Wi-Flex: Reflex Detection with Commodity WiFi

MERT TORUN, University of California Santa Barbara, USA
YASAMIN MOSTOFI, University of California Santa Barbara, USA

In this paper, we are interested in startle reflex detection with WiFi signals. We propose that two parameters related to the received signal bandwidth, *maximum normalized bandwidth* and *bandwidth-intense duration*, can successfully detect reflexes and robustly differentiate them from non-reflex events, even from those that involve intense body motions (e.g., certain exercises). In order to confirm this, we need a massive RF reflex dataset which would be prohibitively laborious to collect. On the other hand, there are many available reflex/non-reflex videos online. We then propose an efficient way of translating the content of a video to the bandwidth of the corresponding received RF signal that would have been measured if there was a link near the event in the video, by drawing analogies between our problem and the classic bandwidth modeling work of J. Carson in the context of analog FM radios (Carson's Rule). This then allows us to translate online reflex/non-reflex videos to an instant large RF bandwidth dataset, and characterize optimum 2D reflex/non-reflex decision regions accordingly, to be used during real operation with WiFi. We extensively test our approach with 203 reflex events, 322 non-reflex events (including 142 intense body motion events), over four areas (including several through-wall ones), and with 15 participants, achieving a correct reflex detection rate of 90.15% and a false alarm rate of 2.49% (all events are natural). While the paper is extensively tested with startle reflexes, it is also applicable to sport-type reflexes, and is thus tested with sport-related reflexes as well. We further show reflex detection with multiple people simultaneously engaged in a series of activities. Optimality of the proposed design is also demonstrated experimentally. Finally, we conduct experiments to show the potential of our approach for providing cost-effective and quantifiable metrics in sports, by quantifying a goalkeeper's reaction. Overall, our results confirm a fast, robust, and cost-effective reflex detection system, without collecting any RF training data, or training a neural network.

CCS Concepts: • **Computing methodologies** → **Modeling and simulation**; **Machine learning**; • **Human-centered computing** → **Ubiquitous and mobile computing**; • **Hardware** → **Wireless devices**.

Additional Key Words and Phrases: Reflex Detection, Computer Vision, Startle Reflexes, WiFi, RF Sensing, Sport Reflexes

1 INTRODUCTION

Imagine these scenarios: touching a hot stove by accident; opening a bottle of champagne that causes a sudden splash in your face; putting on your headset without adjusting the volume, only to be exposed to a very sudden loud noise; your friend showing up as Pennywise right behind you. All these events will most likely result in what is known as a *startle reflex* reaction. More formally, the startle reflex (also referred to as *escape reflex*) is a sudden, involuntary and unsought defensive motor response to an intense sensory stimulus (e.g., sudden noise, sudden motion in the visual field, etc.) [6, 27, 61], and is a sub-category of r eflexes.[1] Fig. 1 shows some sample startle

---

[1]In a different context, startle reflex refers to a particular involuntary motor response newborns exhibit temporarily. But that is different from the subject of this paper.

Authors' addresses: Mert Torun, merttorun@ucsb.edu, University of California Santa Barbara, Santa Barbara, California, USA, 93106-0001; Yasamin Mostofi, ymostofi@ece.ucsb.edu, University of California Santa Barbara, Santa Barbara, California, USA, 93106-0001.

Fig. 1. Sampler of our assembled reflex video dataset, using online videos ([25] and YouTube.) See color PDF for best viewing.

scenarios [25]. A startle response typically involves the activation of a number of muscle groups in a defensive mode. The main distinctive feature of a reflex action is that it does not involve a conscious decision-making process. A startle reflex is thus a considerably important body reaction that is developed through evolution to protect us (as well as other species) from harm.

Startle-type reflexes are important and implicit in a number of physical/mental health conditions and have been studied extensively in medical fields. For instance, startle reflexes are shown to intensify with anxiety disorder, PTSD, phobias, or even stress [26, 39, 49, 50, 52, 56, 62, 71, 75], and further inhibit different characteristics in schizophrenia [8, 88]. As such, startle reflexes have been studied in the context of mental health as part of both a diagnostic and a treatment system (e.g., to measure phobia improvement after a gradual exposure). Studying startle reflexes is also important for the diagnosis and monitoring of a number of physical health conditions, such as post-stroke/brain injuries, autism, Tay-Sachs disease, Urbach-Wiethe disease, Alzheimer's disease, Huntington's disease, and hyperekplexia, among others [31, 34, 38, 40, 51, 63, 68, 88].

Detection of a startle reflex can also be useful for safety/security applications. More specifically, since a startle reflex can be an indication of a state of distress, pain, or harm, one can envision it to be used as a general measure of the safety/well-being of an individual, for instance in a smart home/office space, and in particular for elderly or children. Furthermore, detecting them can also be important for ambient sensing as part of smart spaces, for instance for context inference.

In the medical domain, most related work has focused on using on-body sensors (e.g., electromyography), in order to measure muscle contraction during a reflex, involving expensive bulky setups [9–11, 20]. Such work further use the prior knowledge that a reflex is about to happen and are also not designed to operate in the wild.

In recent years, there has been interest in the area of vision to detect if an action in a given video was intentional, using machine learning [25, 30, 81, 82]. While not directly on reflexes, some of the corresponding videos include startle reflex reactions. However, cameras may not be present in all the environments, need a direct unobstructed view of the participant at all the times, are not privacy-preserving and may thus not be comfortable for everyone [3, 44] (Sec. 7 reports performance of such vision approaches). In this paper, we propose a new approach that enables a WiFi link to **detect startle reflexes and robustly differentiate them from non-reflex events, even from challenging ones that involve fast body motions.** To the best of our knowledge, there is currently no existing approach that can enable robust startle reflex detection with any form of RF transceivers, or in particular with WiFi.

Radio Frequency (RF) signals, such as WiFi, are ubiquitous these days, creating the possibility of using them for sensing and learning about the environment. This sensing potential has created immense interest in the research

community and WiFi signals are utilized for various applications, e.g., localization/tracking [35, 64, 79], smart health [36, 37, 42], occupancy analytic [17, 41, 93], activity recognition [13, 72, 83], and imaging [58, 69, 90], among others.

In this paper, we introduce Wi-Flex, a robust system for startle reflex detection, using only WiFi. In the envisioned setup, a WiFi transmitter (TX) transmits a wireless signal, which bounces off the surroundings and is then received by a WiFi receiver (RX) that measures the Channel State Information (CSI) of the received signal (magnitude or phase difference) for the purpose of reflex detection. One challenge in developing a robust reflex detection system is that there could be non-reflex daily events that would involve rapid body motions, such as certain exercises (e.g., start of a sprint), or simply clapping hands. **Our approach is robust in the sense that it can successfully differentiate a reflex from non-reflex activities, even from challenging non-reflex ones that involve rapid body motions.** It furthermore does not rely on collecting any RF data for analysis, learning, or training purposes, a task that would be laboriously prohibitive. Instead, we show that two received signal bandwidth-related parameters can robustly detect a reflex. More specifically, we propose a methodology for translating freely-available online reflex/non-reflex videos to their corresponding RF received signal bandwidth. This then enables us to create an instant large RF received signal bandwidth dataset, and use it to characterize optimum 2D reflex/non-reflex decision regions for our two bandwidth-related parameters, which can then be used during the real operation with WiFi for decision making.

Remark 1: While we have so far discussed startle reflex detection, **our proposed approach is also applicable to sport reflexes**. By a sport reflex, we refer to a sport-related event that requires a fast body reaction to a visual input. For instance, consider a goalkeeper. As a fast ball approaches, a well-trained goalkeeper would act based on a reflex-type reaction, and without conscious processing. Our proposed methodology for reflex detection can also be applicable for measuring sport reflexes. As such, in this paper we also show experimental results in this context, in addition to extensive experimentation with startle reflexes. We further discuss other potential implications of the proposed work in sports domain.

Remark 2: In this paper, we use the term **reflex event** to describe an event that involves a body reflex. The term **non-reflex event** then denotes otherwise, and is divided into two categories of **normal events** and **challenging normal events**, where **normal event** refers to a non-reflex event that does not involve rapid body motions (e.g., walking, eating), while the term **challenging normal event** denotes non-reflex events that involve rapid body motions (e.g., running, certain exercises, etc.). Testing the system with both categories is important, as we shall do, since challenging normal events can resemble a reflex and a robust system should differentiate them correctly.

Remark 3: For brevity, we use the term *reflex* in the rest of the paper, without any prefix. The goal of the paper is to develop a robust startle reflex detection system, and further demonstrate its applicability to sport-type reflex events, as discussed.

Statement of Contributions:

1. Developing a robust reflex detection system with WiFi requires collecting a massive reflex RF dataset, for analysis and design, which would be prohibitive. We instead propose to translate online available reflex videos to a meaningful instant large RF-based dataset, for the purpose of RF analysis/design, as we elaborate next.

2. As a starting point, we ask if a small number of parameters can capture a reflex in a video. Along this line, we set forth that maximum normalized speed and speed-intense duration suffice to robustly detect a reflex in a video. We then validate this by using a large video dataset of 50 reflex and 82 non-reflex events (including several challenging ones).

3. Speed-related parameters (while the first-intuition choice), however, are not reliably measurable by a wireless link during a reflex due to the corresponding abrupt motions, as we shall see. We then show how we can translate a given video content to metrics that are robustly measurable by a wireless link, yet carry crucial differentiating information on a reflex. More specifically, **we propose a way of translating the content of a given video to**

the bandwidth signal of the corresponding received RF signal that would have been measured if there was a wireless link in the vicinity of the event in the video. We use a bandwidth modeling of the received wireless signal that is inspired by the classic work of J. Carson in the context of analog FM radios [15] **(i.e., Carson's Rule).** Following the design principles learned from Item 2, we then introduce two received RF signal bandwidth-based metrics, **maximum normalized bandwidth** and **bandwidth-intense duration**, and show how they can be measured from a video using Carson's Rule. This then allows us to translate available online reflex/non-reflex videos to an instant large RF received signal bandwidth dataset, and accordingly characterize reflex/non-reflex optimum decision regions for our two bandwidth-based parameters, to be used during the real operation with WiFi for decision making.

4. We extensively test the proposed pipeline with 203 reflex events, 180 normal events and 142 challenging normal events, over four areas (including several through-wall ones) and with 15 participants, achieving an overall probability of correct classification of 90.15% and a false alarm rate of 2.49%. See Fig. 8 for sample reflex experiments. As Section 7 shows, our proposed pipeline considerably outperforms the state-of-the-art in reflex detection. We next show results where a person is engaged in a sequence of events comprised of reflex-inducing, normal, and challenging normal ones. We further show that our proposed system can robustly detect a reflex even when multiple people are present and simultaneously engaged in a variety of activities including both reflex and non-reflex ones. Finally, we show how our bandwidth-based reflex parameters have the potential to provide quantifiable metrics in sports domain to measure an athlete's reaction response. More specifically, we show how increasing the speed of an incoming ball results in the goalkeeper's catch event changing from non-reflex to reflex at some point.

We emphasize that all the events are natural and not acted out. We further note that designing reflex-inducing scenarios that are effective, yet safe, is a challenging task. As such, **the experiments are conscientiously designed, via consultation with our IRB committee, to ensure the safety and well-being of the participants, and were further deemed as not constituting human subject research.** Overall, our results show the potentials of our work for reflex detection with WiFi signals.

2   RELATED WORK

As discussed earlier, startle reflexes are implicit in a number of physical/mental health conditions, and have been utilized for both diagnostic purposes and treatment progress assessment in the medical domain [8, 26, 34, 39, 40, 49–52, 56, 62, 68, 71, 75, 88, 88]. From an academic standpoint, there has also been a considerable interest in studying the impact of factors such as age, anxiety or stress on startle reflexes [47, 49, 60]. Most such work, however, use elaborate, expensive setups involving on-body sensors [77]. For instance, electromyography (EMG) is commonly used in order to directly measure muscle contraction during a reflex [4, 7, 88]. As such, a reliable, transportable, and cost-effective startle reflex measurement system that can also be used as part of a smart home health system is currently lacking in the medical domain.

