time of flight is to transmit a very short pulse and measure the delay between the transmitted pulse and its received echo. Such a design requires sampling the signal at sub-nanosecond intervals – i.e, it requires high speed analog-to-digital converters (ADCs) that operate at multi-GS/s. Such ADCs are high power, expensive, and have low bit resolution, making this approach unattractive in practice.

Instead, WiTrack measures the TOF by leveraging a technique called Frequency-Modulated Carrier Waves (FMCW). We explain FMCW at a high level, and refer the reader to [102] for a more detailed explanation. FMCW transmits a narrowband signal (e.g., a few KHz) whose carrier frequency changes linearly with time. To identify the distance from a reflector, FMWC compares the carrier frequency of the reflected signal to that of the transmitted signal. Since the carrier frequency is changing linearly in time, delays in the reflected signals translate into frequency shifts in comparison to the transmitted wave. Therefore, by comparing the frequency difference between the transmitted signal and the received signal, one can discover the time delay that the signal incurred, which corresponds to the TOF of that signal.

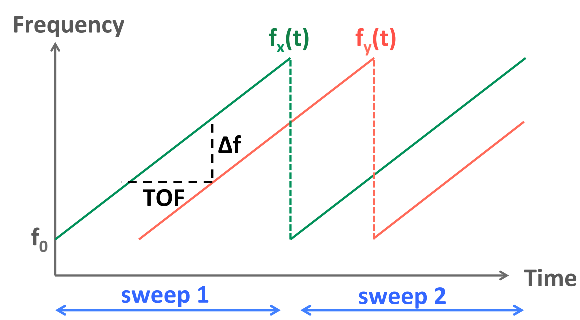


Figure 4-2: **FMCW operation**. The transmitted signal has a carrier frequency $f_x(t)$ that is repeatedly swept in time. Because the received signal is time-shifted with respect to the transmitted signal, its carrier frequency $f_y(t)$ is frequency-shifted with respect to $f_x(t)$.

Fig. 4-2 illustrates this concept. The green line is the carrier frequency of the transmitted signal which sweeps linearly with time. The red line is the carrier frequency of the reflected signal as a function of time. The time shift between the two is the time-of-flight (TOF) for that reflector. The frequency shift Δf between the transmitted and received signals is a function of both the slope of the sweep and the TOF, i.e.:

$$TOF = \Delta f / slope \tag{4.1}$$

Though the above description is for a single reflector, it can be easily generalized to an environment with many reflectors. In this case, the transmitted signal would still consist of a single carrier wave that is linearly swept in time. However, because wireless reflections add up linearly over the medium, the received signal is a linear combination of multiple reflections, each of them shifted by some Δf that corresponds to its own TOF. Hence one can extract all of these TOFs by taking a fourier transform (i.e, an FFT) of the received baseband signal.[4] The output of the FFT gives us the ***TOF profile*** which we define as the reflected power we obtain at each possible TOF between the transmit antenna and receive antenna.

In comparison to transmitting a very short pulse and measuring its sub-nanosecond delay in the time domain, FMCW does not require high speed ADCs because at any point in time, the received baseband signal is narrowband.

FMCW Resolution: It is important to note that the resolution of an FMCW system is a function of the total bandwidth that the carrier frequency sweeps [102]. The resolution is defined by the ability to distinguish between two nearby locations, which depends on the ability to distinguish their TOFs, which itself depends on the resolution in distinguishing frequency shifts Δf. The resolution of identifying frequency shifts is equal to the size of one bin of the FFT. The FFT is typically taken over a duration of one sweep of the carrier frequency (denoted by T_{sweep}) and hence the size of one FFT bin is $1/T_{sweep}$. Since the minimum measurable frequency shift is $\Delta f_{min} = 1/T_{sweep}$, the minimum measurable change in location is:

$$Resolution = C\frac{TOF_{min}}{2} = C\frac{\Delta f_{min}}{2 \times slope}, \tag{4.2}$$

where C is the speed of light and the factor 2 accounts for the fact that the reflected signal traverses the path back and forth.

The $slope$ is equal to the total swept bandwidth B divided by the sweep time T_{sweep}. Hence after substituting for the slope in the above equation we get:

$$Resolution = \frac{C}{2B} \tag{4.3}$$

Since C is very large, obtaining high resolution requires a large B, i.e., the system has to take a narrowband signal and sweep its carrier frequency across a wide bandwidth of multiple GHz.

In our design we chose the following parameter for our FMCW. We have built an FMCW system that sweeps a total bandwidth of 1.69 GHz from 5.56 GHz to 7.25 GHz,

[4] The baseband signal is the received signal after mixing it by the transmitted carrier. The mixing shifts the spectrum of the received signal by the transmitted carrier frequency.

and transmits at sub-milliWatt power. The choice of this bandwidth has been dictated by the FCC regulations for civilian use of spectrum [18]. Specifically, it is the largest contiguous bandwidth below 10 GHz which is available for civilian use at low power.

Based on Eq. 4.3, our sweep bandwidth allows us to obtain a distance resolution of 8.8 cm. Hence the average error in mapping TOF to distance in 1D is about 4.4 cm. Note that the above derivation neglects the impact of noise, and hence provides a lower bound on the achievable resolution. In practice, the system's resolution is affected by the noise level. It also depends on the geometric model that maps TOFs to 3D locations.

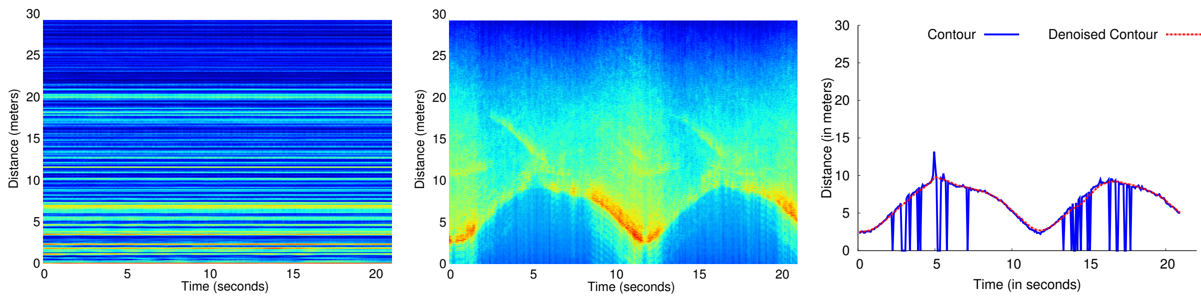
■ 4.3.2 Addressing Static Multi-path

The next step in WiTrack's operation is to distinguish a human's reflections from reflections off other objects in the environment, like furniture and walls. Recall from the previous section that every reflector in the environment contributes a component to the overall received signal, and that component has a frequency shift that is linearly related to the time-of-flight of the reflection based on Eq. 4.1. Typically, reflections from walls and furniture are much stronger than reflections from a human, especially if the human is behind a wall. Unless these reflections are removed, they would mask the signal coming from the human and prevent sensing her motion. This behavior is called the "Flash Effect".

To remove reflections from all of these static objects (walls, furniture), we leverage the fact that since these reflectors are static, their distance to the WiTrack device does not change over time, and therefore their induced frequency shift stays constant over time. Fig. 4-3(a) plots the spectrogram of the received signal as a function of time, for one of the receive antennas of WiTrack. In particular, we take the FFT of the received signal every sweep window, and compute the power in each frequency as a function of time. Note that there is a linear relation between frequency shifts and the traveled distances as follows:

$$distance = C \times TOF = C \times \frac{\Delta f}{slope}. \tag{4.4}$$

Thus, instead of plotting the power in each frequency as a function of time, we can use the above equation to plot the power reflected from each distance as a function of time, as shown in Fig. 4-3(a). The color code of the plot corresponds to a heat-map of the power



(a) **Spectrogram** (b) **After Background Subtraction** (c) **Contour Tracking**

Figure 4-3: **Obtaining the Time-of-Flight (TOF) Estimates.** WiTrack takes an FFT of the received signal in baseband over every sweep period to generate the spectrogram in (a). Then, by subtracting out a given frame from the frame that precedes it, WiTrack eliminates static multipath as in (b). The blue plot in (c) shows how WiTrack can address dynamic multipath by tracking the bottom contour of (b), and then denoise the signal (red plot) to obtain a clean TOF estimate.

in the reflected signal. Strong reflectors are indicated by red and orange colors, weaker reflectors are indicated by yellow and green, and the absence of a reflector is indicated by blue at the corresponding frequency. The figure indicates the presence of very strong static reflectors in the environment. Specifically, it has many horizontal stripes; each of these stripes signifies the presence of a reflector at the corresponding round-trip distance. Because these stripes are horizontal, their corresponding reflectors are stationary over time. Hence, we eliminate the power from these static reflectors by simply subtracting the output of the FFT in a given sweep from the FFT of the signal in the previous sweep. This process is called background subtraction because it eliminates all the static reflectors in the background.

Fig. 4-3(b) is the result of applying background subtraction to Fig. 4-3(a). The figure shows that all static reflectors corresponding to the horizontal lines have been eliminated. This makes it easier to see the much weaker reflections from a moving human. Specifically, we see that the distance of the dominant reflector (the red color signal) is varying with time, indicating that the reflector is moving.

■ 4.3.3 Addressing Dynamic Multi-path

By eliminating all reflections from static objects, WiTrack is left only with reflections from a moving human (see Fig. 4-3(b)). These reflections include both signals that bounce off the human body to the receive antennas, and those that bounce off the human then bounce off other objects in the environment before reaching WiTrack's antennas. We refer to these indirect reflections as dynamic multi-path. It is quite possible that a human reflection that arrives along an indirect path, bouncing off a side wall, is stronger than her direct reflection (which could be severely attenuated after traversing a wall) because the former might be able to avoid occlusion.

Our idea for eliminating dynamic multi-path is based on the observation that, at any point in time, the direct signal reflected from the human to our device has travelled a shorter path than indirect reflections. Because distance is directly related to TOF, and hence to frequency, this means that the direct signal reflected from the human would result in the smallest frequency shift among all strong reflectors after background subtraction.

We can track the reflection that traveled the shortest path by tracing the bottom contour of all strong reflectors in Fig. 4-3(b). The bottom contour can be defined as the closest local maximum to our device. To determine the first local maximum that is caused by human motion, we must be able to distinguish it from a local maximum due to a noise peak. We achieve this distinguishability by averaging the spectrogram across multiple sweeps. In our implementation, we average over five consecutive sweeps, which together span a duration of 12.5 ms. For all practical purposes, a human can be considered as static over this time duration; therefore, the spectrogram would be consistent over this duration. Averaging allows us to boost the power of a reflection from a human while diluting the peaks that are due to noise. This is because the human reflections are consistent and hence add up coherently, whereas the noise is random and hence adds up incoherently. After averaging, we can determine the first local maximum that is substantially above the noise floor and declare it as the direct path to the moving human.

The blue plot in Fig. 4-3(c) shows the output of WiTrack's contour tracking of the signal in Fig. 4-3(b). In practice, this approach has proved to be more robust than tracking the dominant frequency in each sweep of the spectrogram. This is because, unlike the contour which tracks the closest path between a human body and WiTrack's antennas, the point of maximum reflection may abruptly shift due to different indirect paths in the environment or even randomness in the movement of different parts of the human body as a person performs different activities.

4.3.4 Dealing with Noise

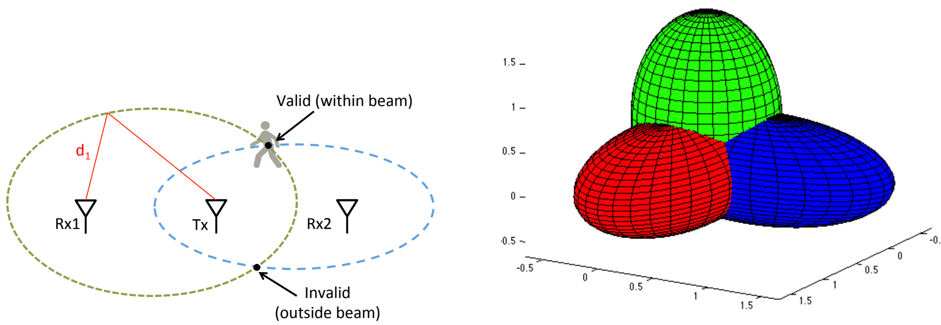
After obtaining the bottom contour of the spectrogram of the signal from each receive antenna, WiTrack leverages common knowledge about human motion to mitigate the effect of noise and improve its tracking accuracy. Specifically, by performing the following optimizations, we obtain the red plot in Fig. 4-3(c):

- *Outlier Rejection:* WiTrack rejects impractical jumps in distance estimates that correspond to unnatural human motion over a very short period of time. For example, in Fig. 4-3(c), the distance from the reflector (the blue line) repeatedly jumps by more than 5 meters over a span of few milliseconds. Such changes in distance are not possible over such small intervals of time, and hence WiTrack rejects such outliers.
- *Interpolation:* WiTrack uses its tracking history to localize a person when she stops moving. In particular, if a person walks around in a room then sits on a chair and remains static, the background-subtracted signal would not register any strong reflector. In such scenarios, we assume that the person is still in the same position and interpolate the latest location estimate throughout the period during which we do not observe any motion, enabling us to track the location of a subject even after she stops moving.
- *Filtering:* Because human motion is continuous, the variation in a reflector's distance to each receive antenna should stay smooth over time. Thus, WiTrack uses a Kalman Filter to smooth the distance estimates.

4.4 Localizing in 3D

After contour tracking and de-noising of the estimate, WiTrack obtains a clean estimate of the distance travelled by the signal from the transmit antenna to the human reflector, and back to one of the receive antennas. Let us call this estimate the round trip distance. At any time, there are three such round trip distances that correspond to the three receive antennas. The goal of this section is to use these three estimates to identify the 3D position of the human, for each time instance.

To do so, WiTrack leverages its knowledge of the placement of the antennas. Recall that the antennas are placed in a T, as in Fig. 4-1(a) where the y-axis is a horizontal line orthogonal to the plane of the T and the z-axis is along its vertical line. WiTrack uses this



(a) **2D Localization** (b) **3D Localization**

Figure 4-4: **WiTrack's Localization Algorithm.** The TOF estimate from a receive antenna defines an ellipse whose foci are the transmit antenna and the receive antenna. (a) shows that WiTrack can uniquely localize a person using the intersection of two ellipses. (b) shows that in 3D, the problem translates into an intersection of three ellipsoids.

reference frame to track the 3D location of a moving target.

Let us focus on identifying the location at a particular time t_i. Also for clarity, let us first assume that we would like to localize the person in the 2D plane defined by the x and y axes. Consider the transmit antenna and the first receive antenna. WiTrack knows the round trip distance from the transmit antenna to the person and back to the first receive antenna. The region of feasible 2D locations for the target need to satisfy this constraint; hence, they fall on the periphery of an ellipse, whose foci are collocated with the Tx and Rx1 antennas and its major axis is equal to the round trip distance. Now consider the second receive antenna. WiTrack knows the round trip distance from the Tx to the person and back to Rx2. Similarly, the feasible solutions to this constraint in 2D are on the periphery of another ellipse whose foci are collocated with the Tx and Rx2 antennas and its major axis is equal to the round trip distance to Rx2. Since the correct location is on both ellipses, it is one of the intersection points, as shown in Fig. 4-4(a). In fact, since our antennas are directional, only one of the two intersection points is feasible, which is the one that yields a location in the direction of the antennas' beams.

It is straightforward to generalize the argument to localizing in 3D. Specifically, in 3D space, the round-trip distance defines an ellipsoid whose two foci are the transmit antenna and one of the receive antennas. In this setting, the intersection of two ellipsoids would define an arc in 3D space, and hence is insufficient to pinpoint the 3D location of a person. However, by adding a third directional antenna, we obtain a unique solution in 3D that is within the beam of all the directional antennas as shown in Fig. 4-4(b). Therefore, our algorithm can localize a person in 3D by using three directional receive antennas.

Finally we note two points:

- The T-shape placement for the antennas is chosen because we assume the user wants to localize motion behind a wall, in which case all the antennas would have to be arranged in one plane facing the wall. We place one antenna below to help determine elevation, while the others are on the same level.
- While the minimum number of Rx antennas necessary to resolve a 3D location is three, adding more antennas would result in more constraints. This would allow us to over-constrain the solution and hence add extra robustness to noise.

■ 4.5 Beyond 3D Tracking

In this section, we build on WiTrack's 3D localization primitive to enable two additional capabilities: estimating a pointing direction from the corresponding arm movement, and detecting a fall.

■ 4.5.1 Estimation of Pointing Angle

We explain how WiTrack provides coarse estimation of body part motion. We consider the following motion: the user starts from a state where her arm is resting next to her body. She raises the arm in a direction of her choice with the intention of pointing toward a device or appliance, and then drops her hand to the first position. The user may move around and at a random time perform the pointing gesture. We require, however, that the user be standing (i.e., not walking) when performing the pointing gesture. The goal is to detect the pointing direction.

To track such a pointing gesture, WiTrack needs to distinguish between the movement of the entire body and the motion of an arm. To achieve this goal, we leverage the fact that the reflection surface of an arm is much smaller than the reflection surface of an entire human body. We estimate the size of the reflection surface from the spectrogram of the received signal at each of the antennas. Fig. 4-5 illustrates the difference between the spectrogram of a whole body motion and that of an arm pointing, as captured by one of WiTrack's receiving antennas. In the figure the human was moving then stopped and performed the pointing gesture. The two bright spots around $t = 18$s and $t = 21$s refer to the

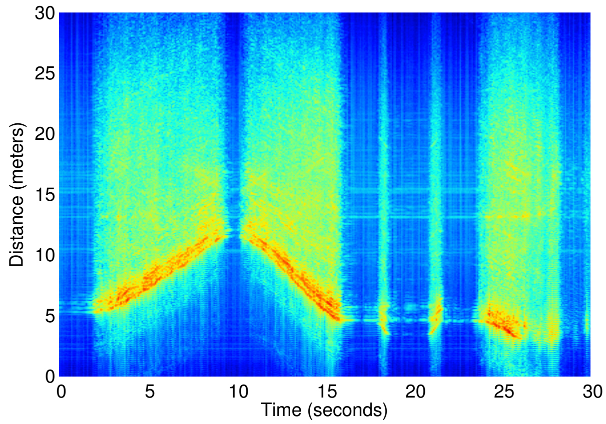


Figure 4-5: **Gestures**. The figure shows a human moving then stopping and pointing with her arm. The small bright regions around $t = 18$s and $t = 21$s correspond to the arm lifting and dropping motions.

arm being lifted and dropped respectively. The figure shows that the signal variance along the vertical axis is significantly larger when the reflector is the entire human body than when it is just an arm motion (note the bright yellow as opposed to the cyan color). If the reflector is large, its parts have slightly different positions from each other; hence, at any point in time the variance of its reflection along the y-axis is larger than that of an arm movement. WiTrack uses this spatial variance to detect body part motion from a whole body motion.

Once we detect it is a body part, WiTrack tries to estimate the direction of the motion to identify the pointing direction, which involves the following steps:

1. *Segmentation:* The goal of segmentation is to determine the start and end of a pointing gesture. Fig. 4-5 shows how WiTrack segments the round trip distance spectrogram obtained from each receive antenna. In our pointing experiments, we ask the user to remain static for one second before performing the pointing gesture. Thus, we are able to detect the start of a pointing gesture since it is always preceded by a period of absence of motion. Similarly, after a person raises her arm in a pointing direction, we ask her to wait for a second before resting her arm back to its initial position. Because WiTrack performs a frequency sweep every 2.5 ms, we can easily distinguish the silence at the start and end of a gesture.
2. *Denoising:* As is the case for a whole body motion, the contour of the segmented spectrogram is denoised and interpolated (see §4.3.4) to obtain a clean estimate of the round trip distance of the arm motion as a function of time, for each receive antenna.
3. *Determining the Pointing direction:* We perform robust regression on the location estimates

of the moving hand, and we use the start and end points of the regression from all of the antennas to solve for the initial and final position of the hand. WiTrack estimates the direction of pointing as the direction from the initial state to the final extended state of the hand. Since the user drops her hand after pointing, WiTrack repeats the above steps for this drop motion obtaining a second estimate of the pointing direction. Then, WiTrack estimates the pointing direction as the middle direction between the two.[5] Being able to leverage the approximate mirroring effect between the arm lifting and arm dropping motions adds significant robustness to the estimation of the pointing angle.

We envision that an application of the estimation of pointing direction can be to enable a user to control household appliances by simply pointing at them. Given a list of instrumented devices and their locations, WiTrack would track the user's hand motion, determine the direction in which she points, and command the device to change its mode (e.g., turn on or off the lights, or control our blinds).

Finally, to demonstrate the pointing gesture within the context of an application, we created a setup where the user can control the operation mode of a device or appliance by pointing at it. Based on the current 3D position of the user and the direction of her hand, WiTrack automatically identifies the desired appliance from a small set of appliances that we instrumented (lamp, computer screen, automatic shades). Our instrumentation is a basic mode change (turn on or turn off). WiTrack issues a command via Insteon home drivers [5] to control the devices. We envision that this setup can evolve to support a larger set of functionalities and be integrated within a home automation systems [54].

■ 4.5.2 Fall Detection

Our objective is to automatically distinguish a fall from other activities including sitting on the ground, sitting on a chair and walking. To do so, we build on WiTrack's elevation tracking along the z dimension. Note that simply checking the person's elevation is not sufficient to distinguish falls from sitting on the floor. To detect a fall, WiTrack requires two conditions to be met: First, the person's elevation along the z axis must change significantly (by more than one third of its value), and the final value for her elevation must be close to the ground level. The second condition is the change in elevation has to occur within a

[5] by zooming on Fig. 4-5 the reader can see how the arm lifting and dropping motions approximately mirror each other's tilt.

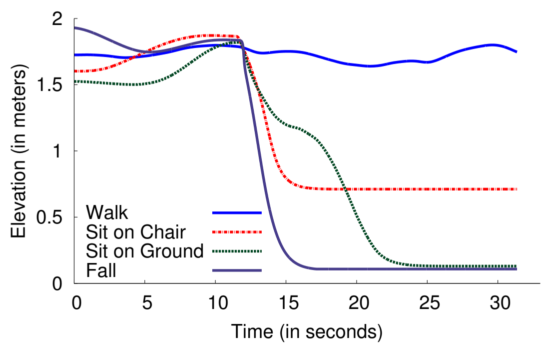


Figure 4-6: **Fall Detection**. WiTrack automatically detects falls by monitoring the absolute value and the change in elevation.

very short period to reflect that people fall more quickly than they sit.

Fig. 4-6 plots WiTrack's estimate of the elevation along the z dimension for four activities: a person walking, sitting on a chair, sitting on the ground, and (simulated) falling on the ground.[6] The figure confirms that walking and sitting on a chair can be identified from falling and sitting on the floor based on elevation because the final elevation is far from $z = 0$. However, to distinguish a fall on the ground from a sitting on the ground, one has to exploit that during a fall the person changes her elevation faster than when she voluntarily sits on the floor.

4.6 Implementation

FMCW Radio Front-End Hardware: We built an FMCW front-end that operates as a daughterboard for the USRP software radio [13]. Below, we describe our design, which is illustrated in the schematic of Fig. 4-7.

The first step of our front end design is the generation of an FMCW signal, which consists of a narrowband signal whose carrier frequency is linearly swept over a large bandwidth. This signal can be obtained by using a voltage-controlled oscillator (VCO). Because the output frequency of a VCO is a linear function of its input voltage, we can generate our desired frequency sweep by feeding a voltage sweep as an input to the VCO. However, small errors in the input voltage can create large non-linearities in the output sweep.

[6] The fall was performed in a padded room as detailed in §4.8.5.

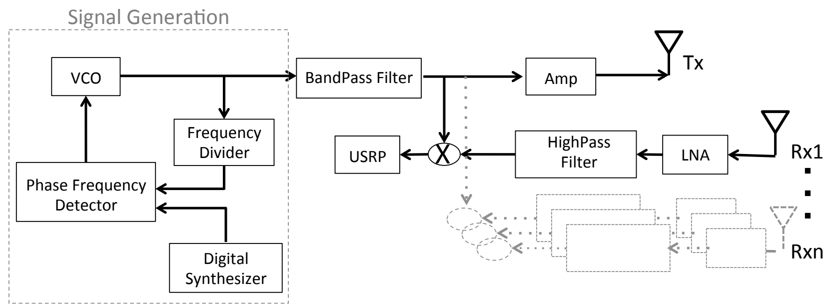


Figure 4-7: **Schematic of the Front End Design.** WiTrack's front end consists of an FMCW signal generation component, and a receive chain that is connected to a USRP.

To obtain a highly linear sweep, we use a feedback mechanism. Specifically, we use a phase frequency detector to compare the output frequency of the VCO with a highly accurate reference signal, and use the offset between the two to control the VCO. Note that even though the reference signal needs to be highly accurate, it does not need to span the same bandwidth as our desired output signal. In particular, rather than directly comparing the output of the VCO to the reference signal, we first use a frequency divider. This allows us to use a reference signal that sweeps from 136.5–181.25 MHz to generate an FMCW signal that sweeps from 5.46–7.25 GHz. This FMCW signal is transmitted over the air using WA5VJB directional antennas [15] after filtering and amplification.

At the receive chain, the transmitted signal is captured using WA5VJB directional antennas and passed through a low-noise amplifier and a high-pass filter to improve its SNR. Recall from §4.3 that an FMCW receiver determines the TOF by measuring the frequency offset between the transmitted and the received signal. This offset can be obtained by downconverting (mixing) the received signal with the transmitted signal. The output of the mixer is then fed to the LFRX-LF daughterboard on USRP2 which samples it at 1 MHz and passes the digitized samples to the UHD driver.

Real-time Software Processing: The implemented prototype performs real-time 3D motion tracking as described in §4.3, §4.4 and §4.5. Tracking is implemented directly in the UHD driver of the USRP software radio. The signal from each receiving antenna is transformed to the Frequency domain using an FFT whose size matches the FMCW sweep period of 2.5ms. To improve resilience to noise, every five consecutive sweeps are averaged creating one FFT frame. Background subtraction is performed by subtracting the averaged FFT frame from the frame that precedes it. The spectrogram is processed for contour tracking by identifying for each time instance the smallest local frequency maximum that is significantly higher than the noise level. Outlier rejection is performed by declaring that the contour should not jump significantly between two successive FFT frames (because a person cannot move much in 12.5ms). The output is smoothed with a Kalman filter.

To locate a person, instead of solving a system of ellipsoid equations in real-time, we leverage that the location of the antennas does not change and is known a priori. Thus, before running our experiments, we use MATLAB's symbolic library to find a symbolic representation of the solutions (x, y, z) as a function of symbolic TOF to each of the receiving antennas. This means that the ellipsoid equations need to be solved only once (for any fixed antenna positioning), independent of the location of the tracked person. After it obtains the 3D location of a person, WiTrack uses python's matplotlib library to output this location in real-time.

Software processing has a total delay less than 75 ms between when the signal is received an a corresponding 3D location is output.

■ 4.7 Evaluation

We empirically evaluate the performance of the WiTrack prototype by conducting experiments in our lab building with 11 human users.

(a) Ground Truth: We determine WiTrack's localization accuracy by testing it against the VICON motion capture system. The VICON is a multi-hundred-thousand dollar system used in filmmaking and video game development to track the human motion and map it to a 3D character animation model. It uses calibrated infrared cameras and records motion by instrumenting the tracked body with infrared-reflective markers. The VICON system has a sub-centimeter accuracy and hence we use it to determine the ground truth location. To track a moving person with the VICON, she is asked to wear a jacket and a hat, which are instrumented with eight infrared markers. To track a subject's hand, she is asked to wear a glove that is also instrumented with six markers. The VICON tracks the infrared markers on the subject's body and fits them to a 3D human model to identify the subject's location.

The VICON system has a built-in capability that can track the center of any object using the infrared-reflective markers that are placed on that object. This allows us to determine the center position of a human subject who is wearing the instrumented jacket and hat. WiTrack however computes the 3D location of the body surface where the signal reflects. In order to compare WiTrack's measurements to those obtained by the VICON, we need to have an estimate of the depth of the center with respect to the body surface. Thus, we use the VICON to run offline measurements with the person standing and having infrared markers around her body at the same height as the WiTrack transmit antenna (about the waist). We use the VICON to measure the average depth of the center from surface for each person. To compare the 3D location computed by the two systems, we first compensate for the average distance between the center and surface for that person and then take the Euclidean distance.

(b) Device Setup: WiTrack is placed behind the wall of the VICON room. The device uses one transmit antenna and three receive antennas. The transmit antenna and two receive antennas are lined up parallel to the wall, and a third receive antenna is placed below the transmit antenna. The distance between the transmit antenna and each receive antenna is 1m, unless otherwise noted.

(c) Human Subjects: The experiments are performed with eleven human subjects: two females and nine males. The subjects are of different heights and builds, and span an age range of 22 to 56 years. In each experiment, the subject is asked to move at will in the VICON room; he/she is tracked using both the VICON system and WiTrack. Note that WiTrack tracks the subject through the wall, from an adjacent room, while the VICON has to be within direct line of sight from the subject.

4.8 Performance Results

4.8.1 Accuracy of 3D Tracking

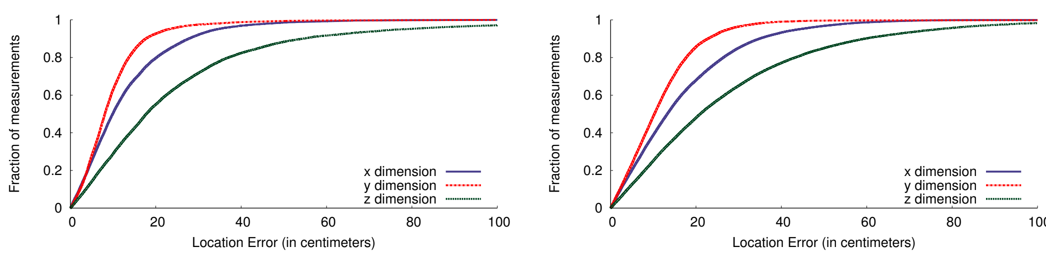
We first focus on the developed 3D tracking primitive and evaluate its accuracy across all three dimensions.

We run 100 experiments each lasting for 1 minute, during which a human subject moves at will in the VICON room. The VICON room has no windows. It has 6-inch hollow walls supported by steel frames with sheet rock on top, which is a standard setup for office buildings. The WiTrack prototype is placed outside the room with all transmit and receive antennas facing one of the walls of the VICON room. Recall that WiTrack's antennas are directional; hence, this setting means that the radio beam is directed toward the wall of the VICON room. In each experiment, we ask the subject to wear the jacket and hat that were instrumented with VICON markers and move inside the VICON-instrumented room. The subject's location is tacked by both the VICON system and WiTrack.

We note that the VICON IR cameras are set to accurately track the target only when she moves in a $6 \times 5\ m^2$ area in the room. Their accuracy degrades outside that area. Since VICON provides the ground truth in our experiment, we ask the target to stay within the $6 \times 5\ m^2$ area where the IR cameras are focused. This area is about 2.5m away from the wall. Thus, the minimum separation between WiTrack and the human subject in these experiments is $3\ m$ and the maximum separation is about $9\ m$.

We perform a total of 100 experiments for this evaluation, each lasting for one minute. Since each FMCW sweep lasts for 2.5ms and we average 5 sweeps to obtain for each TOF measurement, we collect a total of about 480,000 location readings from these 100 experiments.

To show that WiTrack works correctly both in line of sight and through a wall, we repeat the above 100 experiments with one modification, namely we move the WiTrack device inside the room and set it next to the wall from the inside.



