Whole-Home Gesture Recognition Using Wireless Signals

Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel
University of Washington
{qp, sidhant, gshyam, shwetak}@cs.washington.edu

Abstract – This paper presents WiSee, a novel gesture recognition system that leverages wireless signals (e.g., Wi-Fi) to enable whole-home sensing and recognition of human gestures. Since wireless signals do not require line-of-sight and can traverse through walls, WiSee can enable whole-home gesture recognition using few wireless sources. Further, it achieves this goal without requiring instrumentation of the human body with sensing devices. We implement a proof-of-concept prototype of WiSee using USRP-N210s and evaluate it in both an office environment and a two-bedroom apartment. Our results show that WiSee can identify and classify a set of nine gestures with an average accuracy of 94%.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Wireless Communication

Keywords

Gesture Recognition, Wireless Sensing

1. Introduction

As computing moves increasingly away from the desktop, there is a growing need for new ways to interact with computer interfaces. The Xbox Kinect is an example of a commercially available input sensor that enables gesture-based interaction using depth sensing and computer vision. The commercial success of these kinds of devices has spurred interest in developing new user interfaces that remove the need for a traditional keyboard and mouse. Gestures enable a whole new set of interaction techniques for always-available computing embedded in the environment. For example, using a swipe hand motion in-air, a user could control the music volume while showering, or change the song playing on a music system installed in the living room while cooking, or turn up the thermostat while in bed. Such a capability can enable applications in diverse domains including home-automation, elderly health care, and gaming. However, the burden of installation and cost make most vision-based sensing devices hard to deploy at scale, for example, throughout an entire home or building. Given these limitations, researchers have explored ways to move some of the sensing onto the body and reduce the need for environmental sensors [8, 14, 13]. However, even on-body approaches are limited to what people are willing to constantly carry or wear, and may be infeasible in many scenarios (e.g., in a shower).

This paper presents WiSee, the first whole-home gesture recognition system that requires neither user instrumentation nor an infrastructure of cameras. WiSee achieves this by leveraging wireless signals (e.g. Wi-Fi) in an environment. Since these signals do not require line-of-sight and can traverse through walls, very few signal sources need to be present in the space (e.g., a Wi-Fi AP and a few mobile devices in the living room). WiSee works by looking at the minute Doppler shifts and multi-path distortions that occur with these wireless signals from human motion in the environment.

To this end, we address the following two challenges:

(a) How do we capture information about gestures from wireless signals? WiSee leverages the property of Doppler shift [12, 3], which is the frequency change of a wave as its source moves relative to the observer. The canonical example is the change in the pitch of a train's whistle as it approaches and departs from a listener. In the context of wireless signals, if we consider the multi-path reflections from the human body as waves from a source, then a human performing a gesture, results in a pattern of Doppler shifts at the wireless receiver. Thus, a user moving her hand away from the receiver results in a negative Doppler shift, while moving the hand towards the receiver results in a positive Doppler shift.

The challenge, however, is that human hand gestures result in very small Doppler shifts that can be hard to detect from typical wireless transmissions (e.g., Wi-Fi). Specifically, since wireless signals are electromagnetic waves that propagate at the speed of light (c m/sec), a human moving at a speed of v m/sec, results in a maximum Doppler shift of $\frac{2f}{c}v$, where f is the frequency of the wireless transmission [3]. Thus, a 0.5 m/sec gesture results in a 17 Hz Doppler shift on a 5 GHz Wi-Fi transmission. Typical wireless transmissions have orders of magnitude higher bandwidth (20 MHz for Wi-Fi). Thus, for gesture recognition, we need to detect Doppler shifts of a few Hertz from the 20 MHz Wi-Fi signal.

At a high level, WiSee addresses this problem by transforming the received signal into a narrowband pulse with a bandwidth of a few Hertz. The WiSee receiver (which can be implemented on a Wi-Fi AP) then tracks the frequency of this narrowband pulse to detect the small Doppler shifts resulting from human gestures. In §3, we describe our algo-

MobiCom'13, September 30–October 4, Miami, FL, USA.

ACM 978-1-4503-1999-7/13/09 ...$15.00.
10.1145/2500423.2500436

rithm in more detail and show how to make it applicable to existing 802.11 frames.

(b) How can we deal with other humans in the environment? A typical home may have multiple people who can affect the wireless signals at the same time. WiSee uses the MIMO capability that is inherent to 802.11n, to focus on gestures from a particular user. MIMO provides throughput gains by enabling multiple transmitters to concurrently send packets to a MIMO receiver. If we consider the wireless reflections from each human as signals from a wireless transmitter, then they can be separated using a MIMO receiver.

Traditional MIMO decoding, however, relies on estimating the channel between the transmitter and receiver antennas. These channels are typically estimated by sending a distinct known preamble from each transmitter. Such a known signal structure is not available in our system since the human body reflects the same 802.11 transmitter's signals.

Our solution to this problem is inspired by the trigger approach taken by many multi-user games that use Xbox Kinect, in which a user gains control of the interface by performing a specific gesture pattern. In WiSee the target human performs a repetitive gesture, which we use as that person's preamble. A WiSee receiver leverages this preamble to estimate the MIMO channel that maximizes the energy of the reflections from the user. Once the receiver locks on to this channel, the user performs normal (non-repetitive) gestures that the receiver classifies using the Doppler shifts. In §3.3, we explore this idea further and show how to extract the preamble without requiring the human to perform gestures at a pre-determined speed.

The WiSee proof-of-concept is implemented in GNURadio using the USRP-N210 hardware. We classify the gestures from the Doppler shifts using a simple pattern-matching algorithm described in §3.2. We evaluated WiSee with a total of five users in both an office environment and a two-bedroom apartment whose layout is shown in Fig. 7. We performed gestures in a number of scenarios including line-of-sight, non-line-of-sight, and through-the-wall scenarios where the person is in a different room from the wireless transmitter and the receiver. The users perform a total of 900 gestures across the locations.

Our findings are as follows:

- WiSee can classify the nine whole-body gestures shown in Fig. 1, with an average accuracy of 94%. This is promising, given that the accuracy for random guesses is 11.1%.
- Using a 4-antenna receiver and a single-antenna transmitter placed in the living room, WiSee can achieve the above classification accuracy in 60% of the home locations. Adding an additional single-antenna transmitter to the living room achieves the above accuracy in locations across all the rooms. Thus, with a WiSee-enabled Wi-Fi AP acting as a receiver and a couple of mobile devices acting as transmitters, WiSee can enable whole-home gesture recognition.
- Over a 24-hour period, WiSee's average false positive rate—events that detect a gesture in the absence of the target human—is 2.63 events per hour when using a preamble with two gesture repetitions. This goes down to 0.07 events per hour, when the number of repetitions is increased to four.
- Using a 5-antenna receiver and a single-antenna transmitter, WiSee can successfully perform gesture classification, in the presence of three other users performing random gestures. However, the classification accuracy reduces as we further increase the number of interfering users. This is a limitation of WiSee: Given a fixed number of transmitters and receiver antennas, the accuracy reduces with the number of users. However, since typical home scenarios do not have a large number of users in a single room, WiSee can enable a large set of interaction applications for always-available computing in home environments.

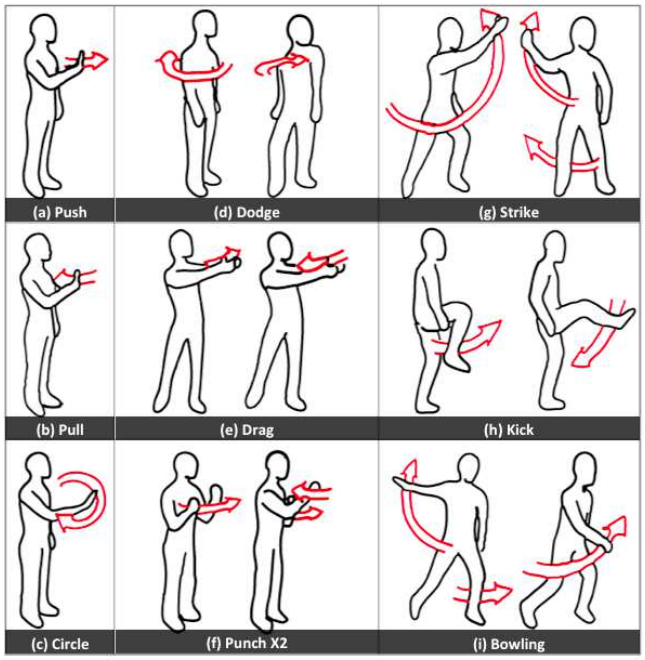


Figure 1—Gesture sketches: WiSee can detect and classify these nine gestures in line-of-sight, non-line-of-sight, and through-the-wall scenarios with an average accuracy of 94%.