In the area of vision, there has been recent interest in identifying if an action in a given video was intentional or not [19, 25, 30, 81, 82]. While not directly on reflex detection (human mistakes are also included in these work), some of the corresponding videos involve a startle reflex. More specifically, [25] presents the first work in the area of vision to detect unintentional actions using machine learning and a large training dataset. This work is then followed by others, who built on it and improved it, while keeping the core approach as machine learning on a large dataset [19, 30, 81, 82]. These work are promising as they suggest that a reflex-like action can be inferred from a video. However, cameras need a direct unobstructed view, are not privacy-preserving, and are thus not a favorable sensing mechanism for everyone [18, 66]. Nevertheless, in this proposed work, we utilize the vast freely-available reflex-related vision dataset (some from these work and some from YouTube) and show how we can translate each video to the bandwidth signal of the corresponding received RF signal. This then allows

us to create an instant large RF bandwidth-related dataset, in order to characterize optimum reflex/non-reflex 2D decision regions for two proposed bandwidth-related parameters, which can then be used during the real operation with WiFi for decision making.

In recent years, there has also been interest in detecting abnormal activities (with different interpretations of the word abnormal). For instance, crime-type abnormalities are detected in a video footage in [24, 65], while most other work in this area focus on detecting actions such as falling (with accelerometer in [43, 59], with motion sensors in [48], with WiFi in [1, 21, 57, 84, 91]). The general area of activity recognition has also received considerable attention in recent years with the main focus on classifying daily normal activities, and typically based on machine learning. For instance, WiFi has been used for gesture recognition [2, 33, 70], exercise classification [13, 32, 45], and driver activity recognition [5, 23]. Similarly, activity recognition has been explored heavily in vision [46, 54, 80]. However, there is no existing work in RF sensing where reflexes are considered as a type of activity. Furthermore, utilizing existing RF sensing methods for reflex detection will not work well since existing work on activity classification is not focused on activities that are abrupt and short-lived. We shall see this in more details in Sections 4 and 7. The intersection of vision and RF has also been explored, for instance, to make concurrent video and RF measurements for labeling purposes [89], or to simulate wave propagation given a video input [13], both for the purpose of training machine learning systems. Overall, there is no existing work on detecting the types of reflexes of interest to this paper with WiFi signals, which is the main motivation for the proposed work of this paper.

Finally, as mentioned earlier, our proposed approach can also have implications for sport-type reflex events, i.e., for situations that involve a fast reaction (without conscious processing) to a fast incoming ball. Along this line, there is a considerable interest in measuring various aspects of the reaction of certain types of athletes for training purposes [16, 53, 55]. Such approaches mainly rely on measuring the response time, with a variety of gadgets, from cheap not so accurate ones [12] to expensive elaborate setups [55, 86]. As we shall see, our proposed approach has the potential to provide a cost-effective, portable, and robust method for characterizing the reaction of an athlete. As such, we not only test our proposed approach with sport-type reflex scenarios, but also show experiments on how our bandwidth-based approach can be used to quantify an athlete's response to an incoming ball.

3 PROBLEM FORMULATION

Consider a pair of wireless transceivers in an area where a person is present. As the person's body parts move, the transmitted signal bounces off different body parts of the person. The complex baseband received signal $s_b(t)$ can then be written as follows [42],

$$s_b(t) = \alpha_s e^{j\theta_s} + \sum_{m \in \Omega} \alpha_m e^{j\left(\frac{2\pi}{\lambda}\psi_m \int v_m(t)dt + \frac{2\pi}{\lambda}d_m\right)}, \tag{1}$$

where $\alpha_s e^{j\theta_s}$ includes the impact of both the direct path from the TX to the RX and the reflections from the static objects, α_m is the amplitude of the signal path reflected off of the mth part of the body, d_m is the length of that path at time $t = 0$, $v_m(t)$ is the speed of the mth body part at time t, and $\psi_m = \cos\phi_{R,m} + \cos\phi_{T,m}$ where $\phi_{R,m}$ and $\phi_{T,m}$ are the angles between the direction of motion of the mth body part and the path to the transmitter and receiver respectively, at time t. The summation in Eq. 1 is over the body parts that are visible to the receiver at that time instant (set Ω), and λ is the wavelength. Off-the-shelf devices, however, can only measure the received power or phase difference (if multiple antennas are available onboard) of this signal. The following can then well approximate the power of the received signal (after DC removal), as was shown in recent work [42],

$$p(t) = \sum_{m \in \Omega} A_m \cos\left(\frac{2\pi\psi_m}{\lambda} \int v_m(t)dt + \Delta\mu_m\right), \tag{2}$$

where $A_m = 2\alpha_s\alpha_m$ and $\Delta\mu_m = \mu_m - \mu_d$ is the difference between the initial phase of the reflected path off of the mth body part and the direct path.

A Note on CSI Phase Measurements: While the absolute phase on off-the-shelf WiFi devices cannot be reliably measured [92], the phase difference of the antennas of the same card can be stably measured with the CSI-Tool [29], as we shall also do in this paper. It is also recently shown that the phase difference between two antennas on-board can be well approximated in a similar form to Eq. 2 (albeit with different values for A_m and $\Delta\mu_m$) [42], allowing us to treat both the power and phase difference under the same unifying umbrella in this paper (see Appendix A for more details).

In this paper, we are then interested in designing a system that can robustly detect reflexes, using only the received signal power or phase difference of a pair of WiFi transceivers. We next lay out the details of our proposed methodology.

4 SYSTEM DESIGN

In this section, we lay out the details of our proposed pipeline for robust reflex detection with WiFi signals. Our proposed approach does not rely on collecting massive RF data for analysis, learning, or training purposes, a task that would be laboriously prohibitive. Instead, it relies on the massive available online videos for characterizing **a small number of key reflex-related parameters** that are **both extractable from a video** while also **robustly measurable by a wireless link**. In other words, our approach is multi-disciplinary, drawing from both areas of vision and wireless systems. Our first step towards achieving our goal is to show that a small number of video-related parameters would suffice for robustly detecting a reflex in a video. More specifically, we first set forth that there are two parameters, **maximum normalized speed** and **speed-intense duration**, that can robustly indicate presence of a reflex in a video, and validate it by using a large video dataset.

This is promising as it indicates that a small number of parameters can capture a reflex in a video. However, speed-related parameters, while an intuitive first-step design choice, are not reliably measurable by a wireless link during a reflex, due to the resulting abrupt motions, as we shall see. We then show how we can translate a video content to metrics that are robustly measurable by a wireless link. More specifically, by using the design principles learned from our speed analysis, we propose two parameters related to the received RF signal bandwidth: maximum normalized bandwidth and bandwidth-intense duration. We then show how we can translate each video to its corresponding RF received signal bandwidth. Towards this goal, our pipeline first extracts the speeds of body parts from each video. It then draws similarity between the form of the received RF signal in front of body motions and the form of an analog FM radio signal, which allows us to mathematically characterize the overall bandwidth of the received signal as a function of the speeds of different body parts, using Carson's Rule from decades-old FM radio literature [15]. In a recent work [42] on nocturnal seizure detection with WiFi, authors drew an analogy between the received power signal in front of a series of movements and the summation of a number of FM radio waves, and successfully characterized the whole Fourier response of the received signal (i.e., the Bessel coefficients), using Carson's Rule [15]. Here we are only interested in the overall bandwidth of the signal. We then design a pipeline to translate our video dataset to a large RF received signal bandwidth dataset, utilizing Carson's Rule [15]. We then show how to characterize the 2D optimum reflex/non-reflex decision regions for the aforementioned two bandwidth-related parameters accordingly, confirming that they can robustly detect a reflex. In the operation phase, a WiFi link then measures these two parameters of the received signal and uses their corresponding video-derived decision regions to robustly detect a reflex. We next start by analyzing a large video dataset and establishing that a small number of parameters can robustly capture a reflex in a video.
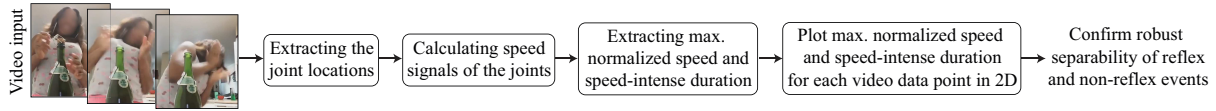
Fig. 2. Steps of initial video dataset analysis to confirm that a reflex can robustly be detected in a video with only two parameters. These two speed-related parameters, while the most intuitive starting point, are not robustly measurable by WiFi during a reflex, necessitating a different design as we shall see.

4.1 Video Data Analysis

As discussed earlier in Section 2, a number of recent work in vision have developed algorithms to detect a non-intentional action in a video. These work are promising as they suggest that reflex-like actions can be inferred from a video. Most of these work, however, use machine learning for extensive training on related video datasets. In this part, we are instead interested in establishing that a few key video-related parameters can robustly detect a reflex.

Consider a video of a person engaged in an activity. We expect that when there is a reflex, the speed of some of the body parts goes up high. As such, instantaneous speed of a body part, normalized by its local average, presents a good potential starting-point metric for capturing the impact of a reflex. We next assemble a large dataset, using freely-available online reflex videos, to formalize the impact of a reflex on the speed of body parts, and characterize related parameters that can separate reflex from non-reflex events.

Video Dataset: We have assembled a large dataset consisting of 50 reflex videos and 50 normal-event videos (no challenging event included for this analysis), using freely-available videos from YouTube as well as existing vision datasets [25]. Fig. 1 shows snapshots of sample reflex videos of our assembled dataset. As can be seen, the dataset includes a diverse set of reflex-inducing incidents. Our normal activity dataset then includes a diverse set of daily activities (e.g., walking, eating, cleaning, cooking, etc.).

Processing Videos: The processing steps are summarized in Fig. 2. As can be seen, frames of each video are fed to OpenPose [14], an open source vision pipeline, in order to extract the joint locations. For all the visible joints, the joint speed as a function of time (in terms of pixels per frame) is calculated by considering the differential over consequent frames. To reduce estimation error, joints for which OpenPose has a low confidence are not considered. Furthermore, only the joints that appear in the majority of the frames are considered for speed calculation, and a low-pass filter is applied to the resulting speed profiles to smooth out the impact of estimation noise. We then define *normalized speed* as the speed signal of each joint normalized at each time instant by its local average over the prior T_{win} seconds (thus it is a function of time). *Maximum normalized speed* is then defined as the maximum of the normalized speed signal over time, and further over the body parts.

Video Histogram Analysis: Fig. 3 (left) shows the histogram of the maximum normalized speeds for the aforementioned assembled reflex-and normal-event video dataset. As can be seen, there is a clear separability between the two events. The figure then confirms that reflex events produce a significantly higher maximum normal- ized speed as compared to normal events, creating a possibility for differentiating the two in a video accordingly.

A Need for a Temporal Metric: So far, we established that the maximum normalized speed can well differentiate reflex from normal events in a video. However, as discussed earlier, there could be non-reflex daily events that involve high body speeds (i.e., challenging normal events in this paper), such as certain exercises (e.g., running), or simply clapping hands. In order to study the differentiability of such events from reflex, based on max. normalized speed, we further assemble a comprehensive video dataset of 32 such fast non-reflex daily events (using freely-available YouTube videos). The dataset is diverse encompassing both sport-related and
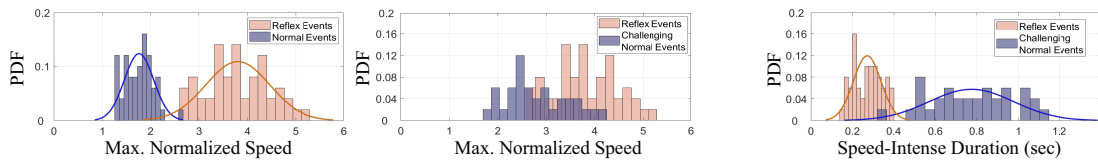
Fig. 3. Video dataset analysis – (left) histogram of maximum normalized speed for reflex and normal events, showing a clear separability, (middle) histogram of maximum normalized speed for reflex and challenging normal events, indicating challenging normal events can have high max. normalized speed, thus not differentiable from reflex based on max. normalized speed, (right) histogram of speed-intense duration for reflex and challenging normal events, showing that speed-intense duration can separate reflex from challenging normal events.

non-sport-related actions.[2] Fig. 3 (middle) shows the histogram of the maximum normalized speed of such events, superposed on the corresponding histogram of the reflex events. As can be seen, the maximum normalized speed of the challenging normal events can become high, making them non-differentiable from reflex events based only on the corresponding maximum normalized speed. This necessitates a second metric that can further differentiate challenging normal events from reflexes.