(a) **CDF in line-of-sight** (b) **CDF through-wall**

Figure 4-8: **Performance of WiTrack's 3D Tracking.** (a) and (b) show the CDF of the location error for WiTrack in line-of-sight and through-wall scenarios respectively.

Fig. 4-8(a) and Fig. 4-8(b) plot the CDFs of the location error along the x, y, and z coordinates. The figure reveals the following findings:

- WiTrack's median location error for the line-of-sight experiments is 9.9 cm, 8.6 cm, and 17.7 cm along the x, y, and z dimensions respectively. In comparison, the median location error in the through-wall experiments is 13.1 cm, 10.25 cm, and 21.0 cm along the x, y, and z dimensions. As expected the location accuracy in line-of-sight is higher than when the device is behind a wall due to the extra attenuation and the reduced SNR. In both cases, however, the median error is fairly small. This is due to the use of an FMCW radio which ensures a highly accurate TOF estimate, and the ability to prevent errors due to multipath and noise, allowing the system to stay accurate as it moves from TOF to a 3D location estimate of the human body.

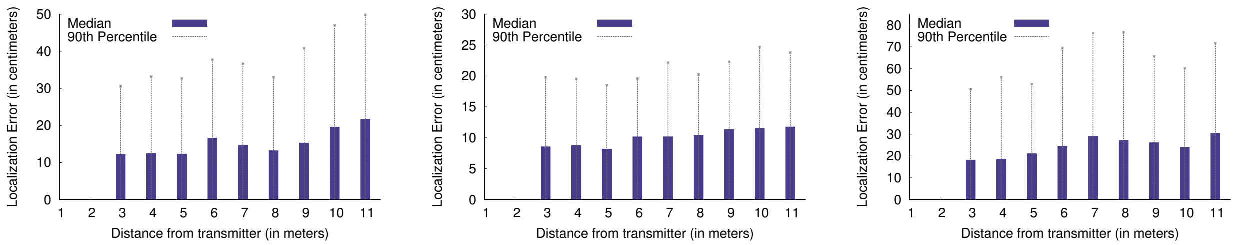
- Interestingly, the accuracy in the y dimension is better than the accuracy in the x dimension. This difference is because the x and y dimensions are not equal from the perspective of WiTrack's antennas. Recall that in the xy-plane, WiTrack's antennas are all along the x-axis. As a result, the two ellipses in the xy-plane, shown in Fig. 4-8, both have their major radius along x and minor radius along y. Hence, the same error in TOF produces a bigger component when projected along the x axis than along the y axis.
- The accuracy along the z-dimension is worse than the accuracy along the x and y dimensions. This is the result of the human body being larger along the z dimension than along x or y.

4.8.2 Accuracy Versus Distance

We are interested in evaluating WiTrack's accuracy as the person gets further away from the device. Thus, we repeat the above through-wall experiments. As mentioned above, VICON requires the human to move in a certain space that is in line of sight of the IR cameras. Thus, to increase the distance from WiTrack to the human we move WiTrack away in the hallway next to the VICON room. Again, we collect 100 experiments, each spanning one minute for a total of 480,000 location measurements.

Fig. 4-9 plots WiTrack's localization error as a function of its distance to the subject. The distance to the subject is determined using the VICON ground-truth coordinates, and rounded to the nearest meter. The figure shows the median and 90^{th} percentile of the estimation error for the x, y, and z coordinates.

The figure shows that the median accuracy changes by 5 to 10 cm for distances that are 3 to 11 m away from the device. As expected, the further the human moves from the device, the larger the estimation error. This increase in error with distance is expected



(a) **Accuracy in x-dimension** (b) **Accuracy in y-dimension** (c) **Accuracy in z-dimension**

Figure 4-9: **3D Localization Accuracy Versus Distance to Device.** (a)-(c) show the location error along the x, y, and z dimensions as a function of how far the subject is from WiTrack. The median and 90^{th} percentile errors increase as the distance from the device to the person increases.

since as the distance gets larger the signal gets more attenuated. However, a second reason stems from the geometry of the ellipsoid-based localization model. Given the equations of the ellipsoid, the TOF multiplied by the speed of light is equal to the major axis of the ellipsoid/ellipse that describes the user's location, and the antenna separation is the distance between the foci. For a fixed antenna separation, as the distance/TOF increases the ellipsoid's surface increases, increasing the overall space of potential locations.

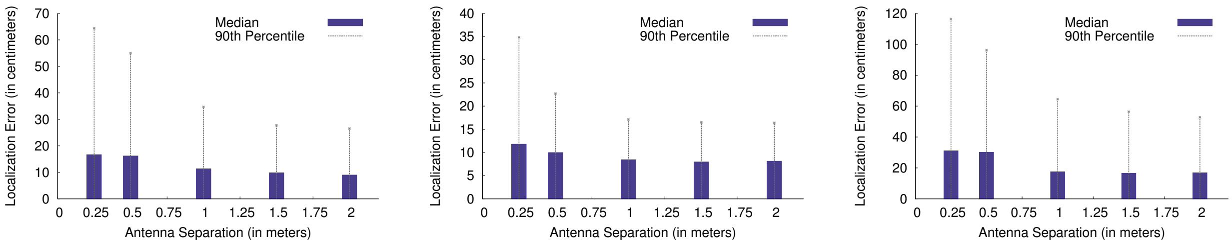
The figure also shows that the accuracy is best along the y dimension, then the x, and finally the z, which is due to the reasons discussed in the previous section.

■ 4.8.3 Accuracy Versus Antenna Separation

Our default setting places the receive antennas 1 m away from the transmit antenna. In this section, we examine the impact of antenna separation on performance.

We evaluate five different configurations. In all of these configurations, the transmit antenna is at an equal distance from all receive antennas, and is placed at the crossing point of a "T" whereas the receive antennas are placed at the edges. We vary the distance between the transmit antenna and each of the receive antennas from 25 cm to 2 m. We run 100 one-minute experiments, 20 for each antenna setting. All experiments are run through a wall. In each experiment, we ask the human subject to move at will inside the VICON room, as we record her location using both the VICON system and WiTrack.

Fig. 4-10 shows WiTrack's localization accuracy as a function of antenna separation. The figure shows that even if one brings the antennas to within 25cm of each other, the median location error stays less than 17 cm, 12 cm, and 31 cm for the x, y, and z dimensions, and 90^{th} percentile becomes 64cm, 35cm, and 116cm respectively. While this is higher than the previous results where the antennas were separated by 1 m, it is still comparable to state of the art localization using a WiFi transmitter (in our case, the user does not need to carry any wireless device).



(a) **Accuracy in x-dimension** (b) **Accuracy in y-dimension** (c) **Accuracy in z-dimension**

Figure 4-10: **3D Localization Accuracy Versus Size of Device.** (a)-(c) show the median and 90^{th} percentile location errors as a function of the antenna separation. Along all three dimensions, a larger separation leads to a decrease in the location error.

The plots show that as the antenna separation increases, the localization accuracy improves along all three dimensions x, y, and z. This behavior is expected, because the further the receive antennas are from each other, the larger the spatial diversity between them. Because of the geometric nature of the algorithm, a spatially diverse setup would lead to a smaller intersection curve between any pair of ellipsoids.[7] For this reason, in a larger setup, the same noise variance in the TOF estimates would be confined to a smaller curve, thus, minimizing estimate error.

Mathematically, for any TOF, the antenna separation is the distance between the foci of the ellipsoid that defines the person's location. Hence for any given TOF, increasing the antenna separation increases the distance between the foci while keeping the ellipsoid's major radius constant. Hence the ellipsoid gets more squashed and its circumference becomes smaller, reducing the region of potential solutions.

■ 4.8.4 Accuracy of Estimating Pointing Direction

In the experiments in this section, the human subjects wear a glove that is instrumented with infrared reflexive markers, and are asked to stand in a given location inside the VI-

[7]For simplicity, consider the 2D case with 1 Tx and 2 Rx antennas. Because of the system's resolution, each ellipse has some fuzzy region about it (i.e., a thickness of $+/\epsilon$, where ϵ is determined by the resolution). Thus, the intersection of two ellipses is a region rather than a single point. This region becomes larger when the Rx antennas are closer to each other, and the larger the region, the larger the ambiguity in localization. In the extreme case where the two receive antennas are co-located, the two ellipses perfectly overlap and the ambiguity region is large.

CON room and point in a direction of their choice. Each pointing gesture consists of raising the subject's hand in the direction of her choice, followed by the subject returning her hand to its original resting position. Across our experiments, we ask the human subjects to stand in random different locations in the VICON room and perform the pointing gesture. We determine the direction in which the subject pointed by using both the VICON recordings and WiTrack's estimates (see §4.5.1).

Fig. 4-11 plots a CDF of the error between the angle as determined by WiTrack and the ground truth angle based on the VICON measurements. The figure shows that the median orientation error is 11.2 degrees, and the 90^{th} percentile is 37.9 degrees. These results suggest that WiTrack provides an enabling primitive to track pointing gestures. We used this capability to enable controlling different household appliances like shades, lamps and computer screens by sending commands to these different appliances over Insteon drivers.

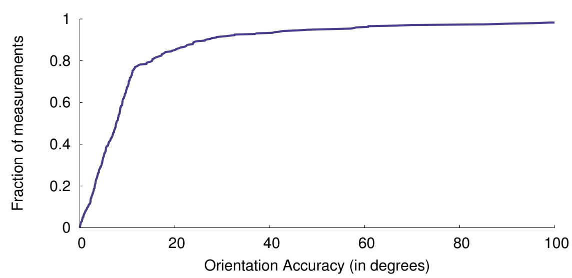


Figure 4-11: **Orientation Accuracy**. The CDF of the orientation accuracy shows that the median orientation error is 11.2 degrees, and the 90^{th} percentile error is 37.9 degrees.

■ 4.8.5 Fall Detection

We test the fall detection algorithm described in §4.5.2 by asking different participants to perform four different activities: walk, sit on a chair, sit on the floor, and simulate a fall. The floor of the VICON room is already padded. We add extra padding to ensure no injury can be caused by simulated falls. We perform 132 experiments in total, 33 for each activity. We log the data files from each of these experiments and process them offline with our fall detection algorithm. We obtain the following results:

- None of the walking or sitting on a chair activities are classified as falls.
- One of the sitting on the floor experiments was classified as a fall.
- Two out of 33 simulated falls were not detected (they were misclassified as sitting on the floor).

Thus, the precision of the fall detection algorithm is 96.9% (since out of the 32 detected falls only 31 are true falls), and the recall is 93.9% (since out of 33 true falls we detected 31). This yields an F-measure of 94.4%.

4.9 Discussion & Limitations

3D motion tracking based purely on RF reflections off a human body is a challenging technical problem. We believe WiTrack has taken an important step toward addressing this problem. However, the version of WiTrack described in this chapter still has limitations, which we address in the remaining chapters of this dissertation:

Tracking one person: Our current design can track only one person at any point in time. This does not mean that WiTrack requires only one person to be present in the environment. Other people can be around, but they have to be behind the directional antennas. We address this limitation in Chapter 5 to enable WiTrack to track multiple persons at the same time.

Requiring motion: A second limitation stems from the fact that WiTrack needs the user to move in order to locate her. This is because WiTrack receives reflections from all static objects in the environment; hence, it cannot distinguish the static user from a piece of furniture. To eliminate these static reflectors, WiTrack subtracts consecutive FMCW sweeps. Unfortunately, that eliminates the reflections of the static user as well. Chapter 5 addresses this limitation as well and shows how we can localize static users in the environment.

Distinguishing between body parts: Currently WiTrack can provide coarse tracking of the motion of one body part. The tracked part has to be relatively large like an arm or a leg. WiTrack however does not know which body part has moved, e.g., it cannot tell whether it is an arm or a leg. In our experiments, the users were pointing with their arms. Extending this basic capability to tracking more general movements of body parts will likely require incorporating complex models of human motion. In particular, Kinect's ability to track body parts is the result of the combination of 3D motion tracking using infrared with complex vision algorithms and advanced models of human motion [132]. Chapter 6 discusses how we can enable distinguish between different moving body from RF reflections.

Often when presenting this research, the issues of privacy and security come up. While not central to this thesis, these issues are central to society, and we have given them considerable thought. Our belief is that addressing these issues is two-fold: the first is legal and the second is technological. From a legal perspective, the use of such technology without the monitored person's consent will likely be deemed illegal under US law. On one hand, law enforcement use will be conditioned by obtaining a search warrant (similar to the Kyllo v United States supreme court ruling in 2001 [19]); on the other hand, use by ordinary consumers will be governed by *Peeping Tom* laws which prevent their use without consent. Technologically, we can ensure our privacy with widespread adoption of such technologies by incorporating challenge-response mechanisms in the user interface, to ensure that the monitored user provides explicit consent. Furthermore, in the face of an adversary, one could potentially deploy RF countermeasure that may discover and jam such technologies. Finally, it is worth noting that WiTrack itself may be used for providing security services (e.g., intrusion detection or home security).

CHAPTER 5

Multi-Person Localization via RF Body Reflections

Recall that WiTrack could only localize a single person in the environment, and it required the person to move in order to be tracked. In this Chapter, we introduce our second version, WiTrack2.0, which transcends both of these limitations. Specifically, WiTrack2.0 accurately localizes multiple users in the environment. It does so by disentangling the reflections of wireless signals that bounce off their bodies. Furthermore, it neither requires prior calibration nor that the users move in order to localize them.

To achieve its goal, WiTrack2.0 has to deal with multiple challenges. As with traditional device-based localization, the most difficult challenge in indoor environments is the multipath effect [86, 160]. Specifically, wireless signals reflect off all objects in the environment making it hard to associate the incoming signal with a particular location. To overcome this challenge, WiTrack (of Chapter 4) focuses on motion to capture signal reflections that change with time. It then assumes that only one person is present in the environment, and hence all motion can be attributed to him. However, if multiple people move in the environment or if the person is static, then this assumption no longer works.

To address this challenge, we observe that the indoor multipath varies significantly when it is measured from different vantage points. Hence, one can address this problem by positioning multiple transmit and receive antennas in the environment, and measuring the time of flight from each of these transmit-receive antenna pairs. However, the signals emitted from the different transmitters will reflect off the bodies of the all the users in the environment, and these reflections interfere with each other leading to wireless collisions. In §5.3, we show how WiTrack2.0 disentangles these interfering reflected signals to localize multiple users in the presence of heavy indoor multipath.

A second challenge that WiTrack2.0 has to address is related to the near-far problem. Specifically, reflections off the nearest person can have much more power than distant reflections, obfuscating the signal from distant people, and preventing their detection or tracking. To address this issue, we introduce Successive Silhouette Cancellation (SSC) an approach to address the near-far problem, which is inspired by successive interference cancellation. This technique starts by localizing the closest person, then eliminates his impact on the received signal, before proceeding to localize further users (who have weaker reflections). It repeats this process iteratively until it has localized all the users in a scene. Note, however, that each user is not a point reflector; hence, his wireless reflection has a complex structure that must be taken into account, as we describe in §5.4.

A third challenge that our system addresses is related to localizing static users. Specifically, recall that WiTrack needs to eliminate reflections off static objects by subtracting consecutive measurements. However, this subtraction also results in eliminating the reflections off static users. To enable us to localize static users, we exploit the fact these users still move slightly due to their breathing. However, the breathing motion is fairly slow in comparison to body motion. Specifically, the chest moves by a sub-centimeter distance over a period of few seconds; in contrast, a human would pace indoors at 1 m/s. Hence, WiTrack2.0 processes the reflected signals at multiple time scales that enable it to accurately localize both types of movements as we describe in §5.5,

We have built a prototype of WiTrack2.0, using USRP software radios and an analog FMCW radio. We run experiments both in line-of-sight (LOS) scenarios and non-line-of-sight (NLOS) scenarios, where the device is in a different room and is tracking people's motion through the wall. Empirical results from over 300 experiments with 11 human subjects show the following:

- *Motion Tracking:* WiTrack2.0 accurately tracks the motion of up to four users simultaneously, without requiring the users to hold or wear any wireless device. In an area that spans $5\ m \times 7\ m$, its median error across all users is 12.1cm in the x/y dimensions.

- *Localizing Static People:* By leveraging their breathing motion, WiTrack2.0 accurately localizes up to five static people in the environment. Its median error is 11.2 cm in the x/y dimensions across all the users in the scene.
- *Tracking Hand Movements:* WiTrack2.0's localization capability extends beyond tracking a user's body to tracking body parts. We leverage this capability to recognize concurrent gestures performed in 3D space by multiple users. In particular, we consider a gesture in which three users point in different directions at the same time. Our WiTrack2.0 prototype detects the pointing directions of all three users with a median accuracy of $10.3°$.

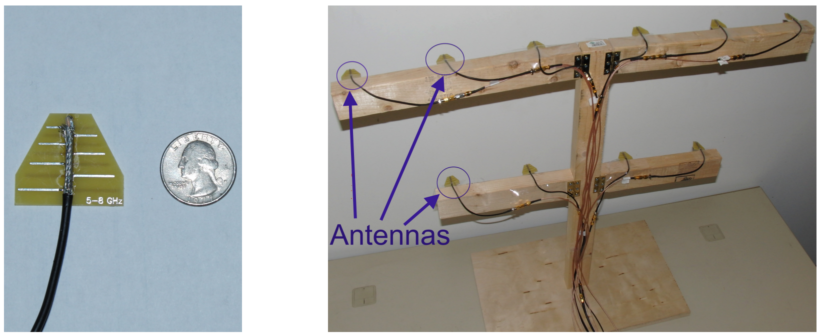
■ 5.1 Related Work

Besides past work highlighted in and §4.1, WiTrack2.0's successive silhouette cancellation algorithm is related to past literature on integrative cancellation frameworks. Specifically, the framework of iteratively identifying and canceling out the strongest components of a signal is widely used in many domains. Naturally, however, the details of how the highest power component is identified and is eliminated varies from one application to another. In the communications community, we refer to such techniques as successive interference cancellation, and they have been used in a large number of applications such as ZigZag [67], VBLAST [167], and full duplex [35]. In the radio astronomy community, these techniques are referred to as CLEAN algorithms and, similarly, have a large number of instantiations [77, 144, 146]. Our work on successive silhouette cancellation also falls under this framework and is inspired by these algorithms. However, in comparison to all the past work, it focuses on identifying the reflections of the humans in the environment and canceling them by taking into account the different vantage points from which the time-of-flight is measured as well as the fact that the human body is not a point reflector.

■ 5.2 WiTrack2.0 Overview

WiTrack2.0 is a wireless system that can achieve highly accurate localization of multiple users in multipath-rich indoor environments, by relying purely on the reflections of wireless signals off the users' bodies. For static users, it localizes them based on their breathing, and can also localize the hand motions of multiple people, enabling a multi-user gesture-based interface.

WiTrack2.0 is a multi-antenna system consisting of five transmit antennas and five receive antennas, as shown in Fig. 5-1. The antennas are directional, stacked in a single plane, and mounted on a foldable platform as shown in Fig. 5-1(b). This arrangement is chosen because it enables see-through-wall applications, whereby all the antennas need to be lined up in a plane facing the wall of interest.



(a) **Antenna** (b) **Antenna Setup**

Figure 5-1: **WiTrack2.0's Antennas and Setup.** (a) shows one of WiTrack2.0's directional antenna ($3cm \times 3.4cm$) placed next to a quarter; (b) shows the antenna setup in our experiments, where antennas are mounted on a $2m \times 1m$ platform and arranged in a single vertical plane.

WiTrack2.0 operates by transmitting RF signals and capturing their reflections after they bounce off different users in the environment. Algorithmically, WiTrack2.0 has two main components: 1) Multi-shift FMCW, a technique that enables it to deal with multipath effects, and (2) Successive Silhouette Cancellation (SSC), an algorithm that allows WiTrack2.0 to overcome the near-far problem. The following sections describe these components.

■ 5.3 Multi-shift FMCW

Multipath is the first challenge in accurate indoor localization. Specifically, not all reflections that survive background subtraction correspond to a moving person. This is because the signal reflected off the human body may also reflect off other objects in the environment before arriving at the receive antenna. As this person moves, this multipath reflection also moves with him and survives the background subtraction step. In single-user localization, one may eliminate this type of multipath by leveraging that these secondary reflections travel along a longer path before they arrive at the receive antenna. Specifically, by electing the smallest TOF after background subtraction, one may identify the round-trip distance to the user.

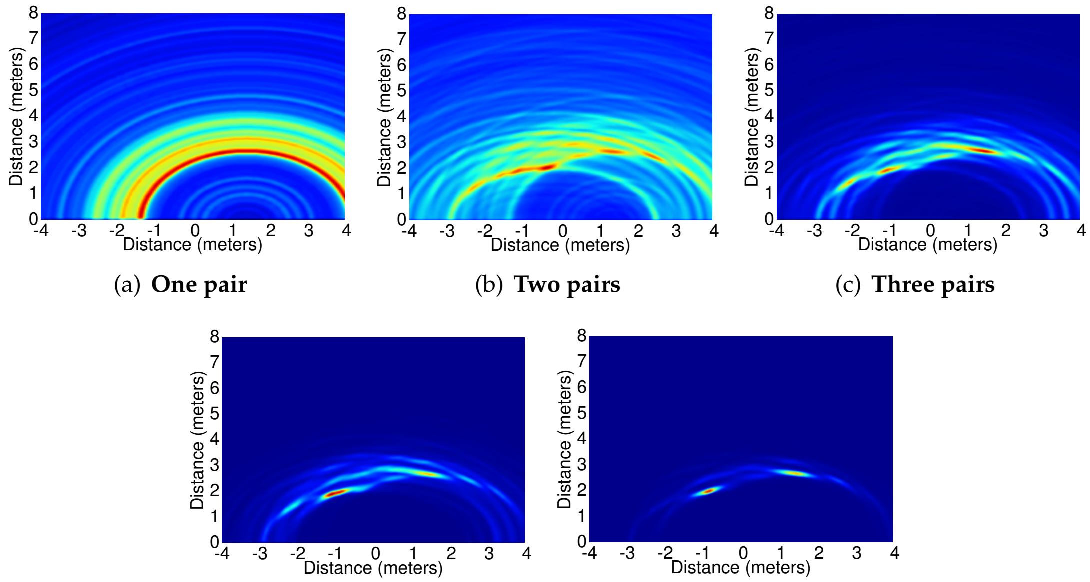
However, the above invariant does not hold in multi-person localization since different users are at different distances with respect to the antennas, and the multipath of a nearby user may arrive earlier than that of a more distant user, or even interfere with it. In this section, we explore this challenge in more details, and show how we can overcome it by obtaining time-of-flight measurements from different vantage points in the environment.

5.3.1 Addressing Multi-path in Multi-User Localization

To explore the above challenge in practice, we run an experiment with two users in a 5×7 m furnished room (with tables, chairs, etc.) in a standard office building. Recall from §4.3 that each transmit-receive antenna pair provides us with a ***TOF profile*** – i.e., it tells us how much reflected power we obtain at each possible TOF between the transmit antenna and receive antenna – and that each such TOF corresponds to an ellipse in 2D (as in Fig. 4-4(a)). We study what happens as we successively overlay ellipses from different transmit-receive pairs.

Now let us map all TOFs in a TOF profile to the corresponding round trip distances using Eq. 4.4, and hence the resulting ellipses. This process produces a heatmap like the one in Fig. 5-2(a), where the x and y axes correspond to the plane of motion. For each ellipse in the heatmap, the color in the image reflects the amount of received power at the corresponding TOF. Hence, the ellipse in red corresponds to a strong reflector in the environment. The orange, yellow, and green ellipses correspond to weaker reflections respectively; these reflections could either be due to another person, multi-path reflections of the first person, or noise. The blue regions in the background correspond to the absence of reflections from those areas.

Note that the heatmap shows a pattern of half-ellipses; the foci of these ellipses are the transmit and receive antennas, both of which are placed along the $y = 0$ axis. The reason we only show the upper half of the ellipses is that we are using directional antennas, and we focus them towards the positive y direction. Hence, we know that we do not receive reflections from behind the antennas.



(a) **One pair** (b) **Two pairs** (c) **Three pairs**

(d) **Four pairs** (e) **Five pairs**

Figure 5-2: **Increasing the Number of Tx-Rx pairs enables Localizing Multiple Users.** The figure shows the heatmaps obtained from combining TOF profiles of multiple Tx-Rx antenna pairs in the presence of two users. The x/y axes of each heatmap correspond to the real world x/y dimensions.

Fig. 5-2(a) shows the ellipses corresponding to the TOF profiles from one Tx-Rx pair. Now, let us see what happens when we superimpose the heatmaps obtained from two Tx-Rx pairs. Fig. 5-2(b) shows the heatmap we obtain when we overlay the ellipses of the first transmit-receive pair with those from a second pair. We can now see two patterns of ellipses in the figure, the first pattern resulting from the TOFs of the first pair, and the second pattern due to the TOFs of the second pair. These ellipses intersect in multiple locations, resulting in red or orange regions, which suggest a higher probability for a reflector to be in those regions. Recall that there are two people in this experiment. However, Fig. 5-2(b) is not enough to identify the locations of these two people.

Figs. 5-2(c) and 5-2(d) show the result of overlaying ellipses from three and four Tx-Rx pairs respectively. The figures show how the noise and multi-path from different antennas averages out to result in a dark blue background. This is because different Tx-Rx pairs have different perspectives of the indoor environment; hence, they do not observe the same noise or multi-path reflections. As a result, the more we overlay heatmaps from different Tx-Rx pairs, the dimmer the multipath effect, and the clearer the candidate locations for the two people in the environment.

Next, we overlay the ellipses from five transmit-receive pairs and show the resulting heatmap in Fig. 5-2(e). We can now clearly see two bright spots in the heatmap: one is red and the other is orange, whereas the rest of the heatmap is mostly a navy blue background indicating the absence of reflectors. Hence, in this experiment, we are able to localize the two users using TOF measurements from five Tx-Rx pairs. Combining these measurements together allowed us to eliminate the multipath effects and localize the two people passively using their reflections.

Summary: As the number of users increases, we need TOF measurements from a larger number of Tx-Rx pairs to localize them, and extract their reflections from multipath. For the case of two users, we have seen a scenario whereby the TOFs of five Tx-Rx pairs were sufficient to accurately localize both of them. In general, the exact number would depend on multipath and noise in the environment as well as on the number of users we wish to localize. These observations motivate a mechanism that can provide us with a large number of Tx-Rx pairs while scaling with the number of people in the environment.

5.3.2 The Design of Multi-shift FMCW

In the previous section, we showed that we can localize two people by overlaying many heatmaps obtained from mapping the TOF profiles of multiple Tx-Rx pairs to the corresponding ellipses. But how do we obtain TOFs from many Tx-Rx pairs? One option is to use one FMCW transmitter and a large number of receivers. In this case, to obtain N Tx-Rx pairs, we would need one transmitter and N receivers. The problem with this approach is that it needs a large number of receivers, and hence does not scale well as we add more people to the environment.

A more appealing option is to use multiple FMCW transmit and receive antennas. Since the signal transmitted from each transmit antenna is received by all receive antennas, this allows us to obtain N Tx-Rx pairs using only $\sqrt{N}$ transmit antennas and $\sqrt{N}$ receive antennas.

However, the problem with this approach is that the signals from the different FMCW transmitters will interfere with each other over the wireless medium, and this interference will lead to localization errors. To see why this is true, consider a simple example where we want to localize a user, and we have two transmit antennas, Tx1 and Tx2, and one receive antenna Rx. The receive antenna will receive two reflections – one due to the signal transmitted from Tx1, and another due to Tx2's signal. Hence, its TOF profile will contain two spikes referring to two time-of-flight measurements TOF_1 and TOF_2.

With two TOFs, we should be able to localize a single user based on the intersection of the resulting ellipses. However, the receiver has no idea which TOF corresponds to the reflection of the FMWC signal generated from Tx1 and which corresponds to the reflection of the FMCW signal generated by Tx2. Not knowing the correct Tx means that we do not know the foci of the two ellipses and hence cannot localize. For example, if we incorrectly associate TOF_1 with Tx2 and TOF_2 with Tx1, we will generate a wrong set of ellipses, and localize the person to an incorrect location. Further, this problem becomes more complicated as we add more transmit antennas to the system. Therefore, to localize the user, WiTrack2.0 needs a mechanism to associate these TOF measurements with their corresponding transmit antennas.

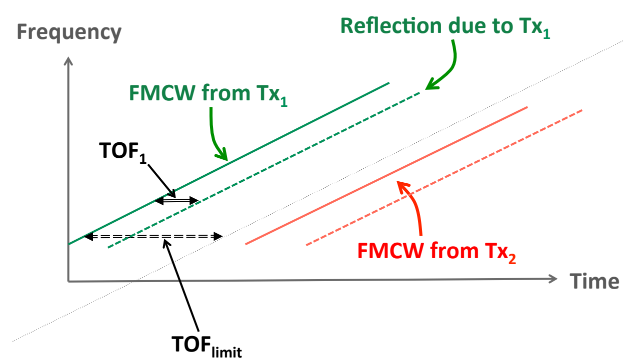


Figure 5-3: **Multi-shift FMCW**. WiTrack2.0 transmits FMCW signals from different transmit antennas after inserting virtual delays between them. Each delay must be larger than the highest time-of-flight (TOF_{limit}) due to objects in the environment.

We address this challenge by leveraging the structure of the FMCW signal. Recall that FMCW consists of a continuous linear frequency sweep as shown by the green line in Fig. 5-3. When the FMCW signal hits a body, it reflects back with a delay that corresponds to the body's TOF. Now, let us say TOF_{limit} is the maximum TOF that we expect in the typical indoor environment where WiTrack2.0 operates. We can delay the FMCW signal from the second transmitter by $\tau > TOF_{limit}$ so that all TOFs from the second transmitter are shifted by τ with respect to those from the first transmitter, as shown by the red line in Fig. 5-3. Thus, we can prevent the various FMCW signals from interfering by ensuring that each transmitted FMCW signal is time shifted with respect to the others, and those shifts are significantly larger than the time-of-flight to objects in the environment. We refer to this design as Multi-shift FMCW.

As a result, the receiver would still compute two TOF measurements: the first measurement (from Tx1) would be TOF_1, and the second measurement (from Tx2) would be $TOF_2' = TOF_2 + \tau$. Knowing that the TOF measurements from Tx2 will always be larger than τ, WiTrack2.0 determines that TOF_1 is due to the signal transmitted by Tx1, and TOF_2' is due to the signal transmitted by Tx2.

This idea can be extended to more than two Tx antennas, as shown in Fig. 6-10. Specifically, we can transmit the FMCW signal directly over the air from Tx1, then shift it by τ and transmit it from Tx2, then shift it by 2τ and transmit it from Tx3, and so on. At the receive side, all TOFs between 0 and τ are mapped to Tx1, whereas distances between τ and 2τ are mapped to Tx2, and so on.

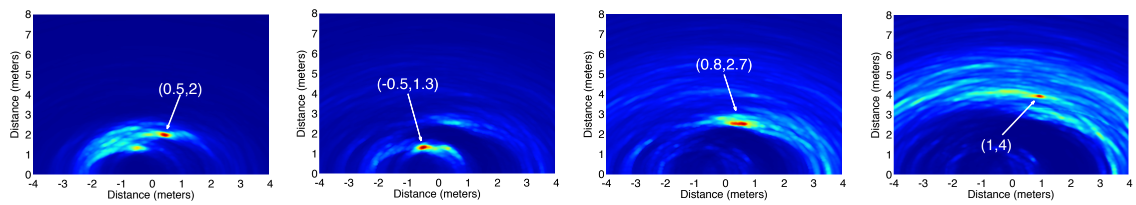
Summary: Multi-shift FMCW has two components: the first component allows us to obtain TOF measurements from a large number of Tx-Rx pairs; the second component operates on the TOFs obtained from these different Tx-Rx pairs by superimposing them into a 2D heatmap, which allows us to localize multiple users in the scene.

