Tracking Synchronous Gestures with WiFi

Zimu Zhou*†, Zheng Yang*, Kun Qian*, Chenshu Wu*, Longfei Shangguan‡, Han Xu§ and Yunhao Liu*
*School of Software and TNList, Tsinghua University
†Computer Engineering and Networks Laboratory, ETH Zurich
‡Princeton University
§CSE, Hong Kong University of Science and Technology
zhou.zimu@tik.ee.ethz.ch, {yang, qian, wu}@greenorbs.com, longfeis@cs.princeton.edu {xu, yunhao}@greenorbs.com

***Abstract*—Tracking synchronous gestures benefits various applications in team sport training, group interactive gaming, and computer-supported cooperative working. We present *WiSync*, a ubiquitous synchronous gesture tracking scheme on a single WiFi link. It reuses the pervasive WiFi networks, imposes no wearable sensors on users and operates in Non-Line-Of-Sight (NLOS) propagation. The idea is that with synchronous gestures from multiple users, the received signals exhibit characteristics as if a single user were performing a different gesture, while in the case of asynchronous gestures, they demonstrate chaotic patterns. To discern synchronous and asynchronous gestures without gesture recognition, we harness their differences in the periodicity during gesture repetitions. We prototype *WiSync* on commodity WiFi infrastructures and evaluate it in two indoor environments. Experimental results show that *WiSync* achieves a balanced synchronous gesture identification accuracy of 90.43% for 2 users performing 8 types of gestures. *WiSync* is also robust to user orientations and can scale to 4 users while retaining a satisfactory performance.**

I. Introduction

Synchronous gestures are common in everyday life and have been ubiquitously utilized in teamwork training, fitness and entertainment. Various team sports such as rowing, require high movement consistency and synchrony among teammates. Multi-player interactive games are increasingly adding cooperation modes to enhance group participation. The Unisoft Just Dance games [1], for example, support a four-player dance crew mode where players dance synchronously to the tempo. Many Computer-Supported Cooperative Work (CSCW) applications promote explicit multi-user coordination to improve social skills [2]. For instance, cooperative gesturing [3] takes simultaneous gestures from multiple users as contributing to one combined command to enrich user experience and increase group cohesion. In addition, tightly synchronous movements are often an eye-catching element in large-scale collective performances such as marching soldiers and dancing groups.

A key enabler for ubiquitous synchronous gesture tracking is to non-invasively monitor gestures of multiple users with minimal infrastructure deployment. Existing gesture monitoring solutions vary in infrastructures [4]–[7]. Wearable gesture tracking solutions are common [7] and commercially available [8]. Yet they require per-person attached devices, thus fail to support multiple users without extra inter-device coordination and communication. Xbox Kinect [5] and Leap Motion [6] enable in-air fine-grained multi-user gesture tracking, yet require a clear LOS view to operate. Alternatively, researchers have explored wireless signals for whole-home activity recognition and tracking. Some leverage radar-like processing for gesture sensing on dedicated software radios [9], [10], while others exploit machine learning to enable activity recognition with commodity WiFi devices [11]–[14]. However, these efforts mostly focus on *single-user* gesture recognition, leaving *multi-user* support largely unexplored.

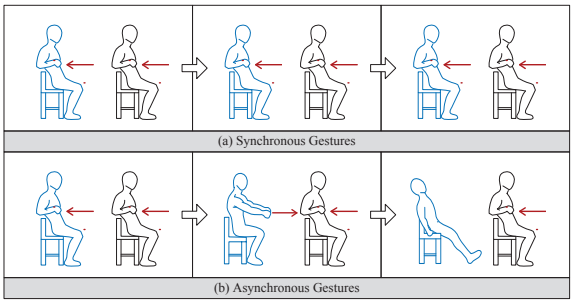


Fig. 1. An illustration of 2 users performing (a) synchronous and (b) asynchronous seated rows.