Towards this goal, we have studied several reflex and challenging normal videos. Consider the joint that resulted in the maximum normalized speed in a given video. We have observed that the time duration, where the normalized speed signal of this joint stays highly correlated with its maximum value, is typically much smaller for a reflex event as compared to challenging normal events. In other words, in a reflex event, the normalized speed typically goes up and comes down abruptly (i.e., the speed changes are abrupt), while in a challenging normal event, the normalized speed can get high but the time duration to the high peak and back down is typically larger. This makes sense as in case of a non-reflex challenging event, the action is intentional and, as such, it takes longer for the motor system to execute the brain's command. On the contrary, in case of a reflex event, there is no conscious planning, resulting in an abrupt increase and decrease of the speed. As such, we propose that the time duration where the normalized speed (of the joint with the max. normalized speed) stays highly correlated with its max. value, can serve as a robust metric to differentiate reflex events from the challenging normal ones. We next formally define our metric and extensively validate our hypothesis with a large video dataset.

Definition – Full Width at Half Maximum: Consider a given signal as a function of a variable. The full width at half maximum (FWHM) is defined as the difference between the two variable values where the signal is equal to half of its maximum value [76]. FWHM is an established metric that has been heavily used in different fields to characterize a duration where a signal stays highly correlated with its maximum value. For instance, it is related to half-power bandwidth (3dB point) in the context of filter design, or half-power beamwidth in the context of antenna design.

Consider the joint with the maximum normalized speed value. Let **speed-intense duration** denote the Full Width at Half Maximum (FWHM) of this joint's normalized speed signal. We then propose to use speed-intense duration as our second metric, to separate reflex from challenging normal events. Fig. 3 (right) shows the histogram of the speed-intense duration of reflex and challenging normal events, indicating a clear separability. Overall, these results are promising as they suggest that the two metrics of maximum normalized speed and speed-intense duration have the potential to robustly separate reflex from non-reflex events, even challenging ones, in a video (See Fig. 2 for the whole pipeline).

---

[2]Note that some sports will involve a reflex reaction, such as reacting to a fast ball as a goalkeeper. Those sport reflex reactions are true reflexes and are considered in the reflex category. On the other hand, other sport-related activities, such as running, can involve high speeds but does not involve a reflex and are thus considered in the challenging normal event category.

Remark 4: Note that while challenging normal events typically have a larger speed-intense duration, general non-challenging normal events can have any intense duration. This is due to the fact that some normal non-challenging events can have relatively small maximum normalized speed, making it possible for the body to achieve the maximum in a short period of time. As such, we can not separate such events from reflex, solely based on the corresponding intense duration and need the maximum normalized speed for this separation. In Section 7, we show how the performance can degrade considerably if only one such parameter is utilized.

4.2 A Need for a New Vision-RF Metric

The previous section established that two speed-related parameters have the potential to robustly detect reflexes in a video. Thus, if we can extract them from WiFi measurements, we can then detect a reflex with WiFi. Robustly extracting these speed-related parameters *in the presence of reflex events*, however, can be challenging, as we discuss next.

In the RF sensing literature, short-time analysis, i.e., processing the signal over short time intervals, has been utilized to extract velocity profiles of some body parts when there is a motion in the vicinity of the link [28, 78, 85]. Such processing approaches include Short-Time-Fourier-Transform (STFT), or other related more sophisticated adaptive multi-widow variations (e.g., Hermite spec- trograms). Ideally, if the time period is short enough for the speeds to be considered constant, short-time analysis can result in useful transient information on the speeds. In practice, however, the short time window can result in a low-resolution in the Fourier domain, which would result in the frequency content mixing up. When the motions are not abrupt, the time window does not have to be as small, resulting in a reasonable frequency-domain resolution. Thus, one can extract useful information on the speeds from the frequency-time signal (i.e, spectrogram), as has been done in the context of gait-based person identification or activity recognition with RF signals [72–74, 83, 87].

When there is a reflex, however, the speeds of body parts can change very abruptly. As such, we cannot robustly extract the aforementioned speed-related parameters of maximum normalized speed and speed-intense duration from the spectrogram (even after using more sophisticated functions). In other words, processing the signal over short time intervals will result in a poor frequency-domain resolution and our speed-related parameters are not robustly extractable. **This then necessitates introducing new metrics for reflex detection that can be characterizable from our video dataset, yet robustly measurable in practice by an RF link,** which motivates our proposed approach of the next part. More specifically, we show how speed signals of a given video can be translated to the bandwidth-time signal of the corresponding RF measurement, which is then easily measurable in practice. By using the design principles learned in Section 4.1 regarding the differentiability of the reflex and non-reflex events, and by extending the same design principle from speed to bandwidth, we then introduce two new bandwidth-related metrics for robust reflex detection. It is noteworthy that **our mathematical characterization of the RF signal bandwidth of the next section explicitly shows why the speed-related parameters are not reliably extractable from the spectrogram during a reflex. In Section 7, we further confirm with experimental data that if velocity-related parameters are chosen for design, instead of the bandwidth-related ones, one loses performance considerably.**

4.3 Translating a Video Input to the Corresponding RF Signal Bandwidth

While the maximum normalized speed and the speed-intense duration are not robustly extractable from a WiFi spectrogram during a reflex, the bandwidth of the received signal can be easily evaluated. Thus, if we can characterize reflex detection metrics from our video dataset, based on the bandwidth of the received signal, we can then utilize them for RF-based reflex detection in practice.

Mathematically characterizing the bandwidth of the received RF signal in the presence of a body motion and relating it to video-related parameters (velocity of different body parts), however, is considerably challenging.
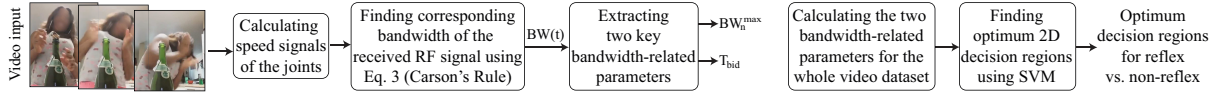
Fig. 4. (left) Proposed pipeline to translate a video input to its corresponding RF received signal bandwidth – (right) The two extracted bandwidth parameters of all the videos are then used to find the optimum 2D reflex/non-reflex decision regions using SVM.

More specifically, this would involve a theoretical characterization of the bandwidth of the signal of Eq. 2, when taking its Fourier transform over a period of time. Along this line, a recent result drew analogies between the signal form of a received RF signal in the presence of a body motion (e.g., Eq. 2) and a classic analog FM signal, and utilized the classic 1922 bandwidth results of J. Carson in the context of FM radios [15, 42], in order to theoretically characterize the bandwidth of the received signal. We next summarize these results, which will allow us to extract robust RF-bandwidth-based metrics from videos for reflex detection.

Frequency Modulation (FM) is a classic analog transmission technique, introduced in 1902 [67], to ensure robust transmissions for radio applications. Characterizing the bandwidth of an FM signal was crucial at that time, yet considerably challenging. In his seminal paper [15], **J. Carson was the first to theoretically characterize the bandwidth of an FM signal (i.e., Carson's rule)**. A typical FM transmitted signal has the form $\cos(2\pi f_c t + k_f \int m(t)dt)$, where $m(t)$ is the signal to be transmitted, f_c is the carrier frequency, and k_f is a constant. By carefully examining Eq. 2, we can think of it as a summation of a number of FM signals, in which the velocity of each body part is the modulating signal and $f_c = 0$. In other words, each moving body part can be thought of as modulating the corresponding body motion into an FM signal that is then received by the WiFi receiver. This then allows one to borrow mathematical tools from the seminal 1922 paper of Carson [15] to characterize the bandwidth of the received signal and relate it to the speed of the body parts. For instance, in a recent work on seizure detection with WiFi, the authors utilized tools from analog FM radio design in order to mathematically characterize the whole Fourier response of the received signal and the corresponding Bessel function coefficients [42]. In this paper, we are only interested in characterizing the overall bandwidth of the received signal. We next present a key result along this line.

Bandwidth of the received RF signal in the presence of a motion: Consider $p(t)$ of Eq. 2 over a moving time window of duration T_{mov}. The bandwidth of this signal, corresponding to its Fourier transform over the window of T_{mov}, can be characterized as follows, as a function of time:

$$BW(t) = \max_m (\psi_m(t) v_{\max,m}(t)/\lambda + f_m(t)), \tag{3}$$

where $f_m(t)$ is the bandwidth of the mth speed ($v_m(t)$) at time t, i.e., the bandwidth when directly taking the Fourier transform of $v_m(t)$ at time t, over the window T_{mov}, and evaluating its maximum spectral content. Moreover, $v_{\max,m}$ is the maximum of $v_m(t)$ over the window of T_{mov} at time t. See [15, 42] for details.

Note that Eq. 3 is applicable to both the received power or phase difference signals since both have a form dictated by Eq. 2, as discussed earlier. As can be seen, Eq. 3 allows us to mathematically characterize the bandwidth of the received RF signal in the presence of body motions, and relate it to speed of different body parts.

Remark 5: As discussed earlier, when body speeds do not change abruptly (not high temporal changes), some speed information can be extracted from the received signal's spectrogram. We can clearly see this scenario as a special case of Eq. 3. Basically, if the speeds do not change that abruptly, the second term, f_m, can become negligible and the maximum spectral content is well characterized by the first speed-based term. However, during a reflex, the speed can change abruptly and the second term is not negligible, motivating using bandwidth instead. As a comparison, the second term is about 6% of first term during a walk, and 10% in a jumping jack, while it is more than 20% for many reflexes (at 5 GHz). In Section 7, we shall further show, with extensive experimental
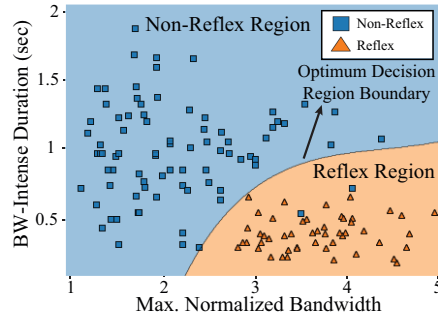
Fig. 5. 2D plot of max. normalized bandwidth and bandwidth-intense duration of the corresponding received RF signal. The bandwidth data points are directly found from our assembled video dataset. The figure shows a clear separability for reflex and non-reflex events. The optimum decision regions (using SVM) are also plotted on the figure.

results, that if speed is instead chosen for design, the performance degrades considerably, since it is not reliably extractable from spectrogram during a reflex.

Remark 6: It is worth emphasizing that here we are merely drawing resemblance between our high-level approximated received signal of Eq. 2 and the general form of a classic analog FM signal. However, this should not be mistaken as the body actively trans- mitting a signal, or the transmitter sending an analog FM signal. Our transmitter is an off-the-shelf transmitter, such as WiFi, and we simply observe that our approximation of the received signal power/phase difference of Eq. 2 resembles a classic analog FM signal (albeit with $f_c = 0$), allowing us to borrow mathematical tools from the seminal work of Carson [15].

4.4 Characterizing RF Bandwidth-Related Reflex Metrics from Video Datasets:

We next show that bandwidth of the received signal can also be used, in a similar manner to the velocity, to differentiate reflex from non-reflex events. In order to establish this, we need to use a large RF dataset of several reflex and non-reflex events, something that would be prohibitive to collect in practice. Instead, we propose to use our large video dataset, discussed in Section 4.1, and our bandwidth characterization of Eq. 3, to directly find the corresponding RF received signal bandwidth that would have been measured if the person in the video was performing the corresponding action near a pair of WiFi transceivers. Fig. 4 shows the steps of the proposed pipeline. For each video input, the speed signal (speed as a function of time) of each joint is extracted using existing vision algorithms, as discussed in Section 4.1. For the mth joint's speed signal, the bandwidth-time signal, i.e., the bandwidth as a function of time, is calculated by using a small moving window and taking the Fourier transform of the corresponding windowed speed signal to get $f_m(t)$. Maximum speed of the mth joint is similarly calculated for each moving window, in order to generate $v_{\max,m}(t)$. Consider a TX-RX pair placed in the area where the motion is occurring in the video. The TX transmits a signal, which will be reflected by the body parts and received by the receiver. We are then interested in finding the bandwidth of the corresponding received RF signal, which we can characterize from Eq. 3. More specifically, we use $f_m(t)$ and $v_{\max,m}(t)$ of each joint, extracted from the video, and calculate BW(t), the bandwidth-time signal of the corresponding received RF signal using Eq. 3. This then allows us to translate each video to the bandwidth of its corresponding RF received signal (BW(t)), resulting in a large RF bandwidth dataset. We next use this dataset to confirm robust reflex/non-reflex separability using bandwidth as a metric, following the same design principle we established from the earlier velocity analysis. More specifically, **we define *maximum normalized bandwidth*, $BW_{\text{n}}^{\text{max}}$, as the maximum of normalized RF bandwidth signal, normalized by its local average over prior T_{win} seconds. We further define *bandwidth-intense duration*, T_{bid}, as the Full Width at Half Maximum of the measured BW(t).** We next show that these two parameters can robustly differentiate reflex from non-reflex events.
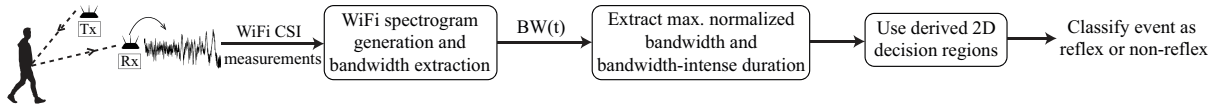
Fig. 6. Reflex detection with WiFi during operation, using our two RF bandwidth-based parameters and their corresponding optimum decision regions, which are derived using a video dataset and our proposed video to RF-bandwidth translation.