■ 5.4 Successive Silhouette Cancellation

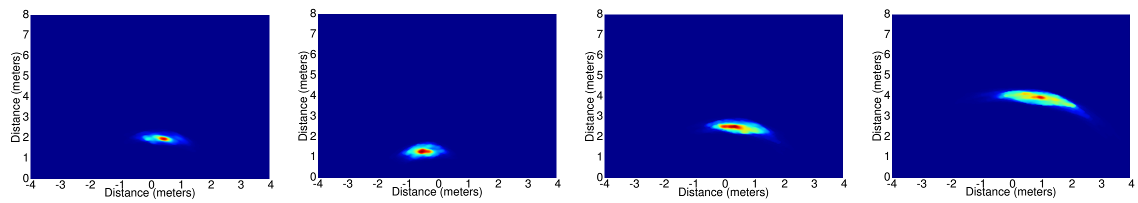
With multi-shift FMCW, we obtain TOF profiles from a large number of Tx-Rx pairs, map them to 2D heatmaps, overlay the heatmaps, and start identifying users' locations. However, in practice this is not sufficient because different users will exhibit the near-far problem. Specifically, reflections of a nearby user are much stronger than reflections of a faraway user or one behind an obstruction.

Fig. 5-4(a) illustrates this challenge. It shows the 2D heatmap obtained in the presence of four persons in the environment. The heatmap allows us to localize only two of these persons: one is clearly visible at $(0.5, 2)$, and another is fairly visible at $(-0.5, 1.3)$. The other two people, who are farther away from WiTrack2.0, are completely overwhelmed by the power of the first two persons.

To deal with this near-far problem, rather than localizing all users in one shot, WiTrack2.0 performs Successive Silhouette Cancellation (SSC) which consists of 4 steps:



(a) **Detect** 1^{st} **Person** (b) **Detect** 2^{nd} **Person** (c) **Detect** 3^{rd}**Person** (d) **Detect** 4^{th} **Person**



(e) **Focus on** 1^{st} **Person** (f) **Focus on** 2^{nd} **Person** (g) **Focus on** 3^{rd} **Person** (h) **Focus on** 4^{th} **Person**

Figure 5-4: **Successive Silhouette Cancellation.** (a) shows the 2D heatmap obtained by combining all the TOFs in the presence of four users. (b)-(d) show the heatmaps obtained after canceling out the first, second, and third user respectively. (e)-(h) show the result of the SSC focusing step on each of the users, and how it enables us to accurately localize each person while eliminating interference from all other users.

1. *SSC Detection:* finds the location of the strongest user by overlaying the heatmaps of all Tx-Rx pairs.
2. *SSC Re-mapping:* maps a person's location to the set of TOFs that would have generated that location at each transmit-receive pair.
3. *SSC Cancellation:* cancels the impact of the person from the TOF profiles of all Tx-Rx pairs.
4. *Iteration:* re-computes the heatmaps using the TOF profiles after cancellation, overlays them, and proceeds to find the next strongest reflector.

We now describe each of these steps in detail by walking through the example with four persons shown in Fig. 5-4.

SSC Detection. In the first step, SSC finds the location of the highest power reflector in the 2D heatmap of Fig. 5-4(a). In this example, the highest power is at $(0.5, 2)$, indicating that there is a person in that location.

SSC Re-mapping. Given the (x, y) coordinates of the person, we map his location back to the corresponding TOF at each transmit-receive pair. Keep in mind that each person is not a point reflector; hence, we need to estimate the spread of reflections off his entire body on the TOF profile of each transmit-receive pair.

To see how we can do this, let us look at the illustration in Fig. 5-5 to understand the effect of a person's body on one transmit-receive pair. The signal transmitted from the transmit antenna will reflect off different points on the person's body before arriving at the receive antenna. Thus, the person's reflections will appear between some TOF_{min} and TOF_{max} in the TOF profile at the Rx antenna.

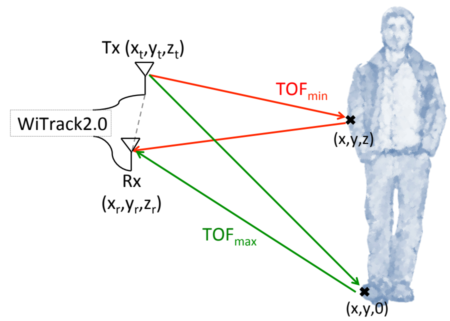


Figure 5-5: **Finding** TOF_{min} **and** TOF_{max}. TOF_{min} is determined by the round-trip distance from the Tx-Rx pair to the closest point on the person's body. Since the antennas are elevated, TOF_{max} is typically due to the round-trip distance to the person's feet.

Note that TOF_{min} and TOF_{max} are bounded by the closest and furthest points respectively on a person's body from the transmit-receive antenna pair. Let us first focus on how we can obtain TOF_{min}. By definition, the closest point on the person's body is the one that corresponds to the shortest round-trip distance to the Tx-Rx pair, where the round-trip distance is the summation of the forward path from Tx to that point and the path from that point back to Rx. Formally, for a Tx antenna at $(x_t, 0, z_t)$, an Rx antenna at $(x_r, 0, z_r)$,[1] we can compute d_{min} as:

$$\min_z \sqrt{(x_t - x)^2 + y^2 + (z_t - z)^2} + \sqrt{(x_r - x)^2 + y^2 + (z_r - z)^2} \tag{5.1}$$

where (x, y, z) is any reflection point on the user's body. One can show that this expression is minimized when:

$$\frac{z - z_t}{z - z_r} = -\sqrt{\frac{(x_t - x)^2 + y^2}{(x_r - x)^2 + y^2}} \tag{5.2}$$

Hence, using the detected (x, y) position, we can solve for z then substitute in Eq. 5.1 to obtain d_{min}.

Similarly, TOF_{max} is bounded by the round-trip distance to point on the person's body that is furthest from the Tx-Rx pair. Again, the x and y coordinates of the furthest point are

[1]Recall that all the antennas are in the vertical plane $y = 0$, which is parallel to a person's standing height.

determined by the person's location from the SSC Detection step. However, we still need to figure out the z coordinate of this point. Since the transmitter and receiver are both raised above the ground (at around 1.2 meters above the ground), the furthest point from the Tx-Rx pair is typically at the person's feet. Therefore, we know that the coordinates of this point are $(x, y, 0)$, and hence we can compute d_{max} as:

$$d_{max} = \sqrt{(x_t - x)^2 + (y)^2 + z_t^2} + \sqrt{(x_r - x)^2 + (y)^2 + z_r^2}.$$

Finally, we can map d_{min} and d_{max} to TOF_{min} and TOF_{max} by dividing them by the speed of light C.

SSC Cancellation. The next step is to use TOF_{min} and TOF_{max} to cancel the person's reflections from the TOF profiles of each transmit-receive pair. To do that, we take a conservative approach and remove the power in all TOFs between TOF_{min} and TOF_{max} within that profile. Of course, this means that we might also be partially canceling out the reflections of another person who happens to have a similar time of flight to this Tx-Rx pair. However, we rely on the fact that multi-shift FMCW provides a large number of TOF profiles from many Tx-Rx pairs. Hence, even if we cancel out the power in the TOF of a person with respect to a particular Tx-Rx pair, each person will continue to have a sufficient number of TOF measurements from the rest of the antennas.

We repeat the process of computing TOF_{min} and TOF_{max} with respect of each Tx-Rx pair and cancelling the power in that range, until we have eliminated any power from the recently localized person.

Iteration. We proceed to localize the next person. This is done by regenerating the heatmaps from the updated TOF profiles and overlaying them. Fig. 5-4(b) shows the obtained image after performing this procedure for the first person. Now, a person at $(-0.5, 1.3)$ becomes the strongest reflector in the scene.

We repeat the same procedure for this user, canceling out his interference, then reconstructing a 2D heatmap in Fig. 5-4(c) using the remaining TOF measurements. Now, the person with the strongest reflection is at $(0.8, 2.7)$. Note that this heatmap is noisier than Figs. 5-4(a) and 5-4(b) because now we are dealing with a more distant person.

WiTrack2.0 repeats the same cancellation procedure for the third person and constructs

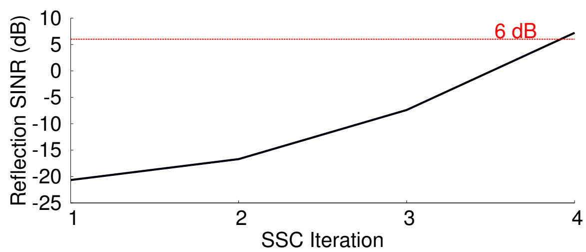


Figure 5-6: **SINR of the Farthest User Throughout SSC Iterations.** The figure shows how the SINR of the farthest user increases with each iteration of the SSC algorithm. After the first, second and third person are removed from the heatmap images in Figs.5-4(a)-(c), the SINR of the fourth person increases to 7dB, allowing us to detect his presence.

the 2D heatmap in Fig. 5-4(d). The figure shows a strong reflection at $(1, 4)$. Recall that our antennas are placed along the $y = 0$ axis, which means that this is indeed the furthest person in the scene. Also note that the heatmap is now even noisier. This is expected because the furthest person's reflections are much weaker. WiTrack2.0 repeats interference cancellation for the fourth person, and determines that the SNR of the maximum reflector in the resulting heatmap does not pass a threshold test. Hence, it determines that there are only four people in the scene.

We note that each of these heatmaps are scaled so that the highest power is always in red and the lowest power is in navy blue; this change in scale emphasizes the location of the strongest reflectors and allows us to better visualize their locations. To gain more insight into the power values and to better understand how SSC improves our detection of further away users, Fig. 5-6 plots the Signal to Interference and Noise Ratio (SINR) of the fourth person during each iteration of SSC. The fourth user's SINR initially starts at -21dB and is not visible in Fig. 5-4(a). Once the first and second users are removed by SSC, the SINR increases to -7dB and we can start detecting the user's presence in the back of Fig. 5-4(c). Performing another iteration raises the fourth person's SINR above the noise floor to 7dB. It also brings it above our threshold of 6dB – i.e., twice the noise floor – making him detectable.

We perform four additional steps to improve SSC:

- *Refocusing Step:* After obtaining the initial estimates of the locations of all four persons, WiTrack2.0 performs a focusing step for each user to refine his location estimate. This is done by reconstructing an interference-free 2D heatmap only using the range in the TOF profiles that corresponds to TOFs between TOF_{min} and TOF_{max} for that Tx-Rx pair.

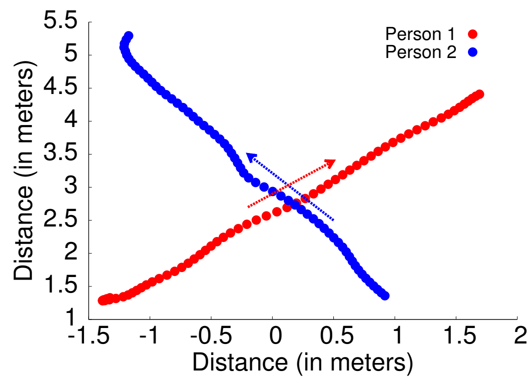


Figure 5-7: **Disentangling Crossing Paths**. When two people cross paths, they typically keep going along the same direction they were going before their paths crossed.

Figs. 5-4(e)- 5-4(h) show the images obtained from this focusing step. In these images, the location of each person is much clearer,[2] which enables higher-accuracy localization.

- *Leveraging Motion Continuity:* After obtaining the estimates from SSC, WiTrack2.0 applies a Kalman filter and performs outlier rejection to reject impractical jumps in location estimates that would otherwise correspond to abnormal human motion over a very short period of time.
- *Disentangling Crossing Paths:* To disentangle multiple people who cross paths, we look at their direction of motion before they crossed paths and project how they would proceed with the same speed and direction as they are crossing paths. This helps us with associating each person with his own trajectory after crossing. Fig. 5-7 shows an example with two people crossing paths and how we were able to track their trajectories despite that. Of course, this approach does not generalize to every single case, which may lead to some association errors after the crossings but not to localization errors.
- *Extending SSC to 3D Gesture Recognition:* Similar to our earlier version of WiTrack presented in Chapter 4, WiTrack2.0 can differentiate a hand motion from a whole-body motion (like walking) by leveraging the fact that a person's hand has a much smaller reflective surface than his entire body. Unlike the earlier version, however, WiTrack2.0 can track gestures even when they are simultaneously performed by multiple users. Specifically, by exploiting SSC focusing, it zooms onto each user individually to track his gestures. In our evaluation, we focus on testing a pointing gesture, where different users point in different directions at the same time. By tracking the trajectory of each moving hand, we can

[2]This is because all other users' reflections are eliminated, while, without refocusing, only users detected in prior iterations are eliminated.

determine its pointing direction. Note that we perform these pointing gestures in 3D and track hand motion by using the TOFs from the different Tx-Rx pairs to construct a 3D point cloud rather than a 2D heatmap.[3] The results in §5.8.3 show that we can accurately track hand gestures performed by multiple users in 3D space.

5.5 Localization Based on Breathing

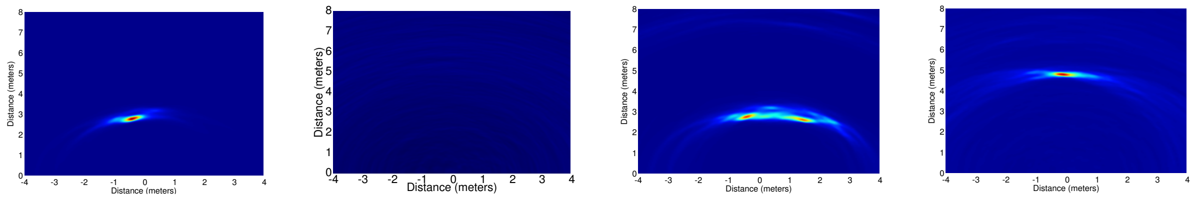
We extend WiTrack2.0's SSC algorithm to localize static people based on their breathing. Recall from Chapter 4 that in order for WiTrack to localize a user based on her radio reflections, we need to eliminate reflections off all static objects in the environment (like walls and furniture). This is typically achieved by performing a background subtraction step, i.e., by taking TOF profiles from adjacent time windows and subtracting them from each other.[4]

Whereas this approach enables us to track moving people, it prevents us from detecting a static person – e.g., someone who is standing or sitting still. Specifically, because a static person remains in the same location, his TOF does not change, and hence his reflections would appear as static and will be eliminated in the process of background subtraction. To see this in practice, we run two experiments where we perform background subtraction by subtracting two TOF profiles that are 12.5 milliseconds apart from each other. The first experiment is performed with a walking person and the resulting heatmap is shown in Fig. 5-8(a), whereas the second experiment is performed in the presence of a person who is sitting at $(0, 5)$ and the resulting heatmap is shown in Fig. 5-8(b). These experiments show how the heatmap of a moving person after background subtraction would allow us to localize him accurately, whereas the heatmap of the static person after background subtraction does not allow us to localize the person.

To localize static people, one needs to realize that even a static person moves slightly due to breathing. Specifically, during the process of breathing, the human chest moves by a sub-centimeter distance over a period of few seconds. The key challenge is that this change does not translate into a discernible change in the TOF of the person. However,

[3]Recall from §4.4 that a given TOF maps to an ellipse in 2D and an ellipsoid in 3D. The intersection of ellipsoids in 3D allow us to track these pointing gestures.

[4]Recall that we obtain one TOF profile by taking an FFT over the received FMCW signal in baseband. Since the FMCW signal is repeatedly swept, we can compute a new TOF profile from each sweep.



(a) **Short subtraction window localizes a walking person.** (b) **Short subtraction window misses a static person.** (c) **Long subtraction window smears a walking person.** (d) **Long subtraction window localizes a static person.**

Figure 5-8: **Need For Multiple Subtraction Windows.** The 2D heatmaps show that a short subtraction window accurately localizes a pacing person in (a) but not a static person in (b). A long subtraction window smears the walking person's location in (c) but localizes a breathing person in (d).

over an interval of time of a few seconds (i.e., as the person inhales and exhales), it would result in discernible changes in the reflected signal. Therefore, by subtracting frames in time that are few seconds apart, we should be able to localize the breathing motion.

In fact, Fig. 5-8(d) shows that we can accurately localize a person who is sitting still by using a subtraction window of 2.5 seconds. Note, however, that this long subtraction window will introduce errors in localizing a pacing person. In particular, since typical indoor walking speed is around 1 m/s [37], subtracting two frames that are 2.5 seconds apart would result in smearing the person's location and may also result in mistaking him for two people as shown in Fig. 5-8(c).

Thus, to accurately localize both static and moving people, WiTrack2.0 performs background subtraction with different subtraction windows. To localize moving users, it uses a subtraction window of 12.5 ms. On the other hand, normal adults inhale and exhale over a period of 3–6 seconds [155] causing their TOF profiles to change over such intervals of time. Hence, we consider the first TOF profile during each 10-second interval, and subtract it from all subsequent TOF profiles during that interval. As a result, breathing users' reflections pop up at different instances, allowing us to detect and localize them.

5.6 Implementation

We built WiTrack2.0 using a single FMCW radio (hardware design was described in §4.6) whose signal is fed into multiple antennas. The FMCW radio generates a low-power (sub-milliWatt) signal that sweeps 5.46-7.25 GHz every 2.5 milliseconds. The range and power are chosen in compliance with FCC regulations for consumer electronics [18].

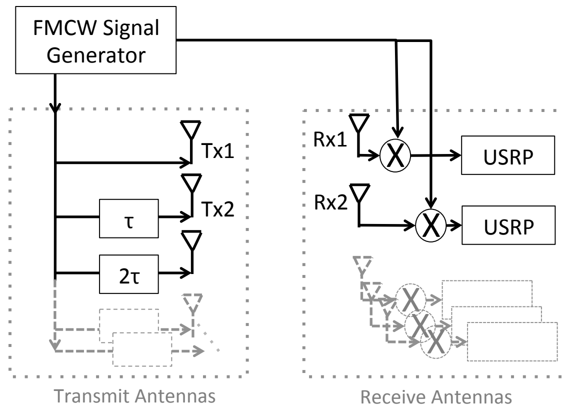


Figure 5-9: **Multi-shift FMCW Architecture**. The generated FMCW signal is fed to multiple transmit antennas via different delay lines. At the receive side, the TOF measurements from the different antennas are combined to obtain the 2D heatmaps.

The schematic in Fig. 5-9 shows how we use this radio to implement Multi-shift FMCW. Specifically, the generated sweep is delayed before being fed to directional antennas for transmission.[5] At the receive side, the signal from each receive antenna is mixed with the FMCW signal and the resulting signal is fed to the USRP. The USRP samples the signals at 2 MHz and transfers the digitized samples to the UHD driver. These samples are processed in software to localize users and recognize their gestures.[6]

The analog FMCW radio and all the USRPs are driven by the same external clock. This ensures that there is no frequency offset between their oscillators, and hence enables subtracting frames that are relatively far apart in time to enable localizing people based on breathing.

■ 5.7 Evaluation

Human Subjects. We evaluate the performance of WiTrack2.0 by conducting experiments in our lab with eleven human subjects: four females and seven males. The subjects differ

[5]The most straightforward option to delay the signal is to insert a wire. However, wires attenuate the signal and introduce distortion over the wide bandwidth of operation of our system, reducing its SNR. Instead, we exploit the fact that, in FMCW, time and frequency are linearly related; hence, a shift τ in time can be achieved through a shift $\Delta f = slope \times \tau$ in the frequency domain. Hence, we achieve this delay by mixing FMCW with signals whose carrier frequency is Δf. This approach also provides us with the flexibility of tuning multi-shift FMCW for different TOF_{limit}'s by simply changing these carrier frequencies.

[6]Complexity-wise, WiTrack2.0's algorithms are linear in the number of users, the number of Tx antennas, and the number of Rx antennas.

in height from 165–185 cm as well as in weight and build and span 20 to 50 years of age. The subjects wear their daily attire. In each experiment, each subject is allowed to move as they wish throughout the room.

Ground Truth. We use the VICON motion capture system to provide us with ground truth positioning information [14]. It consists of an array of infrared cameras that are fitted to the ceiling of a $5\ m \times 7\ m$ room, and requires instrumenting any tracked object with infrared-reflective markers. When an instrumented object moves, the system tracks the infrared markers on that object and fits them into a 3D model to identify the object's location.

We evaluate WiTrack2.0's accuracy by comparing it to the locations provided by the VICON system. To track a user using the VICON, we ask him/her to wear a hard hat that is instrumented with five infrared markers. In addition, for the gestures experiments, we ask each user to wear a glove that is instrumented with six markers.

Experimental Setup. We evaluate WiTrack2.0 in a standard office environment that has standard furniture: tables, chairs, boards, computers, etc. We experiment with two setups: line-of-sight and through-the-wall. In the through-wall experiments, WiTrack2.0 is placed outside the VICON room with all transmit and receive antennas facing one of the walls of the room. Recall that WiTrack2.0's antennas are directional and hence this setting means that the radio beam is directed toward the room. The VICON room has no windows; it has 6-inch hollow walls supported by steel frames, which is a standard setup for office buildings. In the line-of-sight experiments, we move WiTrack2.0 inside the room. In all of these experiments, the subjects' locations are tracked by both the VICON system and WiTrack2.0.

Detection. Recall that WiTrack2.0 uses iterative cancellation to detect different users in the scene. This limits the number of users it can detect because of residual interference from previous iterations. Therefore, we run experiments to identify the maximum number of people that WiTrack2.0 can reliably detect under various conditions. Detection accuracy is measured as the percentage of time that WiTrack2.0 correctly outputs the number of users present in the environment. The number of users in each experiment is known and acts as the ground truth. We run ten experiments for each of our testing scenarios, and plot the accuracies for each them in Fig. 5-10.

We make two observations from this figure. First, the accuracy of detection is higher

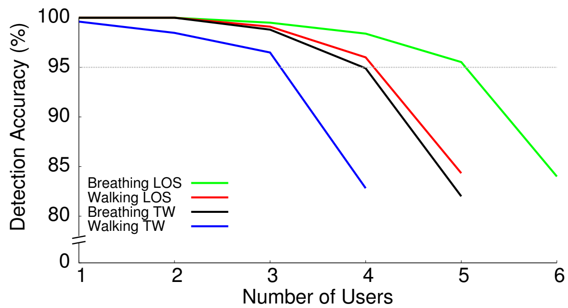


Figure 5-10: **WiTrack2.0 's Detection Accuracy.** The figure shows the percentage of time that WiTrack2.0 accurately determines the number of people in each of our evaluation scenarios.

in line-of-sight than in through-wall settings. This is expected because the wall causes significant attenuation and hence reduces the SNR of the reflected signals. Second, the detection accuracy for breathing-based localization is higher than that of the tracking experiments. While this might seem surprising, it is actually due to the fact that the breathing experiments operate over longer subtraction windows. Specifically, the system outputs the number of people detected and their locations by analyzing the trace over windows of 10 seconds. In contrast, the tracking experiments require outputting a location of each person once every 12.5 ms,[7] and hence they might not be able to detect each person within such a small time window.

For our evaluation of localization accuracy, we run experiments with the maximum number of people that are reliably detectable, where "reliably detectable" is defined as detected an accuracy of 95% or higher. For reference, we summarize these numbers in the table below.

	Line-of-Sight	Through-Wall
Motion Tracking	4	3
Breathing-based Localization	5	4

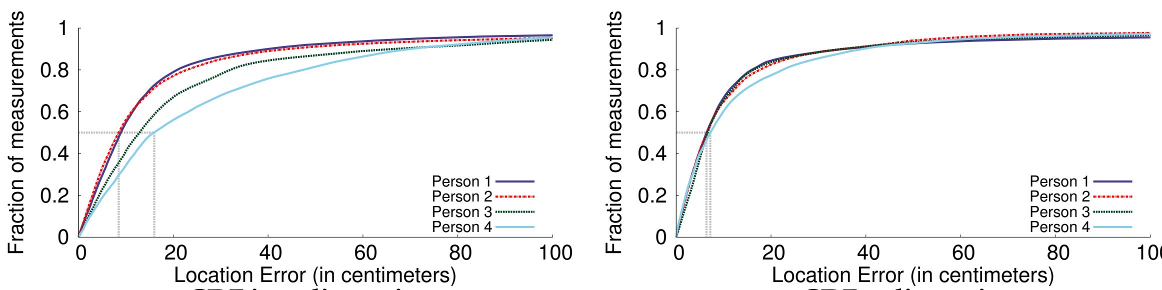
Table 5-1: **Maximum Number of People Detected Reliably.**

[7]Since the user is moving, combining measurements over a longer interval smears his signal.

5.8 Performance Results

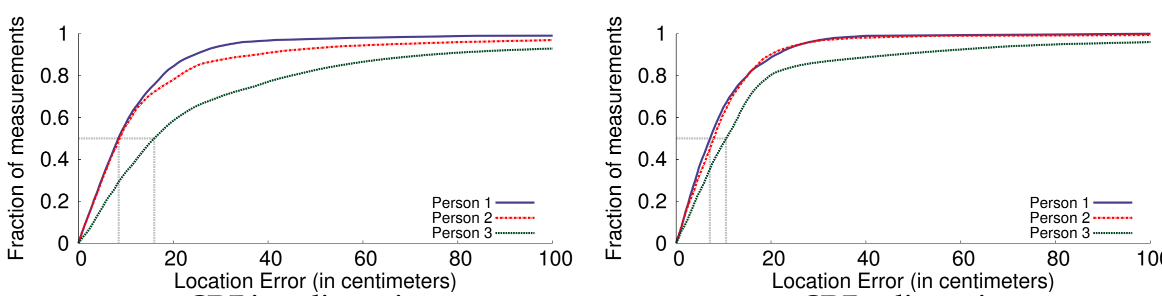
5.8.1 Accuracy of Multi-Person Motion Tracking

We first evaluate WiTrack2.0's accuracy in multi-person motion tracking. We run 100 experiments in total, half of them in line-of-sight and the second half in through-wall settings. In each experiment, we ask one, two, three, or four human subjects to wear the hard hats that are instrumented with VICON markers and move inside the VICON-instrumented room. Each subject's location is tracked by both the VICON system and WiTrack2.0, and each experiment lasts for one minute. Since each FMCW sweep lasts for 2.5ms and we average 5 sweeps to obtain each TOF measurement, we collect around 5,000 location readings per user per experiment.



(a) **CDF in x-dimension** (b) **CDF y-dimension**

Figure 5-11: **Performance of WiTrack2.0's LOS Tracking.** (a) and (b) show the CDFs of the location error in x and y for each of the tracked users in LOS. Subjects are ordered from first to last detected by SSC.



(a) **CDF in x-dimension** (b) **CDF y-dimension**

Figure 5-12: **Performance of WiTrack2.0's Through-Wall Tracking.** (a) and (b) show the CDFs of the location error in x and y for each of the tracked users. Subjects are ordered from first to last detected by SSC.

Figs. 5-11 and 5-12 plot the CDFs of the location error along the x and y coordinates for each of the localized persons in both line-of-sight and through-wall scenarios. The subjects are ordered from the first to the last as detected by SSC. The figures reveal the following findings:

- WiTrack2.0 can accurately track the motion of four users when it is in the same room as the subjects. Its median location error for these experiments is 8.5 cm in x and 6.4 cm in y for the first user detected, and decreases to 15.9 cm in x and 7.2 cm in y for the last detected user.
- In through-wall scenarios, WiTrack2.0 can accurately localize up to three users. Its median location error for these experiments is 8.4 cm and 7.1 cm in x/y for the first detected user, and decreases to 16.1 cm and 10.5 cm in x/y for the last detected user. As expected, the accuracy when the device is placed in the same room as the users is better than when it is placed behind the wall due to the attenuation (reduced SNR) caused by the wall.
- The accuracy in the y dimension is better than the accuracy in the x dimension. This discrepancy is due to the asymmetric nature of WiTrack2.0's setup, where all of its antennas are arranged along the $y = 0$ axis.
- The localization accuracy decreases according to the order the SSC algorithm localizes the users. This is due to multiple reasons: First, a user detected in later iterations is typically further from the device, and hence has lower SNR, which leads to lower accuracy. Second, SSC may not perfectly remove the reflections of other users in the scene, which leads to residual interference and hence lower accuracy.

5.8.2 Accuracy of Breathing-based Localization

We evaluate WiTrack2.0's accuracy in localizing static people based on their breathing. We run 100 experiments in total with up to five people in the room. Half of these experiments are done in line-of-sight and the other half are through-wall. Experiments last for 3-4 minutes. Subjects wear hard hats and sit on chairs in the VICON room.

Figs. 5-13 and 5-14 plot WiTrack2.0's localization error in line-of-sight and through-wall settings as a function of the order with which the subject is detected by the SSC algorithm. The figures show the median and 90^{th} percentile of the estimation error for the x and y coordinates of each of the subjects. The figures show the following results:

- WiTrack2.0's breathing-based localization accuracy goes from a median of 7.24 and 6.3 cm in x/y for the nearest user to 18.31 and 10.85 cm in x/y for the furthest user, in both line-

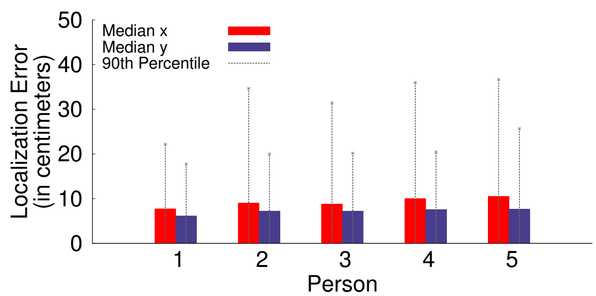


Figure 5-13: **Accuracy for Localizing Breathing People in Line-of-Sight.**. The figure shows show the median and 90^{th} percentile errors in x/y location. Subjects are ordered from first to last detected by SSC.

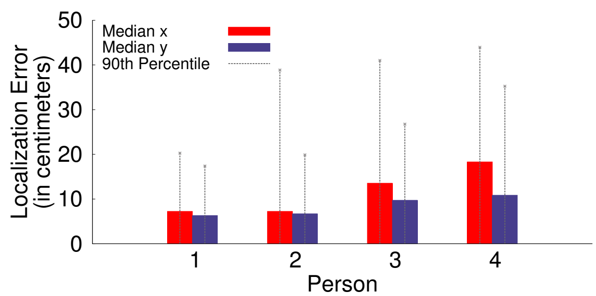


Figure 5-14: **Accuracy for Localizing Breathing People in Through-Wall Experiments.**. The figure shows show the median and 90^{th} percentile errors in x/y location. Subjects are ordered from first to last detected by SSC.

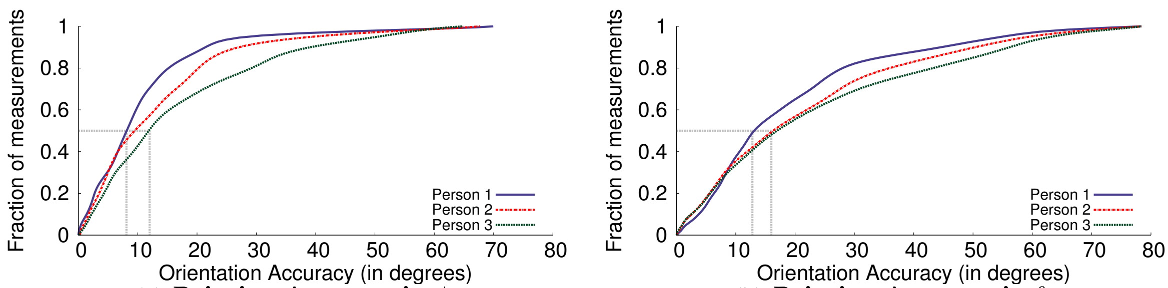
of-sight and through-wall settings.