Contributions: We make the following contributions:

- We introduce the first wireless system that enables gesture recognition in line-of-sight, non-line-of-sight, and through-the-wall scenarios.
- We present algorithms to extract gesture information from communication-based wireless signals. Specifically, we show how to extract minute Doppler shifts from wideband OFDM transmissions that are typical to most modern communication systems including Wi-Fi.
- Finally, using a proof-of-concept prototype, we demonstrate that our system can detect a set of nine whole-body gestures in typical environments.

This paper takes the first step towards leveraging existing wireless networks to enable novel human-computer interaction mechanisms such as whole-home gesture recognition. We hope that this line of work would open up a number of research opportunities at the intersection of wireless networking and HCI, and bridge the two communities.

2. Related Work

Our work is related to prior art in both wireless systems and in-air gesture recognition systems.

(a) Wireless Systems: One can classify the related work in this domain into three main categories: wireless localization, wireless tomography, and through-the-wall radar sys-

tems. Prior work on localization use a variety of techniques to localize wireless devices, including, RSSI [6], fine-grained OFDM channel information [21], and multiple antennas [26]. There has also been recent interest in device-free localization of humans [28, 20] using the RSSI information from wireless devices. WiSee builds on this foundational work but is complementary to it in that it focuses on achieving human gesture recognition using wireless signals.

WiSee is also related to work on wireless tomography that aims to localize humans by deploying a network of sensors throughout the environment [29, 25, 24]. These systems typically use the RSSI value observed at each sensor to localize humans. WiSee builds on this work but significantly differs from it in that it extracts Doppler shifts from wireless signals to perform human gesture recognition. Further, we demonstrate that one can perform whole-home gesture recognition without requiring wireless devices in every room.

Finally, WiSee is related to work on through-the-wall radar systems [4, 18, 3] that can identify objects such as metal pins behind a wall. These systems use expensive ultra-wideband transceivers that use bandwidths on the order of 1 GHz [4]. In contrast, WiSee focuses on gesture recognition and shows how to extract gesture information from wireless transmissions. The closest to our work in this domain is recent work [5] that demonstrates the feasibility of using Wi-Fi signals to detect running in through-the-wall scenarios. However, to the best of our knowledge, none of the prior radar systems have been designed to work for gesture recognition. In contrast, WiSee introduces mechanisms that enable it to go beyond coarse human motion such as running and walking, and delivers the first wireless system that can identify finer-grained human motion such as gestures in LOS, NLOS, and through-the-wall scenarios.

(b) In-Air Gesture Recognition Systems: The commercial success of products like the Xbox Kinect has popularized the use of gestures to control computer systems [22]. Increasingly, in-air gesture recognition is being incorporated into consumer electronics and mobile devices, including laptops [2], smartphones [9, 12], and GPS devices [19]. The related work in this domain use four main techniques: computer vision, ultra-sonic, electric field, and inertial sensing.

Vision-based systems extract gesture information using advances in the hybrid camera technology like pixel-mixed devices (PMDs) [22]. Likewise, ultra-sonic systems leverage Doppler shifts on sound waves to perform gesture recognition [12]. Both these systems, however, require a line-of-sight channel between the sensing device and the human. In contrast, WiSee leverages wireless signals that can operate in non-line-of-sight scenarios and can go through wooden walls and obstacles like curtains and furniture.

Electric Field sensing systems like Magic Carpet [17] instrument the floor with multiple sensors to perform human localization and gesture recognition. However, this imposes heavy instrumentation of the environment and is not practical. Inertial sensing and other on-body sensing methods on the other hand, require the users to wear multiple sensors or carry a device such as a wristband [8, 14, 13]. While attractive, in many instances, such an approach can be inconvenient (for instance, while showering). In contrast, WiSee enables whole-home gesture recognition without the need to instrument the human body.

Finally, prior work has leveraged Doppler shifts to perform gesture recognition in line-of-sight scenarios [12, 15]. In this paper, we present algorithms that allow us to extract Doppler shifts in line-of-sight, non-line-of-sight and through-the-wall scenarios; thus enabling gesture recognition without the need for sensing devices in every room. Further, we introduce algorithms to extract minute Doppler shifts from wide-band OFDM signals that are typically used in communication technologies including Wi-Fi.

3. WiSee

WiSee is a wireless system that enables whole-home gesture recognition. Since wireless signals can typically propagate through walls, and do not require a line of sight channel, WiSee can enable gesture recognition independent of the user's location. To achieve this, we need to answer three main questions: First, how does WiSee extract Doppler shifts from conventional wireless signals like Wi-Fi? Second, how does it map the Doppler shifts to the gestures performed by the user? Third, how does it enable gesture recognition in the presence of other humans in the environment? In the rest of this section, we address each of these questions.

3.1 Extracting Doppler shifts from Wireless Signals

Doppler shift is the change in the observed frequency as the transmitter and the receiver move relative to each other. In our context, an object reflecting the signals from the transmitter can be thought of as a virtual transmitter that generates the reflected signals. Now, as the object (virtual transmitter) moves towards the receiver, the crests and troughs of the reflected signals arrive at the receiver at a faster rate. Similarly, as an object moves away from the receiver, the crests and troughs arrive at a slower rate. More generally, a point object moving at a speed of v at an angle of θ from the receiver, results in a Doppler shift [3] given by:

$$\Delta f \propto \frac{2vcos(\theta)}{c} f \tag{1}$$

where c is the speed of light in the medium and f is the transmitter's center frequency. We note the following:

- The observed Doppler shift depends on the direction of motion with respect to the receiver. For instance, a point object moving orthogonal to the direction of the receiver results in no Doppler shift, while a point object moving towards the receiver maximizes the Doppler shift. Since human gestures typically involve multiple point objects moving along different directions, the set of Doppler shifts seen by a receiver can, in principle, be used to classify different gestures.
- Higher transmission frequencies result in a higher Doppler shift for the same motion. Thus, a Wi-Fi transmission at 5 GHz results in twice the Doppler shift as a Wi-Fi transmission at 2.5 GHz. We note, however, that much higher frequencies (e.g., at 60 GHz) may not be suitable for whole-home gesture recognition since they are more directional and typically not suitable for NLOS scenarios.
- Faster speeds result in larger Doppler shifts, while slower speeds result in smaller Doppler shifts. Thus, it is easier to detect a human running towards the receiver than to detect a human walking slowly. Further, gestures involving full-body motion (e.g. walking towards or away from the receiver) are easier to capture than gestures involving only parts of the body (e.g., hand motion towards or away from the receiver). This is because a full-body motion involves

many more point object moving at the same time. Thus, it creates Doppler signals with much larger energy than when the human uses only parts of her body.

Challenge: Human motion results in a very small Doppler shift that can be hard to detect from a typical wireless transmission (e.g., Wi-Fi, WiMax, LTE, etc.). For instance, consider a user moving her hand towards the receiver at 0.5 m/sec. From Eq. 1, this results in a Doppler shift of about 17 Hertz for a Wi-Fi signal transmitted at 5 GHz ($\theta = 0$). Since the bandwidth of Wi-Fi's transmissions is at least 20 MHz, the resulting Doppler shift is orders of magnitude smaller than Wi-Fi's bandwidth. Identifying such small Doppler shifts from these transmissions can be challenging.

Our Solution: WiSee presents a receiver design that can identify Doppler shifts at the resolution of a few Hertz from Wi-Fi signals. The basic idea underlying WiSee is to transform the received Wi-Fi signal into a narrowband pulse with the bandwidth of a few Hertz. The receiver then tracks the frequency of this narrowband pulse to detect the small Doppler shifts.

WiSee is designed for OFDM-based systems – OFDM is the modulation of choice for most modern wireless systems including 802.11 a/g/n, WiMAX, and LTE. OFDM divides the used RF bandwidth into multiple sub-channels and modulates data in each sub-channel. For instance, Wi-Fi typically divides the 20 MHz channel into 64 sub-channels each with a bandwidth of 312.5 KHz. The time-domain OFDM symbol is generated at the transmitter by taking an FFT over a sequence of modulated bits transmitted in each OFDM sub-channel. Specifically, the transmitter takes blocks of N modulated bits ($N = 64$ in 802.11), and applies an N-point Inverse Fast Fourier Transform (IFFT),

$$\mathbf{x}_k = \sum_{n=1}^{N} \mathbf{X}_n e^{i2\pi kn/N}$$

where $\mathbf{X}_n$ is the modulated bit sent in the n^{th} OFDM sub-channel. Each block of $\mathbf{x}_1, \cdots \mathbf{x}_N$ forms a time-domain OFDM symbol that the receiver decodes by performing the FFT operation, i.e.,

$$\mathbf{X}_n = \sum_{k=1}^{N} \mathbf{x}_k e^{-i2\pi kn/N} \quad (2)$$

To demonstrate how WiSee's receiver works on these OFDM signals, we first consider the scenario where the transmitter repeatedly sends the same OFDM symbol. We then generalize our approach to arbitrary OFDM symbols, making the scheme applicable to existing 802.11 frames.