This paper seeks to advance the design space of wireless gesture sensing by asking the following question: *is it possible to track synchronous gestures from multiple users with a single wireless link on commercial WiFi devices?* The critical insight is that the impact of gestures from multiple users is superposing in the received signals and exhibit distinctive temporal patterns for synchronous and asynchronous gestures. In the case of synchronous gestures, the received signals exhibit characteristics as if a single user were performing a (different) gesture, while those in case of asynchronous gestures tend to demonstrate more chaotic patterns. To capture the subtle received signal dynamics induced by multi-user gestures, we take advantage of the Channel State Information (CSI) [15] on commodity WiFi devices. CSI conveys fine-grained information for gesture tracking. Gestures from multiple users, if unsynchronized, introduce diverse and random time delays and phase shifts of the propagation paths. The superposition of these randomly delayed and shifted paths leads to fast and dramatic CSI fluctuations. We thus explore the idea of assessing the *randomness level* in CSI to infer synchronous gestures from multiple users.

Further, we aim to differ synchronous and asynchronous gestures without gesture recognition to avoid location and user dependency. The idea is to harness the CSI patterns *among gesture repetitions*. Note that synchronous gestures are often required to repeat multiple times in many team sports such as rowing, aerobic exercises, and choreographed dancing,

978-1-5090-2279-3/16/$31.00 ©2016 IEEE

especially during the teaching and training stages of these group activities. Take Fig. 1 as an illustration. The upper figure shows two experienced users synchronously performing three repetitions of *seated rows (thinking of the double sculls in the Olympic Games)* and we expect *repetitive* patterns in the measured CSIs. Conversely, the lower figure depicts an inexperienced user who fails to follow the tempo of the other user, and who eventually gives up. Thus the measured CSIs may contain more irregularity. We utilize such inherent differences in *periodicity* to discern synchronous and asynchronous gesture repetitions without prior gesture recognition.

To deliver the above intuition into a practical solution, triple challenges arise. (1) *How to effectively process CSI data?* Raw CSIs can be noisy due to protocol parameter changes and environmental interference. We leverage outlier removal schemes and wavelet-based denoising to suppress noise in raw CSIs, while retaining the sharp edges introduced by potential gestures. (2) *How to quantify the subtle dynamics in synchronous and asynchronous gesture repetitions?* We exploit recurrence quantification analysis, a non-linear time-series analysis approach [16], to quantify the dynamics in synchronous and asynchronous gesture repetition. (3) *How to harness the frequency diversity in CSI to improve identification robustness?* Due to multipath fading, some CSI subcarriers might be severely attenuated, thus less reliable. On the other hand, some subcarriers may be more sensitive to human gestures than others. Therefore we select the more reliable and informative subcarriers to improve robustness.

We propose *WiSync*, a single-link non-invasive synchronous gesture tracking scheme on a commodity WiFi infrastructure. It works by collecting CSIs when users perform gestures, and analyzing the periodicity and dynamics in the measured CSI sequences to infer synchronous or asynchronous gesture repetition. We implement a prototype of *WiSync* on a mini PC with off-the-shelf Network Interface Cards (NICs) and a commodity wireless router. Experimental results show that *WiSync* achieves an overall balanced synchronous and asynchronous gesture identification accuracy of 90.43% for two users performing eight types of gestures in two indoor environments including LOS, None-Line-Of-Sight (NLOS) and through-wall propagation. *WiSync* is also robust to different relative user orientations and repetition counts and can scale to four users while retaining a satisfactory performance.

The main contributions of this paper are as follows:

- We propose a single-link non-invasive synchronous gesture tracking scheme. It serves as an early step towards general multi-user support for wireless gesture sensing, which will stimulate a new range of teamwork training, cooperative gaming and social interaction applications.
- We design a set of techniques to discern synchronous and asynchronous gesture repetitions without gesture recognition. *WiSync* involves little training overhead and is less location-dependent. It can also be combined with existing wireless gesture recognition solutions to enable fine-grained and automatic activity assessment.
- We prototype *WiSync* with commercial WiFi devices and validate its effectiveness in two indoor environments. Experiments show that *WiSync* achieves synchronous and asynchronous identification accuracies of 90%.