- Localization based on breathing is more accurate than during motion. This is because when people are static, they remain in the same position, providing us with a larger number of measurements for the same location.

5.8.3 Accuracy of 3D Pointing Gesture Detection

We evaluate WiTrack2.0's accuracy in tracking 3D pointing gestures. We run 100 experiments in total with one to three subjects. In each of these experiments, we ask each subject to wear a glove that is instrumented with infrared-reflective markers, stand in a different location in the VICON room, and point his/her hand in a random 3D direction of their choice – as if they were playing a shooting game or pointing at some household appliance to control it. In most of these experiments, all subjects were performing the pointing gestures simultaneously.

Throughout these experiments, we track the 3D location of the hand using the VICON system and WiTrack2.0. We then regress on the 3D trajectory to determine the direction in which each user pointed (similar to the earlier version of WiTrack described in Chapter 4). Fig. 5-15(a) and 5-15(b) plot the CDFs of the orientation error between the angles as measured by WiTrack2.0 and the VICON for the 1st, 2nd and 3rd participant (in the order of detection by SSC). Note that we decompose the 3D pointing gesture along two directions: azimuthal (in the $x - y$ plane), which we denote as ϕ, and elevation (in the $r - z$ plane), which we denote as θ. The accuracy along both of these angles is important since appliances which the user may want to control in a home environment (e.g., lamps, screens, shades) span 3D space.



(a) **Pointing Accuracy in ϕ** (b) **Pointing Accuracy in θ**

Figure 5-15: **3D Gesture Accuracy.** The figure shows the CDFs of the orientation accuracy for the pointing gestures of each participant. Subjects are ordered from first to last detected by the SSC algorithm.

The figure shows that the median orientation error in ϕ goes from 8.2 degrees to 12.4 degrees from the first to the third person, and from 12 degrees to 16 degrees in θ. Note that WiTrack2.0's accuracy in ϕ is slightly higher than its accuracy in θ. This is due to WiTrack2.0's setup, where the antennas are more spread out along the x than along the z, naturally leading to lower robustness to errors along the z axis, and hence lower accuracy in θ. These experiments demonstrate that WiTrack2.0 can achieve high accuracy in 3D tracking of a pointing gesture.

5.9 Discussion & Limitations

WiTrack2.0 marks an important step toward enabling accurate indoor localization that does not require users to hold or wear any wireless device. WiTrack2.0, however, has some limitations which we highlight below. Some of these limitations are left for future work, while others are addressed in the upcoming chapters of this dissertation.

1. *Number of Users:* WiTrack2.0 can accurately track up to 4 moving users and 5 static users. These numbers may be sufficient for in-home tracking. However, it is always desirable to scale the system to track more users.

2. *Coverage Area:* WiTrack2.0's range is limited to 10m due to its low power. To cover larger areas and track more users, one may deploy multiple devices and hand off the trajectory tracking from one to the next, as the person moves around. Managing such a network of devices, coordinating their hand-off, and arbitrating their medium access are interesting problems to explore.

3. *Lack of Identification:* The system can track multiple users simultaneously, but it cannot identify them. Additionally, it can track limb motion (e.g., a hand) but cannot differentiate between different body parts (a hand vs. a leg). Chapter 6 describes how we can overcome both of these limitations and use RF signals to identify humans as well as to determine which body part a person uses for performing gestures.

4. *Limited Gesture Interface:* WiTrack2.0 focuses on tracking pointing gestures; however, the user cannot move other body parts while performing the pointing gesture. In Chapter 6, we overcome this limitation to enable rich gesture-based interfaces from RF reflections.

Overall, we believe WiTrack2.0 pushes the limits of indoor localization and enriches the role it can play in our daily lives. By enabling smart environments to accurately follow our trajectories, it paves way for these environments to learn our habits, react to our needs, and enable us to control the Internet of Things that revolves around our networked homes and connected environments.

CHAPTER 6

Capturing the Human Figure Through a Wall

Capturing the skeleton of a human body, even with coarse precision, enables many applications in computer graphics, ubiquitous computing, surveillance, and user interaction. For example, solutions such as the Kinect allow a user to control smart appliances without touching any hardware through simple gestures, and can customize their behavior by recognizing the identity of the person. Past work on skeletal acquisition has made significant advances in improving precision; however, all existing solutions require the subject to either carry sensors on his/her body (e.g., IMUs, cameras) or be within the line of sight of an external sensor (e.g., structured light, time of flight, markers+cameras). In contrast, in this chapter, we focus on capturing the human figure – i.e., coarse human skeleton – but without asking the subject to wear any sensor, and even if the person is behind a wall.

To achieve this goal, we build on WiTrack (described in Chapters 4 & 5), which uses RF (Radio Frequency) signals to find the location of a person from behind a wall, without requiring the person to hold or wear any device. Recall that WiTrack operates in a fashion similar to Radar and Sonar, albeit at much lower power. It emits wireless signals at very low power (1/1000 of WiFi) in a frequency range that can traverse walls; the signals reflect off various objects in the environment, including the human body, and it uses these reflections to localize the person at any point in time. However, WiTrack captures very limited

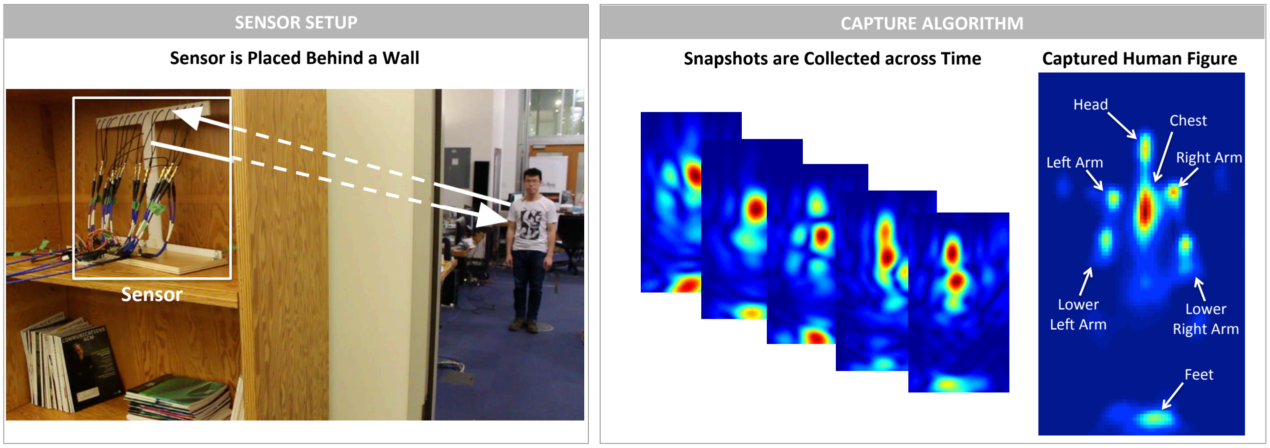


Figure 6-1: **Through-wall Capture of the Human Figure.** The sensor is placed behind a wall. It emits low-power radio signals. The signals traverse the wall and reflect off different objects in the environment, including the human body. Due to the physics of radio reflections, at every point in time, the sensor captures signal reflections from only a subset of the human body parts. We capture the human figure by analyzing multiple reflection snapshots across time and combining their information to recover the various limbs of the human body.

information about the human body. Specifically, it abstracts the whole human body as a single-point reflector, which it tracks.

The challenge in using RF to capture a human figure is that not all body parts reflect the signal back to the sensors. Specifically, at frequency ranges that traverse walls, human limb curves act as ideal reflectors; hence, they may deflect the signal away from the sensors rather than back to them. (This is because RF signals that traverse walls have a wavelength of multiple centimeters, which is larger than the surface roughness of human body parts, causing each part to act as a perfect reflector [34].) At every point in time, the RF sensors capture signals from only a subset of the human body parts, and the sensors lack semantics to understand which body part is reflecting the signal back at that instant. Furthermore, as a person moves, the reflecting limbs vary; for example, at some point, a person's left hand may reflect the signal back but not his right hand or his head, while at other times, his head may reflect the signal back but neither of his hands. To overcome this challenge, past systems that use radar techniques to reconstruct a skeleton require surrounding the human body with a very large antenna array that can capture the reflections off his/her body parts, similar to holographic systems deployed in airports.

In this chapter, we limit ourselves to a compact antenna array that sits in a corner of a room – like a Kinect sensor – and captures the figure of a person behind a wall, as shown in Fig. 6-1. We present RF-Capture, the first system that can capture the human figure when the person is fully occluded (i.e., in the absence of any path for visible light). RF-Capture has two main algorithmic components: The first component is a coarse-to-fine algorithm that efficiently scans 3D space looking for RF reflections of various human limbs and generating 3D snapshots of those reflections. The second component exploits the fact that due to human motion, consecutive RF snapshots tend to expose different body parts and diverse perspectives of the same body part. Thus, this component introduces an algorithm that identifies human body parts from RF snapshots across time, and stitches multiple snapshots together to capture the human figure.

We leverage the captured figure to deliver novel capabilities. First, we show how the captured figure can be incorporated into a classifier to identify different subjects from behind a wall. Our classification accuracy is 95.7% when distinguishing between 5 users, and becomes 88.2% for 15 users. Second, we show that RF-Capture can identify which body part a user moves through a wall with an accuracy of 99.13% when the user is 3 m away and 76.4% when the user is 8 m away. Finally, we show that RF-Capture can track the palm of a user to within a couple of centimeters, tracing letters that the user writes in the air from behind a wall.

We believe the above results present a significant leap towards human figure capture through walls and full occlusion. However, the current system still has limitations. First, our current model assumes that the subject of interest starts by walking towards the device, hence allowing RF-Capture to capture consecutive RF snapshots that expose various body parts. Second, while the system can track individual body parts facing the device, such as a palm writing in the air, it cannot perform full skeletal tracking. This is because not all body parts appear in all RF snapshots. We believe these limitations can be addressed as our understanding of wireless reflections in the context of computer graphics and vision evolves.

6.1 Related Work

RF-Capture is related to past literature in the following areas:

Motion Capture Systems. Past work for capturing the human skeleton relied on motion capture systems that either require instrumenting the human body with markers or operate only in direct line-of-sight to the human body. Specifically, marker-based methods place various types of sensors on the human body – including inertial, infrared, RF, acoustic, or ultrasonic sensors – and capture the human skeleton by tracking these various markers, e.g., [14, 17, 125, 127, 153, 161]. On the other hand, past markerless methods use cameras and infrared-based techniques – including Kinect, multi-view cameras, moving cameras, and time-of-flight cameras – and require a direct line-of-sight from the sensor to the person's body, e.g., [61, 62, 71, 116, 132, 154, 177]. In contrast to all this past work, RF-Capture focuses on capturing coarse human figures without instrumenting the human body with any markers and operates correctly even if the subject is behind a wall or furniture.

Prior art has also investigated motion capture in partial occlusions, e.g., [39, 75, 93, 98, 112, 158]. However, these systems require the majority of the human body to be unoccluded from their sensors, and focus on estimating the positions of occluded limbs or missing markers by fitting a model. In contrast, since RF-Capture uses RF signals that can traverse occlusions, it works even when the person is fully occluded from its sensor, including scenarios where the subject is behind a wall.

Imaging and Reconstruction Algorithms. RF-Capture is related to past work on imaging hidden shapes using light that bounces off corner reflectors in the scene [74, 90, 150]. These past systems operate by estimating the time-of-flight of the object's reflections bouncing off the corner. RF-Capture's reconstruction problem is closely related to such transient imaging techniques; this is because by pointing a time-resolved camera onto a white patch of a wall, that wall essentially becomes a lens-less image sensor with distance information. However, the reconstruction constraints – both in terms of bandwidth and number of sensors – are more stringent in the case of RF-Capture, which limits itself to 20 antennas and less than 2 GHz of bandwidth (while cameras use thousands of pixels and light has hundreds of THz of bandwidth). This allows these transient imaging techniques to achieve higher reconstruction accuracy. Furthermore, in contrast to these systems, RF-Capture only captures specular reflections because of the wavelength of RF signals it uses. However, because it uses RF signals that can traverse occlusions, RF-Capture does not require the placement of corner reflectors in the environment. Furthermore, unlike this past work, it does not require the hidden shape to be fully static during the acquisition time, and hence is evaluated on real human subjects.

Additionally, RF-Capture is related to past work in the Graphics and Vision community on specular object reconstruction [80, 97]. Specifically, for frequencies that traverse walls, reflections off the human body have specular properties. However, past work on specular reconstruction, which operates using visible light, typically assumes the object to be static and non-deformable and aims at recovering surface geometry. In contrast, in RF-Capture, the setting is more complex since the object is moving and deformable, but the goal is simpler since we intend to recover a coarse figure as opposed to surface geometry.

Radar Systems. Radar systems were the first to use RF reflections to detect and track objects. The vast majority of the radar literature focuses on inanimate objects (e.g., planes, metallic structures), as opposed to humans. The radar literature that deals with human subjects can be classified into two categories. The first category is high-frequency imaging radar using terahertz [168], laser [27], or millimeter and sub-millimeter waves [29, 49, 53]. These systems are intrinsically different from ours since they operate at much higher frequencies, where the wavelength is comparable to the roughness of the surface, and hence the human body becomes a scatterer as opposed to a reflector [34]. The advantage of these systems is that they can image the human skeleton at a high accuracy (as in airport terahertz security scanners). However, they operate at much shorter distances, cannot deal with occlusions like wall or furniture, and are expensive and bulky.

The second category uses centimeter-waves, i.e., its carrier frequency is around few GHz, similar to our system. Such designs were discussed in Chapter 2, and we briefly review how they differ from RF-Capture in what follows. These systems have significantly lower resolution than our design. In particular, see-through radar estimates the location of a human but does not reconstruct his/her figure [40, 55, 84, 92, 121, 171]. This includes commercial products, like Xaver-100, Xaver-400, Xaver-800, and Range-R [79]. Unlike RF-Capture, these systems cannot track individual limbs or construct a human figure. On the other hand, the few systems that aim to reconstruct the human body demonstrate their results on a doll covered with foil and require an antenna array larger than the imaged object [181]. In comparison to these systems, RF-Capture provides finer resolution, and allows capturing human figures with a granularity that is sufficient for distinguishing between a set of 15 people. Also, RF-Capture limits itself to a compact array about twice the size of a Kinect, as opposed to a large array that is of the size of the human body. In addition, unlike commercial products that target the military [79], which use restricted frequency bands and transmission powers only available to military and law enforcement, RF-Capture meets the FCC regulations for consumer devices.

Finally, RF-Capture's coarse-to-fine algorithm is inspired by radar lock and track systems of military aircraft, which first identify a coarse location of a target then zoom on its location to track it [60]. In contrast to these systems, however, RF-Capture does not separate searching from tracking into different phases at signal acquisition. Additionally, RF-Capture's goal of reconstructing a human figure differs from these past systems, resulting in differences in the underlying algorithms.

Device-Free Localization and Gesture Recognition. Advances in RF-based indoor localization have led to new systems that can track users without requiring them to carry a wireless transmitter – this includes our own work in the earlier chapters as well as contemporary and follow-up research, e.g., [21, 23, 24, 25, 42, 85, 108, 117, 129, 163, 166, 178]. Some of these systems have demonstrated the potential of using RF signals to recognize a handful of forward-backward gestures by matching them against prior training examples [117]. RF-Capture builds on this literature but extracts finer-grain information from RF signals. In particular, it is the only system that can identify which human limb reflects the signal at any time. It is also the only system that can combine those limbs to generate a human figure from behind a wall.

6.2 Primer

In this section, we review few RF concepts and signal processing techniques which we leverage in RF-Capture. Some of these concepts have been introduced before in this dissertation, but we recap them here for pedagogical reasons.

(a) Phase of RF signals: An RF signal is a wave whose phase is a linear function of the traveled distance. By sampling the signal, we can record both its amplitude and its phase. The sampled signal can be represented as a complex discrete function of time t as follows [147]:

$$s_t = A_t e^{-j2\pi \frac{r}{\lambda} t}, \tag{6.1}$$

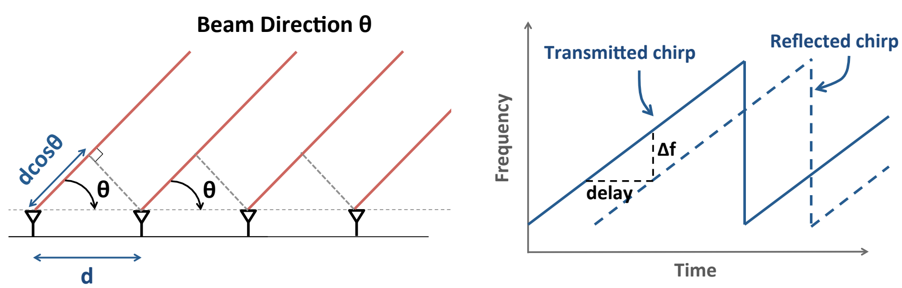
where r is the distance traveled by the signal, λ is its wavelength, and A is its amplitude.

(b) Antenna Arrays: Antenna arrays can be used to identify the ***spatial direction*** from which the RF signal arrives. This process leverages the knowledge of the phase of the received signals to *beamform* in post-processing as shown in Fig. 6-2(a). Mathematically, an N-element antenna array can compute the power P of signals arriving along the direction θ as follows [111]:

$$P(\theta) = \left| \sum_{n=1}^{N} s_n e^{j2\pi \frac{nd\cos\theta}{\lambda}} \right|, \tag{6.2}$$

where s_n is the wireless signal received at the n-th antenna, d is the separation between any two antennas, and λ is the wavelength of the RF signal.

Furthermore, the larger an antenna array is, the stronger its focusing capability is. Specifically, an array of length L has a resolution $\Delta\theta = 0.886\frac{\lambda}{L}$ [111].



(a) Antenna Array **(b) Frequency Chirp**

Figure 6-2: **Measuring Location with RF Signals.** (a) Antenna arrays can be used to focus on signals from a specific direction θ. (b) FMCW chirps can be used to obtain time-of-flight (i.e., depth) measurements.

(c) FMCW Frequency Chirps: Frequency Modulated Carrier Wave (FMCW) is a technique that allows a radio device to measure the ***depth*** of an RF reflector. The technique is described thoroughly in §4.3, and we briefly review it in what follows. An FMCW device transmits a frequency chirp –i.e., a periodic RF signal whose frequency linearly increases in time, as shown in Fig. 6-2(b). The chirp reflects off objects in the environment and travels back to the device after the time-of-flight. The device can measure the time-of-flight and use it to infer the depth of the reflector. To do so, the device leverages the linear relationship between time and frequency in chirps. Specifically, it measures the time-of-flight (and its associated depth) by measuring the frequency shift between the transmitted and received signal. Mathematically, a frequency chirp of slope k can be used to compute the signal power P emanating from a particular depth r as follows [102]:

$$P(r) = \left| \sum_{t=1}^{T} s_t e^{j2\pi \frac{kr}{c} t} \right|, \tag{6.3}$$

where s_t is the baseband time signal, c is the speed of light, and the summation is over the duration T of each chirp.

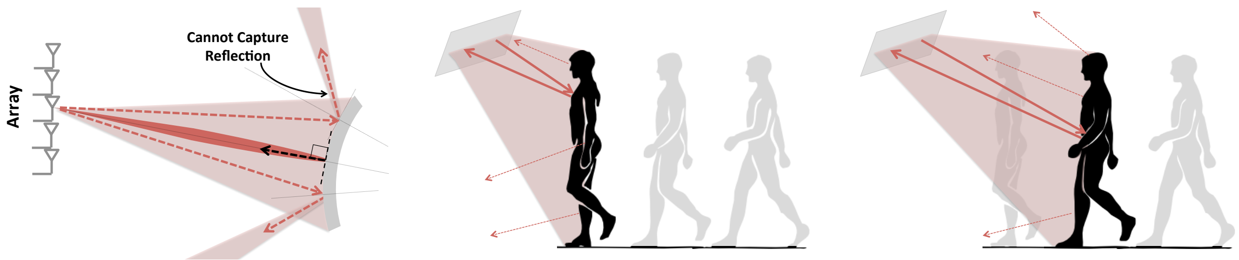
Furthermore, by increasing the bandwidth of the chirp signal, one can achieve finer depth resolution. Specifically, a frequency chirp of bandwidth B has a depth resolution $\Delta r = \frac{c}{2B}$ [102].

(d) Eliminating Static Reflectors: To capture the human figure, we first need to separate human reflections from the reflections of other objects in the environment (e.g., walls and furniture). To do so, we use standard background subtraction, where subtraction is performed in the complex domain since an RF signal is a sequence of complex numbers (with magnitude and phase). Specifically, reflections of static objects remain constant over time and can be eliminated by subtraction. Hence, we collect the reflections of static objects before any human enters the room and subtract them from the received chirps at later times. Of course, this requires knowing whether there are humans in the room or not, which we achieve by leveraging our work on WiTrack (in Chapters 4 & 5).

6.3 RF-Capture Overview

The device: RF-Capture is a system that captures the human figure – i.e., a coarse human skeleton – through walls. It operates by transmitting low-power RF signals (1/1000 the power of WiFi), capturing their reflections off different objects in the environment, and processing these reflections to capture the human figure. RF-Capture's prototype consists of a T-shaped antenna array, as shown in Fig. 6-1. The vertical segment of the "T" consists of transmit antennas and the horizontal segment of the "T" consists of receive antennas. The antennas are connected to an FMCW transceiver which time-multiplexes its transmission between the different transmit antennas, and which can be operated from a computer using a USB cable. The total size of the antenna array is 60×18 cm^2.

In contrast to typical techniques for imaging humans such as visible light, X-ray, terahertz, and millimeter-wave, RF-Capture operates at lower frequencies between 5.46GHz and 7.24GHz. The advantage of operating at such relatively low RF frequencies is that



(a) Simple Reflector (b) Reflections off Human Body (c) Exploiting Motion to Capture Various Body Parts

Figure 6-3: **RF Reflections.** (a) Only signals that fall along the normal to the surface are reflected back toward the device. (b) The human body has a complex surface, but at any point in time only signals close to the normal to the surface are reflected back toward the device. (c) As the person walks, different body parts reflect signals toward the device and become visible to the device.

they traverse walls. Additionally, operating at these frequencies allows us to leverage the low-cost massively-produced RF components in those ranges.

The challenge: The key challenge with operating at this frequency range (5-7GHz) is that the human body acts as a reflector rather than a scatterer. As a result, at any point in time, our antenna array can capture only a subset of the RF reflections off the human body. To see why this is the case, consider the simplified example in Fig. 6-3(a). Recall the basic reflection law: reflection angle is equal to the angle of incidence. Thus, while an antenna array can transmit signals towards the reflecting body, only signals that fall close to the normal to the surface are reflected back toward the array. In contrast, signals that deviate from the normal to the surface are deflected away from our array, making those parts of the reflector invisible to our device. The human body has a much more complex surface; however the same principle still applies, as illustrated in Fig. 6-3(b).

The solution idea: Our solution to the above problem exploits user motion to capture his figure. Specifically, while the antenna array receives reflections only from very few points on the user's surface, these points vary as the person moves, and trace the person's body. Fig. 6-3(b) and (c) illustrate this concept. The figures show that as the person walks, the relation between the incident signal and the normal to the surface for his various body parts naturally changes, providing opportunities for capturing the signals reflected from various body parts. Hence, we could capture the instantaneous RF reflections over consecutive time frames, relate them to each other to identify which reflections are coming from which body part, and combine their information across time and motion to capture the human figure.

In order to transform the above idea into a practical system, we need a design that satisfies two requirements: on one hand, the system needs to achieve spatial resolution sufficient for constructing the human figure; on the other hand, the system should process the signals in real-time at the speed of its acquisition (as in Kinect).

The design of RF-Capture harnesses the above idea while satisfying our design requirements. Specifically, the system has two key components:

- *Coarse-to-fine 3D Scan:* This component generates 3D snapshots of RF reflections by combining antenna arrays with FMCW chirps. A key consideration in designing this algorithm is to ensure low computational complexity. Specifically, directly scanning each point in 3D space to collect its reflections is computationally intractable. Thus, this component introduces a coarse-to-fine algorithm that starts by scanning 3D reflections at coarse resolution, then zooms in on volumes with high power and recursively refines their reflections. The implementation of this algorithm is based on computing FFTs which allows it to achieve low computational complexity.
- *Motion-Based Figure Capture:* This component synthesizes consecutive reflection snapshots to capture the human figure. It operates by segmenting the reflections according to the reflecting body part, aligning them across snapshots while accounting for motion, and then stitching them together to capture the human figure. In §6.8, we demonstrate that this approach can deliver a spatial resolution sufficient for capturing the human figure and its limbs through walls and occlusions.

Next, we describe these components in detail.

■ 6.4 Coarse-to-Fine 3D Scan

RF-Capture uses a combination of a 2D antenna array and FMCW chirps to scan the surrounding 3D space for RF reflections. However, since much of 3D space is empty, it would be highly inefficient to scan every point in space. Thus, RF-Capture uses a coarse-to-fine algorithm that first performs a coarse resolution scan to identify 3D regions with large reflection power. It then recursively zooms in on regions with large reflected power to refine its scan. Below, we explain how this coarse-to-fine scan can be integrated with the operation of antenna arrays and FMCW.

Each voxel in 3D space can be uniquely identified by its spherical coordinates (r, θ, ϕ) as shown in Fig. 6-4. By projecting the received signals on θ and ϕ using the 2D antenna array and on r using the frequency chirp, we can measure the power from a particular 3D voxel. Mathematically, the power arriving from a voxel (r, θ, ϕ) can be computed as:

$$P(r, \theta, \phi) = \left| \sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{t=1}^{T} s_{n,m,t} e^{j2\pi \frac{kr}{c} t} e^{j \frac{2\pi}{\lambda} \sin\theta (nd \cos\phi + md \sin\phi)} \right|, \tag{6.4}$$

where N is the number of receive antennas, and M is the number of transmit antennas, and $s_{n,m,t}$ is the signal received by receive antenna n from transmit antenna m at time t.

Equation 6.4 shows that the algorithmic complexity for computing the reflection power is cubic for every single 3D voxel. Thus, we want to minimize the number of 3D voxels that we scan while maintaining high resolution of the final 3D reflection snapshot. To do so, we refine the resolution of our antenna array and FMCW chirps recursively as described below.

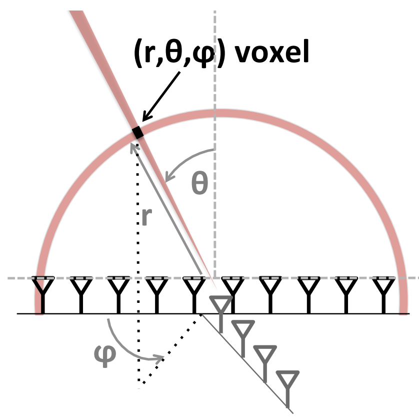


Figure 6-4: **Scanning.** A 2D antenna array with FMCW ranging can focus on any (r, θ, ϕ) voxel in 3D.

Coarse-to-Fine Angular Scan: RF-Capture exploits an intrinsic property of antenna arrays, namely: the larger an array is, the narrower its beam, and the finer its spatial resolution. Thus, RF-Capture starts with a small array of few antennas, and uses more antennas only to refine regions that exhibit high reflection power. Fig. 6-5 illustrates this design. The figure uses a 1D array for clarity. In the first iteration of the algorithm, RF-Capture computes power using signals from only the two middle antennas of the array, while ignoring the signal from the other antennas. This results in a small aperture, and hence a very wide beam. Using this wide beam, RF-Capture localizes the person to a wide cone as shown by the red region in Fig. 6-5(a). In the next iteration, it incorporates two more antennas in the array. However, in this iteration, it does not need to scan the entire angular space, but rather only the space where it had detected a person in the previous iteration (i.e., the red region in Fig. 6-5(a)). The algorithm proceeds in the same manner until it has incorporated all the antennas in the array and used them to compute the finest possible direction as shown in Fig. 6-5(b).

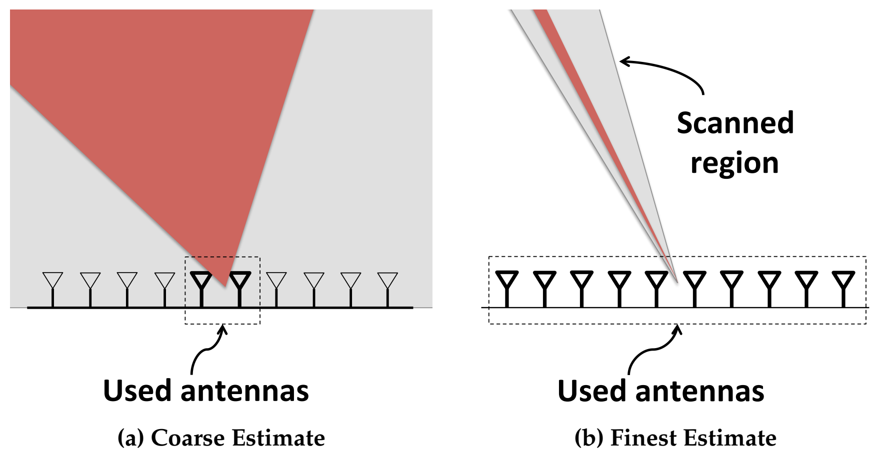


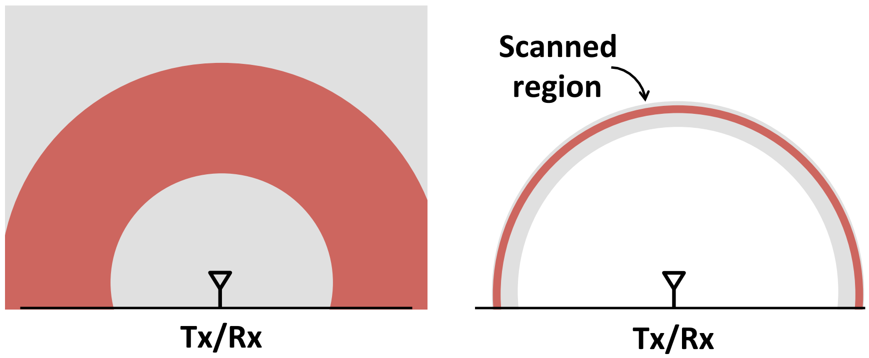
Figure 6-5: **Coarse-to-Fine Angular Scan.** We start by using a small number of antennas which gives us a wide beam and coarse angular resolution. Then, we refine our estimate by using more antennas to achieve a narrower beam and finer resolution, but use that beam only to scan regions of interest.

While the above description uses a 1D array for illustration, the same argument applies to 2D arrays. In particular, our 2D array has a T-shape. Thus, in each iteration, we refine the resolution by including an extra antenna from the vertical segment and two antennas from the horizontal segment.

Coarse-to-Fine Depth Scan: Recall that the depth resolution of FMCW is inversely proportional to the bandwidth of the signal (see §6.2(c)). Hence, RF-Capture can recursively refine its depth focusing by gradually increasing the amount of bandwidth it uses.

Specifically, it starts by using a small chunk of its bandwidth, which would result in very coarse resolution as shown in Fig. 6-6(a). It then localizes the person to a wide spherical ring. In the following iteration, it uses a larger amount of bandwidth but scans only the spherical ring where it identified the reflector. It proceeds iteratively until it has used all of its bandwidth, as shown in Fig. 6-6(b).