Case 1: Transmitter sends the same OFDM symbol. In this case, instead of performing an FFT over each OFDM symbol, WiSee's receiver performs a large FFT over M consecutive OFDM symbols. As a consequence of this operation, the bandwidth of each OFDM sub-channel is reduced by a factor of M. To see this, say the receiver performs a $2N$-point FFT over two consecutive, identical OFDM symbols. The output of the FFT can be written as,[1]

$$\mathbf{X}_n = \sum_{k=1}^{N} \mathbf{x}_k e^{-i2\pi kn/2N} + \sum_{k=N+1}^{2N} \mathbf{x}_k e^{-i2\pi kn/2N}$$

[1]For simplicity, we ignore the noise term in the above equation. However, as with standard OFDM decoding, these linear equations hold even in the presence of noise.

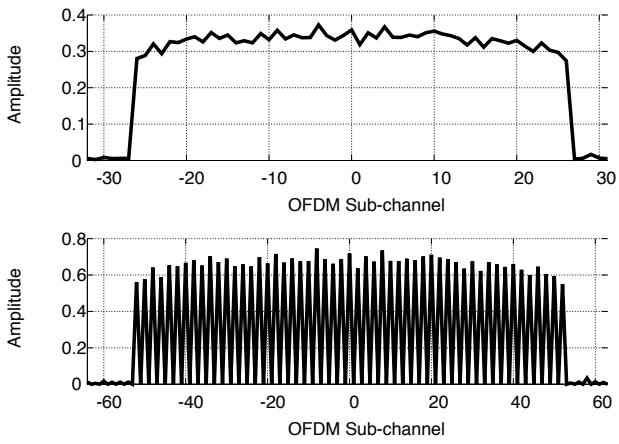


Figure 2—Creating a narrowband signal using WiSee: The first subplot shows the output of an FFT taken over an OFDM symbol. The second subplot shows the output of an FFT taken over two identical OFDM symbols. They show that taking a larger FFT over identical OFDM symbols, reduces each subchannel's bandwidth.

Since the first N transmitted samples are identical to the last N samples, i.e., $\mathbf{x}_k = \mathbf{x}_{k+N}$, for $k = 1$ to N, we can re-write the above equation as,

$$\mathbf{X}_n = \sum_{k=1}^{N} \mathbf{x}_k e^{-i2\pi kn/2N} + \sum_{k=1}^{N} \mathbf{x}_k e^{-i2\pi (k+N)n/2N}$$

After simplification, we get:

$$\mathbf{X}_n = \sum_{k=1}^{N} \mathbf{x}_k e^{-i2\pi kn/2N}(1 + e^{-i\pi n})$$

Now, when n is an even number, $(1+e^{-i\pi n}) = 2$, but when n is an odd number, $(1+e^{-i\pi n}) = 0$. Thus, the above equation can be re-written as,

$$\mathbf{X}_{2l} = 2\sum_{k=1}^{N} \mathbf{x}_k e^{-i2\pi kl/N}, \quad \mathbf{X}_{2l+1} = 0$$

Thus, as shown in Fig. 2, the odd sub-channels are zero and the even sub-channels capture the output (Eq. 2) of an N-point FFT on a single OFDM symbol. Intuitively, this happens because in each sub-channel, the same modulated information is transmitted in both the OFDM symbols. Thus, the bandwidth used by each sub-channel effectively halves. More generally, when the receiver performs an MN-point FFT over an OFDM symbol that is repeated M times, the bandwidth of each sub-channel is reduced by a factor of M. Thus, WiSee can create multiple narrowband signals centered at each sub-channel by repeating an OFDM symbol and performing a large FFT operation.

Now, by performing a large FFT over an one-second duration, the WiSee receiver can create a one-Hertz wide narrowband signal. The WiSee receiver tracks this narrowband signal to capture the Doppler shift (see §3.2). Note that one can average the Doppler shifts observed across all the OFDM sub-channels to significantly reduce the noise in the Doppler measurements.

Case 2: Transmitter sends arbitrary OFDM symbols. Our description so far assumes that the transmitter repeatedly sends the same OFDM symbol. Typical 802.11 transmitters however send different data across symbols. We

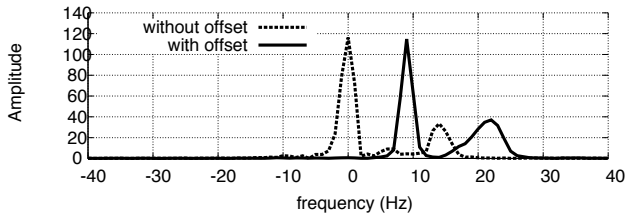


Figure 3—Dealing with Residual Frequency offset: The figure plots the frequency profile of the narrowband signal centered at one of the subchannels. It shows the profile in the presence of a gesture, both with and without residual offsets. WiSee accounts for the residual offset by tracking the peak with the maximum energy (the other peaks correspond to Doppler shifts).

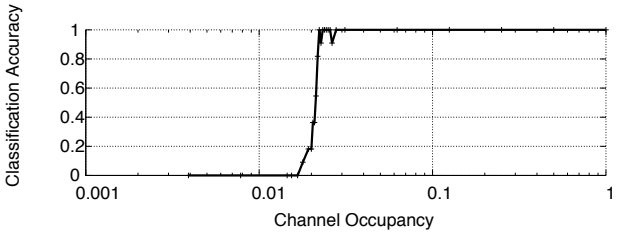


Figure 4—Classification accuracy versus transmitter occupancy: The accuracy of distinguishing between two gestures is high even when the transmitter transmits only 3% of the time.

show how to extract Doppler shifts from such transmissions. At a high level, WiSee achieves this by designing a *data-equalizing re-encoder* at the receiver that transforms each received OFDM symbol into the same symbol. To do this, the receiver first decodes the symbols using the standard 802.11 decoder. Specifically, the receiver performs an FFT on each time-domain OFDM symbol and transforms it into the frequency-domain. The receiver, then, decodes the modulated bits in each sub-channel, passes the modulated bits through the demodulator and the convolutional/Viterbi decoder to get the transmitted bits.[2]

Now that the WiSee receiver knows the modulated bits, it transforms every received symbol into the first OFDM symbol. Say $\mathbf{X}_n^i$ denotes the modulated bit in the n^{th} sub-channel of the i^{th} OFDM symbol. The WiSee receiver equalizes the i^{th} OFDM symbol with the first symbol. In particular, it multiplies the n^{th} frequency sub-channel in the i^{th} symbol with $\frac{\mathbf{X}_n^1}{\mathbf{X}_n^i}$.

Now that all the received symbols are data-equalized to the first symbol, the receiver performs an IFFT on each of these equalized symbols to get the corresponding time-domain samples. Since these data-equalization operations only modify the data in each sub-channel, it does not change either the wireless channel or the Doppler shift information. Now, the receiver effectively has repeated OFDM symbols and we are back to Case 1.

We note that human gestures change the phase and amplitude of the received symbols. A traditional decoder accounts for these changes by using the pilot bits that are present in every OFDM symbol. In particular, the receiver decodes by removing these phase and amplitude changes that encode the gesture information. To avoid this, during the re-encoding phase before computing the IFFT, the WiSee receiver re-introduces the phase and amplitude changes that were removed by the decoder. This ensures that the gesture information is not lost in decoding.

3.1.1 Practical Issues

We answer the following questions:

(a) How does WiSee deal with frequency offsets? A frequency offset between the transmitter and the receiver creates a shift in the center frequency, which can be confused for a Doppler shift. To address this issue, a WiSee receiver takes a two-pronged approach. First, it leverages prior work [23, 16] to get a coarse estimate of the frequency offset using the preamble at the beginning of the transmission; it compensates the estimated frequency offset in the rest of the transmission. Second, to account for any residual frequency offset, WiSee leverages the fact that in the absence of residual offsets, the energy in the DC frequency (center frequency of each OFDM sub-channel) corresponds to the paths from the transmitter to the receiver that do not involve the human. This shows up as the large peak in Fig. 3 (dotted line). However, with residual offsets, both the DC and the Doppler frequencies are shifted by the same amount (the solid line in the figure). Since the DC energy (i.e., the energy from the transmitter to the receiver minus the human reflections) is typically much higher than the Doppler energy (reflections from the human), the WiSee receiver tracks the frequency corresponding to the maximum energy and corrects for the residual frequency offset.