Fig. 7. Experiment locations - An office area (through-wall setting), a kitchen, a living room, and a semi-covered outdoor area.

Practical aspects of translating a video to the bandwidth of the corresponding RF signal: The temporal speed signal of the joints in a video will naturally be given in pixels per frame (by the vision algorithms). However, in order to translate them to the corresponding received signal bandwidth, using Eq. 3, we need the speeds in meters per second (to be consistent with λ). In order to solve this practical challenge, we propose to use an adult's interpupillary distance. More specifically, let d_{ip} denote the interpupillary distance of the person in the video, which will be in pixels. Studies have shown that the interpupillary distance of an adult is in the following range: 63.36 ± 3.83mm [22]. As such, we use the mean distance of 63.36mm, resulting in a pixel-to-meter conversion of $d_{ip}/0.06336$. In summary, for each video input, we take the maximum interpupillary distance over all the frames as d_{ip}. Any video where the person is never facing the camera is not used to make sure the pixel-to-m conversion is accurate. We can then convert the pixel/frame speed signal of each joint to m/s, using this pixel-to-m conversion and the fps (frames per second) of the video.

While our successful experimental classification results of the future sections implicitly confirm the validity of the proposed video to RF-bandwidth pipeline, we further explicitly confirm this pipeline in Section 7 via concurrent video-WiFi measurements.

RF bandwidth analysis and optimum decision regions: Consider our assembled video dataset of 50 reflex and 82 non-reflex events, discussed earlier in Section 4.1, samples of which are shown in Fig. 1. For each video input, we then implement the steps of the proposed pipeline of Fig. 4, using a moving window of length 0.4 seconds, and for a given TX/RX position pair (more on TX/RX locations later), in order to get the corresponding RF signal bandwidth as a function of time, and further extract the corresponding maximum normalized bandwidth (BW_n^{max}) and bandwidth-intense duration (T_{bid}). This results in a large RF bandwidth dataset, where each data point is comprised of the corresponding two parameters of BW_n^{max} and T_{bid}, for both reflex and non-reflex events. This then allows us to characterize the differentiability of reflex and non-reflex events based on these bandwidth-related parameters. More specifically, Fig. 5 shows the 2D plot for all the resulting data points (BW_n^{max} and T_{bid}) of all the videos. **It can be seen that there is a clear differentiability between reflex and non-reflex events, confirming that our proposed bandwidth-related parameters can be used to robustly detect a reflex.** In order to formally characterize the optimum reflex/non-reflex decision regions, we apply Support Vector Machine (SVM) to the data points of Fig. 5 (class size imbalance is accounted for), resulting in the marked decision boundary, which we shall use in the rest of the paper.

Remark 7 (Optimum decision regions' independence of TX/RX locations): In principle, TX/RX locations affect the bandwidth through $0 \leq \psi_m(t) \leq 2$ (see Eq. 1 for definition). However, **we emphasize that the derived decision regions have little dependency on locations of inserted TX/RX in the videos**. In general, if the

link is not too close to the person, and a motion is not close to parallel to the link, the resulting ψ will be close to the maximum value of two. From studying several reflex videos, for instance, we observed that when there is a reflex, several muscles exhibit intense motions and as such the ψ of the maximizing m will be close to two in Eq. 3. As such, we simply fixed our TX/RX locations for all the videos, which randomized the initial location/orientation of the person with respect to the link. But the resulting ψ of the maximizing m, once there was an event, was indeed near 2 for all the videos (the same was the case for non-reflex events). The derived decision regions will then become independent of TX/RX placement, which is important. During real WiFi experiments, we have no control over the person's body motions. However, our successful classification results of the next section further confirm that the decision regions are independent of TX/RX locations. In Section 7, we also confirm this from a different angle by deriving decision regions when TX/RX are individually placed in each video such that the ψ of maximizing m is exactly two. As we shall see, this has little impact on the performance.

4.5 Reflex Detection with WiFi

During the operation, the optimum decision regions derived in Section 4.4 will be used for reflex detection based on the received wireless measurements of a pair of transceivers. The overall WiFi pipeline is shown in Fig. 6. More specifically, the received WiFi CSI data is the input to the WiFi pipeline. The spectral content of the received signal (power or phase difference), and the associated bandwidth is then generated as a function of time, using a small moving window. The two key parameters of maximum normalized bandwidth and the bandwidth-intense duration are then extracted and compared with the optimum decision regions of Section 4.4 (Fig. 5), to make a reflex/non-reflex decision. In the next section, we extensively test the performance of the proposed approach.

5 EXPERIMENTAL SETUP

We next discuss our experimental details, including our subjects, the experiment locations, extensive description of all reflex, and non-reflex activities (including the challenging ones), and our WiFi data collection process, before presenting extensive results.

Experiment Subjects: To test our pipeline, we recruited 15 subjects.[3] During recruitment, each subject was informed (in a formal letter) that she/he will take part in a series of activities, which are all safe. The informing letter was phrased to inform the subjects that they may have different responses to different activities, but without explicitly mentioning reflex responses. The letter further indicated that all the activities are safe. The letter and the corresponding participant agreement form were written in consultation with our IRB committee.

Experiment Settings: The experiments took place in 4 different locations: an office, a living room, a kitchen and a semi-covered outdoor area, as shown in Fig. 7. **The experiments further include extensive through-wall reflex detection in the office area.** Overall, the experimental areas are diverse in order to show the robustness of the proposed approach in different settings.

Experiment Activities: We next discuss the reflex, normal, and challenging normal activities that took place during the experiments, a summary of which can be found in Table 1. **We emphasize that all the events are naturally occurring and not acted out. In particular, all the reflexes are naturally-induced responses.** Furthermore, the type of experimental reflex activities are majorly different from the reflex activities of the online videos that were used to find the optimum decision regions in Section 4.1. While a few of the non-reflex normal activities are similar in their high-level nature to the ones in the non-reflex videos (e.g., walking, eating, sitting down), the setup, execution, and details are completely different.

Reflex Events: In general, it is challenging to design events that can induce a reflex, while ensuring that they are safe, causing no harm or pain. For instance, touching a hot stove can induce a reflex but is a pain-causing

---

[3]The Institutional Review Board (IRB) committee has reviewed/approved this research and determined that it does not constitute human subject research.

Table 1. Our WiFi experimental activities including 21 reflex-inducing, 20 normal, and 17 challenging normal events. Activities are majorly different from those of online videos that were used to find the optimum reflex decision regions in Section 4.1.

Reflex Events		Non-Reflex Events	
		Normal Events	Challenging Normal Events
RE1) Holding a suddenly-vibrating item	RE2) Unexpected shoulder contact	NE1) Eating NE2) Drinking	CNE1) Dancing CNE2) Jumping
RE3) Watching a video with a jump scare	RE4) Empty cup splash prank[4]	NE3) Studying NE4) Lying down	CNE3) Clapping
RE5) An unexpected nearby water splash	RE6) Sudden loud noise	NE5) Reading NE6) Watching TV	CNE4) Dumbbell exercises
RE7) Unexpected item falling down nearby	RE8) Acting as a goalkeeper	NE7) Tying shoes NE8) Drawing	CNE5) Doing lunges
RE9) Dodging an unexpected plastic bat swing	RE10) Playing dodgeball game	NE9) Opening a door/cabinet	CNE6) Shaking a bottle
RE11) Reacting to catch a fast-approaching ball	RE12) Playing keepie uppie	NE10) Preparing a drink	CNE7) Sit-ups
RE13) Reacting to re-balance an item on hand before losing it		NE11) Applying hand sanitizer	CNE8) Hammering a nail
RE14) Playing a video game with a gradual close-up jump scare		NE12) Wearing a jacket	CNE9) Running
RE15) Reacting to prevent a fall while walking on a narrow wooden plank		NE13) Solving crossword puzzle	CNE10) Throwing an object
RE16) Plastic spider jumping out of a box unexpectedly		NE14) Picking up an item	CNE11) Sitting down
RE17) Reacting to prevent a fall after holding chicken pose for a while		N15) Playing video games	CNE12) Jump rope
RE18) A soft item thrown unexpectedly at participant		NE16) Having a conversation	CNE13) Scrubbing a surface
RE19) Attempting to catch an item that is unexpectedly rolling off table		NE17) Walking around	CNE14) Rock paper scissors
RE20) Reacting to maintain balance after losing it standing on one foot		NE18) Vacuum cleaning	CNE15) Jumping Jacks
RE21) Moving hand out of the way of a (soft)		NE19) Walking into/out of an area	CNE16) Intense yoga poses
object unexpectedly falling down towards it		NE20) Opening a package	CNE17) Push-ups

incident. As such, we have carefully designed a set of reflex-inducing experiments, with a particular attention to the safety and well-being of the participants. More specifically, our reflex category includes 21 events, as summarized in Table 1. Sample reflex events are also shown in Fig. 8. We crafted the events such that a variety of different sensory-motor reflex responses are included. For instance, some events have an unexpected visual component (e.g., an unexpected object appearing in the visual field) that would induce a motor response, while some have an unexpected audio component (e.g., an unexpected loud noise). While majority of the events have an unexpected sensory component that can induce a reflex response in the participant, there are a few sport-related events, such as acting as a goalkeeper or playing dogdeball, that would naturally include a reflex response to a fast incoming ball.

Normal Events: This category includes several normal daily activities, as listed in Table 1.

Challenging Normal Events: As mentioned earlier, some normal activities may involve high-speed body motions. In order to ensure the robustness of our approach, it is important to test it with such challenging normal events, to confirm that they are classified as non-reflex. Table 1 shows a detailed list of these events.

We note that, in general, if a participant had pauses while engaged in a non-reflex activity (e.g., chewing for a bit, then pausing for a bit, before proceeding to chew again), we only report the performance once, and for the worst-performing chunk.

Remark 8: We note that for each participant, we designed the sequence of events very carefully so they do not anticipate a reflex after their first encounter. In other words, once a participant experiences his/her first reflex-inducing event, he/she may anticipate another immediate unexpected event, which can reduce/eliminate their reflex response. In order to prevent this, we carefully planned the sequence of events such that several

---

[4]In this event, after offering to make tea for a participant, with two empty cups nearby, hot boiling water is poured into one cup during a casual chat. After the first cup is filled, instead of filling the 2nd cup, the 2nd cup is suddenly splashed at them, creating an immediate reflex response to avoid what the participant thinks holds boiling water.

Fig. 8. Snapshots of sample reflex events from our WiFi experiments. See Table 1 for details and color PDF for better viewing.

normal and challenging normal activities occur between the reflex events, resulting in a successful induction of a reflex throughout the experimentation.

Remark 9: As mentioned earlier, our goal for explicitly separating the non-reflex events to *normal* and *challenging normal* was to motivate the need for the *two* proposed reflex parameters. During experimentation, we need to include extensive events from both categories, as summarized in Table 1. However, we note that we have no control over how a participant executes such an event. For instance, a person may sit down very rapidly, making it a challenging event, or may sit slowly, making it more of a normal event (based on our observations, most participants sat down rapidly, and so we put it under challenging normal events in Table 1). In other words, the boundary between the two categories can be blurry in practice. This, however, does not impact our reflex detection performance since it is focused on robust classification between reflex and non-reflex categories.

WiFi Data Collection During Operation: We use the internal antennas of two laptops equipped with Intel 5300 NICs as transceivers. The laptops are placed 1 m above the ground, and 1 m apart (see Fig. 7). Furthermore, the distance of an event to the link varies between 1.5 m and 4 m for all the experiments since the participants can move freely and unexpectedly. The CSI information is collected at 5.32 GHz on 30 sub-carriers. We use one internal antenna of the TX laptop and all 3 of the RX laptop for CSI data collection, resulting in 90 data streams.