But, what does it mean for us to iteratively increase the bandwidth? Similar to our antenna array iterative approach, we still collect all the data, but process it selectively. Specif-



(a) Coarse Estimate (c) Finest Estimate

Figure 6-6: **Coarse-to-Fine Depth Scan.** We start by using a small chunk of bandwidth which gives us coarse depth resolution. Then, we refine our estimate by adding more bandwidth to achieve finer resolution, but use that bandwidth only to scan regions of interest.

ically, recall that a frequency chirp consists of a signal whose frequency linearly increases over a sweep as shown in Fig. 6-2(b). Whereas all the samples of a sweep collectively cover the entire bandwidth, a subset of those samples covers a subset of the sweep's bandwidth. Similar to iteratively adding more antennas to our processing, RF-Capture iteratively adds chirp samples to achieve finer depth resolution.

Additional Points: A few points are worth noting:

- RF-Capture performs the above iterative refinement in both FMCW bandwidth and antenna arrays simultaneously as shown in Fig. 6-7.

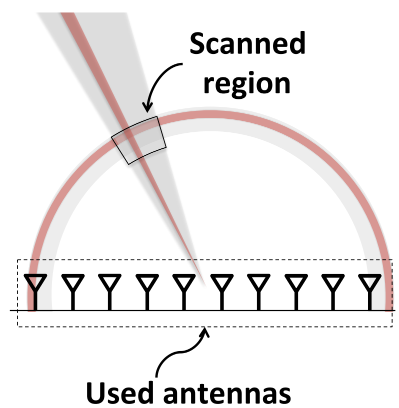


Figure 6-7: **Coarse-to-Fine 3D Scan.** We can partition and iterate jointly using chirps and antenna arrays. In any given iteration, we only scan the small region identified by the previous iteration.

- Standard antenna array equations (as described in §6.2(b) and Fig. 6-2(a)) rely on an approximation which assumes that the signals received by the different antennas are all parallel. To improve the final accuracy of reconstruction and achieve higher focusing capabilities, we use a more complex model in the final iteration of the coarse-to-fine algorithm [126]. Specifically, the power from an (x, y, z) voxel in 3D space can be expressed as a function of the round-trip distances $r_{(n,m)}(x, y, z)$ to each transmit-receive pair (m, n) as follows:[1]

$$P(x,y,z) = \left| \sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{t=1}^{T} s_{n,m,t} e^{j2\pi \frac{kr_{(n,m)}(x,y,z)}{c} t} e^{j2\pi \frac{r_{(n,m)}(x,y,z)}{\lambda}} \right| \tag{6.5}$$

- Finally, the coarse-to-fine algorithm (in our current implementation) allows RF-Capture to generate one 3D snapshot (a 3D frame) of the reflection power every 75ms on Nvidia Quadro K4200 GPU. This represents a speedup of $160,000\times$ over a standard non-linear projection of Eq. 6.5 which requires on average 200 minutes for rendering a single time snapshot on the same GPU platform. Furthermore, because the switched antenna array has a signal acquisition time of 80ms, the 75 ms rendering time allows RF-Capture to generate a new 3D snapshot within the same signal acquisition period. In addition, it results in a frame rate that is sufficient to continuously track human motion across time. Being able to assume that reflecting bodies smoothly move across a sequence of 3D frames is important for identifying human body parts and tracking them, as we explain in the next section.

6.5 Motion-based Figure Capture

Now that we have captured 3D snapshots of radio reflections of various human body parts, we need to combine the information across consecutive snapshots to capture the human figure. This process involves the following four steps:

- *Compensation for Depth:* Since RF-Capture collects 3D snapshots as the user moves, the subject's body is at different depths in different snapshots. Therefore, RF-Capture needs to compensate for differences in depth before it can combine information across consecutive snapshots.
- *Compensation for Swaying:* As the person walks, his body naturally sways. To combine information across consecutive snapshots, RF-Capture has to compensate for this swaying

[1] The inverse square law is implicit in $s_{n,m,t}$ and doesn't need to be inverted in the phased array formulation. This is a standard approximation in antenna arrays since the phase varies by 2π every wavelength, which is a much bigger effect than changes in amplitude. Accounting for the minute variations in amplitude can produce minor sidelobe reductions, but is often negligible [126].

and re-align the 3D voxels across snapshots.

- *Body Part Segmentation:* Recall that each of the 3D snapshots reveals a small number of body parts. In the next step, RF-Capture segments each snapshot to extract the body parts visible in it and label them (e.g., head, chest, left arm, left hand, etc.).
- *Skeletal Stitching:* In the final step, RF-Capture uses a simple model of the human skeleton to combine the detected body parts across a few consecutive snapshots and capture the human figure.

In what follows, we describe each of these steps in detail. To make the exposition clearer, we describe these steps by applying them to the output of an experiment collected with RF-Capture. In this experiment, the RF-Capture sensor is behind a wall. We ask a user to walk toward the RF-Capture device starting from a distance of about 3 m from the sensor. The antennas of RF-Capture are positioned at 2 m above the ground, so that reflections from humans arrive along upward directions.

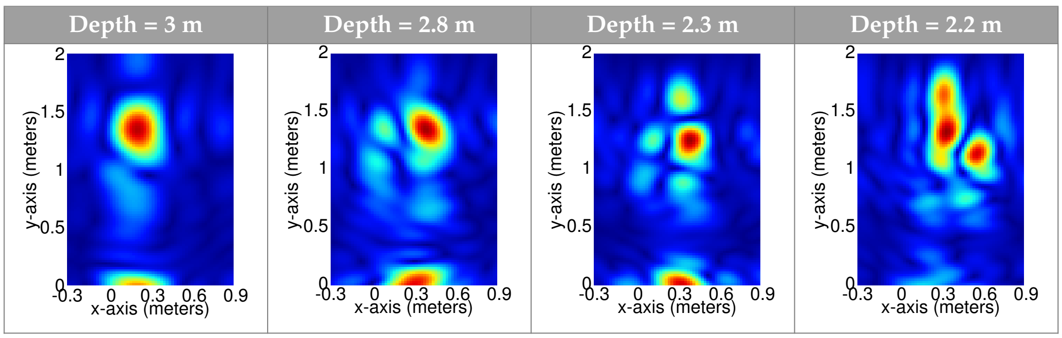
6.5.1 Compensating for Depth

When imaging with an antenna array, an object looks more blurry as it gets farther away from the array. This is because the beam of an antenna array has the shape of a cone, and hence is wider at larger distances. Since our 3D snapshots are taken as the subject walks towards the array, the subject is at different depths in different snapshots, and hence experiences different levels of blurriness across snapshots. Thus, before we can combine a subject's reflections across RF snapshots, we need to compensate for his change in depth.

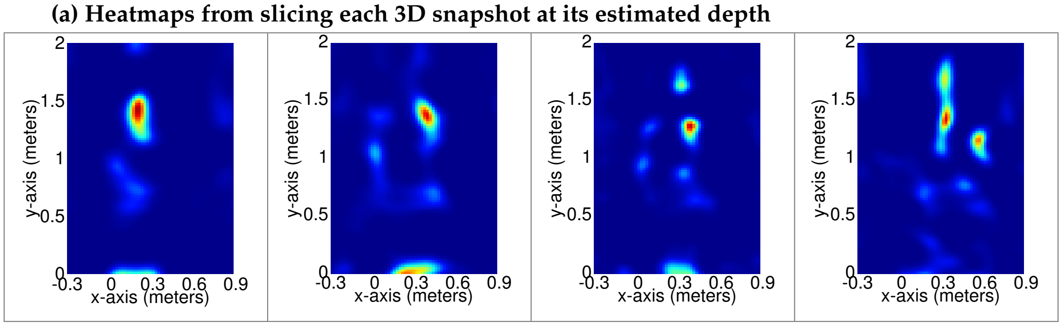
To do so, we first need to know the subject's depth in each snapshot. This is easy since our snapshots are three-dimensional by construction –i.e., we know the depth of each voxel that reflects power. Of course, the human body is not flat and hence different body parts exhibit differences in their depth. However, these differences are relatively small. Thus, for our purpose, we take the median depth of the RF reflections in each 3D snapshot, and consider it as the person's depth in that snapshot.

Next, we compensate for depth-related distortion by deconvolving the power in each snapshot with the point spread function caused by the antenna-array beam at that depth. The point spread function is computed directly from the array equation, Eq. 6.5, and the deconvolution is done using the Lucy-Richardson method [100].

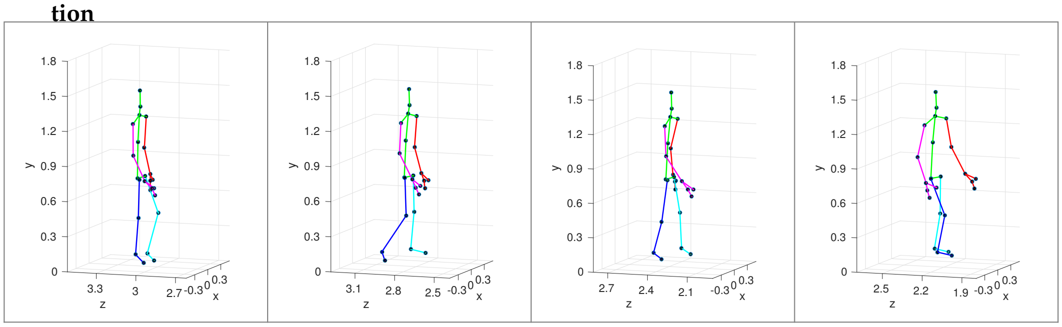
Fig. 6-8 illustrates this process. The top row shows different RF snapshots as the person walks towards the antenna array. The snapshots are plotted by slicing the 3D snapshot at the median depth for the reflected signals, and showing the power as a heat map, where red refers to high reflected power and dark blue refers to no reflection. It is clear from this row that reflected bodies look wider and more blurry at larger depths. The second row shows the same snapshots after compensating for depth distortion. These snapshots are less blurry and more focused on the actual reflection points.



(a) Heatmaps from slicing each 3D snapshot at its estimated depth



(b) Output after deconvolving the images with the depth-dependent point-spread function



(c) Kinect skeletal tracking as a baseline for comparison. We rotate the Kinect output by 45° to visualize the angles of the limbs as the user moves.

Figure 6-8: **RF-Capture's heatmaps and Kinect skeletal output as a user walks toward the deployed sensors.** As the user walks toward the device, RF-Capture captures different parts of his body at different times/distances since its antennas' perspective changes with respect to his various body parts.

6.5.2 Compensating for Sway

Next, RF-Capture compensates for the user's sway as he walks by using the reflection from his chest as a pivot. Specifically, because the human chest is the largest convex reflector in the human body, it is expected to be the dominant reflector across the various snapshots, enabling us to identify it and use it as a pivot to center the snapshots. From an RF perspective, the human chest is said to have the largest radar cross section [56]. Indeed, the heatmaps in Fig. 6-8(b) show a dominant reflection (dark red) around the height of the subject's chest ($z = 1.4m$).

To align the snapshots, we first determine the dominant reflection point in each snapshot. In most snapshots, this would correspond to the human chest. We then perform robust regression on the heights of these maxima across snapshots, and reject the outliers.[2] This allows us to detect snapshots in which the chest is not the most dominant reflection point and prevent them from affecting our estimate of the chest location. Once we have identified the chest location in each snapshot, we compensate for minor sways of the human body by aligning the points corresponding to the chest across snapshots.

Note that aligning the human body across snapshots makes sense only if the human is walking in the same direction in all of these snapshots. Thus, RF-Capture considers the trajectory of the point with the highest reflection power on the human body, and performs the above alignment only for periods during which the human is walking toward the device without turning around.

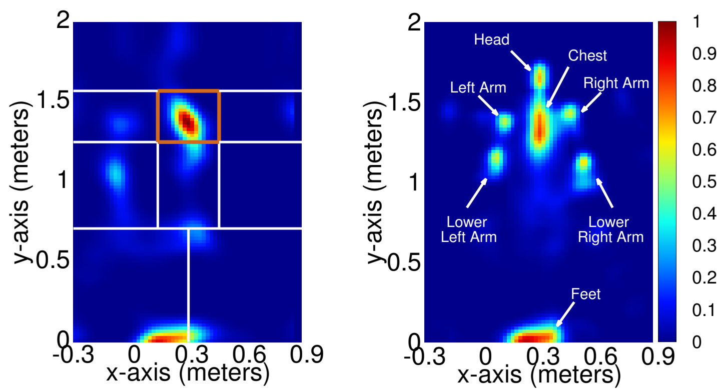
6.5.3 Body Part Segmentation

After identifying the human chest as a pivot and aligning the consecutive snapshots, we segment the areas around the chest to identify the various human body parts.

Specifically, RF-Capture defines a bounding region centered around the subject's chest. For example, Fig. 6-9(a) shows the rectangle in orange centered around the detected subject's chest. (This is the second image from Fig. 6-8(b) after sway compensation.) Using the chest as a pivot, RF-Capture automatically segments the remainder of the heatmap into 8

[2]To perform robust regression, we use MATLAB's default parameters, i.e., bisquare weighting function with a tuning constant of 4.685, and eliminate outliers whose heights are more than two standard deviations away from the mean.

regions, each corresponding to a different body part of interest. The first region constitutes the rectangle below the chest, which corresponds to the user's lower torso, while the region above the chest corresponds to the subject's head. The regions to the left and right of the chest correspond to the arms and the hands. Finally, the regions below the torso correspond to the subjects' legs and feet. In our implementation, we specify the width of the torso region to 35 cm, and the height of the upper torso (chest) to 30 cm, while the lower torso is 55 cm. These numbers work well empirically for 15 different adult subjects with different ages, heights, builds, and genders. We envision that exploiting more powerful segmentation and pose estimation algorithms – such as those that employ recognition or probabilistic labeling, e.g., [70, 107, 132] – would capture better human figures. Such techniques are left for future work.



(a) Segmentation **(b) Captured Figure**

Figure 6-9: **Body Part Segmentation and Captured Figure.** (a) shows the different regions used to identify body parts, and (b) shows the captured synthesized from 25 time frames.

Once RF-Capture performs this segmentation, the blobs in the heatmaps of Fig. 6-8(b) become more meaningful, and can be automatically assigned body part labels. For example, for the heatmap generated at 2.8m, it can now automatically detect that the blob to the left of the chest is the right arm, and the blob below it is the right hand. On the other hand, the heatmap at 2.2m shows the subject's left hand and his head, but none of his right limbs.

To gain a deeper understanding into the segmented images, we use a Kinect sensor as a baseline. The Kinect is placed in the same room as the moving subject, while the RF-Capture sensor is outside the room. Both devices face the subject. We plot in Fig. 6-8(c) the output of Kinect skeletal tracking that corresponds to the RF snapshots in Fig. 6-8(b).

We rotate the Kinect skeletal output by $45°$ in Fig. 6-8(c) so that we can better visualize the angles of the various limbs. We also perform a coordinate transformation between the RF-Capture's frame of reference and the Kinect frame of reference to account for the difference in location between the two devices. Comparing Kinect's output with that of RF-Capture, we note the following observations:

- RF-Capture can typically capture reflections off the human feet across various distances. This is because the feet reflect upward in all cases, and hence they reflect toward RF-Capture's antennas.
- It is difficult for RF-Capture to capture reflections from the user's legs. This is because even as the legs move, they deflect the incident RF signals away from the antenna array (toward the ground) rather than reflecting them back to the array since the normal to the surface of the legs stays almost parallel to the ground. (Note that placing the antenna array on the ground instead would enable it to capture a user's legs but would make it more difficult for the array to capture his head and chest reflections.)
- The tilt of a subject's arm is an accurate predictor of whether or not RF-Capture can capture its reflections. For example, in the third snapshot of Fig. 6-8(c) (i.e., at 2.3m), the subject's right arm (color-coded in pink) is tilted upward; hence, it reflects the incident signal back to RF-Capture's antennas allowing it to capture the arm's reflection. Indeed, this matches RF-Capture's corresponding (third) heatmap in Fig. 6-8(b). On the other hand, the subject's left arm (color-coded in red) is tilted upward in the fourth snapshot (i.e., at 2.2m), allowing RF-Capture to capture its reflections in the corresponding heatmap.

■ 6.5.4 Skeletal Stitching

After segmenting the different images into body parts, RF-Capture stitches the various body parts together across multiple snapshots to capture the human figure. We distinguish between two types of body reflectors: rigid parts and deformable parts:

- *Rigid Parts, i.e., head and torso:* Once RF-Capture compensates for depth and swaying, these structures do not undergo significant deformations as the subject moves. Hence, RF-Capture sums up each of their regions across the consecutive snapshots (i.e., sum up their reflected power). Doing so provides us with a more complete capture of the user's torso since we collect different reflection points on its surface as the user walks. Furthermore, we

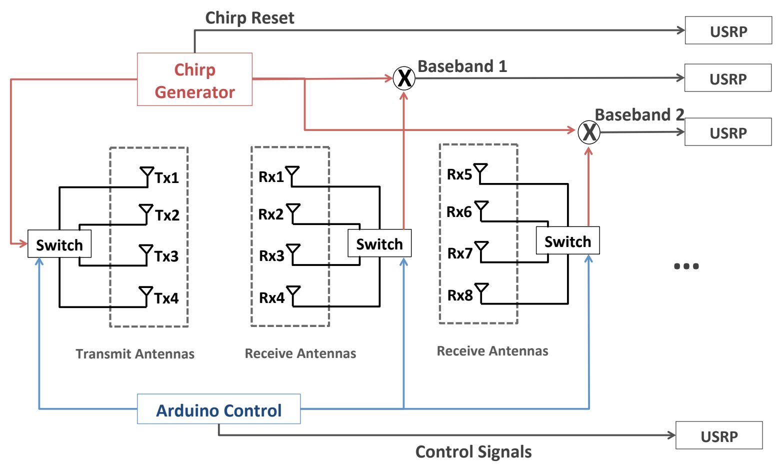


Figure 6-10: **RF-Capture's Hardware Schematic.** The setup consists of a chirp generator connected to a 2D antenna array via a switching platform. The figure shows the Tx chain, and two of the Rx chains.

found that such a more complete capture of the torso is very helpful in identifying users as we show in §6.8.

- *Deformable parts, i.e., arms and feet:* RF-Capture cannot simply add the segments corresponding to the human limbs across snapshots. This is because as the human moves, his arms sway back and forth, and adding the different snapshots together results in smearing the entire image and masking the form of the hand. Instead, our approach is to identify the highest-SNR (signal-to-noise ratio) segment for each body part, and select it for the overall human figure. This is because a higher SNR indicates less sensitivity to noise and hence higher reliability.

Finally, to ensure that the resultant figure is smooth, we perform alpha blending [141]. Fig. 6-9(b) shows the result of synthesizing 25 frames together, collected over a span of 2 seconds as the user walks towards our antenna setup. The figure shows that by combining various snapshots across time/distance, RF-Capture is capable of capturing a coarse skeleton of the human body.

6.6 Implementation

Our prototype consists of hardware and software components.

Hardware: A schematic of RF-Capture's hardware is presented in Fig. 6-10. It has the following components:

- *FMCW Chirp Generator:* We built an FMCW radio on a printed circuit board (PCB) using off-the-shelf circuit components, and based on the design in §4.6. The resulting radio can be operated from a computer via the USB port. It generates a frequency chirp that repeatedly sweeps the band $5.46 - 7.24\ GHz$ every $2.5\ ms$. The radio has an average power of 70μWatts, which complies with the FCC regulations for consumer electronics in that band [18].
- *2D Antenna array:* (shown in Fig. 6-1): The antenna array consists of 16 receive antennas (horizontal section of the T) and 4 transmit antennas (vertical section of the T); the antennas are log-periodic with 6dBi gain. This multiple-transmit multiple-receive architecture is equivalent to a 64-element antenna array. The overall array dimension is $60\ cm \times 18\ cm$.[3]
- *Switching Platform:* We connect all four transmit antennas to one switch, so that at any point in time, we transmit the chirp from only one antenna. Similarly, we connect every four receive antennas to one switch and one receive chain. Each receive chain is implemented using a USRP software radio equipped with an LFRX daughterboard. The sampled signals are sent over an Ethernet cable to a PC for processing. This design allows us to use a single transmit chain and only four receive chains for the entire 2D antenna array.

Software: RF-Capture's algorithms are implemented in software on an Ubuntu 14.04 computer with an i7 processor, 32GB of RAM, and a Nvidia Quadro K4200 GPU. We implement the hardware control and the initial I/O processing in the driver code of the USRP. The coarse-to-fine algorithm in §6.4 is implemented using CUDA GPU processing to generate reflection snapshots in real-time. In comparison to C processing, the GPU implementation provides a speedup of $36\times$.

Calibration: FMCW and antenna array techniques rely on very accurate phase and frequency measurements. However, various hardware components – including filters, wires, switches, and amplifiers – introduce systematic phase and frequency offsets. To make sure these offsets do not introduce errors for our system, we perform a one-time calibration of the system where we connect each of the Tx and Rx chains over the wire and estimate these offsets. We then invert these offsets in software to eliminate their effect.

[3] The antenna separation is 4 cm, which is around λ. Such separation is standard for UWB arrays since the interference region of grating lobes is filtered out by the bandwidth resolution [128].

6.7 Evaluation Environment

(a) Participants: To evaluate the performance of RF-Capture we recruited 15 participants. Our subjects are between 21–58 years old ($\mu = 31.4$), weigh between 53–93 kg ($\mu = 78.3$), and are between 157–187 cm tall ($\mu = 175$). During the experiments, the subjects wore their daily attire, including shirts, hoodies, and jackets with different fabrics. The experiments were conducted over a span of 5 months; the same subject had different clothes in different experiments. These experiments were approved by our IRB.

(b) Experimental Environment: All experiments are performed with the RF-Capture sensor placed behind the wall as shown in Fig. 6-1. The experiments are performed in a standard office building; the interior walls are standard double dry walls supported by metal frames. The evaluation environment contains office furniture including desks, chairs, couches, and computers. The antennas are located 2m above the ground level, ensuring that the device is higher than the tallest subject.

(c) Baseline: We use Kinect for baseline comparison. In our experiments, both Kinect and RF-Capture face the subject, but Kinect is in line-of-sight of the subject, while the RF-Capture sensor is behind the room's wall. We use Kinect's skeletal output to track the subject, and we perform a coordinate transformation between RF-Capture's frame of reference and Kinect's frame of reference.

6.8 Results

RF-Capture delivers two sets of functions: the ability to capture the human figure through walls, and the ability to identify and track the trajectory of certain body parts through walls. Below, we evaluate both functions in detail.

6.8.1 Body Part Identification and Tracking

Body Part Identification

We first evaluate RF-Capture's ability to detect and distinguish between body parts. We run experiments where we ask each of our subjects to walk toward the device (as shown in Fig. 6-1), stop at her chosen distance in front of it, then move one of the following body parts: left arm, right arm, left foot, right foot, and head. The subject can stop at any distance between 3m and 8m away from RF-Capture. We perform 100 such experiments. Throughout these experiments, the subjects performed different movements such as: nodding, waving an arm, sliding a leg, or rotating a hand in place.

Classification: We would like to identify which body part the subject moved by mapping it to the segmented 3D snapshots. Hence, in each experiment, we collect the reflection snapshots as the user walks and process them according to the algorithms in §6.4 and §6.5 to capture the segmented body parts. Then, we focus on the snapshots after the user stops walking, and moves one limb while standing still. We determine the location of the body part that the user has moved. We compare the identified body part against the user's reported answer for which body part she/he moved after she/he stopped walking and was standing still.

Results: Fig. 6-11 plots the classification accuracy among the above 5 body parts as a function of the user's distance to the RF-Capture sensor. When the user is at 3 m from the antenna setup, the classification accuracy is 99.13%. The accuracy gradually decreases with distance, and reaches 76.48% when the user is 8 m away.

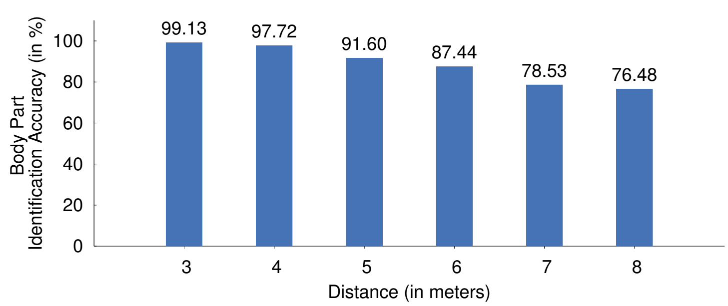


Figure 6-11: **Body Part Identification Accuracy with Distance.** The figure shows RF-Capture's accuracy in identifying the moving body part as a function of the user's distance to the device.

To better understand the source of the errors, we show the confusion matrix in Table 6-1 for the case where one of our subjects stands 5 m away from RF-Capture. The table shows that most errors come from RF-Capture being unable to detect a user's body part motion. This is because while the user did move his limbs, some of these motions may have not altered the reflection surface of the limb to cause a change detectable by the antennas. The other main source of classification errors resulted from misclassifying an upper limb as a lower limb, as opposed to confusing a left limb with a right limb. For example, the right leg and right hand are confused in 5.6% of the experiments, while the right hand and the left hand are never confused. The reason is that our antenna array is wider along the horizontal axis than the vertical axis, as can be seen in Fig. 6-1. Hence, the antenna has a narrower beam (i.e., higher focusing capability) when scanning horizontally than when it scans vertically.

		Estimated					
		Left Hand	Right Hand	Left Leg	Right Leg	Head	Undetected
Acutal	Left Hand	91.6	0.0	5.6	0.0	2.8	0.0
	Right Hand	0.0	90.2	0.0	9.4	0.4	0.0
	Left Leg	0.0	0.0	89.7	0.0	0.0	10.3
	Right Leg	0.0	0.0	0.0	86.8	0.0	13.2
	Head	0.0	0.0	0.0	0.0	90.5	9.5

Table 6-1: **Confusion Matrix of Body Part Identification.** The table shows the classification accuracy of the various body parts at 5 m.

Body Part Tracking

Next, we would like to evaluate the accuracy of localizing a detected body part in RF-Capture's 3D snapshots. Recall however that human body parts appear in a 3D snapshot only when the incident signal falls along a direction close to the normal to the surface. To ensure that the body part of interest remains visible in the 3D snapshots during the experiment, we focus on localizing the human palm as the user moves his/her hand in front of the device, as in Fig. 6-12. In particular, the user is asked to raise his hand as in Fig. 6-12, and write an English letter of his/her choice in mid-air.

Figure 6-12: **Tracking the Human Hand Through Walls.** RF-Capture tracks the subject's palm through a wall while the Kinect tracks it in line-of-sight.

Note that in each of the 3D snapshots, RF-Capture detects multiple body parts. Hence, we only focus on reflections that change over time and ignore static reflections from static body parts. Once we localize the moving reflection, we attribute it to the location of the subject's palm and define our error as the difference between this location and the Kinect-computed location for the subject's hand.[4]

Results: We plot the CDF (cumulative distribution function) of the 3D tracking error across 100 experiments in Fig. 6-13. The figure shows that the median tracking error is 2.19cm and that the 90^{th} percentile error is 4.84cm. These results demonstrate that RF-Capture can track a person's body part with very high accuracy. To gain further insight into these results, we show two of the letters written by our subjects in Fig. 6-14. The figure shows the trajectory traced by by RF-Capture (in blue) and Kinect (in red), as the subject wrote the letters "S" and "U".

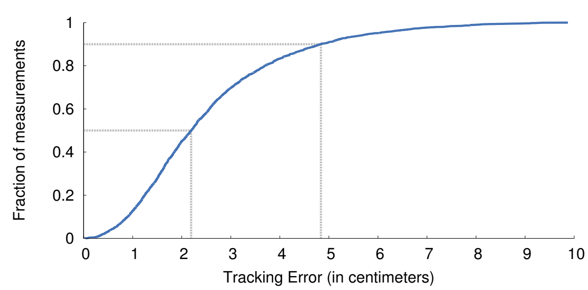


Figure 6-13: **Body Part Tracking Accuracy.** The figure shows the CDF of RF-Capture's accuracy in tracking the 3D trajectory of the subject's hand.

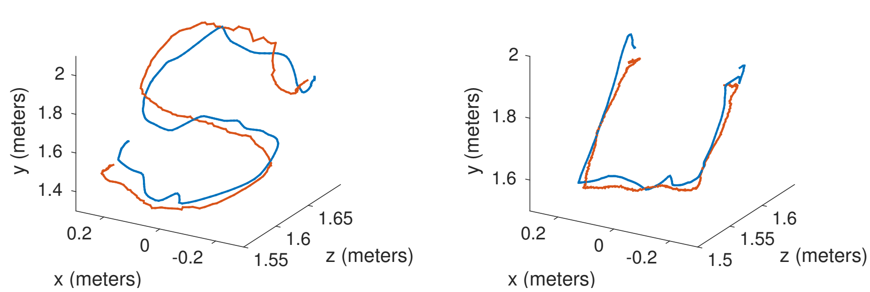


Figure 6-14: **Writing in the Air.** The figure shows the output of RF-Capture (in blue) and Kinect (in red) for two sample experiments were the subject wrote the letters "S" and "U" in mid-air.

[4] We perform a coordinate transformation between RF-Capture's frame of reference and that of Kinect to account for their different locations.

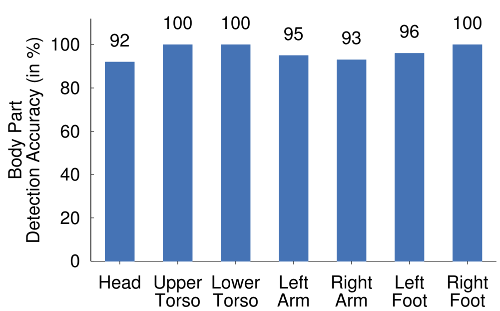
6.8.2 Human Figure Capture and Identification

In this section, we focus on evaluating the quality of the figures captured by RF-Capture, as well as the amount of motion required to capture such figures.

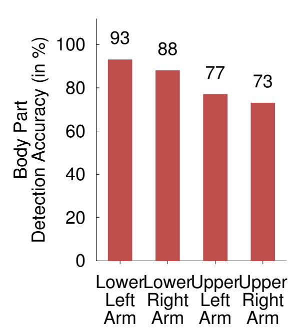
Amount of Motion Required for Figure Capture

We would like to understand how much walking is needed for our figure capture. Thus, we ask users to walk towards the device, and we divide each experiment into windows during which the subject walks by only two steps. Our intuition is that two steps should be largely sufficient to capture reflections off the different body parts of interest because as a human takes two steps, both his left and right limbs sway back and forth, providing RF-Capture with sufficient perspectives to capture their reflections.

Results: Fig. 6-15(a) shows the results from 100 experiments performed by our subjects. The x-axis denotes the body parts of interest, and the y-axis shows the percentage of experiments during which we detected each of those body parts. The figure shows that the human torso (both the chest and lower torso) is detected in all experiments; this matches our initial observation that the chest is a large convex reflector that appears across all frames. The other human body parts are detected in more than 92% of the experiments.