(b) Does the transmitter have to continuously transmit on the wireless medium? So far, we assume that the transmitter transmits continuously and the receiver uses the signal to compute the Doppler shifts. This is, however, not feasible since 802.11 packets are typically less than a few milliseconds. Further, while transmitting continuously might be feasible in an un-utilized network, it can significantly affect the throughput of other devices in a busy network. WiSee instead performs the following procedure: it linearly interpolates the received OFDM symbols to fill the time slots where no transmission happens. The interpolation is done per sub-channel after the OFDM symbols are transformed into the frequency domain. After interpolation, the receiver transforms all OFDM symbols, both original and interpolated, back to the time domain and forms a synthesized time-continuous trace. The underlying assumption here is that during the short time-period between two transmissions, the user's motion does not dis-continuously change the wireless channel. To see the effects of this interpolation, we perform an experiment where the user performs two simple gestures—either move her arm towards the receiver or away from the receiver. The user is ten feet away from the receiver. We apply our gesture classification algorithm, which we describe in the next section. Fig. 4 shows the results with the transmissions evenly spread out in time. The plot shows that the classification accuracy is high even when the transmission occupancy is as low as 3%.

(c) What about the cyclic prefix? The above discussion assumes that the transmitter sends the OFDM symbols back-to-back. However, an 802.11 transmitter sends a cyclic prefix (CP) between every two OFDM symbols to prevent inter-symbol interference. The CP is created by taking the last k samples from each OFDM symbol. We consider the CP

[2]For the reasons described in §3.1.1, discarding occasional erroneous packets does not significantly affect Doppler shift computation.

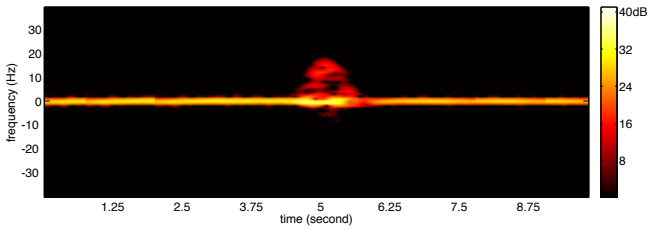


Figure 5—Frequency-time Doppler profile of an example gesture. The user moves her hand towards the receiver.

to be a specific kind of discontinuity between the OFDM symbols. Thus, we can perform interpolation between the OFDM symbols as described earlier. We note, however, that since all the CPs have a fixed length, such an interpolation is equivalent to resampling the OFDM symbols at a constant rate given by $\frac{Symbol\_length+CP\_length}{Symbol\_length}$, where *Symbol_length* and *CP_length* denote the length of the OFDM symbol and CP respectively. Since such resampling of the symbols does not change the doppler pattern, in practice we simply skip the CPs to reduce the computation.

3.2 Mapping Doppler Shifts to Gestures

So far we described how to transform the wideband 802.11 transmissions into a narrowband signal at the receiver. In this section, we show how to extract the Doppler information and map it to the gestures. Specifically, we describe the following three steps: (1) Doppler extraction which computes the Doppler shifts from the narrowband signals, (2) Segmentation which identifies a set of segments that correspond to a gesture, and (3) Classification which determines the most likely gesture amongst a set of gestures. We describe how WiSee performs each of these steps. We focus on the single user case; in §3.3, we extend our design to work in the presence of other users.

(1) Doppler Extraction: WiSee extracts the Doppler information by computing the frequency-time Doppler profile of the narrowband signal. To do this, the receiver computes a sequence of FFTs taken over time. Specifically, it computes an FFT over samples in the first half-a-second interval. Such an FFT give a Doppler resolution of 2 Hertz. The receiver then moves forward by a 5 ms interval and computes another FFT over the next overlapping half-a-second interval. It repeats this process to get a frequency-time profile.

Fig. 5 plots the frequency-time Doppler profile (in dB) of a user moving her hand towards the receiver. The plot shows that, at the beginning of the gesture most of the energy is concentrated in the DC (zero) frequency. This corresponds to the signal energy between the transmitter and the receiver, on paths that do not include the human. However, as the user starts moving her hand towards the receiver, we first see increasing positive Doppler frequencies (corresponding to hand acceleration) and then decreasing positive Doppler frequencies (corresponding to hand deceleration).

We note that the WiSee receiver is only interested in the Doppler shifts produced by human gestures. Since the speeds at which a human can typically perform gestures are between 0.25 m/sec and 4 m/sec [12], the Doppler shift of interest at 5 GHz is between 8 Hz and 134 Hz. Thus, the WiSee receiver reduces its computational complexity by analyzing the FFT output corresponding to only these frequencies.

(2) Segmentation: To do this, WiSee leverages the structure of the Doppler profiles, shown in Fig. 6. These correspond to the gestures in Fig 1. The plots show that the profiles are a combination of positive and negative Doppler shifts. Further, each gesture comprises of a set of segments that have positive and negative Doppler shifts. For example, the profile in Fig. 6(a) has just one segment with positive Doppler shift. However, Fig. 6(b) has two segments each of which has a positive and a negative Doppler shift. Further, within each segment, the Doppler energy first increases and then decreases (which correspond to acceleration and deceleration of human body parts).

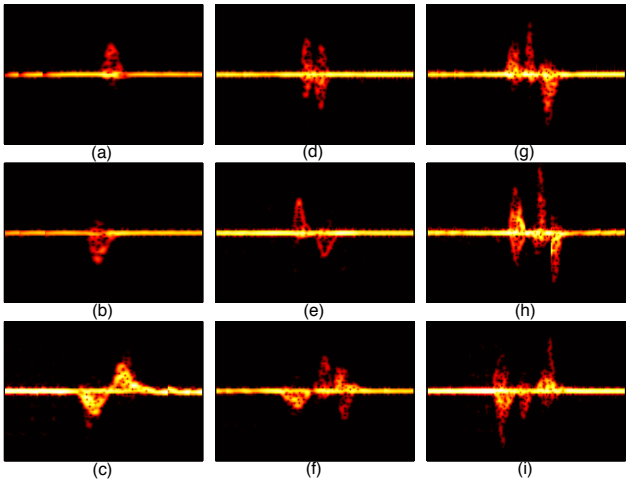


Figure 6—Frequency-time Doppler profiles of the gestures in Fig. 1. WiSee segments the profiles into sequences of positive and negative Doppler shifts, which uniquely identify each gesture.

A WiSee receiver leverages these properties to first find segments and then cluster segments into a gesture. Our process of finding segments is intuitively similar to packet detection in wireless communication systems. In communication, to detect the beginning of a packet, the receiver computes the average received energy over a small duration. If the ratio between this energy and noise level is greater than a threshold, then the receiver detects the beginning of a packet. Similarly, if this ratio falls below a threshold, the receiver detects the end of the packet. Likewise, in our system, the energy in each segment first increases and then decreases. So the WiSee receiver computes the average energy in the positive and negative Doppler frequencies (other than the DC and the four frequency bins around it). If the ratio between this average energy and the noise level is greater than 3 dB, the receiver detects the beginning of a segment. When this ratio falls below 3 dB, the receiver detects the end of the segment.[3] To cluster segments into a single gesture, WiSee's receiver uses a simple algorithm: if two segments are separated by less than one second, we cluster them into a single gesture.

(3) Gestures Classification: As described earlier, the Doppler profiles in Fig. 6 can be considered as a sequence of positive and negative Doppler shifts. Further, from the plots, we see that the patterns are unique and different across the nine gestures. Thus, the receiver can classify gestures by matching the pattern of positive and negative Doppler shifts. Specifically, there are three types of segments: segments with only positive Doppler shifts, segments with only

[3]The noise level is calibrated at the receiver by computing the energy in the non-DC frequencies, in the absence of gestures.

negative Doppler shifts, and segments with both positive and negative Doppler shifts. These can be represented as three numbers, '1', '-1', and '2'. Each gesture in Fig. 6 can now be written as a unique sequence of these three numbers. Now, gesture classification can be performed by comparing and matching the received number sequence with the set of pre-determined sequences. We note that our classification algorithm works with different users performing gestures at different speeds. This is because, different speeds only change the duration of each segment and the specific Doppler frequencies that have energy, but do not change the pattern of positive and negative shifts. Thus, the gestures performed at different speeds result in the same pattern of numbers and hence can be classified.

We briefly comment on the selection of the gestures in Fig. 1. In this paper, we picked gestures that can be encoded by a sequence of positive and negative Doppler shifts. As shown in Fig. 1, this covers a variety of interesting gesture patterns. In principle, one can imagine extending WiSee to more general gesture patterns by modeling the human body motion and leveraging additional features from the signal; this, however, is not in the scope of this paper.