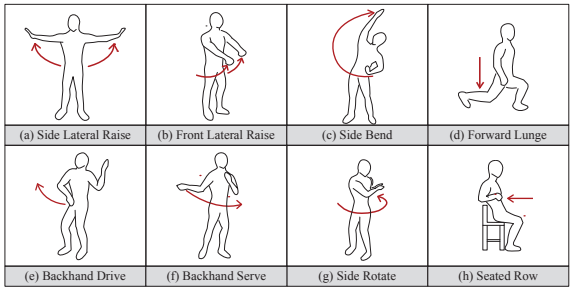


Fig. 2. Sketches of the eight tested gestures.

The subsequent sections begin with measurements and observations, followed by the design and evaluation of *WiSync*. Finally we review related work and conclude this work.

II. Measurements and Observations

In this section, we conduct measurements to show the feasibility to detect synchronous gestures using WiFi signals.

A. Measurements

As illustrative measurements, we deploy a three-meter WiFi link in an office environment. We use a commercial wireless router as the transmitter and a mini PC with Intel WiFi Link 5300 NIC as the receiver to collect CSIs [15]. The transmitter and the receiver are placed 0.8m above the floor.

We conduct the following measurements. (1) One user performs eight workout and gaming gestures (Fig. 2) following the tempo generated by a metronome application. The tempo is set to 40 beats per second (bpm) and each gesture is performed five repetitions. A smartphone is mounted on the user's arm to collect acceleration during his movements via an acceleration API. The clocks of the mini PC and the smartphone are synchronized before the measurements. (2) Two users perform the same gestures as Fig. 2. Each user has an arm-mounted smartphone to record acceleration. To simulate synchronous gestures, the two users listen to the tempos broadcast by the metronome application. For asynchronous gestures, one user follows the tempo using earphones, while the other performs the gestures at his own pace and is free to perform arbitrary actions or stop in the middle. We collect CSIs and accelerations and compare their temporal patterns with the metronome-generated tempo.

B. Observations

Observation 1: The CSI amplitudes of the same gesture remain stable during each gesture repetition, while different gestures exhibit distinctive temporal patterns.

Fig. 3 plots the CSI amplitudes and the 3-axis acceleration traces of one user performing the eight gestures in Fig. 2. For different gestures, both the CSI amplitudes and the acceleration traces exhibit distinctive patterns. Such pattern differences across gestures have been exploited in wireless [12] or inertial [7] based gesture recognition. However, one primary drawback of wireless-based solutions is that the CSI-based gesture signatures may be location-dependent [14]. To avoid location dependency in CSI-based solutions, we only leverage the temporal patterns among gesture repetitions to assess the gesture consistency among multiple users. As shown

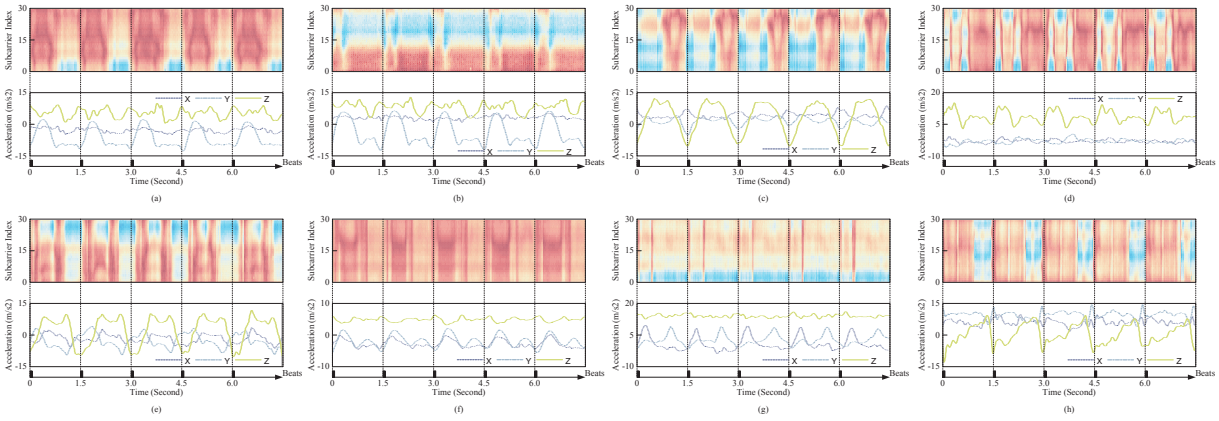


Fig. 3. Measurements of one user performing the eight gestures in Fig. 2. Each gesture is performed five times at the same location following a metronome-generated tempo. Each figure shows the CSI amplitudes (the upper sub-figure) and the 3-axis acceleration traces (the lower sub-figure).