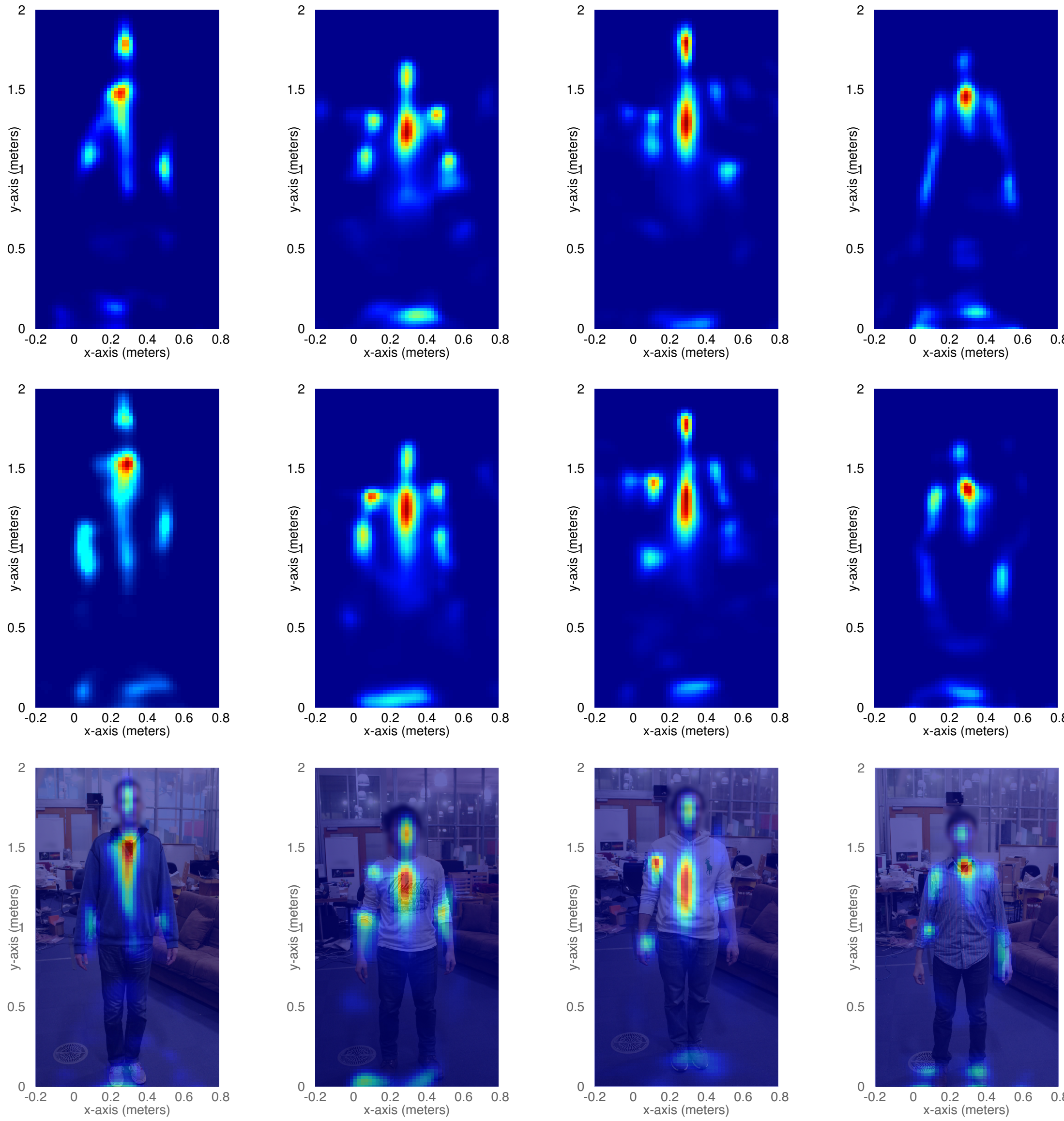
(a) Basic Figure



(b) Dividing Arms

Figure 6-15: **Body Part Detection Accuracy.** In these experiments, the user walks in exactly two steps. The figure shows the percentage of experiments during which RF-Capture was able to capture a particular body part in the human figure, using only two steps of motion.

To understand detection accuracy for finer human figures, we segment each arm into upper and lower parts and show their corresponding detection accuracy in Fig. 6-15(b). The plot shows that the upper arm is detected in a smaller number of experiments, which is also expected because humans usually sway the lower segments of their arms more than the upper segments of their arms as they walk.



(a) Subject A (b) Subject B (c) Subject C (d) Subject D

Figure 6-16: **Human Figures Obtained with RF-Capture.** The graphs show examples of the human figures generated by RF-Capture. Each column shows a different human subject, while each row shows figures of the same subject across different experiments.

Sample Captured Figures

Next, we would like to gain a deeper understanding of the figures captured by RF-Capture, and how they relate to the human's heights and builds. Thus, we plot in Fig. 6-16 the figures of four of our subjects as output by RF-Capture. Each of the columns corresponds to a different subject, and each of the rows corresponds to the output of an experiment performed on a different day. In the final row of Fig. 6-16, we overlay the obtained heatmaps over the subject's photo. Based on these plots, we make the following observations:

- Figures of the same subject show resemblance across experiments and differ from figures of a different subject. This indicates that RF-Capture can be useful in differentiating between people when they are occluded or behind a wall.
- RF-Capture can capture the height of a subject. Fig. 6-16 shows that subject A's head is higher than the rest, while subjects B and D are around the same height. In reality, subject A is 187cm tall, while subjects B and D are 170cm and 168cm respectively.
- Depending on the subject and the experiment, the feet may appear separated or as a single blob at the bottom of the heatmap. This is typically due to whether the subject is walking with his feet separated or closer to each other.

Human Identification

We want to evaluate whether the human figures generated by RF-Capture reveal enough information to differentiate between people from behind a wall. Hence, we ask our subjects to walk towards RF-Capture from behind a wall, as described in §6.7(b), and use RF-Capture to synthesize their figures. We run experiments with 15 subjects. In each experiment, we ask one of the subjects to walk toward the device from a distance of 3 m to a distance of 1 m. We run four experiments with each subject across a span of 15 days.

Classification: We divide our experiments into a training set and a testing set. In particular, out of each user's four experiments, three are used for training and one is used for testing. To obtain our feature vectors for classification, we transform the 2D normalized reconstructed human to a 1D-feature vector by concatenating the rows. For dimensionality reduction, we apply PCA on the feature vectors and retain the principal components that cover 99% of the variance. We then use the PCA features to train an SVM model. The SVM model is a multi-class classifier, with a cost of 10, and a first-order polynomial kernel of $\gamma = 1$ and $coefficient = 1$. The classification is performed in MATLAB on the skeleton generated from our C++/CUDA code.

Results: Fig. 6-17 shows RF-Capture's classification accuracy as a function of the number of users it is trained on. The results show that when RF-Capture is used to classify between only two users, the accuracy is 98.1%. We note that this accuracy is the average accuracy resulting from tests that consist of randomly choosing two of our fifteen subjects, and repeating for different pairs. The standard deviation of this classification across all possible pairs of subjects is 8.1%. As the number of users we wish to classify increases, RF-Capture's classification accuracy decreases. In particular, looking back at Fig. 6-17, we see that the accuracy decreases to 92% for classifying 10 subjects, and 88% for 15 subjects.

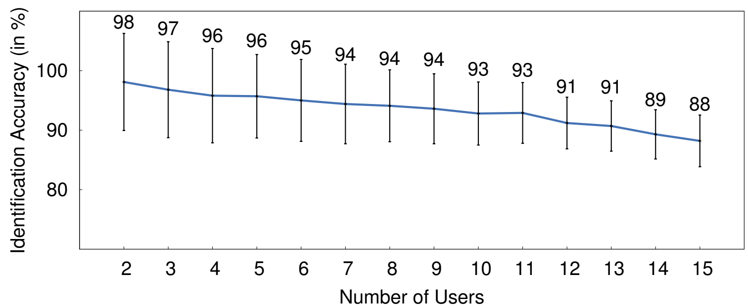


Figure 6-17: **Identification Accuracy as a Function of the Number of Users.** RF-Capture uses the captured figures to identify users through classification. The accuracy decreases as the number of users we wish to classify increases.

To gain a deeper understanding into the classification errors, we show the confusion matrix of the 10-subjects experiments in Table 6-2. Among these subjects, subject 7 corresponds to subject A in Fig. 6-16. This subject has been misclassified often as subject 9 in the table. In fact, these two subjects were the tallest among all of our volunteers. Subject 4 is the shortest and is never misclassified as anyone else. Generally, as one would expect, the more distinctive one's height and build are, the easier it is to classify him.

Acutal \ Estimated	1	2	3	4	5	6	7	8	9	10
1	99.7	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	96.0	0.4	1.3	0.3	0.0	0.9	0.0	0.4	0.7
3	0.0	0.0	99.9	0.0	0.0	0.1	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	2.1	0.0	90.8	0.0	4.8	2.3	0.0	0.0
6	0.0	0.0	5.1	0.0	0.0	94.2	0.0	0.7	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.1	86.9	0.0	13.0	0.0
8	0.0	1.0	0.0	0.0	0.0	0.0	0.8	97.6	0.0	0.6
9	0.0	0.0	0.0	0.0	0.0	0.0	8.8	0.0	91.2	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0	1.1	0.8	0.0	98.1

Table 6-2: **Confusion Matrix of Human Identification.** The table shows the classification accuracy for each of our subjects.

6.9 Discussion

We present RF-Capture, a system that can capture the human figure through walls, and identify users and body parts even if they are fully occluded. However, the system exhibits some limitations:

- It assumes that the subject starts by walking towards the device, hence allowing RF-Capture to obtain consecutive RF snapshots that expose his body parts. Future systems should expand this model to a more general class of human motion and activities.
- The current method captures the human figure by stitching consecutive snapshots, and hence cannot perform fine-grained full skeletal tracking across time. Future work may consider combining information across multiple RF-Capture sensors to refine the tracking capability.
- Our implementation adopts a simple model of the human body for segmentation and skeletal stitching. Future work can explore more advanced models to capture finer-grained human skeleta.

Despite these limitations, we believe that RF-Capture marks an important step towards motion capture that operates through occlusions and without instrumenting the human body with any markers. It also motivates a new form of motion capture systems that rely on, or are augmented with, RF sensing capabilities. We envision that as our understanding of human reflections in the context of Vision and Graphics evolve, these capabilities would extend human pose capture to new settings. For example, they can expand the reach of gaming consoles, like the Xbox Kinect, or gesture recognition sensors, like those embedded in smart TVs, to operate through obstructions and cover multiple rooms. They would also enable a new form of ubiquitous sensing which can understand users' activities, learn their habits, and monitor/react to their needs. In addition, they can provide more representative motion capture models in biomechanics, ergonomics, and character animation.

Chapter 7
Smart Homes That Monitor Breathing and Heart Rate

The past few years have witnessed a surge of interest in ubiquitous health monitoring [73, 82]. Today, we see smart homes that continuously monitor temperature and air quality and use this information to improve the comfort of their inhabitants [106, 175]. As health-monitoring technologies advance further, we envision that future smart homes would not only monitor our environment, but also monitor our vital signals, like breathing and heartbeats. They may use this information to enhance our health-awareness, answering questions like "Do my breathing and heart rates reflect a healthy lifestyle?" They may also help address some of our concerns by answering questions like "Does my child breathe normally during sleep?" or "Does my elderly parent experience irregular heartbeats?" Furthermore, if non-intrusive in-home continuous monitoring of breathing and heartbeats existed, it would enable healthcare professionals to study how these signals correlate with our stress level and evolve with time and age, which could have a major impact on our healthcare system.

Unfortunately, typical technologies for tracking vital signals require body contact, and most of them are intrusive. Specifically, today's breath monitoring sensors are inconvenient: they require the person to attach a nasal probe [64], wear a chest band [152], or lie on a special mattress [4]. Some heart-rate monitoring technologies are equally cumber-

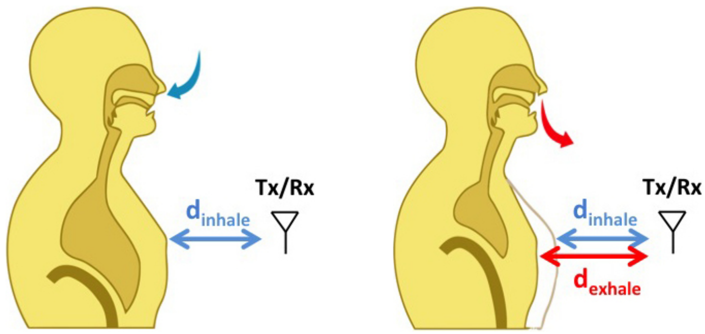


Figure 7-1: **Chest Motion Changes the Signal Reflection Time.** (a) shows that when the person inhales, his chest expands and becomes closer to the antenna, hence decreasing the time it takes the signal to reflect back to the device. (b) shows that when the person exhales, his chest contracts and moves away from the antenna, hence the distance between the chest and the antenna increases, causing an increase in the reflection time.

some since they require their users to wear a chest strap [63], or place a pulse oximeter on their finger [68]. The more comfortable technologies such as wristbands do not capture breathing and have lower accuracy for heart rate monitoring [46]. Additionally, there is a section of the population for whom wearable sensors are undesirable. For example, the elderly typically feel encumbered or ashamed by wearable devices [65, 138], and those with dementia may forget to wear them. Children may remove them and lose them, and infants may develop skin irritation from wearable sensors [148].

In this chapter, we ask whether it's possible for smart homes to monitor our vital signs remotely – i.e., without requiring any physical contact with our bodies. While past research has investigated the feasibility of sensing breathing and heart rate without direct contact with the body [52, 57, 58, 59, 88, 113, 179], the proposed methods are more appropriate for controlled settings but unsuitable for smart homes: They fail in the presence of multiple users or extraneous motion. They typically require the user to lie still on a bed facing the device. Furthermore, they are accurate only when they are within close proximity to the user's chest.

We introduce Vital-Radio, a new input device for tracking breathing and heartbeats without physical contact with the person's body. Vital-Radio works correctly in the presence of multiple users in the environment and can track the vital signs of the present users simultaneously. Also, Vital-Radio does not require the user to face the device or be aware of its presence. In fact, the user can be sleeping, watching TV, typing on her laptop, or checking her phone. Furthermore, Vital-Radio can accurately track a user's breathing and heart rate even if she is 8 meters away from the device, or in a different room.

Vital-Radio works by using wireless signals to monitor the minute movements due to inhaling, exhaling, and heartbeats. Specifically, it transmits a low-power wireless signal and measures the time it takes for the signal to reflect back to the device. The reflection time depends on the distance of the reflector to the device, and changes as the reflector moves. Fig. 7-1 illustrates the impact of breathing on the signal's reflection time. When the person inhales, his chest expands and moves forward, reducing the reflection time. In contrast, when the person exhales, his chest contracts moving away from the device, hence increasing the reflection time. Generally, even when the person is not directly facing our device, the wireless signal traverses his body and his vital signs cause periodic changes in the signal's reflection time. Vital-Radio measures these changes and analyzes them to extract breathing and heartbeats.

A key feature of Vital-Radio is its ability to monitor the vital signs of multiple people and operate robustly without requiring the users to lie still. The main challenge in delivering this feature is that any motion in the environment can affect the wireless signal and hence interferes with tracking breathing or heartbeats. Past work addresses this challenge by requiring that only one person be present in front of the device and that the person remains still. In contrast, Vital-Radio recognizes that one can address this problem by building on WiTrack (described in Chapters 4 & 5), which localizes users using wireless signals. Specifically, Vital-Radio first localizes each user in the environment, then zooms in on the signal reflected from each user and analyzes variations in his reflection to extract his breathing and heart rate. By isolating a user's reflection, Vital-Radio also eliminates other sources of interference including noise or extraneous motion in the environment, which may otherwise mask the minute variations due to the user's vital signs. This enables Vital-Radio to monitor multiple users' breathing and heart rates, and to operate at distances up to 8 m from the user or even from behind a wall.

We built a real-time prototype of Vital-Radio and validated its capabilities by conducting experiments with 14 subjects. For baselines, we use FDA-approved breathing and heart rate monitors; these include chest straps for monitoring the inhale-exhale motion and pulse oximeters placed on the subject's finger to monitor their heart rate. In our benchmark evaluation, we ask the users to wear the baseline monitors, while Vital-Radio monitors them remotely without any body contact. We compare the output of Vital-Radio with the ground truth from the FDA-approved baselines, demonstrating that Vital-Radio accurately tracks breathing patterns and heartbeats. Over more than 200 two-minute experiments, our results show that:

- Vital-Radio can accurately track a person's breathing and heart rate without body contact, even when the user is up to 8 meters away from the device, or behind a wall.
- Vital-Radio's median accuracy for breathing is 99.3% (error of 0.09 breath/minute) and for heart rate is 98.5% (0.95 beat/minute) when the person is 1 m away from the device. The accuracy decreases to 98.7% (error of 0.15 breath/minute) and 98.3% (1.1 beat/minute) when the person is 8 m away from the device.
- In an area that spans $8\ m \times 5\ m$, Vital-Radio can monitor the vital signs of up to three individuals with the same accuracy as for one person.

We also perform activity-focused experiments to explore Vital-Radio's monitoring capabilities. Specifically, we demonstrate that Vital-Radio can accurately measure users' breathing and heart rates while they are typing on their computer or using their cell phones. We also demonstrate that Vital-Radio can track sharp changes in vital signs. Specifically, we perform experiments where users are asked to exercise, and show how Vital-Radio accurately tracks the change in breathing and heart rates after exercising.

We believe Vital-Radio takes a significant step toward enabling smart homes that allow people to monitor their vital signals, and that its capabilities can have a significant impact on our health awareness and our health-care system.

■ 7.1 Related Work

The desire for non-contact monitoring of vital signs has occupied researchers since the late 70's [96]. Early work presented a proof of concept that the wireless signal is affected by movements of the chest. In these experiments, the person lies still on a bed and the sensor is placed only 3 cm away from the apex of the heart. The results are qualitative with no evaluation of accuracy.

Subsequently, military research explored the potential of building radars that can detect human presence through walls or under rubble by relying on the fact that breathing impacts wireless signals [84, 151, 171, 176]. Specifically, because wireless signals traverse obstacles, they could be used to sense the chest movements of a trapped victim through rubble or enable SWAT teams to sense movement from behind an obstacle and avoid being ambushed. However, since these systems target the military, they typically transmit at excessive power and use military-reserved spectrum bands [171, 176], which is not feasible for consumer devices. More importantly, this line of work generally focused on the detection of users by sensing motion due to their vital signs rather than estimating or monitoring the vital signs themselves.

Recently, the mounting interest in technologies for well-being has led researchers to investigate non-contact methods for analyzing vital signs. Current research on this topic can be divided into two areas: vision-based techniques and wireless systems. Specifically, advances in image processing allowed researchers to amplify visual patterns in video feeds (such as color changes due to blood flow) to detect breathing and heart rate [31, 169]; however, such video-based techniques require the user to face the camera and do not work when he/she turns around or is outside the camera's field of view.

Similarly, advances in wireless transmission systems and signal processing have enabled researchers to detect and analyze human vital signs. Past proposals use one of the following techniques: Doppler radar [57, 58, 59], WiFi [88, 113], or ultra-wideband radar [28, 52, 179]. The key challenge in using wireless signals to extract vital signs is that any motion in the environment affects the signal. Since breathing and heartbeats are minute movements, they can be easily masked by interference from any other source of movement in the environment. Furthermore, the presence of multiple users – even if none of them moves – prevents these systems from operating correctly since the wireless signal will be affected by the combination of their vital signs, making it hard to disentangle the vital signs of each individual. Past proposals deal with this problem by ensuring that there is only one source of motion in the environment: namely, the vital signs of the monitored individual. Hence, their experimental setup has one person, who typically lies still in close proximity to the device [6, 28, 52, 57, 58, 59, 88, 113, 179].

In contrast to these past systems, Vital-Radio has an intrinsic mechanism that enables it to separate different sources of motion in the environment. To do so, Vital-Radio builds on state-of-the-art wireless localization techniques [24], which can identify the distance between the device and different moving objects. Vital-Radio, however, uses these methods to disentangle the incoming signals based on distance, rather than estimate the actual location. This allows it to separate signals reflected off different bodies and body parts. It then analyzes their motion independently to estimate the breathing and heart rate of potentially multiple individuals.

7.2 Context and Scope

We envision that Vital-Radio can be deployed in a smart home to monitor its inhabitants' breathing and heart rates, without body instrumentation. The device can monitor multiple users' vital signs simultaneously, even if some of them are occluded from the device by a wall or a piece of furniture. A single device can monitor users' vital signs at distances up to 8 meters, and hence may be used to cover a studio or a small apartment. One can cover a larger home by deploying multiple Vital-Radio devices in the environment.

Vital-Radio's algorithms run continuously, separating signals from different users, then analyzing the signal from each user independently to measure his/her vital signs. However, when a user walks (or performs a large body motion), the chest motion is mainly impacted by the walk and no longer representative of the breathing and heart rate.[1] At home, there are typically sufficient intervals when a user is quasi-static; these include scenarios where the user is watching TV, typing on a laptop, reading a newspaper, or sleeping. Vital-Radio can use all of these intervals to monitor a user's vital signs, and track how they vary throughout the day.

7.3 Theory of Operation

Vital-Radio transmits a low power wireless signal and measures the time it takes its signal to travel to the human body and reflect back to its antennas. Knowing that wireless signals travel at the speed of light, we can use the reflection time to compute the distance from the

[1] The vast majority of vital signs monitors, including chest bands that monitor breathing and pulse oximeters that monitor heart rate, cannot provide accurate estimates when the user walks or moves a major limb [43, 91, 114]. To prevent such motion from causing errors in its vital-signs estimates, Vital-Radio automatically detects periods during which the user is quasi-static and computes estimates only during such intervals.

device to the human body. This distance varies slightly and periodically as the user inhales and exhales and his heart beats. Vital-Radio captures these minute changes in distance and uses them to extract the user's vital signs.

However, natural environments have a large number of reflectors, such as walls and furniture as well as multiple users whose bodies all reflect the wireless signal. To address these issues, Vital-Radio's operation consists of three steps:

- Isolate reflections from different users and eliminate reflections off furniture and static objects.
- For each user, identify the signal variations that are due to breathing and heartbeats, and separate them from variations due to body or limb motion.
- Analyze signal variations to extract breathing and heart rates.

In what follows, we describe how these steps enable us to monitor users' vital signs using Vital-Radio.

7.3.1 Step 1: Isolate Reflections from Different Users and Eliminate Reflections off Furniture and Walls

To understand the operation of Vital-Radio, let us consider the scenario in Fig. 7-2, where the device is placed behind the wall of a room that has two humans and a table. When Vital-Radio transmits a wireless signal, part of that signal reflects off the wall; the other part traverses the wall, reflects off the humans and the table inside the room, and then traverses the wall back to the device.

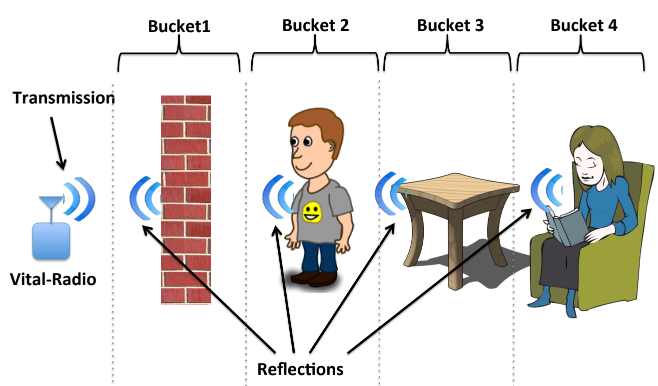


Figure 7-2: **Separating Reflectors into Different Buckets.** Vital-Radio uses a radar technique called FMCW to separate the reflections arriving from objects into different buckets depending on the distance between these objects and the device.

To isolate signals reflected off different objects, Vital-Radio uses a radar technique called FMCW (Frequency Modulated Carrier Waves). We refer the reader to §4.3 for a detailed description of how FMCW works. A key property of FMCW that we exploit in this chapter is that it enables separating the reflections from different objects into buckets based on their reflection times. Since wireless signals travel at the speed of light, signals reflected off objects at different distances would fall into different buckets.

However, in contrast to WiTrack (of Chapters 4 & 5), which uses FMCW to sense the amount of power arriving from different distances to localize the users, Vital-Radio uses the FMCW technique as a filter –i.e., it uses it to isolate the reflected signals arriving from different distances in the environment into different buckets, before it proceeds to analyze the signals in each of these buckets to extract the vital signs (step 2 below).

We use the same implementation of FMCW as the earlier chapters in this dissertation; in this implementation, the resolution of FMCW buckets is about 8 cm. This has two implications:

- Reflections from two objects that are separated by at least 8 cm would fall into different buckets. Hence, two users that are few feet apart would naturally fall into different buckets. For example, in Fig. 7-2, the wall, Bob, the table, and Alice are at different distances from our device, and hence FMCW isolates the signals reflected from each of these entities into different buckets, allowing us to focus on each of them separately.
- Using FMCW as a filter also allows us to isolate some of the limb motion from chest movements due to breathing and heartbeats. For example, the signal reflected off the user's feet will be in a different bucket from that reflected off the user's chest. Thus, having the user move his feet (in place) does not interfere with Vital-Radio's ability to extract the breathing and heart rate.

After bucketing the reflections based on the reflector's distance, Vital-Radio eliminates reflections off static objects like walls and furniture. Specifically, since static objects don't move, their reflections don't change over time, and hence can be eliminated by subtracting consecutive time measurements.

At the end of this step, Vital-Radio would have eliminated all signal reflections from static objects (e.g., walls and furniture), and is left with reflections off moving objects separated into buckets.[2]

7.3.2 Step 2: Identifying Reflections Involving Breathing and Heart Rate

After Vital-Radio isolates reflections from different moving users into separate buckets, it proceeds by analyzing each of these buckets to identify breathing and heart rate. For example, in Fig. 7-2, we would like to identify whether the user in bucket 2 is quasi-static and his motion is dominated by his vital signs, or whether he is walking around or moving a limb.

To do that, Vital-Radio zooms in on the signal reflection which it isolated in the corresponding bucket. This wireless reflection is a wave; the phase of the wave is related to the distance traveled by the signal as follows [147]:

$$\phi(t) = 2\pi \frac{d(t)}{\lambda}, \tag{7.1}$$

where λ is the wavelength of the transmitted signal, and $d(t)$ is the traveled distance from the device to the reflector and back to the device. The above equation shows that one can identify variations in $d(t)$ due to inhaling, exhaling, and heartbeats, by measuring the resulting variations in the phase of the reflected signal.

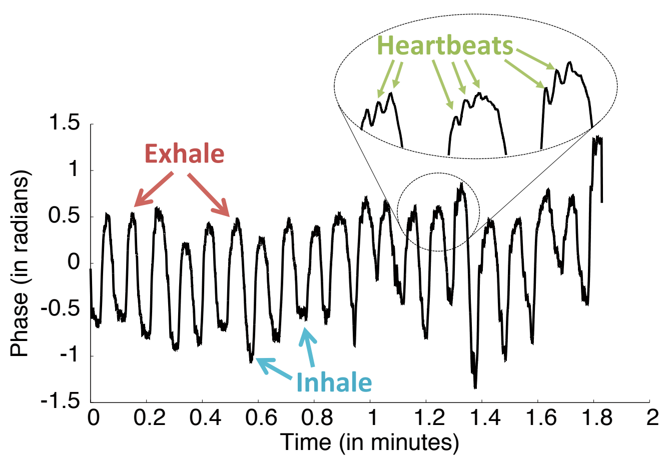


Figure 7-3: **Phase variation due to vital signs.** The figure shows the variations in phase due to breathing and heartbeats, where peaks and valleys in the phase correspond to exhale and inhale motions respectively; also, zooming in on the signal allows us to observe the heartbeats modulated on top of the breathing motion.

To illustrate how the phase varies with vital signs, let us consider the example in Fig. 7-

[2]While unlikely, it is possible that multiple users are at the same distance from the device and hence fall into the same bucket. To deal with such cases, one may deploy multiple devices so that if two users are at the same distance with respect to one device, they are at different distances with respect to another device.

1, where a user sits facing the device. When the person inhales, his chest expands and gets closer to the device; and when he exhales, his chest contracts and gets further away from the device. Because the phase and the distance to a reflector are linearly related, Vital-Radio can track a person's breathing. Fig. 7-3 shows the phase of the captured reflection as a function of time. Specifically, a peak in the phase corresponds to an exhale (highest distance from the device), and a valley in the phase corresponds to an inhale (smallest distance from the device). We note that our implementation uses a wavelength λ around 4.5 cm. According to the above equation, sub-centimeter variations in the chest distance due to breathing cause sub-radian variations in the phase, which is what we observe in the figure.

Similarly, a person's heartbeats cause minute movements of different parts of his body. Specifically, the physiological phenomenon that allows Vital-Radio to extract heart rate from signal reflections is ballistocardiography (BCG). BCG refers to movements of the body synchronous with the heartbeat due to ventricular pump activity [115]. Past work has documented BCG jitters from the head, torso, buttock, etc. [20, 31]. Periodic jitters cause periodic variations in the wireless signal allowing us to capture the heart rate. These movements translate to smaller fluctuations on top of the breathing motion in the wireless reflection as we can see from local peaks in Fig. 7-3. Note that the periodicity of breathing and heartbeats is independent of the user's orientation. For example, if the user has his back to the device, the valleys become peaks and vice versa, but the same periodicity persists.

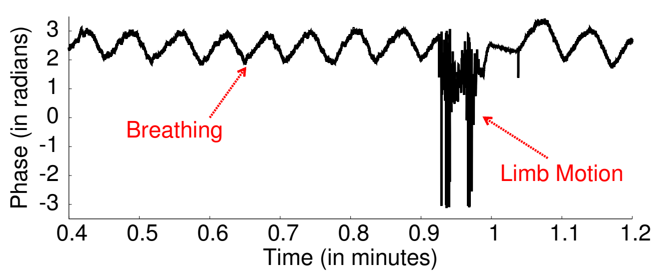


Figure 7-4: **Limb motion affects vital sign monitoring.** The figure shows the subject breathing until right before the 1 minute mark where he waves his hand. The device eliminates time intervals when such motion happens.

Still, an important question to answer is: what happens when a person moves around or moves a limb, and how can Vital-Radio distinguish such motions from breathing and heartbeats? To help answer this question, we show in Fig. 7-4 a scenario where the user waves his hand before the one minute mark resulting in aperiodic phase variations of the signal.

To deal with such scenarios, Vital-Radio exploits that motion due to vital signs is periodic, while body or limb motion is aperiodic. It uses this property to identify intervals of time where a user's whole body moves or where she performs large limb movements and discards them so that they do not create errors in estimating vital signs. To achieve this, Vital-Radio operates on time windows (30 seconds in our implementation). For each window, it measures the periodicity of the signal. If the periodicity is above a threshold, it determines that the dominant motion is breathing and heart rate; otherwise, it discards the window. A typical approach to measure a signal's periodicity is evaluating the sharpness of its *Fourier transform* (or FFT) [38]. Hence, we perform an FFT on each window, choose the FFT's peak frequency, and determine whether the peak's value is sufficiently higher than the average power in the remaining frequencies.[3]

This metric allows us to maintain intervals where a user does not perform large limb movements, including scenarios where the user types on her laptop or checks her phone. This is because, while these movements are indeed aperiodic, they do not mask the breathing or the heart rate since their power does not overwhelm the repetitive movements due to our vital signs.[4] Additionally, in some of these scenarios, the user's hands are stretched out to the laptop and away from his chest as he is typing. As a result, the major part of his typing motion falls into a separate FMCW bucket than the user's chest. Naturally, because the human body is connected, hand movements would still result in muscle stretches and minor shoulder jitters that are close to the user's chest; however, because such movements are weak and aperiodic, they are diluted at the output of the FFT. In contrast, periodic movements due to vital signs are enforced in the FFT operation, which results in maintaining intervals of such quasi-static scenarios.

The above steps allow us to filter out extraneous motion and focus on time windows where the dominant motion for each user is the breathing and heart rate. In the following section, we show how Vital-Radio extracts breathing and heart rate from these intervals.

[3] In our implementation, we choose this peak to be at least $5\times$ above the average power of the remaining frequencies.

[4] Mathematically, these signals would appear as "white noise" in low frequencies, and are filtered out in Step 3 of Vital-Radio's operation.