We note that it is unlikely that random gestures such as eating, stretching, etc. would be confused with the specific gestures used to control devices. This is because, as we describe in the next section, a user gains control of the system by performing a special, hard-to-confuse gesture sequence that acts as a preamble. We also note that one can leverage techniques like Hidden Markov Models (HMMs) and Dynamic Time Warping (DTW) to increase the gesture space. Exploring these algorithms however is not in the scope of this paper.

3.3 Multiple Humans

WiSee leverages MIMO to improve the accuracy and robustness of the system, and to enable it to work in the presence of multiple humans. However, as described in §1, MIMO decoding requires a known preamble to compute the MIMO channel of the target user. WiSee uses a repetitive gesture as a preamble. Specifically, the user pushes her hand towards and away from the receiver, and repeats this gesture to form the preamble. This creates a sequence of alternating positive (+1) and negative (-1) Doppler shifts, i.e., an alternating sequence of +1 and -1 symbols. The WiSee receiver uses this sequence to correlate and detect the presence of a target human. Note that, similar to communication systems [11], this correlation works even in the presence of interfering users, since their motion is uncorrelated with the preamble's alternating sequence of positive and negative Doppler shifts. Next, WiSee finds the MIMO channel that maximizes the Doppler energy from the target user. At a high level, it runs an iterative algorithm (similar to [10]) on each segment of the preamble gesture to find the MIMO direction that maximizes the Doppler energy. It then averages the MIMO direction across segments to improve the estimation accuracy. Using this estimated MIMO direction, a WiSee receiver mitigates interference from other users and locks onto the target user by projecting the received signal on the desired direction.[4]

[4]Prior AoA algorithms [27] typically require fine-grained calibration of the MIMO systems, including measuring the distance between antennas, and tracking phase offset. Our system, however, avoids such calibration since we are not interested in the physical direction of the signal, but the MIMO direction that maximizes the Doppler energy.

Specifically, say the WiSee receiver has N antennas and the preamble gesture has M segments. Our objective is to find complex weights, $\mathbf{W}_n$, for each of the antennas, such that the Doppler energy for each of the segments is maximized. Specifically, if $\mathbf{D}_m$ is the Doppler energy in the mth segment, then we want to find the set of directions, $\mathbf{W}_n$, $n = 1, \cdots, N$, that maximizes:

$$\mathbf{D}_m = \sum_{n=1}^{N} \mathbf{W}_n \mathbf{D}_{nm},$$

for all $n = 1, 2, 3...N$, where $\mathbf{D}_{nm}$ is the Doppler energy corresponding to the mth segment on the nth antenna.

WiSee applies gradient descent that iterates over the amplitudes and phases of all $\mathbf{W}_n$s to find a optimal set of weights. Note that while the phases can span values between 0 and 2π, the dynamic range of amplitudes is primarily determined by the antenna gains and the different receive gains on the radios (in our implementation this range is 6 dB).

We note the following: Firstly, the iterative algorithm occasionally gets stuck in local minima where the Doppler energy does not considerably increase with iterations. To mitigate this problem, we select multiple initial points that are evenly spaced and repeat the algorithm starting from these points. Secondly, since WiSee's receiver uses up to five antennas, the search space significantly increases with the number of receive antennas. To minimize complexity, we run the iterative algorithm pair-wise on each antenna with respect to the first antenna. Specifically, we run the iteration algorithm to find the weights on the N-1 antennas independently where the weights are computed with respect to the first antenna.

Finally, using the repetitive gesture in the preamble, the receiver can improve the estimation accuracy by averaging across gestures. WiSee further improves this accuracy by tracking this channel as the user performs gestures. Specifically, the WiSee receiver applies the iterative algorithm on every gesture performed by the target user. This allows it to adapts the MIMO direction as different users interfere with the target user.[5] Our results in §4.3 show that, in the presence of three other interfering users, WiSee can classify the first two gestures in Fig. 1 with an average accuracy of 90% using a 5-antenna receiver.

We note that WiSee not only enables a user to perform gestures in presence of other humans, but also enables multiple users to concurrently interact with the system. Specifically, the WiSee receiver can track the MIMO direction of each user to classify the gestures from multiple users.

3.3.1 Further Discussion

We discuss how one may augment WiSee's current design to make it more robust and secure.

(1) Tracking a mobile target user: Our description and evaluation assume that the target user performs gestures from a fixed location and that she performs the repetitive motion (preamble) when she moves to a new location. However, in principle, one can reduce the need for repeating the pattern by tracking the user as she moves in the environment.

[5]As the interfering users change, the optimal MIMO direction that maximizes the Doppler energy also changes. By applying the iterative algorithm on every gesture, WiSee can track these changes.

Specifically, human motion (e.g., walking and running) creates significant Doppler shifts, which as explained in §3 have a higher energy than human gestures. Thus, the receiver can, in principle, track the MIMO channel as the target user moves, reducing the need to perform the repetitive gesture again.

(2) Providing security: One of the risks of using a whole-home gesture recognition system is enabling an unauthorized user outside the home to control the devices within. To address this problem, one may use a secret pattern of gestures as a secret key to get access to the system. Once the access is granted, the receiver can track the authorized user and perform the required gestures. Evaluating the potential of such an approach, however, is outside the scope of this paper.

3.4 Addressing Multi-path Effects

So far we assumed that the reflected signals from the human body arrive at the wireless receiver along a single direction. In practice, however, the reflections, like typical wireless signals, arrive at the receiver along multiple paths.

Extracting general Doppler profiles in the presence of multi path is challenging. However, the use of only the positive and negative Doppler shifts for gesture classification simplifies our problems. Specifically, we have to address two main problems: First, due to multi-path, a user performing a gesture in the direction of the receiver from an adjacent room, can create both positive and negative Doppler shifts at the receiver. Second, strong reflectors like metallic surfaces can flip the positive and negative Doppler shifts. For example, the receiver can observe stronger negative Doppler shifts from a user moving her hand towards the receiver, if the user is standing close to a metallic surface behind her.

The iterative algorithm used by WiSee intrinsically addresses the first problem. Specifically, as shown in recent work on Angle-of-Arrival (AoA) systems [27], multiple antennas can be used to separate multi-paths by adjusting the phase on each antenna. Since WiSee's iteration algorithm can adapt both the amplitude and the phase to maximize either the positive or the negative Doppler energy, it automatically finds a MIMO direction that can focus on multi paths that result in similar Doppler shifts. We note however that unlike AoA systems, computing Doppler shifts does not require distinguishing between the multi-paths in the system. Instead WiSee only needs to distinguish between sets of paths that all create either a positive or a negative Doppler shift. Thus, one can perform gesture recognition using a lower number of antennas than is required in AoA systems.

To address the flipping problem between positive and negative Doppler shifts, WiSee leverages the preamble. Specifically, since the repetitive gesture in the preamble always starts with the user moving her hand towards the receiver, the receiver can calibrate the sign of the subsequent Doppler shifts. Specifically, if it sees a negative Doppler shift when it expects a positive shift, the receiver flips the sign of the Doppler shift. This allows WiSee to perform gesture recognition independent of the user location.

4. Evaluation

We implement a prototype of WiSee on the software radio platform and evaluate it on the USRP-N210 hardware. Each USRP is equipped with a XCVR2450 daughterboard, and communicates on a 10 MHz channel at 5 GHz. Since USRP-N210 boards cannot support multiple daughterboards, we built a MIMO receiver by combining multiple USRP-N210s using an external clock [1]. In our evaluation, we use MIMO receivers that have up to five antennas. We use single-antenna USRP-N210s as transmitters.[6]

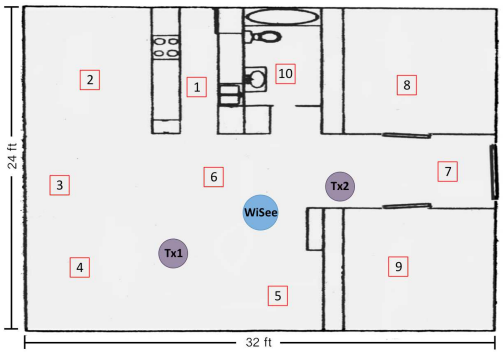


Figure 7—Floor plan of the two-bedroom apartment: The WiSee receiver, Tx1, and Tx2 are placed in the living room. The layout has 4 LOS, 4 NLOS, and 2 through-the-wall scenarios.