We then extract the complex CSI from these streams, using the CSI-Tool [29], and generate two phase difference signals between successive receiver antennas, per subcarrier. We further denoise the data using PCA. More specifically, we generate the spectrograms of the first 5 PCA components, using a time window $T_{mov} = 0.4$ sec, with a shift of 0.01 sec. We then average these 5 spectrograms to obtain the final one. The bandwidth signal (BW(t)) is then generated by finding the frequency where the power is at 95% of the total power in the spectrogram, at time t. For event detection, we use bandwidth of 5 Hz as a cut-off for determining both the start and end of an event. We note that all the events were successfully captured. Finally, for each detected event, our two classification parameters, maximum normalized bandwidth and bandwidth-intense duration, are generated using a local average window of $T_{win} = 1$ second.

Table 2. Classification accuracy of the proposed pipeline (top) through walls, (middle) in non through-wall settings, and (bottom) over all the experiments.

	Reflex Events	Non-Reflex Events
Probability of correct classification through-walls	89.57%	96.91%
Probability of correct classification in non through-wall settings	92.50%	98.94%
Overall correct classification prob.	90.15%	97.51%

6 EXPERIMENTAL EVALUATION

In this section, we present the results of our proposed WiFi-based reflex detection framework, in different environments, with people engaged in different kinds of reflex-inducing activities, and through walls. Our experimentation also includes several challenging normal events, in addition to normal events, to show system's robustness. We further show the variability of reflex detection performance per activity as well as per individual. Moreover, we show results where a person is engaged in a sequence of events comprised of reflex-inducing ones, normal activities, and challenging normal ones. We also show that our proposed system can robustly detect a reflex even when multiple people are present and simultaneously engaged in a variety of activities, from normal, challenging normal, to reflex-inducing ones. Finally, we show how increasing the speed of an incoming ball results in the goalkeeper's catch event changing from non-reflex to reflex, showing the potential of our introduced reflex parameters to provide quantifiable metrics in sports.

6.1 Overall Performance of Wi-Flex

We have conducted a total of 203 reflex events, 180 normal events and 142 challenging normal events over the four areas of Fig. 7 (including several through-wall ones) with 15 participants. Table 1 shows a list of the 21 reflex-inducing events, 20 normal events, and 17 challenging normal events that occurred in our experiments, as discussed earlier. Fig. 8 further shows sample snapshots of the reflex-inducing events. We next discuss the results.

Through-wall Reflex Detection: We conducted extensive experiments (163 reflex, 126 normal, and 101 challenging normal events) through walls in the office area of Fig. 7 (left), with the transceivers behind the wall. The experiments spanned 10 partic- ipants, performing the events listed in Table 1. Table 2 (top-row) shows the results in this through-wall area. As can be seen, our system can robustly detect 89.57% of reflexes through walls, with a very small probability of false alarm of 3.09%.

Reflex Detection in non through-wall settings: We conducted extensive experiments (40 reflex, 54 normal, and 41 challenging normal events) over the three non through-wall areas of Fig. 7 (right). Table 2 (middle row) shows the results in these areas. As can be seen, our system can robustly detect 92.50% of reflexes in these areas, with a very small probability of false alarm of 1.06%.

The overall classification accuracy of our pipeline, over all the experiments, is shown in the last row of Table 2. As can be seen, Wi-Flex can correctly classify 90.15% (183/203) of the reflex events as reflex and 97.51% (314/322) of non-reflex events as non-reflex, resulting in a very small false alarm rate of 2.49%.

6.2 Performance Variability per Activity

Our experiments include several reflex and non-reflex activities as summarized in Table 1. As can be seen, the activities are diverse in each group and, as such, may induce different responses. In other words, some reflex activities may be more challenging to detect, while some non-reflex activities may have a higher chance of resembling reflex. Fig. 9 explicitly shows the probability of correct classification for 13 sample reflex events as well as the probability of false alarm for 19 non-reflex events. All the activities are performed through walls, in
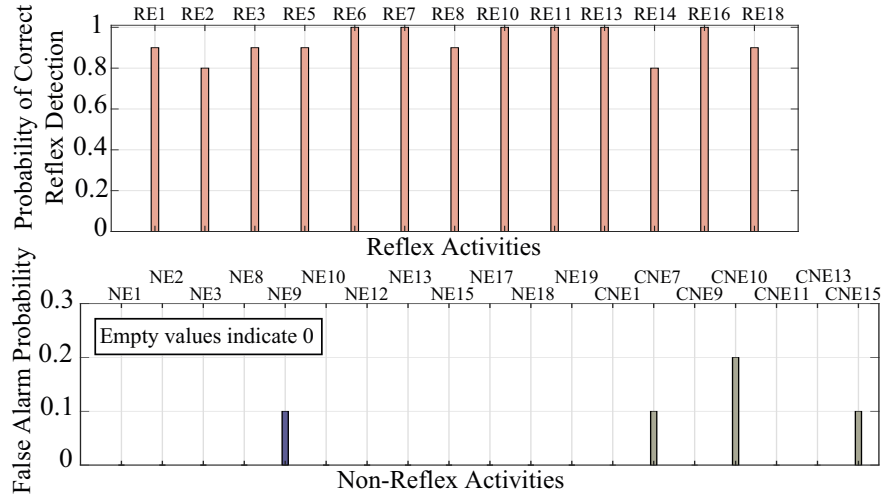
Fig. 9. Performance per activity (through walls) – (top) correct classification probability of 13 reflex activities, (bottom) false alarm probability of 19 non-reflex activities (12 normal, 7 challenging normal), averaged over 10 participants. See Table 1 for the description of activities.

the office area of Fig. 7 (left) and with 10 of our participants (i.e., each result is averaged over all 10). To show the impact of the activity on the performance, the ones with the most diverse performance are chosen. As can be seen, the performance can vary depending on the activity, as expected. For instance, unexpected shoulder contact event resulted in a correct reflex detection rate of 80%, while events such as an unexpected loud noise or the plastic spider jumping out of a box, resulted in 100% correct classification rate. On the other hand, in the non-reflex category, throwing an object resulted in the highest false alarm rate, due to its high-intensity nature. Overall, however, we can see that Wi-Flex can robustly detect reflexes across different activities.

6.3 Performance Variability per Individual

Different people can have different levels of reflex reactions. Furthermore, people can have different ways of performing a challenging normal event. We next study the impact of an individual on the reflex detection results. Fig. 10 shows the probability of correct reflex detection averaged over 13 reflex events, as well as the false alarm probability averaged over 19 non-reflex events (12 normal and 7 challenging normal), for 10 participants. As can been seen, the performance varies among the participants, as expected. However, our pipeline can still detect reflexes robustly in different individuals.

6.4 Impact of the Different Environments on the Performance

We next study the impact of different environments on the performance of Wi-flex. Table 3 shows Wi-flex performance in four different areas: office space, kitchen, living room and a semi-covered outdoor area. The areas further have different dimensions, allowing us to study the impact of the size of the area on the performance as well. As can be seen, Wi-flex performs robustly in all the areas. Nevertheless, we can see some performance variability over the areas, as they experience different levels of multipath. For instance, the kitchen and the office areas experience more clutter and thus more multipath, resulting in a slightly lower probability of correct classification, as compared to the semi-covered outdoor area, which experiences the least amount of multipath. Overall, however, the table confirms that Wi-flex is robust to environmental changes.
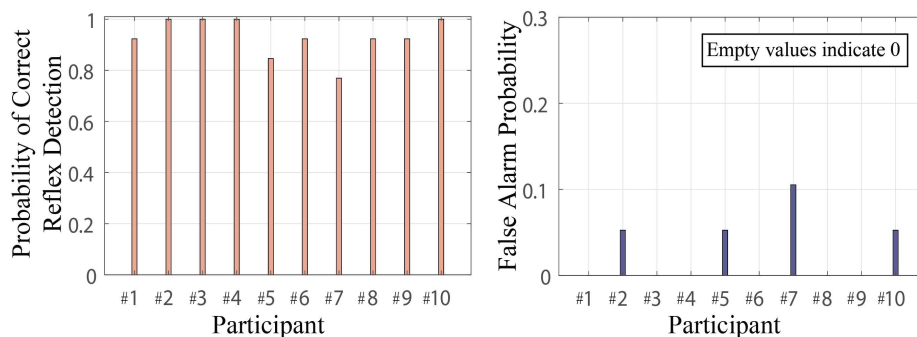
Fig. 10. Performance per individual (through walls) – (left) correct classification probability for 10 participants, each averaged over 13 reflex activities and (right) false alarm probability for 10 participants, each averaged over 19 non-reflex activities (12 normal, 7 challenging normal).
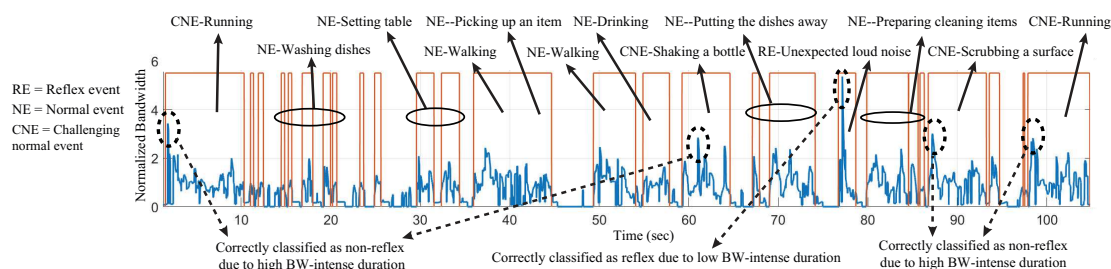

Fig. 11. Reflex Detection for a sequence of events in the kitchen area while a participant takes part in a series of activities including normal, challenging normal and reflex. Events are marked and reflex/non-reflex events are correctly classified.

6.5 Explicit Analysis of a Sequence of Events

In many of our experiments, an action of a participant that constitutes a particular type of event was naturally preceded/followed by an action that constitutes another type of action (e.g., a person grabbing something before having a reflex reaction). In previous parts, we focused on reporting performance for an individual event. In this part, we explicitly run experiments involving a sequence of events that includes all the three categories of events. Fig. 11 shows the normalized bandwidth signal for 105 seconds of WiFi data collected in the kitchen area of Fig. 7, while a participant takes part in a series of activities. The detected events are also shown in the figure. As can be seen, the events are diverse, embodying normal, challenging normal and reflex events, which are all correctly classified. In particular, we can see that the participant takes part in challenging normal events at 2, 63, 86, and 97 second marks, which all result in a high maximum normalized bandwidth, but are correctly classified as non-reflex due to their high bandwidth-intense duration. Finally, the participant has a reflex reaction to a sudden loud noise at 77 second mark, which is also correctly classified.

6.6 Reflex Detection with Multiple People

In some of our reflex-inducing events, we had to help set up the event, for instance by throwing a ball at the participant or by swinging a bat. While we tried to make sure that our impact on the received signal is minimal in such cases (e.g., we are far enough) in order to focus on the performance with one person, some of the reported results were already in a multiple-people setting. In this part, however, we explicitly show the performance when multiple people are present and simultaneously engaged in a variety of activities. For instance, multiple people may be all doing exercises, one having a reflex while others engaged in other activities, or in general all involved

Table 3. Impact of the environment on Wi-flex performance – As can be seen, Wi-flex performs robustly in different areas.

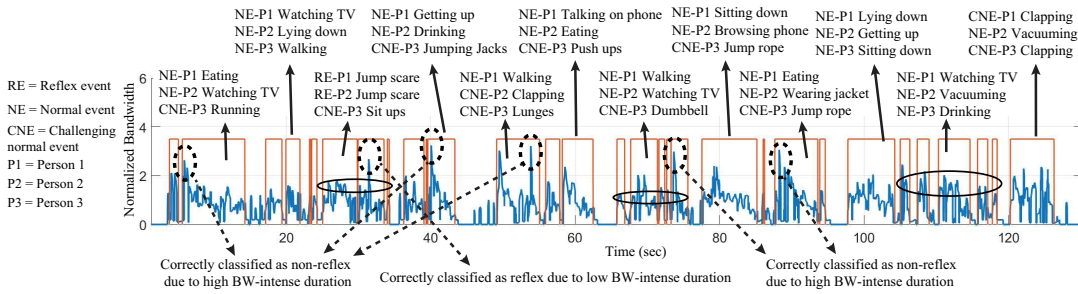| Experiment Location	Area Dimensions (width x length)	Reflex Events Correct Classification Prob.	Non-Reflex Events Correct Classification Prob.
Office Area (through-wall)	2.9m x 3.8m	89.57%	96.91%
Kitchen (non-through wall)	1.3m x 2.8m	90.90%	97.29%
Living Room (non-through wall)	3.5m x 4.1m	91.66%	100.00%
Semi-Covered Outdoor (non-through wall)	N/A	94.11%	100.00%

Fig. 12. Reflex Detection with Multiple People in the living room area while three people simultaneously engage in a series of activities, including normal, challenging normal and reflex. Events are marked on the figure and are correctly classified.

in any combination of reflex/normal/challenging normal events at the same time. Fig. 12 shows an example case where three people are engaged in a series of activities (marked on the figure) in the living room area of Fig. 7, while WiFi is collecting data for 129 seconds. As can be seen, at several instances, multiple people are doing a variety of activities at the same time, but reflex/non-reflex cases are all correctly classified. For instance, at 52 and 123 second marks, two people were both doing a challenging normal event, but were all classified as non-reflex. Furthermore, two people were having a reflex at 31 second mark (due to a video jump scare) while another person was engaged in a challenging normal event (exercising), but the reflex was correctly detected. Overall, the results show the potential of Wi-Flex for operation in multiple-people settings.