7.3.3 Step 3: Extracting Breathing and Heart Rate

Breathing Rate Extraction

Because breathing is a periodic motion, we can extract the frequency (rate) of breathing by performing a Fourier transform (an FFT). The peak at the output of the FFT will correspond to the dominant frequency, which in our case is the breathing rate. Specifically, we perform an FFT of the phase signal in Fig. 7-3 over a 30 second window and plot the output in Fig. 7-5. The peak of this signal gives us an initial estimate of the person's breathing rate.

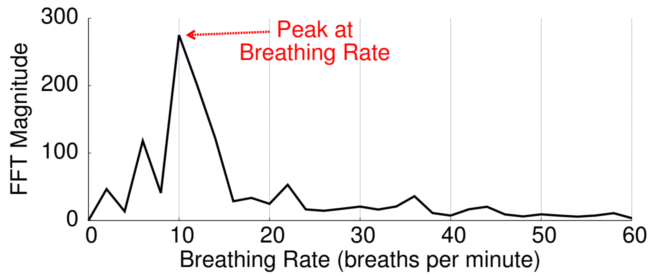


Figure 7-5: **Output of Fourier Transform for Breathing.** The figure shows the output of the FFT performed on the phase of the signal of Fig. 7-3. The FFT exhibits a peak around 10 breaths/minute, providing a coarse estimate of the breathing rate.

However, simply taking the peak of the FFT does not provide an accurate estimate of breathing rate. Specifically, the frequency resolution of an FFT is $1/window\ size$. For a window size of 30 seconds, the resolution of our breath rate estimate is $\approx 0.033Hz$, i.e., 2 breaths/minute. Note that a larger window size provides better resolution, but is less capable of tracking changes in breathing rate. To obtain a more precise measurement, we exploit a well-known property in signal processing which states that: if the signal contains a single dominant frequency, then that frequency can be accurately measured by performing a linear regression on the phase of the complex time-domain signal [110]. Hence, we perform an additional optimization step, whereby we filter the output of the FFT, keeping only the peak and its two adjacent bins; this filtering allows us to eliminate noise caused by extraneous and non-periodic movements. Then, we perform an inverse FFT to obtain a complex time-domain signal $s'(t)$. The phase of $s'(t)$ will be linear and its slope will correspond to the breathing frequency, i.e., the breathing rate. Mathematically, we can compute an accurate estimate of the breathing rate (in terms of breaths per minute) from the following equation:

$$Estimate = 60 \times \frac{slope\{\angle s'(t)\}}{2\pi}, \tag{7.2}$$

where the factor of 60 transforms this frequency from Hz (i.e., $1/second$) to breaths/minute.

Heart Rate Extraction

Similar to breathing, the heartbeat signal is periodic, and is modulated on top of the breathing signal, as shown in Fig. 7-3. However, the breathing signal is orders of magnitude stronger than the heartbeat. This leads to a classical problem in FFT's, where a strong signal at a given frequency leaks into other frequencies (i.e., leaks into nearby bins at the output of the FFT) and could mask a weaker signal at a nearby frequency.

To mitigate this leakage, we filter the frequency domain signal around [40-200] beats per minute; this allows us to filter out breathing, which is typically between 8 and 16 breaths per minute [149][5] as well as high frequency noise (which is higher than 200 beats per minute).

We plot the output of this obtained frequency domain signal in Fig. 7-6, and pick the maximum *peak* of this output as the frequency that corresponds to the heart rate. Note that we do not simply pick the *absolute* maximum of the FFT, because this absolute maximum is typically the first bin after filtering (i.e., around 40 beats/minute), and is due to the leakage from the breathing. In contrast, in this example, the peak occurs at 66 beats/minute.

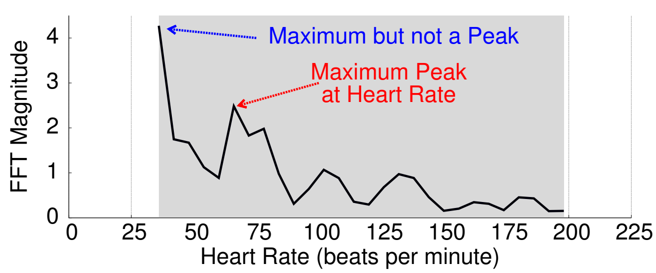


Figure 7-6: **Output of Fourier Transform for Heart Rate.** The figure shows the output of the FFT after applying a hanning window and filtering between [40-200] beats/minute. The highest peak (i.e., "local maximum") provides a coarse estimate of the heart rate.

Similar to breathing, simply taking the peak of the FFT leads to poor resolution. To obtain a more precise estimate of the heart rate, we take an inverse FFT of the signal in the FFT bin corresponding to the heart rate peak and the two adjacent FFT bins. We then regress on the phase of this signal using equation 7.2. After this regression step, the obtained heart rate is 66.7 beats/minute, whereas the ground truth heart rate obtained from

[5]In fact, this filtering allows us to filter out the breathing signal and its first harmonic.

a pulse oximeter is around 66.5 beats/minute.

Finally, we note that for computing heart rate, we use an FFT window of 10 seconds only. This window is long enough to capture the periodicity of heartbeats but it is short enough to quickly react to an increase/decrease in heart rate. Also note that the FFT is computed over overlapping windows that are shifted by 30ms, hence providing a new estimate every 30ms.

7.4 Implementation

Our implementation consists of the following components:

Hardware: We use the FMCW radio described in §4.6. The device generates a signal that sweeps from 5.46 GHz to 7.25 GHz every 2.5 milliseconds, transmitting sub-mW power. These parameters are chosen such that the transmission system is compliant with FCC regulations for consumer electronics.

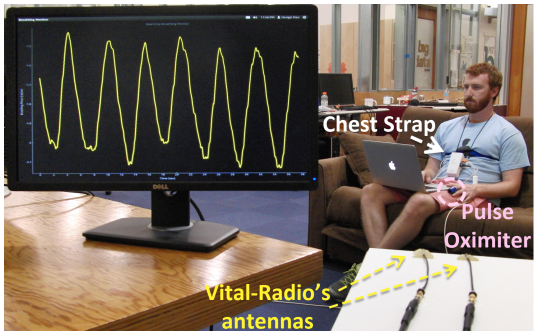
The FMCW radio connects to a computer over Ethernet. The received signal is sampled (digitized) and transmitted over the Ethernet to the computer for real-time processing.

Software: We implement the signal processing algorithms described in the previous sections in C++. The code runs in realtime, plotting on the screen the breathing and heart rate as function of time and at the same time logs them to a file. The code operates on shifted overlapping FFT windows and generates new estimates every 30ms. The output also shows user motion –i.e., the code tags every 30ms window to show whether the user is quasi-static or performing a major motion.

7.5 Experimental Evaluation

Participants: To evaluate the performance of Vital-Radio we recruited 14 participants (3 females). These participants were between 21 and 55 years old ($\mu = 31.4$), weighed between 52 and 95 kg ($\mu = 78.3$), and were between 164 and 187 cm tall ($\mu = 175$). During the experiments, the subjects wore their daily attire, including shirts, T-shirts, hoodies, and jackets with different fabric materials.

Ground Truth: To determine Vital-Radio's accuracy, we compare its output against the



(a) **Setup.**

(b) **Antennas.**

Figure 7-7: **Experimental Setup.** (a) shows a user sitting about 2.5m away from Vital-Radio's antennas; the user also wears a chest strap and a pulse oximeter, which are connected to the Alice PDx for obtaining ground truth measurements. (b) shows one of Vital-Radio's antennas placed next to a quarter.

Alice PDx [1], an FDA approved device for monitoring breathing and heart rate. The Alice PDx is equipped with a chest band and a pulse oximeter. The chest band is strapped around each subject's chest to monitor breathing, and the pulse oximeter is placed on his/her finger to monitor heart rate during the experiment.

Experimental Environment: We perform our experiments in a standard office building; the interior walls are standard double dry walls supported by metal frames with sheet rock on top. The evaluation environment contains office furniture including desks, chairs, couches, and computers.

Throughout the experiments, Vital-Radio's antennas are placed on a table, about 3 feet above the ground as shown in Fig. 7-7. The user sits at some distance from these antennas and wears the Alice PDx's chest band and pulse oximeter to obtain the ground truth measurements as shown in the figure. In our evaluation, we vary the distance and orientation of the user with respect to Vital-Radio to determine its accuracy in different scenarios as we show in the next section.

■ 7.5.1 Core Experiment: Accuracy Versus Distance

We would like to validate Vital-Radio's ability to monitor a subject's breathing and heart rate at different distances from our device. In this experiment, we place the device in the corner of a large room, whose floor plan is shown in Fig. 7-8. The device's antennas are pointed toward the center of the room to ensure that they capture motion inside that room. We ask the subject to sit on a chair at marked locations whose distances range from 1 m to 8 m away from the device. In each experiment, the subject sits on a chair facing Vital-Radio's antennas and wears the Alice PDx, as shown in Fig. 7-7.

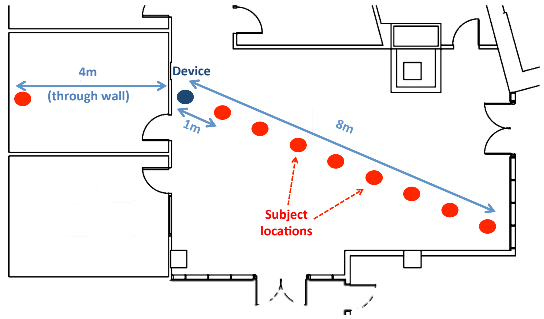


Figure 7-8: **Testbed.** The figure shows a layout of our experimental setup, marking the location of Vital-Radio in navy blue, and the different locations where our monitored subjects sat down in red.

We run a total of 112 experiments, where we ask each of the 14 subjects to sit at the marked locations from 1m to 8 m.[6] Each experiment lasts for two minutes, during which the user sits facing the device in each of these locations. We extract the breathing and heart rate in real-time using Vital-Radio and log these vital signs using the AlicePDx. During each two minute experiment, Vital-Radio outputs a vital sign estimate every 30 milliseconds, leading to a total of 448,000 measurements across all experiments and all locations.

Based on the ground truth measurements using the Alice PDx, the subjects' breathing rates range from 5 to 23 breaths/minute, while their heart rates vary from 53 to 115 beats/minute. These rates span the range of adult breathing and heart rates [104, 149].[7]

Breathing Rate Accuracy

We compare the output of Vital-Radio with that of the Alice PDx, and plot the median and 90^{th} percentile accuracy of breathing as a function of distance from 1 to 8 meters in Fig. 7-9. The figure shows that our median accuracy is 99.3% at 1 m and remains as high as 98.7% at 8 m from the device. It also shows that our 90^{th} percentile accuracy is higher than 90% across all these distances.

[6] We limit the experiments to distances of 8 m because the localization accuracy of FMCW-based systems for consumer applications drops beyond this range (see §4.8.2).

[7] We note that one of our subject has a significantly high heart rate of 115 beats/minute. We confirmed with the subject that this value, which was measured with the Alice PDx, is compatible with his medical records. Also one of our subjects has a low breathing rate of 5 breaths/minute, as measured by the Alice PDx. This subject practices yoga on a daily basis.

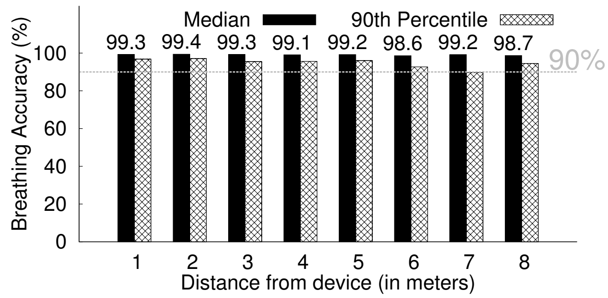


Figure 7-9: **Breathing Accuracy vs Distance.** The figure shows Vital-Radio's breathing accuracy versus its distance to the subject.

Heart Rate Accuracy

We plot the median and 90^{th} percentile accuracy of heart rate as a function of distance from 1 to 8 meters in Fig. 7-10. The figure shows that our median accuracy is 98.5% at 1 meter and drops to 98.3% at 8 meters from the device. It also shows that our 90^{th} percentile accuracy remains higher than 90%, even with the subject is 8 m away from the device.

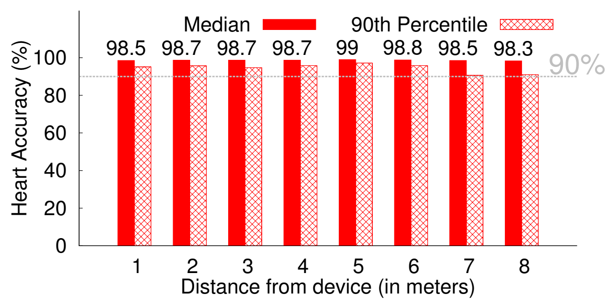


Figure 7-10: **Heart Rate Accuracy vs Distance.** The figure shows Vital-Radio's heart rate accuracy versus its distance to the subject.

7.5.2 Accuracy in Various Scenarios

Accuracy versus Orientation

To validate that Vital-Radio operates correctly even when subjects do not directly face the device, we run experiments where we ask our subjects to orient themselves in different directions with respect to the device. Specifically, we ask each subject to sit at the 4 m distance from Vital-Radio and we run experiments in four different orientations: subject faces the device, subject has his back to the device, and the subject is facing left or right (perpendicular) to the device.

We plot the median accuracies for breathing and heart rate for these four different orientations in Fig. 7-11. The figure shows that, indeed, when the user faces the device, the median accuracy of breathing and heart rate is highest (99.1% and 98.7% respectively). However, this accuracy only slightly drops by at most 3% across all the different orientations. Note that the device can detect chest motion even when that motion is perpendicular. This is because when one inhales, his chest expands in all directions, and Vital-Radio can detect a chest side expansion though it is minute.[8]

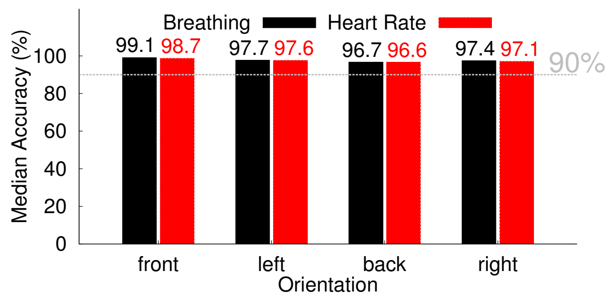


Figure 7-11: **Accuracy versus Orientation.** The figure shows Vital-Radio's median accuracy for breathing and heart rate for a user sitting 4 m from the device and facing different directions.

Next, we validate that Vital-Radio does not require the user to be along a straight line facing the antenna. Specifically, we place the antennas at the center of the room, and ask users to sit at a distance of 4 m from the antennas and at angles ranging from $-90°$ to $+90°$ with respect to the pointing direction of the antennas. We perform 20 one-minute experiments with different subjects at different angles. The results show that Vital-Radio can capture the user's vital signs as long as she is at an angle between $-75°$ and $+75°$ with respect to the antenna's pointing direction. Specifically, the median accuracy is above 98% when the user is on a straight line with respect to the antenna, and decreases to 96% at the far edge (i.e., $\pm 75°$).[9]

Through-Wall Accuracy

In order to test the ability of Vital-Radio to measure user's vital signs even when they are in a different room, we run a set of through-wall experiments where the device is placed in a different room than our subjects. Specifically, we use the experimental setup in Fig. 7-8.

[8]Such expansion is no smaller than variations due to heartbeats.
[9]This result is expected since Vital-Radio uses log-periodic antennas whose directionality is around $150°$.

The device is kept in the larger room, while the subject sits in an adjacent room behind a wall. The subject faces the device and is about 4 m from it.

Across all experiments, our median accuracies are 99.2% and 90.1% respectively for breathing and heart rate. These results indicate that the breathing rate remains almost the same both in the presence and absence of the wall (at the same distance of 4 m). However, the median heart rate accuracy drops due to the fact that the wall attenuates the heart rate signal significantly (which was already a very minute signal), hence, reducing our signal-to-noise ratio. Still, the heart rate accuracy remains around 90% even in such through-wall scenarios.

Multi-User Accuracy

We are interested in evaluating Vital-Radio's accuracy for multi-user vital sign monitoring. Hence, we perform controlled experiments, where we ask three of our users to sit on a chair at the 2 m, 4 m, and 6 m marks in Fig. 7-8. In each experiment, Vital-Radio determines that there are 3 users, each at his respective distance from the device, and outputs the vital signs of each; however, the baseline (AlicePDx) can only monitor a single user at any point in time. Hence, to evaluate accuracy, we first connect the baseline to the first user and compare its output to the output of VitalRadio for the user at that distance and for that moment. Then, we move the baseline to the remaining users in succession.

We run 20 experiments with different sets of subjects and plot the accuracies in Fig. 7-12. The figure shows that Vital-Radio's breathing and heart rate monitoring accuracy is around 98% for all three users. Note also that the median accuracy of the nearest user is higher than that of the further two users because of the increase in distance between these users and the device. These results verify that Vital-Radio can monitor multiple users' vital signs, and that its monitoring accuracy for multiple users is the same as that for a single user.

Next, we would like to confirm that Vital-Radio can accurately capture the vital signs of a quasi-static user while other users are moving in the environment. In principle, Vital-Radio should still operate correctly since FMCW separates reflections from different users based on their distance to the device. Hence, we run experiments where we ask one of our subjects to sit on a chair at the 3 m mark in Fig. 7-8 while asking another subject to walk

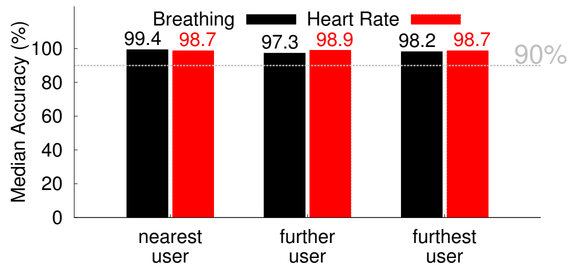


Figure 7-12: **Multi-User Accuracy.** The figure shows Vital-Radio's median accuracy in monitoring the vital signs of 3 users simultaneously. The users are sitting at 2 m, 4 m, and 6 m from the device.

around in the room. Over 20 experiments with different subjects, the median accuracy of breathing and heart rate remains above 98% for the monitored user as long as the moving user is at a distance of at least 1.5 m away from him. This accuracy drops below 75% when the moving user gets closer than 1 m to the monitored person. This is because when the two users are closer than 1 m, the reflections off their bodies interfere with each other, preventing Vital-Radio from isolating the signal variations due to the monitored user's vital signs.

7.5.3 Activity-Focused Experiments

Daily In-Place Activities

We would like to evaluate Vital-Radio's accuracy in monitoring users as they perform in-place day-to-day activities, such as typing on their laptops, watching TV, or sleeping. Thus, we divided the subjects into two groups: one interacting with their laptops and another interacting with their smartphones. In each experiment, we ask the subject to sit at 4 m from the device and naturally use his/her phone or laptop. Each experiment lasts for 5 minutes; the user reports at the end the activities he/she performed with their laptop or phone. The reports show that the users performed various tasks ranging from checking and responding to their emails to reading news on the web or logging on to Facebook or Instagram. Some users were texting using their phones and one user worked on a problem set on his laptop.

Throughout the experiments where our subjects used their phones, the median accuracies for breathing and heart rate were 99.4% and 98.9% respectively. These accuracies slightly dropped to 99.3% and 98.7% when our subjects were using their laptops. This minor drop in accuracy is expected because using a laptop typically involves slightly larger movements than using a phone, leading to a slight reduction in Vital-Radio's accuracy. Note, however, that these accuracies are almost the same as those when subjects were sitting still at the same distance with respect to our antenna. Hence, Vital-Radio was able to monitor users' breathing and heart rate as they perform day-to-day activities that do not require them to move around their apartments.

Exercising and Health-Awareness

Heart rate recovery – which corresponds to how fast a person's heart rate decreases after exercising – is an important metric for determining how healthy a person's heart is. Specifically, a stronger heart has a fast heart rate recovery, and that recovery rate is a predictor of mortality [48, 81]. Hence, an important way in which Vital-Radio may contribute to a smart home inhabitants' well-being is by accurately measuring the heart rate after users exercise.

To evaluate this capability, we evaluate the accuracy of Vital-Radio's real-time vital sign monitoring after asking our subjects to exercise. Specifically, each of the subjects jumps rope for 2 minutes then sits down on a chair, 4 m away from the device, and breathes normally. During these experiments, we also ask our subjects to wear the Alice PDx chest band and oximeter to obtain ground truth measurements for their vital signs. While both Vital-Radio and Alice PDx cannot accurately measure heart rate during the excessive motion of jumping, they can both measure the vital signs when a user sits down after exercising.

Fig. 7-13 overlays the heart rate estimated by Vital-Radio (in red) on top of the the ground-truth heart rate of a subject as monitored by the pulse oximeter (in black), throughout a two-minute period after the subject stops exercising. The figure shows that Vital-Radio can effectively track the variations in the heart rate. Note that, throughout this two-minute period, there's a downward trend in the heart rate (from about 93 beats/minute to around 70 beats/minute), which is expected since the subject is in a resting state after exercising. Also, note that throughout this period, the heart rate varies continuously about that trend, and that both the ground truth and Vital-Radio are able to capture these variations.

Across the experiments with multiple subjects, Vital-Radio's median accuracy in mea-

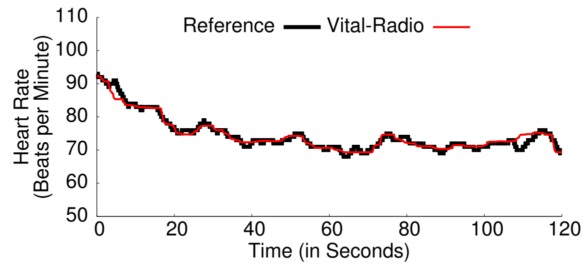


Figure 7-13: **Tracking Heart Rate after Exercising.** The figure shows how Vital-Radio' can accurately and in real-time track a user's heart rate as it decreases after exercising.

suring breathing and heart rate is 99.4% and 99% respectively, and the 90^{th} percentile is 91.7% and 96.8%. These figures are similar to the accuracy achieved in our previous experiments, where subjects were fully rested, indicating that Vital-Radio can indeed capture our vital signs and track them accurately even as they vary.

7.6 Limitations

In this section, we elaborate on the limitations of Vital-Radio:

- *Minimum Separation between Users:* Vital-Radio uses FMCW to separate reflections from different users before extracting the per-user vitals. For ideal point reflectors, FMCW can separate reflections from two objects if they are at least C/2B apart (see §4.3), where B is the bandwidth and C is the speed of light. For Vital-Radio, this translates to a theoretical minimum separation of 8 cm. However, because a human is not a point reflector, our experiments show that a separation of 1–2 m is needed for high accuracy.
- *Monitoring Range:* Since Vital-Radio is a wireless system, it requires a minimum signal-to-noise ratio (SNR) to extract the signal from the noise, and this SNR bounds its range and accuracy. Specifically, the maximum distance at which Vital-Radio detects users is 8m. This is because the SNR drops with user distance from the device.
- *Quasi-static Requirement:* Our implementation measures the vital signs only for quasi-static users (e.g., typing, watching TV). This is because signal variations due to full body motion would otherwise overwhelm the small variations due to vital signs, and prevent Vital-Radio from capturing the minute movements.
- *Non-human Motion:* Vital-Radio uses FMCW to separate reflections from different objects in space; hence, it can separate the reflection of various moving objects (e.g., humans, fans,

pets). It then analyzes the reflections of each moving object to detect breathing. Since the periodicity of breathing is much lower than fans, the device never confuses a fan as a human. Even if the device confuses a fan for a human, it will not affect the vital signs of the real humans since their signals are separated from the fans by FMCW. However, it may still identify the presence of a pet and output its breathing and heart rate assuming it is another user in the environment.

7.7 Future Opportunities

The HCI community has significant literature on the use of physiological sensing for various applications [33, 76, 103, 143]. In particular, HCI researchers have used physiological sensing to *evaluate user experience* including emotional reactions, stress levels, cognitive performance, and user engagement. But, a key concern with past sensors (e.g., oximeters, EEG, FNIRs, GSR) is that they require direct contact with the user's body, and hence may affect a user's response. In contrast, Vital-Radio doesn't require users to to be aware of its presence, and hence doesn't interfere with user experience.

Additionally, Vital-Radio enables *new interface and interaction capabilities*. For example, it may be incorporated into user interfaces to adapt to a user without requiring him to wear sensors. Also, it can enable environments to adapt the music or lighting by sensing the user's vital signs and inferring his mood. Further, a user walking up to a Vital-Radio-enabled kiosk in an unfamiliar location (such as an airport) might receive customized assistance based on his stress level.

Beyond these applications, we believe that Vital-Radio can impact a wide array of areas in HCI including quantified self, smart homes, elderly care, personal health and well-being, and mobile emotional sensing.

CHAPTER 8
Conclusion

Can wireless signals extend our senses? This dissertation answers the question affirmatively. It introduces new algorithms and software-hardware systems that use radio frequency (RF) signals to extend our senses along multiple dimensions. The presented technologies demonstrate that we can "see" through walls, track the exact location of a person in a closed room, and identify the person by relying purely on the reflections of RF signals off his/her body. These systems can also track our gestures in mid-air and enable us to control appliances by simply pointing at them. They can even monitor our breathing and heart rate without any physical contact with our bodies, even if we are behind a wall.

Summary of Contributions: The contributions of our research can be viewed through various lenses:

- From a ubiquitous sensing perspective, we present a fundamentally new approach for sensing the human body. In particular, in contrast to traditional approaches which required instrumenting the human body with sensors, our research does not require any physical contact with the human body.
- From a networking perspective, this dissertation expands the role that wireless networks can play in our daily lives. Specifically, in contrast to today's networks which use wireless signals to communicate, we show that future networks may use these signals for sensing, gesture control, and health monitoring.
- Our contributions may also be viewed from graphics/vision perspective, whereby we

demonstrate how RF signals can be used to gain visual access to new information, which we could not otherwise see.

Broader Impact: The research in this dissertation has already borne fruit, particularly in the health-care domain. First, it is the core underlying technology for a device that is currently deployed in elderly homes to enhance elderly safety. In particular, in contrast to today's fall detection solutions, which require the elderly to wear pendants, our developed technology on 3D motion tracking can detect falls without requiring the elderly to hold or wear any device. This device is being developed by a recent startup, Emerald. Second, this research is currently being used in a clinical study for sleep apnea monitoring at Massachusetts General Hospital. Specifically, in contrast to today's solutions for sleep apnea diagnosis require instrumenting the patient's body with many sensors, our developed technology on remote monitoring of breathing and heart rate can be used to detect sleep apnea without any physical contact with the patient's body. Finally, multiple medical institutions – including the medical centers at Boston University and University of California at San Francisco – are studying applications for this research in domains that range from pediatrics to geriatrics.

8.1 Looking Forward

We have only started scratching the surface of possibilities for RF as a sensing modality. While this dissertation has taken major steps in unlocking some of these possibilities, the presented designs exhibit limitations which would be interesting to explore in the future. In this section, we revisit these limitations and highlight some of the exciting avenues for future research.

We start by recapitulating the limitations highlighted in the earlier chapters of this dissertation:

- *Scale:* Our implementations can accurately track up to 4 moving users and 5 static users (as demonstrated in Chapter 5). These numbers may be sufficient for in-home tracking. However, it is always desirable to scale these systems to track more users. Furthermore, our sensing range is limited to 10m due to the low transmitted power. To cover larger areas and track more users, one may deploy multiple devices and hand off the trajectory tracking from one to the next, as the person moves around. Managing such a network of devices, coordinating their hand-off, and arbitrating their medium access are interesting problems to explore.

- *Reconstruction Resolution:* Our current method for reconstructing a human figure (described in Chapter 6) captures the human figure by stitching consecutive snapshots. In addition, our implementation adopts a simple model of the human body for segmentation and skeletal stitching. Future work can explore more advanced models to capture finer-grained human skeleta and over finer time resolutions.

- *Uninterrupted Vital-Sign Monitoring:* Our implementation of Vital-Radio measures the vital signs only for quasi-static users (e.g., typing, watching TV). This is because signal variations due to full body motion would otherwise overwhelm the small variations due to vital signs, and prevent Vital-Radio from capturing the minute movements. Our current implementation identifies such events and discards the corresponding measurements to prevent them from creating errors. Overcoming this limitation to enable uninterrupted vital sign monitoring is an interesting avenue for future work.

- *Non-human Motion:* Our designs in Chapters 5 & 7 use FMCW to separate reflections from different objects in space; hence, they can separate the reflection of various moving objects (e.g., humans, fans, pets). However, they do not try to distinguish the type of moving objects. Future work may combine the vital-sign estimates (from Chapter 7) with a model of a human body (as in Chapter 6) to overcome this limitation.

Beyond overcoming these limitations, we envision that future wireless systems will use RF as a sensing modality along multiple avenues:

- *Emotion Recognition:* RF sensing can enable machines to interact with us at deeper levels than today's interfaces. In particular, past research on affective computing has shown that our vital signs are correlated with our emotions. Thus, by bridging Vital-Radio (which can track breathing and heart rate) with affective computing, we can enable smart environments and interfaces to sense and adapt to our emotional reactions, stress levels, and cognitive performance.

- *Health Diagnosis and Prediction:* Non-invasive monitoring and diagnosis are active areas of research, and we believe that wireless sensing can deliver powerful solutions. Vital-

Radio has taken an initial step towards this vision by enabling remote sensing of breathing and heart rate using RF. While powerful, this technology cannot capture critical vital signs that do not cause body movements, such as blood pressure, oxygen saturation, and glucose levels. Extending RF to capture such vitals can render ICU vital sign monitors completely non-invasive and enable continuous monitoring of diabetes patients.

- *Robotic Systems:* Wireless sensing can empower robots with new capabilities in home and industrial environments. Our research has shown how we can detect, track, and reconstruct human figures from RF signals through occlusions. Such capabilities can be used by autonomous vehicles to detect pedestrians, by drones in search-and-rescue scenarios, and by personal robots for more natural human-robot interaction.

- *Smart Environments:* The systems we built can accurately track humans, recognize their gestures, and monitor their vital signs; however, they lack higher-level semantics in understanding human activities – i.e., they cannot understand exactly what a person is doing. By composing gestures and movements into higher level tasks, we can enable smart environments to understand and adapt to our behavior and to actively contribute to our lifestyles and well-being.

In sum, we believe that future wireless systems will use RF for sensing to deliver services that touch our everyday lives, similar to how their use of RF for communications has made Wi-Fi and cellular indispensable. This dissertation has made multiple strides in that direction. To do so, it bridges state-of-the-art concepts and tools from diverse areas including networking, signal processing, HCI, and Graphics. It also builds on a deep understanding of RF signals, operates across software-hardware boundaries, and introduces new systems and new algorithms that require redesigning the entire computing stack, from the hardware to the applications. We believe that this approach will become a necessity as wireless devices become ever-more ubiquitous and as their services keep expanding beyond communications in the coming decades.

APPENDIX A

Convergence of Iterative Nulling in WiVi

We prove why iterative nulling proposed in §3.3 converges. Vital-Radio models the channel estimate errors as additive (in line with common practice of modeling quantization error [110]). Hence, by substituting $\hat{h_1}$ with $h_1 + \Delta_1$, and $\hat{h_2}$ with $h_2 + \Delta_2$, in Eq. 3.1, we obtain:

$$h_{res} = h_1 + h_2 \left(-\frac{h_1 + \Delta_1}{h_2 + \Delta_2} \right) \approx \frac{h_1}{h_2} \Delta_2 - \Delta_1 + \frac{\Delta_1 \Delta_2}{h_2} \tag{A.1}$$

which follows from the first order Taylor series approximation of $\frac{1}{1-x}$ since $\Delta_2 << h_2$.