The transmitter and the receiver are not connected to the same clock. We build on the UHD code base to continuously transmit OFDM symbols over a 10 MHz wide channel. The transmitter uses different 802.11 modulations (BPSK, 4QAM, 16QAM, and 64QAM) and coding rates. The transmit power we use in our implementation is 10 mW which is lower than the maximum power allowed by USRP-N210s (and Wi-Fi devices). This is because USRP-N210s exhibit significant non-linearities at higher transmit powers, which limits the ability to decode OFDM signals. We note, however, that, with higher transmission powers, one can in principle perform gesture recognition at larger distances.

We evaluate our prototype design in two environments:

- *An office building.* The UW CSE building has an external structure that is primarily metal and concrete. We run the experiments in offices that are separated by double sheet-rock (plus insulation) walls with a thickness of approximately 5.7 inches. The building has a number of other Wi-Fi access points and devices operating in the frequency of interest.
- *A two-bedroom apartment.* The layout of the apartment is shown in Fig. 7. It consists of a living room connected to the kitchen and dining area, and two bedrooms and a bathroom. The walls are hollow and have a thickness of 5.5 inches; the doors are made of wood and have a thickness of 1.4 inches.

We run our experiments with a total of five users. In addition to evaluating WiSee's ability to achieve whole-home gesture recognition, we extensively evaluate it in the six different scenarios, shown in Fig. 8.

(a) LOS-txrxcloseby: Here a receiver and a transmitter are placed next to each other in a room. The user performs gestures in line-of-sight to the receiver.

(b) LOS-txrxwall: Here a receiver and a transmitter are placed in adjacent rooms separated by a wall. The user performs the gestures in the room with the transmitter.

[6]Our implementation does not use off-the-shelf Wi-Fi transmitters since we don't have a configuration that will allow current USRPs to reliably operate at a bandwidth of 20 MHz.

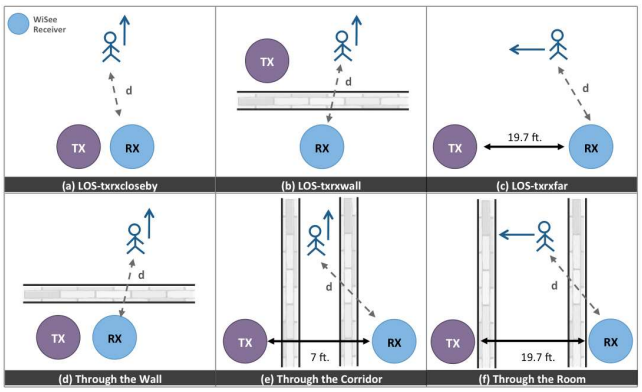


Figure 8—Scenario layouts.

(c) LOS-txrxfar: Here a receiver and a transmitter are placed 19.7 feet away from each other. The user performs gestures in line-of-sight to the receiver.

(d) Through-the-Wall: Here a receiver and a transmitter are placed next to each other close to a wall. The user performs gestures in the room adjacent to the wall.

(e) Through-the-Corridor: Here a receiver and a transmitter are placed in different rooms separated by a corridor. The user performs the gestures in the corridor.

(f) Through-the-Room: Here a receiver and a transmitter are placed in different rooms separated by a room. The user performs the gestures in the middle room.

In scenarios (b), (c), (e), and (f) both the transmitter and the receiver use omnidirectional antennas. However, in scenarios (a) and (d) where the transmitter and the receiver are placed next to each other, to prevent the transmitter's signal from overwhelming the receiver's hardware, the transmitter uses a Ettus LP0965 directional antenna that is placed in a direction orthogonal to the receive antennas. The receiver, however, still uses omnidirectional antennas. In principle, we can further reduce the transmitter's interference in these two scenarios by leveraging techniques like full-duplex [7] and interference nulling [16]. However, this is not in the scope of this paper.

In the rest of this section, we first evaluate the feasibility of gesture recognition using wireless signals in these six scenarios. We then evaluate WiSee's performance in a whole-home scenario. Finally, we demonstrate WiSee's ability to work in the presence of other humans.

4.1 Feasibility of Wireless Gesture Detection

We start by evaluating the feasibility of gesture detection using wireless signals in various scenarios.

Experiments: We run experiments in the six scenarios depicted in Fig. 8. We pick the office building to run these experiments since it has more rooms and also allows us to evaluate the system at larger distances than in our two-bedroom apartment. To evaluate how well WiSee can detect the presence of a gesture, we compute the Doppler SNR from the frequency-time Doppler profile. Specifically, Doppler SNR is the ratio between the average energy in the non-DC frequencies in the profile, with and without the gesture. We ask the user to move her hand towards the receiver, i.e., the first gesture in Fig. 1. The user performs the gesture in the general direction of the receiver independent of its location. Our intuition behind this choice is that the user would naturally gesture in the direction of the device she wants to control. We compute the average Doppler SNR at each location by having each user repeat the gesture ten times.

Results: Figs. 9(a)-(f) plot the average Doppler SNR as a function of distance, for the six scenarios. The plots show the results for different number of antennas at the receiver. They show the following:

(a) Versus distance: In scenarios (a), (b), (d), and (e), as the distance between the user and the receiver increases, the average Doppler SNR reduces. This is expected because the strength of the signal reflections from the human body reduces with distance. However, the received Doppler SNR is still about 3 dB at 12 feet, which is sufficient to identify gestures. In scenarios (c) and (f), however, the Doppler SNR does not significantly reduce with the distance from the receiver. This is because in both these scenarios, as the user moves away from the receiver, she gets closer to the transmitter. Thus, while the human reflections get weaker as the user moves away from the receiver; since the user moves closer to the transmitter, the transmitted signals arrive at the user with a higher power, thus, increasing the energy in the reflected signals. As a result, the Doppler SNR is as high as 15 dB at distances of about 25 feet.

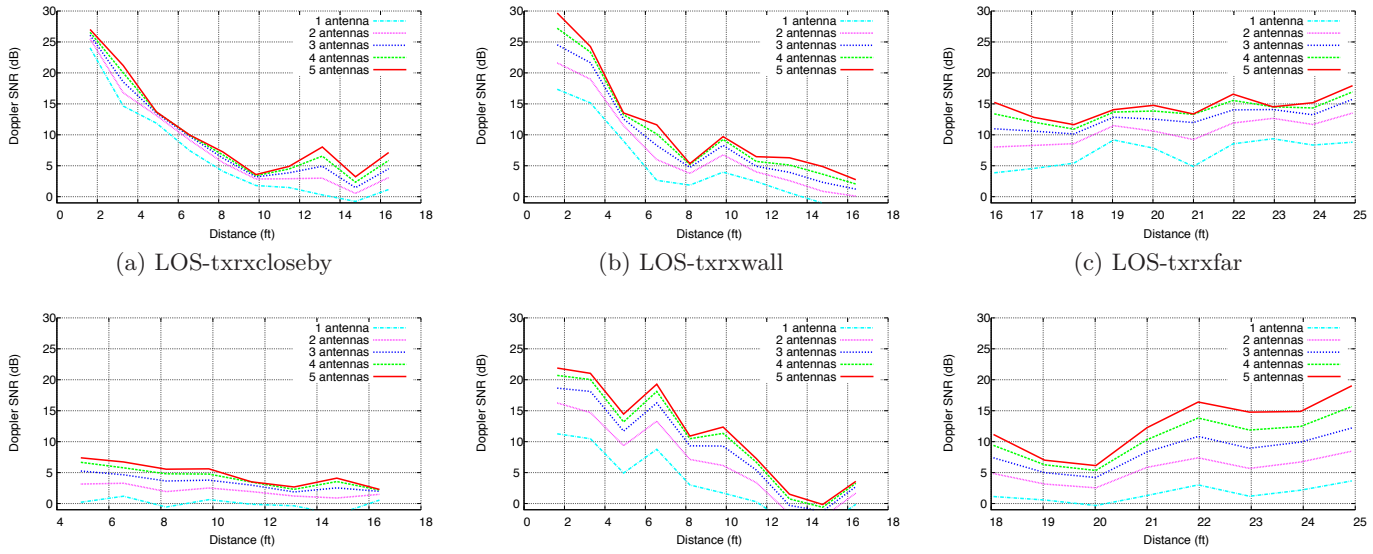
(b) Versus number of antennas: Across all the scenarios, using more antennas at the WiSee receiver increases the Doppler SNR. This is expected because additional antennas provide diversity gains that are particularly helpful in the low SNR regime. Further the gains in the Doppler SNR are higher at large distances, and through-the-* scenarios—the gains are as high as 10 dB in some locations. Also note that in scenarios (d) and (f), the Doppler SNR at a single-antenna receiver is as low as 1 dB across many positions; such low SNRs are not sufficient to classify gestures. Additional antennas significantly increase the Doppler SNR, enabling gesture detection in these scenarios. As described in §3.4, this is because of the wireless multi-path effects. In line-of-sight scenarios there is a strong direct path between the human body and the WiSee receiver or the transmitter. However, the through-the-* scenarios experience significant multi-path that effectively adds additional interference at the receiver. With more antennas at the receiver, the algorithm in §3.2 can reduce the multi-path interference and hence can significantly improve the Doppler SNR. We note that across all the scenarios, using 3-4 antennas at the receiver is sufficient to achieve most of the MIMO benefits.