6.7 Goalkeeper-Catch with Varying Ball Speeds

We next show how our proposed two bandwidth-based parameters have the potential to provide quantifiable metrics in sports domain to measure an athlete's reaction response. More specifically, we run four goalkeeper-type experiments with varying incoming ball speeds, while WiFi is making measurements. We expect that for low ball speeds, the goalkeeper will not have a reflex, while fast incoming balls will induce a reflex. Fig. 13 shows the results for 4 speeds of 2.92, 4.37, 5.83, 7.78 m/s, thrown at the participant from a distance of 3.5 meters, and at directions that require a lateral jump. As can be seen, the catch event is classified as reflex once the ball speed reaches 7.78 m/s, while the first three lower speeds are classified as non-reflex as the ball speed was too low for the person to have a reflex-type response. The results are also consistent with the participant's description of their experience and show the potential of the proposed pipeline in sports domain by introducing new quantifiable metrics for reflex detection, using only cheap WiFi signals.

7 DISCUSSION AND FUTURE DIRECTIONS

We next discuss several aspects of the proposed work, and motivate future directions.
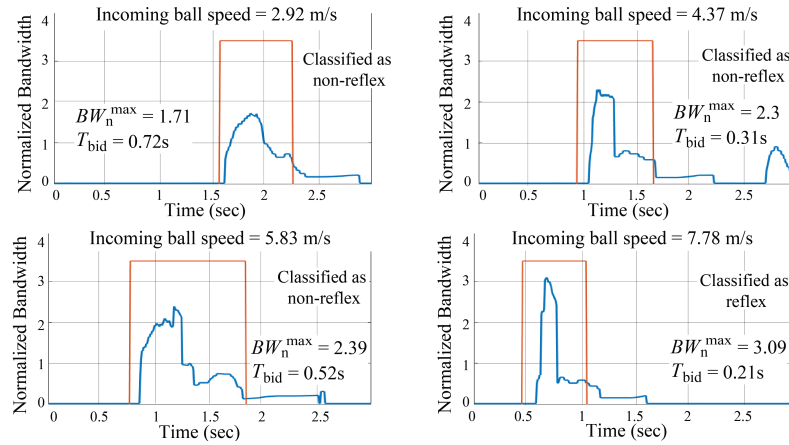
Fig. 13. Goalkeeper's response is classified by our system as the incoming ball speed increases. The results are consistent with the participant's description of their experience.

The need for two metrics: We established that we need two parameters to characterize a reflex. We next show the performance if only one parameter is used. If only normalized bandwidth is used, correct classification prob. becomes 69.43% for reflex and 82.29% for non-reflex events. On the other hand, if bandwidth-intense duration is only used, correct classification prob. becomes 73.89% for reflex and 76.08% for non-reflex events, showing the need for two parameters, after comparison with Table 2.

Bandwidth vs. Speed: We established that due to abrupt speed changes during a reflex, extracting speed-related parameters from spectrogram will not be as reliable (Remark 5), while the bandwidth can be easily evaluated. We then proposed two bandwidth-related parameters and showed how we can translate each video to the corresponding RF signal bandwidth, to generate a large RF bandwidth dataset and find the optimum 2D decision regions accordingly. We next show the performance if the speed was used instead. More specifically, we use the two speed-related parameters of Section 4.1 and find their optimum decision regions, using our video dataset, to be used during testing. This results in correct classification prob. of 72.90% for reflex events and 94.41% for non-reflex events, which is considerably worse than Table 2 (90.15% and 97.51% respectively), confirming that bandwidth is the right design choice.

Using both bandwidth and speed parameters: For the sake of completeness, we next show the performance if both bandwidth and speed parameters are utilized for the classification task. More specifically, consider the case where the two speed-related parameters of section 4.1 as well as the two bandwidth-related parameters of section 4.4 are used for classification (total of 4 parameters). We then find the corresponding 4D decision regions, using SVM and our assembled video dataset. During the operation, these four parameters are extracted from the received signal for classification. This results in the overall correct classification prob. of 80.29% for reflex events and 95.34% for non-reflex events (over all areas), which are still considerably worse than Table 2 (90.15% and 97.51% respectively). These results further confirm that speed-related parameters are not reliably extractable during a reflex due to its short-lived abrupt nature, making the introduced bandwidth-parameters the right choice.

Comparison with the state-of-the-art: As discussed earlier, there are work in the area of vision on detecting an unintentional act in a video, based on extensive training using machine learning. While not directly on reflexes, we nevertheless report their best performance for completeness. In [82], an overall classification accuracy of 81.6% is achieved. We, on the other hand, have an overall classification accuracy of 94.66%, without the use of RF training data or deep learning, showing the great potential of the proposed framework.
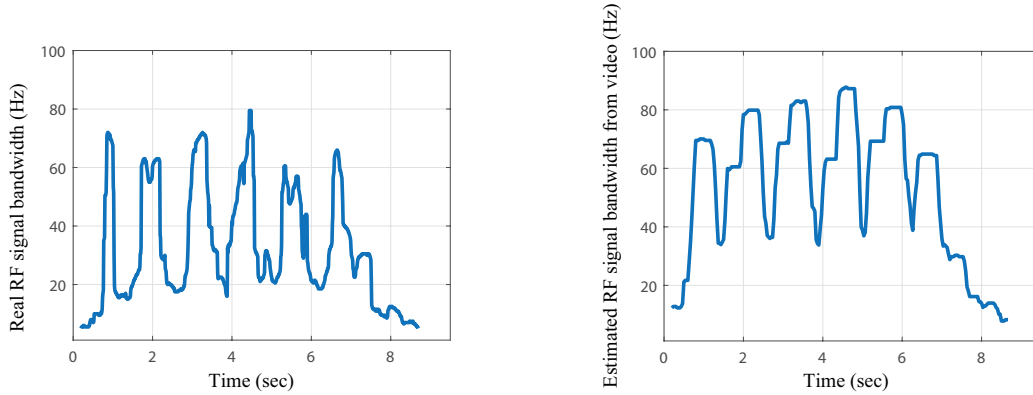
Fig. 14. Concurrent video-WiFi measurements during a jumping jack activity – (left) real bandwidth from the received WiFi signal measurement and (right) estimated bandwidth from the video signal via our proposed method.

Can existing state-of-the-art WiFi sensing work be extended for reflex detection? As discussed earlier, there is no existing work on reflex detection based on RF sensing. Activity recognition is the closest work in the area of RF sensing to the problem of interest in this paper (see sample reviewed papers in Section 2). Here we show how to extend existing RF sensing work for reflex detection and compare with the performance. Existing approaches are typically based on extracting speed-related features from the spectrogram. As such, we first need to find speed-related thresholds for reflex detection, something that is not done in the literature, necessitating our proposed vision-based approach of Section 4 for speed-related thresholds. Once we extract the speed-related thresholds, the previous discussion point titled "Bandwidth vs. Speed" explicitly showed the performance if speed-related parameters are instead utilized for reflex detection. It can be seen that the performance drops significantly (from 90.15% to 72.90% for instance). This analysis confirms that directly extending existing work in RF sensing to reflex detection will not work properly, due to the abrupt and short-lived nature of reflexes, motivating the proposed work of this paper.

Impact of TX/RX Locations: In Section 4.4 we discussed how the derived decision regions have little dependency on the location of the inserted TX/RX in the videos. As such, we simply fixed the TX/RX locations for all the videos. We next further confirm this by deriving decision regions when TX/RX are placed in each video such that the ψ of maximizing m is exactly two. This will then result in the correct classification prob. of 89.66% for reflex events and 97.83% for non-reflex events, which shows a negligible impact on our results, as compared to Table 2.

To see the impact of the TX/RX separation, we next run a series of experiments, with both reflex and non-reflex activities, with the TX/RX separation of 0.4, 0.6 and 0.75 m, in addition to the original distance of 1 meter. This results in a less than 5.67% difference in extracted feature values, confirming robustness to the distance between the transmitter and receiver. Once the distance becomes larger than 1 meter, the angle dependency can become more dominant, and should be taken into account as part of future work.

Concurrent video-WiFi measurements to explicitly confirm the proposed video to RF-bandwidth pipeline: In this paper, we have proposed how a video footage can be translated to the corresponding RF signal bandwidth, using tools from vision and based on a mathematical modeling inspired by Carson's rule (i.e., Eq. 3). This has then allowed us to generate a large synthetic RF bandwidth dataset, for both reflex and non-reflex events, from available online videos, and utilize them to derive key bandwidth-based metrics for reflex detection. While the successful performance of our extensive experimentation with real data implicitly confirms the validity of the proposed video to RF-bandwidth pipeline, we have also explicitly confirmed the validity of the proposed approach by collecting simultaneous video and WiFi measurements. For instance, Fig. 14 (left) shows the real

Table 4. Comparison of different classification methods (RBF SVM (utilized), kNN and MLP) in terms of accuracy and total computation time.

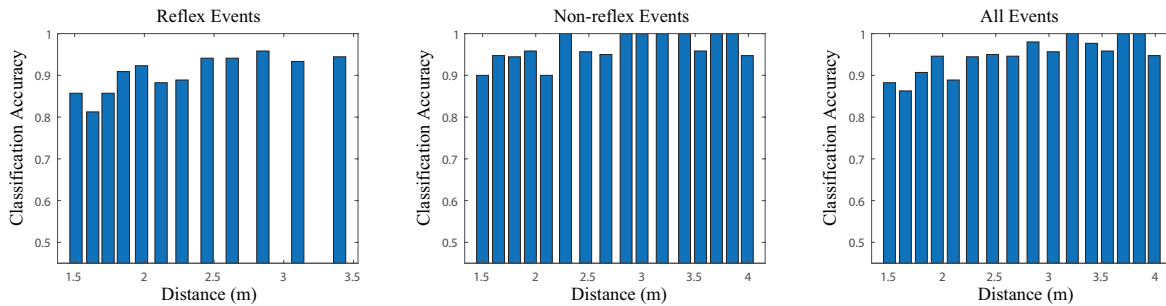| Classification Method	Reflex Events Correct Classification Prob.	Non-Reflex Events Correct Classification Prob.	Total Computation Time
kNN	88.66%	96.27%	2.351 s
MLP	90.64%	97.20%	36.342 s
RBF SVM (utilized)	90.15%	97.51%	0.758 s

Fig. 15. Impact of the distance between the wireless link and the subject for (left) reflex events, (middle) non-reflex events, and (right) all events.

bandwidth from a received WiFi signal measurement while a person is involved in a jumping jack activity, while the right figure shows the estimated bandwidth from the collected video via the proposed pipeline. As can be seen, the two plots match well.

Performance comparison with other classification methods: So far, we utilized the radial basis function (RBF) kernel SVM to come up with the 2D decision regions, once we translated the online videos to their corresponding two bandwidth-related parameters (Fig. 5). We next show the performance if other well-known classification methods are used to devise the decisions. More specifically, Table 4 shows the performance for 3 methods: RBF SVM (utilized), k-nearest neighbor (kNN), and multi-layer perceptron (MLP) feed forward neural network. As can be seen, the results are very comparable, with the perceptron having the slight edge in correct reflex classification and SVM in correct non-reflex classification. However, utilizing a neural network has a much higher total computation time, in comparison to simpler statistical learning methods such as SVM and kNN, and not necessary in this case. In other words, since we have shown that we only need two bandwith-related parameters to capture the essence of a reflex, a simple strategy such as SVM presents a solid choice for classification.

Impact of the distance between the wireless link and the subject: We next analyze the impact of the distance between the subject and the link on the performance. Fig. 15 provides the histogram of the classification accuracies for reflex and non-reflex events, based on their distance to the wireless link. We can see that if the subject is too close to the link, the classification accuracy slightly drops, as expected. Overall, however, we see that Wi-flex achieves a high accuracy for all the distances.