Iterating on h_1 alone. We first analyze how the algorithm converges if it were iterating only on Step 1. According to Algorithm 1, $\hat{h_1}$ is refined to $h_{res} + \hat{h_1}$. By updating the precoding vector, the new received channel after nulling h'_{res} is $h_{res} \frac{\Delta_2}{h_2}$ by applying the first order Taylor series approximation of $\frac{1}{1+\Delta_2/h_2}$ since $\Delta_2 << h_2$. Hence, $|h'_{res}| << |h_{res}|$. Therefore, after the i-th iteration, $h_{res}^{(i)}$ becomes $h_{res}^{(0)} \left(\frac{\Delta_2}{h_2} \right)^i$.

Iterating on h_2 alone. We now analyze how the algorithm converges if it were iterating only on Step 2. According to Algorithm 1, $\hat{h_2}$ is refined to $\left(1 - \frac{h_{res}}{\hat{h_1}}\right) \hat{h_2}$. By updating the precoding vector, the new received channel after nulling is:

$$h'_{res} \approx h_1 - \frac{\hat{h_1}}{\hat{h_2}} h_2 \left(1 + \frac{h_{res}}{\hat{h_1}} \right) = h_{res} \frac{\Delta_2}{h_2} \tag{A.2}$$

which follows from the first order Taylor series approximation of $\frac{1}{1-h_{res}/\hat{h_1}}$ since $h_{nulling} <<$

h_1. Hence, $|h'_{res}| << |h_{res}|$, and $h_{res}^{(i)}$ converges as above.

Iterative nulling on h_1 and h_2. By the above arguments, after i iterations on h_1 and j iterations on h_2, the nulled channel becomes:

$$h_{res}^{(i,j)} = h_{res}^{(0)} \left(\frac{\Delta_2}{h_2} \right)^{i+j} \tag{A.3}$$

References

[1] Alice PDx. http://www.healthcare.philips.com/main/homehealth/sleep/alicepdx/default.wpd. Phillips.

[2] Fitbit Flex. http://www.fitbit.com/flex.

[3] How Signal is affected. www.ci.cumberland.md.us/. City of Cumberland Report.

[4] iBabyGuard. http://www.ibabyguard.com/.

[5] Insteon ApplianceLinc. http://www.insteon.com. Insteon.

[6] Kai Medical. http://www.kaimedical.com/.

[7] LAN/MAN CSMA/CDE (ethernet) access method. IEEE Std. 802.3-2008.

[8] Leap Motion. https://www.leapmotion.com.

[9] Nintendo Wii. http://www.nintendo.com/wii.

[10] RadarVision. http://www.timedomain.com. Time Domain Corporation.

[11] Seeing through walls - MIT's Lincoln Laboratory. http://www.youtube.com/watch?v=H5xmo7iJ7KA.

[12] Urban Eyes. https://www.llnl.gov. Lawrence Livermore National Laboratory.

[13] USRP N210. http://www.ettus.com. Ettus Inc.

[14] VICON T-Series. http://www.vicon.com.

[15] WA5VJB antenna. http://www.wa5vjb.com. Kent Electronics.

[16] X-box Kinect. http://www.xbox.com. Microsoft.

[17] Zebra MotionWorks. https://www.zebra.com/us/en/solutions/location-solutions/sports-tracking.html. Zebra.

[18] *Understanding the FCC Regulations for Low-power, Non-licensed Transmitters*. Office of Engineering and Technology Federal Communications Commission, 1993.

[19] Kyllo v. United States, US Supreme Court, 533 U.S. 27, 121 S. Ct. 2038, 150 L. Ed. 2d 94, 2001.

[20] Oeyvind Aardal, Yoann Paichard, Sverre Brovoll, Tor Berger, Tor Sverre Lande, and S-E Hamran. Physical working principles of medical radar. *Biomedical Engineering, IEEE Trans. on*, 2013.

[21] Heba Abdelnasser, Moustafa Youssef, and Khaled A Harras. Wigest: A ubiquitous wifi-based gesture recognition system. In *IEEE INFOCOM*, 2015.

[22] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. Capturing the human figure through a wall. *ACM SIGGRAPH Asia*, 2015.

[23] Fadel Adib, Zachary Kabelac, and Dina Katabi. Multi-Person Localization via RF Body Reflections. In *Usenix NSDI*, 2015.

[24] Fadel Adib, Zachary Kabelac, Dina Katabi, and Robert C. Miller. 3D Tracking via Body Radio Reflections. In *Usenix NSDI*, 2014.

[25] Fadel Adib and Dina Katabi. See through walls with Wi-Fi! In *ACM SIGCOMM*, 2013.

[26] Fadel Adib, Hongzi Mao, Zachary Kabelac, Dina Katabi, and Robert C. Miller. Smart Homes That Monitor Breathing and Heart Rate. In *ACM CHI*, 2015.

[27] Brett Allen, Brian Curless, and Zoran Popović. The space of human body shapes: reconstruction and parameterization from range scans. In *ACM Transactions on Graphics (TOG)*, 2003.

[28] L. Anitori, A de Jong, and F. Nennie. Fmcw radar for life-sign detection. In *IEEE Radar Conference*, 2009.

[29] Roger Appleby and Rupert N Anderton. Millimeter-wave and submillimeter-wave imaging for security and surveillance. *Proceedings of the IEEE*, 2007.

[30] Paramvir Bahl and Venkata N Padmanabhan. Radar: An in-building RF-based user location and tracking system. In *IEEE INFOCOM*, 2000.

[31] Guha Balakrishnan, Fredo Durand, and John Guttag. Detecting pulse from head motions in video. In *IEEE CVPR*, 2013.

[32] Edward J Baranoski. Urban operations, the new frontier for radar. *DARPA Tech*, 2005.

[33] Javier A Bargas-Avila and Kasper Hornbæk. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *ACM CHI*, 2011.

[34] Petr Beckmann and Andre Spizzichino. *The scattering of electromagnetic waves from rough surfaces*. Artech House, Inc., 1987.

[35] Dinesh Bharadia, Emily McMilin, and Sachin Katti. Full duplex radios. In *SIGCOMM*. ACM, 2013.

[36] Samuel S Blackman. Multiple-target tracking with radar applications. *Dedham, MA, Artech House, Inc.*, 1986.

[37] R.W. Bohannon. Comfortable and maximum walking speed of adults aged 20-79 years: reference values and determinants. *Age and ageing*, 1997.

[38] James Callahan, David Cox, Kenneth Hoffman, Donal O'shea, Harriet Pollatsek, and Lester Senechal. *Calculus in context: the five college calculus project, Chapter 7 (Periodicity)*. WH Freeman, 1995.

[39] Jinxiang Chai and Jessica K Hodgins. Performance animation from low-dimensional control signals. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 686–696. ACM, 2005.

[40] G.L. Charvat, L.C. Kempel, E.J. Rothwell, C. Coleman, and EJ Mokole. An ultra-wideband (UWB) switched-antenna-array radar imaging system. In *IEEE ARRAY*, 2010.

[41] G.L. Charvat, L.C. Kempel, E.J. Rothwell, C.M. Coleman, and E.L. Mokole. A through-dielectric radar imaging system. *IEEE Trans. Antennas and Propagation*, 2010.

[42] K. Chetty, G.E. Smith, and K. Woodbridge. Through-the-wall sensing of personnel using passive bistatic wifi radar at standoff distances. *IEEE Trans. Geoscience and Remote Sensing*, 2012.

[43] Y.M. Chi, Tzyy-Ping Jung, and G. Cauwenberghs. Dry-contact and noncontact biopotential electrodes: Methodological review. *Biomedical Engineering, IEEE Reviews in*, 2010.

[44] Krishna Chintalapudi, Anand Padmanabha Iyer, and Venkata N Padmanabhan. Indoor localization without the pain. In *ACM MobiCom*. ACM, 2010.

[45] J.I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti. Achieving single channel, full duplex wireless communication. In *ACM MobiCom*, 2010.

[46] CNET. Do wristband heart trackers actually work? A checkup. http://www.cnet.com/news/how-accurate-are-wristband-heart-rate-monitors/.

[47] G. Cohn, D. Morris, S. Patel, and D. Tan. Humantenna: using the body as an antenna for real-time whole-body interaction. In *ACM CHI*, 2012.

[48] Christopher R Cole, Eugene H Blackstone, Fredric J Pashkow, Claire E Snader, and Michael S Lauer. Heart-rate recovery immediately after exercise as a predictor of mortality. *New England Journal of Medicine*, 1999.

[49] Ken B Cooper, Robert J Dengler, Nuria Llombart, Tomas Bryllert, Goutam Chattopadhyay, Erich Schlecht, John Gill, Choonsup Lee, Anders Skalare, Imran Mehdi, et al. Penetrating 3-d imaging at 4-and 25-m range using a submillimeter-wave radar. *Microwave Theory and Techniques, IEEE Transactions on*, 2008.

[50] T.M. Cover and J.A. Thomas. *Elements of information theory*. Wiley-interscience, 2006.

[51] Jiangpeng Dai, Xiaole Bai, Zhimin Yang, Zhaohui Shen, and Dong Xuan. Perfalld: A pervasive fall detection system using mobile phones. In *IEEE PERCOM*, 2010.

[52] Philip de Chazal, Emer O'Hare, Niall Fox, and Conor Heneghan. Assessment of sleep/wake patterns using a non-contact biomotion sensor. In *IEEE EMBS*, 2008.

[53] RJ Dengler, KB Cooper, G Chattopadhyay, I Mehdi, E Schlecht, A Skalare, C Chen, and PH Siegel. 600 ghz imaging radar with 2 cm range resolution. In *Microwave Symposium, 2007. IEEE/MTT-S International*, 2007.

[54] Colin Dixon, Ratul Mahajan, Sharad Agarwal, A Brush, Bongshin Lee, Stefan Saroiu, and Victor Bahl. An operating system for the home. In *Usenix NSDI*, 2012.

[55] Traian Dogaru and Calvin Le. Validation of xpatch computer models for human body radar signature. Technical report, Army Research Laboratory, 2008.

[56] Traian Dogaru, Lam Nguyen, and Calvin Le. Computer models of the human body signature for sensing through the wall radar applications. Technical report, Army Research Laboratory, 2007.

[57] Amy D Droitcour, Olga Boric-Lubecke, and Gregory TA Kovacs. Signal-to-noise ratio in doppler radar system for heart and respiratory rate measurements. *Microwave Theory and Techniques, IEEE Trans.*, 2009.

[58] Droitcour, Amy D et. al. Range correlation and i/q performance benefits in single-chip silicon doppler radars for noncontact cardiopulmonary monitoring. *Microwave Theory and Techniques, IEEE Trans.*, 2004.

[59] Rich Fletcher and Jing Han. Low-cost differential front-end for doppler radar vital sign monitoring. In *IEEE MTT*, 2009.

[60] Forbes. How Does A Fighter Jet Lock Onto And Keep Track Of An Enemy Aircraft?, 2013. http://www.forbes.com/sites/quora/2013/07/17/how-does-a-fighter-jet-lock-onto-and-keep-track-of-an-enemy-aircraft/.

[61] Juergen Gall, Carsten Stoll, Edilson De Aguiar, Christian Theobalt, Bodo Rosenhahn, and H-P Seidel. Motion capture using joint skeleton tracking and surface estimation. In *IEEE CVPR*, 2009.

[62] Varun Ganapathi, Christian Plagemann, Daphne Koller, and Sebastian Thrun. Real time motion capture using a single time-of-flight camera. In *IEEE CVPR*, 2010.

[63] Garmin. Heart Rate Monitor. https://buy.garmin.com/en-US/US/shop-by-accessories/fitness-sensors/heart-rate-monitor/prod10996.html.

[64] GigaOm. Could a breath-monitoring headset improve your health? http://gigaom.com/2013/09/20/could-a-breath-monitoring-headset-improve-your-health/.

[65] Hulya Gokalp and Malcolm Clarke. Monitoring activities of daily living of the elderly and the potential for its use in telecare and telehealth: a review. *TELEMEDICINE and e-HEALTH*, 2013.

[66] S. Gollakota, F. Adib, D. Katabi, and S. Seshan. Clearing the RF smog: Making 802.11 robust to cross-technology interference. In *ACM SIGCOMM*, 2011.

[67] Shyamnath Gollakota and Dina Katabi. Zigzag decoding: combating hidden terminals in wireless networks. In *ACM SIGCOMM*, 2008.

[68] Clinical Guard. Wireless Wristband Pulse Oximeter CMS-50FW With Software For Sleep Study. http://www.clinicalguard.com/wireless-wristband-pulse-oximeter-cms50fw-with-software-for-sleep-study-p-455.html.

[69] Chris Harrison, Desney Tan, and Dan Morris. Skinput: appropriating the body as an input surface. In *ACM CHI*, 2010.

[70] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H.-P. Seidel. A statistical model of human pose and body shape. In *EUROGRAPHICS*, 2009.

[71] Nils Hasler, Bodo Rosenhahn, T Thormahlen, Michael Wand, Jürgen Gall, and H-P Seidel. Markerless motion capture with unsynchronized moving cameras. In *IEEE CVPR*, 2009.

[72] Haitham Hassanieh, Fadel Adib, Dina Katabi, and Piotr Indyk. Faster gps via the sparse fourier transform. In *ACM MobiCom*, 2012.

[73] Fierce Mobile Healthcare. Market for embedded health monitoring-gadgets to hit 170M devices by 2017. http://www.fiercemobilehealthcare.com/story/market-embedded-health-monitoring-gadgets-hit-170m-devices-2017/2012-08-03.

[74] Felix Heide, Lei Xiao, Wolfgang Heidrich, and Matthias Hullin. Diffuse mirrors: 3d reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. In *IEEE CVPR*, 2014.

[75] Lorna Herda, Pascal Fua, Ralf Plankers, Ronan Boulic, and Daniel Thalmann. Skeleton-based motion capture for robust reconstruction of human motion. In *Computer Animation*. IEEE, 2000.

[76] Hirshfield, L. et. Al. Brain measurement for usability testing and adaptive interfaces: an example of uncovering syntactic workload with functional near infrared spectroscopy. In *ACM CHI*, 2009.

[77] JA Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplement Series*, 1974.

[78] S. Hong, J. Mehlman, and S. Katti. Picasso: full duplex signal shaping to exploit fragmented spectrum. In *ACM SIGCOMM*, 2012.

[79] Chad Huffman, Jon Hayes, and Lars Ericson. Through-the-wall sensors (ttws) for law enforcement: Test & evaluation. In *ManTech Advanced Systems International*, 2014.

[80] Ivo Ihrke, Kiriakos N Kutulakos, Hendrik Lensch, Marcus Magnor, and Wolfgang Heidrich. Transparent and specular object reconstruction. In *Computer Graphics Forum*, 2010.

[81] Imai, K. et. Al. Vagally mediated heart rate recovery after exercise is accelerated in athletes but blunted in patients with chronic heart failure. *Journal of the American College of Cardiology*, 1994.

[82] InformationWeek. Health-Monitoring Devices Market Outpaces Telehealth. http://www.informationweek.com/mobile/health-monitoring-devices-market-outpaces-telehealth/d/d-id/1104636.

[83] M. Jain, J.I. Choi, T. Kim, D. Bharadia, S. Seth, K. Srinivasan, P. Levis, S. Katti, and P. Sinha. Practical, real-time, full duplex wireless. In *ACM MobiCom*, 2011.

[84] Yong Jia, Lingjiang Kong, Xiaobo Yang, and Kunde Wang. Through-wall-radar localization for stationary human based on life-sign detection. In *IEEE RADAR*, 2013.

[85] Kiran Joshi, Dinesh Bharadia, Manikanta Kotaru, and Sachin Katti. Wideo: Fine-grained device-free motion tracing using rf backscatter. In *Usenix NSDI*, 2015.

[86] Kiran Joshi, Steven Hong, and Sachin Katti. Pinpoint: Localizing interfering radios. In *Usenix NSDI*, 2013.

[87] H. Junker, P. Lukowicz, and G. Troster. Continuous recognition of arm activities with body-worn inertial sensors. In *IEEE ISWC*, 2004.

[88] Kaltiokallio, O. et. Al. Non-invasive respiration rate monitoring using a single cots tx-rx pair. In *ACM/IEEE IPSN*, 2014.

[89] Y. Kim and H. Ling. Human activity classification based on micro-doppler signatures using a support vector machine. *IEEE Trans. Geoscience and Remote Sensing*, 2009.

[90] Ahmed Kirmani, Tyler Hutchison, James Davis, and Ramesh Raskar. Looking around the corner using transient imaging. In *IEEE CVPR*, 2009.

[91] Lanata, A et. Al. Comparative evaluation of susceptibility to motion artifact in different wearable systems for monitoring respiratory rate. 2010.

[92] C. Le, T. Dogaru, N. Lam, and M. Ressler. Ultrawideband (uwb) radar imaging of building interior: Measurements and prediction. In *IEEE Transactions on Geoscience and Remote Sensing*, 2009.

[93] Lei Li, James McCann, Nancy Pollard, and Christos Faloutsos. Bolero: a principled technique for including bone length constraints in motion capture occlusion filling. In *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, 2010.

[94] X Rong Li and Yaakov Bar-Shalom. Design of an interacting multiple model algorithm for air traffic control tracking. *IEEE Transactions on Control Systems Technology*, 1993.

[95] K.C.J. Lin, S. Gollakota, and D. Katabi. Random access heterogeneous MIMO networks. In *ACM SIGCOMM*, 2010.

[96] Lin, J. et. Al. Microwave apexcardiography. *Microwave Theory and Techniques, IEEE Trans.*, 1979.

[97] Ding Liu, Xida Chen, and Yee-Hong Yang. Frequency-based 3d reconstruction of transparent and specular objects. In *IEEE CVPR*, 2014.

[98] Guodong Liu and Leonard McMillan. Estimation of missing markers in human motion capture. *The Visual Computer*, 22(9-11):721–728, 2006.

[99] Xuefeng Liu, Jiannong Cao, Shaojie Tang, Jiaqi Wen, and Peng Guo. Contactless respiration monitoring via wifi signals. *IEEE Transactions on Mobile Computing*, 2015.

[100] Leon B Lucy. An iterative technique for the rectification of observed distributions. *The astronomical journal*, 79:745, 1974.

[101] B. Lyonnet, C. Ioana, and M.G. Amin. Human gait classification using microdoppler time-frequency signal representations. In *IEEE Radar Conference*, 2010.

[102] Bassem R Mahafza. *Radar systems analysis and design using MATLAB*. Chapman & Hall, 2013.

[103] Sascha Mahlke and Manfred Thüring. Studying antecedents of emotional experiences in interactive contexts. In *ACM CHI*, 2007.

[104] Mayo Clinic. What's a normal resting heart rate? http://www.mayoclinic.org/healthy-living/fitness/expert-answers/heart-rate/faq-20057979.

[105] B. Michoud, E. Guillou, and S. Bouakaz. Real-time and markerless 3D human motion capture using multiple views. *Human Motion–Understanding, Modeling, Capture and Animation*, 2007.

[106] Min, J. et. Al. Toss nturn: Smartphone as sleep and sleep quality detector. In *ACM CHI*, 2014.

[107] Greg Mori, Xiaofeng Ren, Alexei A Efros, and Jitendra Malik. Recovering human body configurations: Combining segmentation and recognition. In *IEEE CVPR*, 2004.

[108] S. Nannuru, Yunpeng Li, Yan Zeng, M. Coates, and Bo Yang. Radio-frequency tomography for passive indoor multitarget tracking. *Mobile Computing, IEEE Transactions on*, 2013.

[109] Norbert Noury, Anthony Fleury, Pierre Rumeau, AK Bourke, GO Laighin, V Rialle, and JE Lundy. Fall detection-principles and methods. In *IEEE EBMS*, 2007.

[110] A.V. Oppenheim, R.W. Schafer, J.R. Buck, et al. *Discrete-time signal processing*. Prentice hall Englewood Cliffs, NJ:, 1989.

[111] Sophocles J Orfanidis. *Electromagnetic waves and antennas*. Rutgers University New Brunswick, NJ, 2002.

[112] Sang Il Park and Jessica K Hodgins. Capturing and animating skin deformation in human motion. In *ACM Transactions on Graphics (TOG)*, volume 25, pages 881–889. ACM, 2006.

[113] Patwari, N. et. Al. Breathfinding: A wireless network that monitors and locates breathing in a home. *Selected Topics in Signal Processing, IEEE Journal of*, 2014.

[114] Petterson, M. et. Al. The effect of motion on pulse oximetry and its clinical significance. *Anesthesia & Analgesia*, 2007.

[115] Eduardo Pinheiro, Octavian Postolache, and Pedro Girão. Theory and developments in an unobtrusive cardiovascular system representation: ballistocardiography. *The open biomedical engineering journal*, 2010.

[116] Ronald Poppe. Vision-based human motion analysis: An overview. *Computer vision and image understanding*, 108(1):4–18, 2007.

[117] Qifan Pu, Siyu Jiang, Shyamnath Gollakota, and Shwetak Patel. Whole-home gesture recognition using wireless signals. In *ACM MobiCom*, 2013.

[118] H.S. Rahul, S. Kumar, and D. Katabi. JMB: scaling wireless capacity with user demands. In *ACM SIGCOMM*, 2012.

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

[120] Swati Rallapalli, Aishwarya Ganesan, Krishna Chintalapudi, Venkat N Padmanabhan, and Lili Qiu. Enabling physical analytics in retail stores using smart glasses. In *ACM MobiCom*, 2014.

[121] T.S. Ralston, G.L. Charvat, and J.E. Peabody. Real-time through-wall imaging using an ultrawideband multiple-input multiple-output (MIMO) phased array radar system. In *IEEE ARRAY*, 2010.

[122] S.S. Ram, C. Christianson, Y. Kim, and H. Ling. Simulation and analysis of human micro-dopplers in through-wall environments. *IEEE Trans. Geoscience and Remote Sensing*, 2010.

[123] S.S. Ram, Y. Li, A. Lin, and H. Ling. Doppler-based detection and tracking of humans in indoor environments. *Journal of the Franklin Institute*, 2008.

[124] S.S. Ram and H. Ling. Through-wall tracking of human movers using joint doppler and array processing. *IEEE Geoscience and Remote Sensing Letters*, 2008.

[125] Ramesh Raskar, Hideaki Nii, Bert Dedecker, Yuki Hashimoto, Jay Summet, Dylan Moore, Yong Zhao, Jonathan Westhues, Paul Dietz, John Barnwell, et al. Prakash: lighting aware motion capture using photosensing markers and multiplexed illuminators. In *ACM Transactions on Graphics (TOG)*, volume 26, page 36. ACM, 2007.

[126] Mark A Richards. *Fundamentals of radar signal processing*. Tata McGraw-Hill Education, 2005.

[127] Daniel Roetenberg, Henk Luinge, and Per Slycke. Xsens mvn: full 6dof human motion tracking using miniature inertial sensors. *Xsens Motion Technologies BV, Tech. Rep*, 2009.

[128] Jodi L Schwartz and Bernard D Steinberg. Ultrasparse, ultrawideband arrays. *Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on*, 1998.

[129] M. Seifeldin, A. Saeed, A.E. Kosba, A. El-Keyi, and M. Youssef. Nuzzer: A large-scale device-free passive localization system for wireless environments. *IEEE Transactions on Mobile Computing*, 2013.

[130] Souvik Sen, Bozidar Radunovic, Romit Roy Choudhury, and Tom Minka. Spot localization using phy layer information. In *ACM MobiSys*, 2012.

[131] Tie-Jun Shan, M. Wax, and T. Kailath. On spatial smoothing for direction-of-arrival estimation of coherent signals. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 1985.

[132] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124, 2013.

[133] Stephan Sigg, Ulf Blanke, and Gerhard Tröster. The telepathic phone: Frictionless activity recognition from wifi-rssi. In *IEEE PerCom*, 2014.

[134] Merrill Ivan Skolnik. Radar handbook. 1970.

[135] Graeme E Smith and Bijan G Mobasseri. Robust through-the-wall radar image classification using a target-model alignment procedure. *Image Processing, IEEE Transactions on*, 2012.

[136] F. Soldovieri and R. Solimene. Through-wall imaging via a linear inverse scattering algorithm. *IEEE Geoscience and Remote Sensing Letters*, 2007.

[137] R. Solimene, F. Soldovieri, G. Prisco, and R. Pierri. Three-dimensional through-wall imaging under ambiguous wall parameters. *IEEE Trans. Geoscience and Remote Sensing*, 2009.

[138] Robert Steele, Amanda Lo, Chris Secombe, and Yuk Kuen Wong. Elderly persons perception and acceptance of using wireless sensor networks to assist healthcare. *International journal of medical informatics*, 2009.

[139] Petre Stoica and Randolph L. Moses. *Spectral Analysis of Signals*. Prentice Hall, 2005.

[140] William C. Stone. Nist construction automation program report no. 3: Electromagnetic signal attenuation in construction materials. In *NIST Construction Automation Workshop 1995*.

[141] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[142] Kun Tan, He Liu, Ji Fang, Wei Wang, Jiansong Zhang, Mi Chen, and Geoffrey Voelker. SAM: Enabling Practical Spatial Multiple Access in Wireless LAN. In *ACM MobiCom*, 2009.

[143] Tan, C. et. Al. Combining think-aloud and physiological data to understand video game experiences. In *ACM CHI*, 2014.

[144] A Richard Thompson, James M Moran, and George W Swenson Jr. *Interferometry and synthesis in radio astronomy*. John Wiley & Sons, 2008.

[145] DJ Titman. Applications of thermography in non-destructive testing of structures. *NDT & E International*, 2001.

[146] Jenho Tsao and Bernard D Steinberg. Reduction of sidelobe and speckle artifacts in microwave imaging: the clean technique. *Antennas and Propagation, IEEE Transactions on*, 1988.

[147] D. Tse and P. Vishwanath. *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.

[148] U.S. National Library of Medicine MedLine Plus. Heart-respiratory monitor - infantsShare. http://www.nlm.nih.gov/medlineplus/ency/article/007236.htm.

[149] U.S. National Library of Medicine MedLine Plus. Rapid Shallow Breathing. http://www.nlm.nih.gov/medlineplus/ency/article/007198.htm.

[150] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 3:745, 2012.

[151] S. Venkatesh, C.R. Anderson, N.V. Rivera, and R.M. Buehrer. Implementation and analysis of respiration-rate estimation using impulse-based uwb. In *IEEE MILCOMM*, 2005.

[152] Vivonoetics. Respiration: Single or Dual-Band. http://vivonoetics.com/products/vivosense/analyzing/respiration-single-or-dual-band/.

[153] Daniel Vlasic, Rolf Adelsberger, Giovanni Vannucci, John Barnwell, Markus Gross, Wojciech Matusik, and Jovan Popović. Practical motion capture in everyday surroundings. In *ACM Transactions on Graphics (TOG)*, volume 26, page 35. ACM, 2007.

[154] Daniel Vlasic, Ilya Baran, Wojciech Matusik, and Jovan Popović. Articulated mesh animation from multi-view silhouettes. In *ACM Transactions on Graphics (TOG)*, volume 27, page 97. ACM, 2008.

[155] H Kenneth Walker, W Dallas Hall, and J Willis Hurst. *Respiratory Rate and Pattern*. Butterworths, 1990.

[156] H. Wang, R.M. Narayanan, and Z.O. Zhou. Through-wall imaging of moving targets using uwb random noise radar. *IEEE Antennas and Wireless Propagation Letters*, 2009.

[157] 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 *ACM UbiComp*, 2016.

[158] Jack M Wang, David J Fleet, and Aaron Hertzmann. Gaussian process dynamical models for human motion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):283–298, 2008.

[159] Jue Wang, Fadel Adib, Ross Knepper, Dina Katabi, and Daniela Rus. RF-Compass: Robot Object Manipulation Using RFIDs. In *ACM MobiCom*, 2013.

[160] Jue Wang and Dina Katabi. Dude, where's my card? RFID positioning that works with multipath and non-line of sight. In *ACM SIGCOMM*, 2013.

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

[162] Wei Wang, Alex X Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. Understanding and modeling of wifi signal based human activity recognition. In *ACM MobiCom*, 2015.

[163] 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 *ACM MobiCom*, 2014.

[164] Teng Wei and Xinyu Zhang. mtrack: High-precision passive tracking using millimeter wave radios. In *ACM MobiCom*, 2015.

[165] Joey Wilson and Neal Patwari. Radio tomographic imaging with wireless networks. In *IEEE Transactions on Mobile Computing*, 2010.

[166] Joey Wilson and Neal Patwari. See-through walls: Motion tracking using variance-based radio tomography networks. In *IEEE Transactions on Mobile Computing*, 2011.

[167] Peter W Wolniansky, Gerard J Foschini, GD Golden, and R Valenzuela. V-blast: An architecture for realizing very high data rates over the rich-scattering wireless channel. In *IEEE ISSSE*, 1998.

[168] Ruth M Woodward, Bryan E Cole, Vincent P Wallace, Richard J Pye, Donald D Arnone, Edmund H Linfield, and Michael Pepper. Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue. *Physics in Medicine and Biology*, 2002.

[169] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM SIGGRAPH*, 2012.

[170] Jie Xiong and Kyle Jamieson. ArrayTrack: a fine-grained indoor location system. In *Usenix NSDI*, 2013.

[171] Yanyun Xu, Shiyou Wu, Chao Chen, Jie Chen, and Guangyou Fang. A novel method for automatic detection of trapped victims by ultrawideband radar. *Geoscience and Remote Sensing, IEEE Transactions on*, 2012.

[172] Lei Yang, Qiongzheng Lin, Xiangyang Li, Tianci Liu, and Yunhao Liu. See through walls with cots rfid system! In *ACM MobiCom*, 2015.

[173] Y. Yang and A. Fathy. Design and implementation of a low-cost real-time ultra-wide band see-through-wall imaging radar system. In *IEEE/MTT-S International Microwave Symposium*, 2007.

[174] Y. Yang and A.E. Fathy. See-through-wall imaging using ultra wideband short-pulse radar system. In *IEEE Antennas and Propagation Society International Symposium*, 2005.

[175] Yang, R. et. Al. Making sustainability sustainable: Challenges in the design of eco-interaction technologies. In *ACM CHI*, 2014.

[176] Yarovoy, A. et. Al. Uwb radar for human being detection. *Aerospace and Electronic Systems Magazine, IEEE*, 2006.

[177] Mao Ye, Huamin Wang, Nianchen Deng, Xubo Yang, and Ruigang Yang. Real-time human pose and shape estimation for virtual try-on using a single commodity depth camera. *IEEE transactions on visualization and computer graphics*, 20(4):550–559, 2014.

[178] Moustafa Youssef, Matthew Mah, and Ashok Agrawala. Challenges: device-free passive localization for wireless environments. In *ACM MobiCom*, 2007.

[179] Alberto Zaffaroni, Philip de Chazal, Conor Heneghan, Patricia Boyle, Patricia Ronayne Mppm, and Walter T McNicholas. Sleepminder: an innovative contact-free device for the estimation of the apnoea-hypopnoea index. In *IEEE EMBS*, 2009.

[180] Yang Zhao, Neal Patwari, Jeff M Phillips, and Suresh Venkatasubramanian. Radio tomographic imaging and tracking of stationary and moving people via kernel distance. In *ACM ISPN*, 2013.

[181] X Zhuge, TG Savelyev, AG Yarovoy, LP Ligthart, J Matuzas, and B Levitas. Human body imaging by microwave uwb radar. In *IEEE EuRAD*, 2008.