Summary: The key takeaways from the above results are: First, using 3-4 antennas at the WiSee receiver is sufficient to achieve gesture detection in all the above scenarios. Second, gesture detection is feasible as long as the user is in the "range" of the receiver. This range can, however, be increased by spatially separating the transmitter and the receiver. Another option that we explore in the next section, is to leverage multiple transmitters (e.g., mobile phone in the living room, and laptop in the kitchen) to increase this range.

4.2 Gesture Recognition in Whole-Home Scenario

Next, we evaluate WiSee in a whole-home scenario.

Experiments: We run experiments in the two-bedroom apartment shown in Fig. 7. We use a 4-antenna WiSee receiver to compute the frequency-time Doppler profile using the al-



(a) LOS-txrxcloseby (b) LOS-txrxwall (c) LOS-txrxfar

(d) Through-the-Wall (e) Through-the-Corridor (f) Through-the-Room

Figure 9—Doppler SNR versus distance from the receiver. The plots show the Doppler SNR in various scenarios as a function of both distance and the number of receiver antennas.

gorithm described in §3.2. We then map the profile to the gestures using our pattern-mapping algorithm. We use two single-antenna transmitters, as shown in Fig. 7, to maximize the range of our system. The transmitter and receiver use omni-directional antennas in our experiments. Since it is difficult to implement carrier sense on software radios, the two transmitters time-share the medium every 10 ms. The receiver and the two transmitters are all placed in the living room, with all the doors closed. We note that performing gesture recognition with the doors closed is more challenging since the signal has to traverse through the doors and walls and hence experiences significant attenuation. We pick ten locations (marked in the layout) spanning all the rooms in the apartment. These locations include line-of-sight, non-line-of-sight, and through-the-wall settings. In each location,the users perform the nine gestures shown in Fig. 1 in the direction of the receiver, at a speed that was not predetermined. Before each experiment, the users were shown how to perform each gesture. The users were also able to check the gesture sketches (Fig. 1) during the experiment. Each gesture is performed a total of 100 times across all the locations.

Results: Fig. 10 plots the confusion matrix for the ten gestures across all the locations. Each row denotes the actual gesture performed by the user and each column the gestures it was classified into. The last column counts the fraction of gestures that were not detected at the receiver. Each element in the matrix corresponds to the fraction of gestures in the row that were classified as the gesture in the column. The table shows the following:

- The average accuracy is 94% with a standard deviation of 4.6% when classifying between our nine gestures. This is in comparison to a random guess, which has an accuracy of 11.1% for nine gestures. This shows that one can extract rich information about gestures from wireless signals. We note that despite wireless signals typically being noisy, since Doppler shifts are detected over the duration that is of the order of a second, WiSee achieves the above high accuracy by averaging across time. Further, the multiple antennas also provide spatial diversity that also increases the accuracy.
- Only 2% of all the gestures (18 out of 900) were not detected at the receiver. Further investigation revealed that these mis-detections occurred when the user was in the kitchen and one of the bedrooms. In these locations, the reflected signals are weak and hence the Doppler SNR for these specific gestures was close to 0 dB.

We note that when only tx1 was used to perform gesture recognition, the accuracy was greater than 90% only in six of the ten considered locations. This shows that each transmitter provides a limited range for gesture recognition. Adding more transmitters increases this effective range. We, however, note that the two-bedroom apartment scenario only required two transmitters placed in the living room to successfully classify gestures across all the locations.

4.3 Gestures in the Presence of Other Humans

Finally, we evaluate WiSee in the presence of other humans. We first measure the false detection rate in the absence of the target human. Then, we compute the accuracy of gesture recognition for the target human, in the presence of other humans. Finally, we stress test WiSee to see where it fails.

False detection rate in the presence of other humans: As described in §3.3, WiSee detects the target human by using a repetitive gesture as a preamble. The repetitive gesture provides a protection against confusing other humans for the target user. We compute the average number of false detection events, i.e., when the WiSee receiver detects the target user (repetitive gesture), in her absence. To do this, we place our WiSee receiver and transmitter in the middle of an office room (with dimensions 32 feet by 30 feet) occupied by 12 people, over a 24-hour period. The receiver looks for a repetitive gesture where the user moves her hand towards and away from the receiver; thus, each repetition results in a positive Doppler shift followed by a negative Doppler shift.

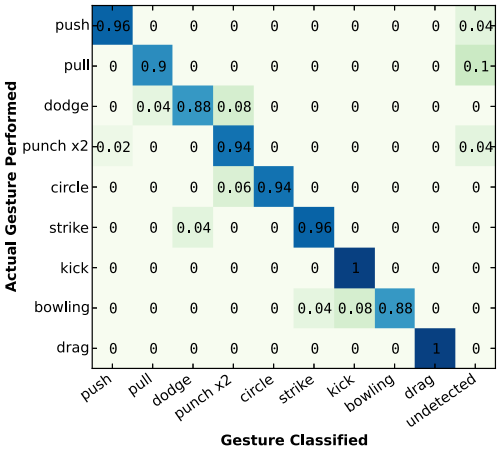


Figure 10—Confusion matrix for gestures in the home scenario: The figure shows that the average detection and classification accuracy is 94% across the nine gestures. In contrast, random guesses have an accuracy of 11.1%. This shows that WiSee can extract rich information about gestures from wireless signals.

The occupants move about and have meetings and lunches in the office as usual. We believe that, given the higher density of people, the office room is a worse scenario compared to our two-bedroom apartment. We note that there are other scenarios in which WiSee can be evaluated, which are, however, not in the scope of this paper.

Fig. 11 plots the number of false detection events per hour as a function of time. The figure shows results for different number of repetitions in the preamble. The plot shows that when the receiver uses a preamble with only one repetition (i.e., perform the gesture once), the number of false events is, on the average, 15.62 per hour. While this is low, it is expected because typical human gestures do not frequently result in a positive Doppler shift followed by a negative Doppler shift. For example, in our experiments, walking caused a continuous monotone Doppler shift that was not confused with alternating positive and negative Doppler shifts. Also, as the number of repetitions in the preamble increases, the false detection rate significantly reduces. Specifically, with three repetitions, the average false detection rate reduces to 0.13 events per hour; with more than four repetitions, the false detection rate is zero. This is expected because it is unlikely that typical human motion would produce a repetitive pattern of positive and negative Doppler shifts. Further, since the WiSee receiver requires repetitive positive and negative Doppler shifts to occur at a particular range of speeds (0.25 m/s to 4 m/s), it is unlikely that even typical environmental and mechanical variations would produce them.

Classifying the target human gestures in the presence of other humans: As described in §3.3, WiSee computes the MIMO channel for the target user that minimizes the interference from the other humans. We would like to evaluate the use of MIMO in classifying a target user's gestures, in the presence of other moving humans. We run experiments in a 13 feet by 19 feet room with our WiSee receiver and transmitter. We have the target user perform the two gestures in Fig. 1(a) and Fig. 1(b). Our experiments have up to four interfering users in random locations in the room. The users were asked to perform arbitrary gestures using their arms.

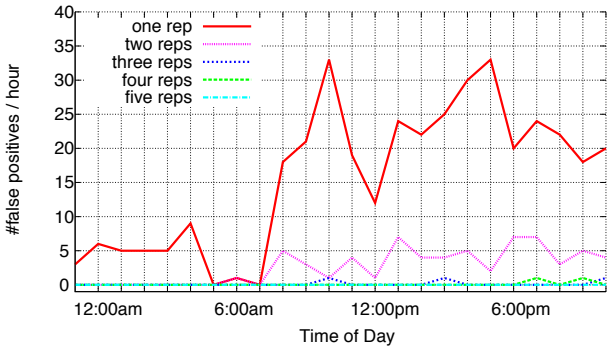


Figure 11—False Detection Rate from a 24-Hour Trace: The figure plots the false detection rate in an office room with 12 people over a 24-hour period on a weekday.