Execution Time: On an Intel Core i7-9700K processor, our pipeline takes 107 ms, on average, to process one second of data.

Beyond Reflex Detection: In this paper, we brought a foundational understanding to the impact of a reflex on an RF link and proposed a robust reflex detection system accordingly. Our bandwidth-related parameters can

also provide quantifiable metrics for characterizing the intensity of a reflex response, which is of importance for a number of applications, as discussed next.

Application in Smart Health: As discussed earlier, startle reflex assessment, in particular measuring reflex intensity, is important in the medical domain. Our proposed approach not only has the potential to detect a reflex but can provide quantifiable metrics for measuring the intensity of the reflex, and can thus be tested with/tailored to the relevant medical conditions as part of future work. It can further enable a smart home health system accordingly.

Application in Sports: As discussed, measuring the reflex reaction of athletes in certain sports/positions is important for training purposes. Wi-Flex can provide a reliable, yet cost-effective metric for reflex characterization, without a need for an elaborate setup, and can thus be tested in sports-domain as part of future work.

8 CONCLUSIONS

We considered reflex detection with WiFi, while considering startle reflexes as well as sport-type reflexes. We proposed that the maximum normalized bandwidth and bandwidth-intense duration of the received signal can successfully detect reflexes and robustly differentiate them from non-reflex events that involve intense body motions. We proposed an efficient way of translating the content of a video to the bandwidth of the corresponding received RF signal that would have been measured if there was a link near the event in the video, by drawing analogies to the classic work of Carson in the context of analog FM radios (i.e., Carson's Rule). This then allowed us to translate online reflex/non-reflex videos to an instant RF bandwidth dataset, and characterize 2D optimum reflex/non-reflex decision regions accordingly. We extensively tested the pipeline with 203 reflex and 322 non-reflex events, over four areas (including several through-wall ones), and with 15 participants, achieving a correct reflex detection rate of 90.15% and a false alarm rate of 2.49%. We further showed reflex detection results with multiple people, and tested several different aspects of the work. Finally, we conducted experiments to show the potential of our approach for providing cost-effective and quantifiable metrics in sports domain.

ACKNOWLEDGMENTS

REFERENCES

[1] 2016. RT-Fall: A real-time and contactless fall detection system with commodity WiFi devices, author=Wang, Hao and Zhang, Daqing and Wang, Yasha and Ma, Junyi and Wang, Yuxiang and Li, Shengjie. *IEEE Transactions on Mobile Computing* 16, 2 (2016), 511–526.

[2] Heba Abdelnasser, Moustafa Youssef, and Khaled A Harras. 2015. Wigest: A ubiquitous wifi-based gesture recognition system. In *2015 IEEE conference on computer communications (INFOCOM)*. IEEE, 1472–1480.

[3] Taslima Akter, Bryan Dosono, Tousif Ahmed, Apu Kapadia, and Bryan Semaan. 2020. " I am uncomfortable sharing what I can't see": Privacy Concerns of the Visually Impaired with Camera Based Assistive Applications. In *29th USENIX Security Symposium (USENIX Security 20)*. 1929–1948.

[4] Diana Armbruster, Tobias Grage, Clemens Kirschbaum, and Alexander Strobel. 2018. Processing emotions: Effects of menstrual cycle phase and premenstrual symptoms on the startle reflex, facial EMG and heart rate. *Behavioural brain research* 351 (2018), 178–187.

[5] Yunhao Bai and Xiaorui Wang. 2020. CARIN: Wireless CSI-based driver activity recognition under the interference of passengers. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–28.

[6] Mirte J Bakker, J Gert Van Dijk, Arn MJM van den Maagdenberg, and Marina AJ Tijssen. 2006. Startle syndromes. *The Lancet Neurology* 5, 6 (2006), 513–524.

[7] Georgios Balogiannis, Xanthi Stachtea, Dido Yova, and Charalabos Papageorgiou. 2020. A novel optical non-intrusive method for measuring acoustic startle reflex in humans. *Physiological Measurement* 41, 6 (2020), 065003.

[8] David L Braff, Christian Grillon, and Mark A Geyer. 1992. Gating and habituation of the startle reflex in schizophrenic patients. *Archives of general psychiatry* 49, 3 (1992), 206–215.

[9] P Brown, BL Day, JC Rothwell, PD Thompson, and CD Marsden. 1991. The effect of posture on the normal and pathological auditory startle reflex. *Journal of Neurology, Neurosurgery & Psychiatry* 54, 10 (1991), 892–897.

[10] P Brown, JC Rothwell, PD Thompson, TC Britton, BL Day, and CD Marsden. 1991. New observations on the normal auditory startle reflex in man. *Brain* 114, 4 (1991), 1891–1902.

[11] P Brown, JC Rothwell, PD Thompson, TC Britton, BL Day, and CD Marsden. 1991. The hyperekplexias and their relationship to the normal startle reflex. *Brain* 114, 4 (1991), 1903–1928.

[12] Daniel Burke, Susan Linder, Joshua Hirsch, Tanujit Dey, Daniel Kana, Shannon Ringenbach, David Schindler, and Jay Alberts. 2017. Characterizing information processing with a mobile device: measurement of simple and choice reaction time. *Assessment* 24, 7 (2017), 885–895.

[13] Hong Cai, Belal Korany, Chitra R Karanam, and Yasamin Mostofi. 2020. Teaching RF to Sense Without RF Training Measurements. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–22.

[14] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019).

[15] J.R. Carson. 1922. Notes on the Theory of Modulation. *Proceedings of the Institute of Radio Engineers* 10, 1 (1922), 57–64. https://doi.org/10.1109/JRPROC.1922.219793

[16] Carlos Castellar, Francisco Pradas, Luis Carrasco, Ana De La Torre, and José Antonio González-Jurado. 2019. Analysis of reaction time and lateral displacements in national level table tennis players: are they predictive of sport performance? *International Journal of Performance Analysis in Sport* 19, 4 (2019), 467–477.

[17] Emmanuel Cecchet, Amrita Acharya, Tergel Molom-Ochir, Amee Trivedi, and Prashant Shenoy. 2020. WiFiMon: A mobility analytics platform for building occupancy monitoring and contact tracing using wifi sensing. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 792–793.

[18] Zhenghua Chen, Chaoyang Jiang, and Lihua Xie. 2018. Building occupancy estimation and detection: A review. *Energy and Buildings* 169 (2018), 260–270.

[19] Dipayan Das, Saumik Bhattacharya, Umapada Pal, and Sukalpa Chanda. 2021. PLSM: A Parallelized Liquid State Machine for Unintentional Action Detection. *arXiv preprint arXiv:2105.09909* (2021).

[20] Carlo De Luca. 2006. Electromyography. *Encyclopedia of medical devices and instrumentation* (2006).

[21] Jianyang Ding and Yong Wang. 2020. A WiFi-based smart home fall detection system using recurrent neural network. *IEEE Transactions on Consumer Electronics* 66, 4 (2020), 308–317.

[22] Neil A Dodgson. 2004. Variation and extrema of human interpupillary distance. In *Stereoscopic displays and virtual reality systems XI*, Vol. 5291. International Society for Optics and Photonics, 36–46.

[23] Shihong Duan, Tianqing Yu, and Jie He. 2018. Widriver: Driver activity recognition system based on wifi csi. *International Journal of Wireless Information Networks* 25, 2 (2018), 146–156.

[24] Shikha Dubey, Abhijeet Boragule, and Moongu Jeon. 2019. 3d resnet with ranking loss function for abnormal activity detection in videos. In *2019 International Conference on Control, Automation and Information Sciences (ICCAIS)*. IEEE, 1–6.

[25] Dave Epstein, Boyuan Chen, and Carl. Vondrick. 2019. Oops! Predicting Unintentional Action in Video. *arXiv preprint arXiv:1911.11206* (2019).

[26] Christian Grillon. 2002. Startle reactivity and anxiety disorders: aversive conditioning, context, and neurobiology. *Biological psychiatry* 52, 10 (2002), 958–975.

[27] Christian Grillon and Johanna Baas. 2003. A review of the modulation of the startle reflex by affective states and its application in psychiatry. *Clinical neurophysiology* 114, 9 (2003), 1557–1579.

[28] Xiaonan Guo, Bo Liu, Cong Shi, Hongbo Liu, Yingying Chen, and Mooi Choo Chuah. 2017. WiFi-Enabled Smart Human Dynamics Monitoring. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems* (Delft, Netherlands) *(SenSys '17)*. Association for Computing Machinery, New York, NY, USA, Article 16, 13 pages. https://doi.org/10.1145/3131672.3131692

[29] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. 2011. Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM computer communication review* 41, 1 (2011), 53–53.

[30] Keita Hara, Kazuaki Nakamura, and Noboru Babaguchi. 2015. Temporal spotting of human actions from videos containing actor's unintentional motions. In *2015 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1–6.

[31] Robert J Harvey, Maya Topf, Kirsten Harvey, and Mark I Rees. 2008. The genetics of hyperekplexia: more than startle! *Trends in Genetics* 24, 9 (2008), 439–447.

[32] Jing He and Wei Yang. 2022. IMar: Multi-user Continuous Action Recognition with WiFi Signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–27.

[33] Wenfeng He, Kaishun Wu, Yongpan Zou, and Zhong Ming. 2015. WiG: WiFi-based gesture recognition system. In *2015 24th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 1–7.

[34] Zahra Jafari, Bryan E Kolb, and Majid H Mohajerani. 2020. Prepulse inhibition of the acoustic startle reflex and P50 gating in aging and alzheimer's disease. *Ageing Research Reviews* 59 (2020), 101028.

[35] C. R. Karanam, B. Korany, and Y. Mostofi. 2019. Tracking from One Side: Multi-person Passive Tracking with WiFi Magnitude Measurements. In *Proceedings of ACM/IEEE International Conference on Information Processing in Sensor Networks.* 181–192.

[36] Usman Mahmood Khan, Zain Kabir, and Syed Ali Hassan. 2017. Wireless health monitoring using passive WiFi sensing. In *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*. IEEE, 1771–1776.

[37] Usman Mahmood Khan, Zain Kabir, Syed Ali Hassan, and Syed Hassan Ahmed. 2017. A deep learning framework using passive WiFi sensing for respiration monitoring. In *GLOBECOM 2017-2017 IEEE Global Communications Conference.* IEEE, 1–6.

[38] Floris Klumpers, Barak Morgan, David Terburg, Dan J Stein, and Jack van Honk. 2015. Impaired acquisition of classically conditioned fear-potentiated startle reflexes in humans with focal bilateral basolateral amygdala damage. *Social cognitive and affective neuroscience* 10, 9 (2015), 1161–1168.

[39] Michael Koch. 1999. The neurobiology of startle. *Progress in neurobiology* 59, 2 (1999), 107–128.

[40] Markus Kofler, Jörg Müller, Laura Reggiani, and Josep Valls-Solé. 2001. Influence of age on auditory startle responses in humans. *Neuroscience letters* 307, 2 (2001), 65–68.

[41] Belal Korany and Yasamin Mostofi. 2021. Counting a stationary crowd using off-the-shelf wifi. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services.* 202–214.

[42] Belal Korany and Yasamin Mostofi. 2022. Nocturnal Seizure Detection Using Off-the-Shelf WiFi. *IEEE Internet of Things Journal* 9, 9 (2022), 6996–7008. https://doi.org/10.1109/JIOT.2021.3115505

[43] Min-Seok Lee, Jong-Gwan Lim, Ki-Ru Park, and Dong-Soo Kwon. 2009. Unsupervised clustering for abnormality detection based on the tri-axial accelerometer. In *2009 ICCAS-SICE*. IEEE, 134–137.

[44] Jinyang Li, Zhenyu Li, Gareth Tyson, and Gaogang Xie. 2020. Your privilege gives your privacy away: An analysis of a home security camera service. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 387–396.

[45] Shengjie Li, Xiang Li, Qin Lv, Guiyu Tian, and Daqing Zhang. 2018. WiFit: Ubiquitous bodyweight exercise monitoring with commodity wi-fi devices. In *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, 530–537.

[46] Weiyao Lin, Ming-Ting Sun, Radha Poovandran, and Zhengyou Zhang. 2008. Human activity recognition for video surveillance. In *2008 IEEE international symposium on circuits and systems*. IEEE, 2737–2740.

[47] Katja Ludewig, Stephan Ludewig, Antonia Seitz, Martina Obrist, Mark A Geyer, and Franz X Vollenweider. 2003. The acoustic startle reflex and its modulation: effects of age and gender in humans. *Biological psychology* 63, 3 (2003), 311–323.