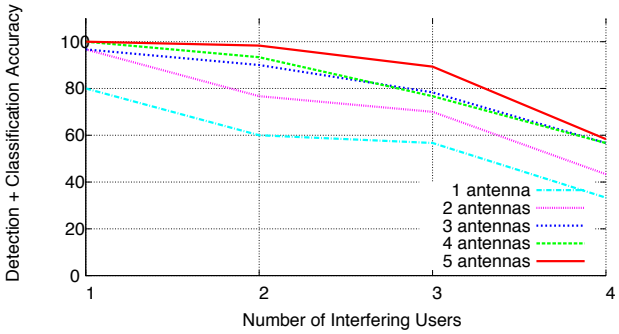


Figure 12—WiSee in the presence of other interfering users: The figure plots the detection and classification accuracy of the target user in the presence of other users in a 13 × 19 sq. feet room. The plots show that, given a fixed number of antennas, as the number of interfering users increases, the accuracy decreases. However, with three interfering users, the accuracy is still as high as 90% with a five-antenna receiver.

Fig. 12 plots the average recognition accuracy of the target user's gestures as a function of the number of interfering users. The figure shows results for different number of antennas at the WiSee receiver. The plots show that using a five-antenna receiver, the accuracy is as high as 90% with three interfering users in the room. Further, using additional antennas significantly improves this accuracy in the presence of multiple interfering users. We note however, that for a fixed number of transmitters and antennas at the receiver, the classification accuracy degrades with the number of users (e.g., a conference room setting or a party scenario). For example, in our experiments, the accuracy is less than 60% with four interfering users. However, since typical home scenarios do not have a large number of users in a room, WiSee can enable a significant set of interaction applications for always-available computing embedded in the environment.

Stress-testing WiSee: Since WiSee leverages MIMO to cancel the signal from the interfering human, it suffers from the near-far problem that is typical to interference cancellation systems. Specifically, reflections from an interfering user closer to the receiver, can have a much higher power than that of the target user. To evaluate WiSee's classification accuracy in this scenario, we run the following experiment: We fix the location of the target user six feet away from the WiSee receiver. We then change the interfering user's location between three feet and ten feet from the receiver. The target user performs the two gestures shown in Fig. 1(a) and

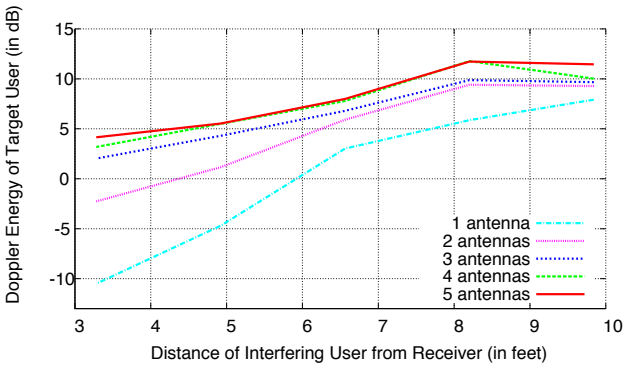


Figure 13—Stress testing WiSee in a near-far scenario: The figure plots the Doppler energy of a target user in the presence of another interfering user at various distances. The target user is fixed 6 feet away from the receiver. The plot shows that adding antennas at the receiver helps address the near-far problem.

Fig. 1(b). In each experiment, the interfering user adversarially performs the opposite gesture to the one performed by the target user. The WiSee receiver computes the MIMO channel for the target user using her preamble.

Fig. 13 plots the target user's Doppler SNR at the WiSee receiver as a function of the distance between the interfering user and the receiver. The figure shows multiple plots corresponding to different number of antennas at the receiver. The plots show that, smaller the distance between the interfering user and the receiver, lower the Doppler SNR. Specifically, the Doppler SNR is as low as -10 dB when the interfering user is about 3.4 feet from the receiver. However, the figure also shows that adding antennas at the receiver significantly improves the Doppler SNR. Specifically, with four-antennas, the Doppler SNR increases from -10 dB to 4.7 dB; which is sufficient to classify gestures. Thus, we conclude that adding additional receive antennas can help mitigate the near-far problem.

5. Conclusion

In this paper, we take the first step towards transforming Wi-Fi into a gesture-recognition sensor. We present WiSee, a novel gesture recognition system that leverages wireless signals to enable whole-home sensing and recognition of human gestures. Since wireless signals do not require line-of-sight and can traverse through walls, WiSee can enable whole-home gesture recognition using few signal sources. Our results in a 2-bedroom apartment show that WiSee can extract a rich set of gesture information from wireless signals and enable whole-home gesture recognition using only two wireless sources placed in the living room.

Acknowledgements: We thank the members of the UW Networks and Wireless group, our shepherd Moustafa Youssef, and the anonymous MOBICOM reviewers for their helpful comments and Lilian de Greef for help with the gesture sketches in Fig. 1.

6. References

[1] ACQUITEK Inc. Fury GPS Disciplined Frequency Standard.
[2] R. Block. Toshiba Qosmio G55 features SpursEngine, Visual Gesture Controls.
[3] W. Carrara, R. Goodman, and R. Majewski. Spotlight Synthetic Aperture Radar: Signal Processing Algorithms. Artech House, 1995.
[4] G. Charvat, L. Kempel, E. Rothwell, C. Coleman, and E. Mokole. A Through-dielectric Radar Imaging System. In Trans. Antennas and Propagation, 2010.
[5] K. Chetty, G. Smith, and K. Woodbridge. Through-the-wall Sensing of Personnel Using Passive Bistatic WiFi Radar at Standoff Distances. In Trans. Geoscience and Remote Sensing, 2012.
[6] K. Chintalapudi, A. Iyer, and V. Padmanaban. Indoor Localization without the Pain. In NSDI, 2011.
[7] J. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti. Achieving single channel full duplex wireless communication. In Mobicom, 2010.
[8] G. Cohn, D. Morris, S. Patel, and D. Tan. Humantenna: using the body as an antenna for real-time whole-body interaction. In CHI 2012.
[9] M. Fisher. Sweet Moves: Gestures and Motion-Based Controls on the Galaxy S III.
[10] S. Gollakota, F. Adib, D. Katabi, and S. Seshan. Clearing the RF Smog: Making 802.11 Robust to Cross-Technology Interference. In SIGCOMM,2011.
[11] S. Gollakota and D. Katabi. Zigzag decoding: combating hidden terminals in wireless networks. In ACM SIGCOMM, 2008.
[12] S. Gupta, D. Morris, S. Patel, and D. Tan. Soundwave: using the doppler effect to sense gestures. In HCI 2012.
[13] C. Harrison, D. Tan, and D. Morris. Skinput: appropriating the body as an input surface. In CHI 2010.
[14] D. Kim, O. Hilliges, S. Izadi, A. D. Butler, J. Chen, I. Oikonomidis, and P. Olivier. Digits: freehand 3d interactions anywhere using a wrist-worn gloveless sensor. In UIST 2012.
[15] Y. Kim and H. Ling. Human Activity Classification Based on Micro-Doppler Signatures Using a Support Vector Machine. In Trans. Geoscience and Remote Sensing, 2012.
[16] K. Lin, S. Gollakota, and D. Katabi. Random Access Heterogeneous MIMO Networks. In SIGCOMM, 2011.
[17] J. Paradiso, C. Abler, K.-y. Hsiao, and M. Reynolds. The magic carpet: physical sensing for immersive environments. In CHI, 1997.
[18] T. Ralston, G. Charvat, and J. Peabody. Real-time through-wall imaging using an ultrawideband MIMO phased array radar system. In Array, 2010.
[19] A. Santos. Pioneer's latest Raku Navi GPS units take commands from hand gestures.
[20] M. Scholz, S. Sigg, H. R. Schmidkte, and M. Beigl. Challenges for device-free radio-based activity recognition. In CoSDEO workshop, 2007.
[21] S. Sen, B. Radunovic, R. R. Choudhury, and T. Minka. Spot localization using PHY layer information . In Mobisys, 2012.
[22] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In CVPR, 2011.
[23] K. Tan, J. Fang, Y. Zhang, S. Chen, L. Shi, J. Zhang, and Y. Zhang. Fine-grained Channel Access in Wireles LAN. In SIGCOMM, 2010.
[24] J. Wilson and N. Patwari. Radio Tomographic Imaging with Wireless Networks. In Trans. Mobile Computing, 2009.
[25] J. Wilson and N. Patwari. Through-Wall Motion Tracking Using Variance-Based Radio Tomography Networks. In ARXIV, 2009.
[26] J. Xiong and K. Jamieson. Towards fine-grained radio-based indoor location. In HotMobile, 2012.
[27] J. Xiong and K. Jamieson. ArrayTrack: A Fine-Grained Indoor Location System. In NSDI, 2013.
[28] M. Youssef, M. Mah, and A. Agrawala. Challenges: Device-free passive localization for wireless environments. In Mobicom, 2007.
[29] Y. Zhao and N. Patwari. Robust Estimators for Variance-Based Device-Free Localization and Tracking. In ARXIV, 2011.