[48] Xiaomu Luo, Huoyuan Tan, Qiuju Guan, Tong Liu, Hankz Hankui Zhuo, and Baihua Shen. 2016. Abnormal activity detection using pyroelectric infrared sensors. *Sensors* 16, 6 (2016), 822.

[49] Larissa N Maslova, Veta V Bulygina, and Nina K Popova. 2002. Immediate and long-lasting effects of chronic stress in the prepubertal age on the startle reflex. *Physiology & behavior* 75, 1-2 (2002), 217–225.

[50] Anna Marie Medina, Veronica Y Mejia, Anne M Schell, Michael E Dawson, and Gayla Margolin. 2001. Startle reactivity and PTSD symptoms in a community sample of women. *Psychiatry research* 101, 2 (2001), 157–169.

[51] KM Miller, MS Okun, M Marsiske, EB Fennell, and D Bowers. 2009. Startle reflex hyporeactivity in Parkinson's disease: an emotion-specific or arousal-modulated deficit? *Neuropsychologia* 47, 8-9 (2009), 1917–1927.

[52] CA Morgan Iii, Christian Grillon, Steven M Southwick, Michael Davis, and Dennis S Charney. 1995. Fear-potentiated startle in posttraumatic stress disorder. *Biological psychiatry* 38, 6 (1995), 378–385.

[53] Milos Mudric, Ivan Cuk, Aleksandar Nedeljkovic, Srecko Jovanovic, and Slobodan Jaric. 2015. Evaluation of Video-based method for the measurement of reaction time in specific sport situation. *International Journal of Performance Analysis in Sport* 15, 3 (2015), 1077–1089.

[54] Ronald Mutegeki and Dong Seog Han. 2020. A CNN-LSTM approach to human activity recognition. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. IEEE, 362–366.

[55] Marco Obetko, Matej Babic, and Pavol Peráček. 2019. Changes in disjunctive reaction time of soccer goalkeepers in selected training load zones. *Journal of Physical Education and Sport* 19 (2019), 420–426.

[56] Edward M Ornitz and Robert S Pynoos. 1989. Startle modulation in children with posttraumatic stress disorder. *The American Journal of Psychiatry* (1989).

[57] Sameera Palipana, David Rojas, Piyush Agrawal, and Dirk Pesch. 2018. FallDeFi: Ubiquitous fall detection using commodity Wi-Fi devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–25.

[58] Anurag Pallaprolu, Belal Korany, and Yasamin Mostofi. 2022. Wiffract: a new foundation for RF imaging via edge tracing. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking.* 255–267.

[59] Cuong Pham, Linh Nguyen, Anh Nguyen, Ngon Nguyen, and Van-Toi Nguyen. 2021. Combining skeleton and accelerometer data for human fine-grained activity recognition and abnormal behaviour detection with deep temporal convolutional networks. *Multimedia Tools and Applications* 80, 19 (2021), 28919–28940.

[60] Eleonora Poli and Alessandro Angrilli. 2015. Greater general startle reflex is associated with greater anxiety levels: a correlational study on 111 young women. *Frontiers in behavioral neuroscience* 9 (2015), 10.

[61] David Fernando Ramirez-Moreno and Terrence Joseph Sejnowski. 2012. A computational model for the modulation of the prepulse inhibition of the acoustic startle reflex. *Biological cybernetics* 106, 3 (2012), 169–176.

[62] William J Ray, Christine Molnar, Deane Aikins, Alissa Yamasaki, Michelle G Newman, Louis Castonguay, and Thomas D Borkovec. 2009. Startle response in generalized anxiety disorder. *Depression and anxiety* 26, 2 (2009), 147–154.

[63] Larry Schneck, Jerome Maisel, and Bruno W Volk. 1964. The startle response and serum enzyme profile in early detection of Tay-Sachs' disease. *The Journal of pediatrics* 65, 5 (1964), 749–756.

[64] Elahe Soltanaghaei, Avinash Kalyanaraman, and Kamin Whitehouse. 2018. Multipath triangulation: Decimeter-level wifi localization and orientation with a single unaided receiver. In *Proceedings of the 16th annual international conference on mobile systems, applications, and services.* 376–388.

[65] Waqas Sultani, Chen Chen, and Mubarak Shah. 2018. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 6479–6488.

[66] Kailai Sun, Qianchuan Zhao, and Jianhong Zou. 2020. A review of building occupancy measurement systems. *Energy and Buildings* 216 (2020), 109965.

[67] D.G. Tucker. 1970. The invention of frequency modulation in 1902. *The Radio and Electronic Engineer* 40 (1970), 33–37.

[68] Akinori Ueki, Kyoko Goto, Noriko Sato, Hiroyuki Iso, and Yoshio Morita. 2006. Prepulse inhibition of acoustic startle response in mild cognitive impairment and mild dementia of Alzheimer type. *Psychiatry and clinical neurosciences* 60, 1 (2006), 55–62.

[69] Stavros Vakalis, Liang Gong, and Jeffrey A Nanzer. 2019. Imaging with WiFi. *IEEE Access* 7 (2019), 28616–28624.

[70] Raghav H Venkatnarayan, Griffin Page, and Muhammad Shahzad. 2018. Multi-user gesture recognition using WiFi. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services.* 401–413.

[71] Scott R Vrana, John A Constantine, and James S Westman. 1992. Startle reflex modification as an outcome measure in the treatment of phobia: Two case studies. *Behavioral Assessment* (1992).

[72] Fangxin Wang, Wei Gong, and Jiangchuan Liu. 2018. On spatial diversity in WiFi-based human activity recognition: A deep learning-based approach. *IEEE Internet of Things Journal* 6, 2 (2018), 2035–2047.

[73] Wei Wang, Alex X Liu, and Muhammad Shahzad. 2016. Gait recognition using wifi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* 363–373.

[74] Wei Wang, Alex X Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2015. Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the 21st annual international conference on mobile computing and networking.* 65–76.

[75] Allison M Waters, Michelle G Craske, R Lindsey Bergman, Bruce D Naliboff, Hideki Negoro, and Edward M Ornitz. 2008. Developmental changes in startle reactivity in school-age children at risk for and with actual anxiety disorder. *International Journal of Psychophysiology* 70, 3 (2008), 158–164.

[76] Eric W Weisstein. 2002. Gaussian function. *https://mathworld. wolfram. com/* (2002).

[77] DE Wilkins, M Hallett, and MM Wess. 1986. Audiogenic startle reflex of man and its relationship to startle syndromes: a review. *Brain* 109, 3 (1986), 561–573.

[78] Chenshu Wu, Feng Zhang, Yuqian Hu, and K. J. Ray Liu. 2021. GaitWay: Monitoring and Recognizing Gait Speed Through the Walls. *IEEE Transactions on Mobile Computing* 20, 6 (2021), 2186–2199. https://doi.org/10.1109/TMC.2020.2975158

[79] Dan Wu, Ruiyang Gao, Youwei Zeng, Jinyi Liu, Leye Wang, Tao Gu, and Daqing Zhang. 2020. FingerDraw: Sub-wavelength level finger motion tracking with WiFi signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–27.

[80] Kun Xia, Jianguang Huang, and Hanyu Wang. 2020. LSTM-CNN architecture for human activity recognition. *IEEE Access* 8 (2020), 56855–56866.

[81] Jinglin Xu, Guangyi Chen, Jiwen Lu, and Jie Zhou. 2022. Unintentional action localization via counterfactual examples. *IEEE transactions on image processing* (2022).

[82] Jinglin Xu, Guangyi Chen, Nuoxing Zhou, Wei-Shi Zheng, and Jiwen Lu. 2022. Probabilistic Temporal Modeling for Unintentional Action Localization. *IEEE Transactions on Image Processing* (2022).

[83] Huan Yan, Yong Zhang, Yujie Wang, and Kangle Xu. 2019. WiAct: A passive WiFi-based human activity recognition system. *IEEE Sensors Journal* 20, 1 (2019), 296–305.

[84] Xu Yang, Fangyuan Xiong, Yuan Shao, and Qiang Niu. 2018. WmFall: WiFi-based multistage fall detection with channel state information. *International Journal of Distributed Sensor Networks* 14, 10 (2018), 1550147718805718.

[85] Bohan Yu, Yuxiang Wang, Kai Niu, Youwei Zeng, Tao Gu, Leye Wang, Cuntai Guan, and Daqing Zhang. 2021. WiFi-Sleep: Sleep Stage Monitoring Using Commodity Wi-Fi Devices. *IEEE Internet of Things Journal* 8, 18 (2021), 13900–13913. https://doi.org/10.1109/JIOT.2021.3068798

[86] Erika Zemková. 2016. Differential contribution of reaction time and movement velocity to the agility performance reflects sport-specific demands. *Human movement* 17, 2 (2016), 94–101.

[87] Yunze Zeng, Parth H Pathak, and Prasant Mohapatra. 2016. WiWho: WiFi-based person identification in smart spaces. In *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN).* IEEE, 1–12.

[88] Junfeng Zhang, Meng Wang, Baoyu Wei, Jiangwei Shi, and Tao Yu. 2022. Research Progress in the Study of Startle Reflex to Disease States. *Neuropsychiatric Disease and Treatment* 18 (2022), 427.

[89] Mingmin Zhao, Yonglong Tian, Hang Zhao, Mohammad Abu Alsheikh, Tianhong Li, Rumen Hristov, Zachary Kabelac, Dina Katabi, and Antonio Torralba. 2018. RF-based 3D skeletons. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication.* 267–281.

[90] Wei Zhong, Kai He, and Lianlin Li. 2019. Through-the-wall imaging exploiting 2.4 GHz commodity WiFi. *arXiv preprint arXiv:1903.03895* (2019).

[91] Dali Zhu, Na Pang, Gang Li, and Shaowu Liu. 2017. NotiFi: A ubiquitous WiFi-based abnormal activity detection system. In *2017 International Joint Conference on Neural Networks (IJCNN).* 1766–1773. https://doi.org/10.1109/IJCNN.2017.7966064

[92] Yiwei Zhuo, Hongzi Zhu, and Hua Xue. 2016. Identifying a New Non-Linear CSI Phase Measurement Error with Commodity WiFi Devices. In *2016 IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS).* 72–79. https://doi.org/10.1109/ICPADS.2016.0019

[93] Han Zou, Yuxun Zhou, Jianfei Yang, and Costas J Spanos. 2018. Device-free occupancy detection and crowd counting in smart buildings with WiFi-enabled IoT. *Energy and Buildings* 174 (2018), 309–322.

A   APPROXIMATED FORM FOR THE PHASE DIFFERENCE AND RECEIVED POWER SIGNALS

Both received power and phase difference signals of the baseband equivalent signal of Eq. 1 will have the form presented in Eq. 2, albeit with different coefficients for A_m and $\Delta\mu_m$, as recently shown in [42]. Here we briefly summarize the proof. Consider Eq. 1, it is easy to confirm that we have the following squared magnitude signal (after DC removal):

$$|c(t)|^2 = c(t)c^*(t) = A_m cos(\frac{2\pi}{\lambda}\psi \int v(t)dt + \Delta\mu_m) \tag{4}$$

where $A_m = 2\alpha_s\alpha_m$ and $\Delta\mu_m = \mu_m - \mu_s$ is the initial phase difference between the reflected path and the direct path.

Next, consider the phase difference of the signal of Eq. 1. For ease of computation, and without loss of generality, lets consider the phase of the scaled signal $e^{-j\mu_s}c(t)/\alpha_s$. It can be easily confirmed that the phase of this scaled signal can be approximated as,

$$\angle c(t) \approx \frac{\alpha_m}{\alpha_s}sin(\frac{2\pi}{\lambda}\psi d(t) + \Delta\mu_m) \tag{5}$$

when the direct path is stronger than the reflect path i.e. $\frac{\alpha_m}{\alpha_s} \ll 1$. Then, the phase difference between receivers i and j (e.g., different antennas of a WiFi card) can be written as

$$s_p(t) = A_p cos(\frac{2\pi}{\lambda}\psi \int v(t)dt + \Delta\mu_p) \tag{6}$$

where $A_p = 2(\alpha_m/\alpha_s)sin(1/2(\Delta\mu_{m,i} - \Delta\mu_{m,j}))$ and $\Delta\mu_p = 1/2(\Delta\mu_{m,i} + \Delta\mu_{m,j})$.

Equations 4 and 6 show that received signal power and phase difference share the same signal form, albeit with different coefficients. As such, the proposed approach of this paper is valid for both signals. During our experiments, we found the phase difference signal to be more robust to hardware noise and as such, we used phase difference in our experiments.