in Fig. 3, for the eight gestures, the temporal patterns of CSI amplitudes remain similar between repetitions. This similarity stems from the fact that the user is repeating the same gestures according to the metronome-generated tempo. The repetitive patterns are independent of different types, thus mitigating location dependency.

Observation 2: The CSI amplitudes of synchronous gesture repetitions from two users exhibit notable periodic patterns as if performed by one single user.

Fig. 4a depicts the CSI amplitudes and the 3-axis acceleration traces of two users synchronously performing five repetitions of side lateral raises (Fig. 2a). As Fig. 4a shows, the CSI amplitudes for the synchronous gestures still demonstrate notable repetitive patterns as if only one user were performing the gestures. The timings are also consistent with the tempo with high accuracy. However, the temporal patterns differ from those for a single user performing the same gesture (Fig. 3a). This is because human presence within the monitored range can alter the propagation paths, and therefore the wireless channel may change dramatically with user numbers and locations. The difference in CSI patterns of one user and two users performing the same gesture indicates that location-specific training might be indispensable if we were to decode the superposition of multiple gestures by successively resolving gestures performed by one user. To avoid any labor-intensive training, we track synchronous gestures by analyzing only the *quality of repetitive patterns* from CSI.

Observation 3: The CSI amplitudes of asynchronous gestures from two users suffer dramatic variation when one user fails to perform the gesture at the same pace, leading to disturbances in the periodic patterns.

Fig. 4b shows the CSI amplitudes and the 3-axis acceleration traces of two users asynchronously performing side lateral raises (Fig. 2a). During the measurements, User 1 performs the gestures according to the tempo, as shown in his acceleration traces. In contrast, User 2 tries to follow User 1 but fails to keep up. As recorded in his acceleration traces, his movements exhibit arbitrary patterns in the middle (during 3.0s to 4.5s). We observe that the superposition of the repetitive gestures of User 1 and the arbitrary gestures of User 2 leads to irregularity in the CSI amplitudes. Especially in the

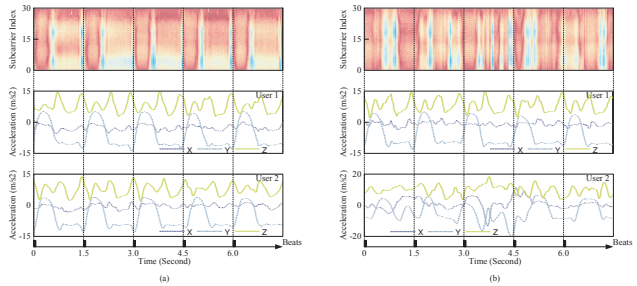


Fig. 4. Measurements of two users performing side lateral raises. Each figure shows the CSI amplitudes and the 3-axis acceleration traces for (a) synchronous gesture repetitions and (b) asynchronous gestures.

middle (from 3.0s to 4.5s), the CSIs fluctuate dramatically when User 2 performs arbitrary motions rather than side lateral raises. This is because asynchronous movements alter different propagation paths at different time intervals, thus introducing various and random time delays and phase shifts of the propagation paths. These time delays and phase shifts of propagation paths can lead to random yet dramatic constructive and destructive phase superposition, which results in fast and significant fluctuations in CSI amplitudes. Therefore, we may assess the extent of randomness in CSI fluctuation to infer asynchronous gestures from multiple users. This approach may eliminate the need of per-person worn inertial sensors and the extra communication overhead to synchronize, schedule and collect data from multiple sensors.

Summary: According to the measurements, CSI amplitudes can capture the motion patterns of different gestures. Since humans are only part of the wireless channels, the CSI patterns of one user may differ from those of two users performing the same gesture. However, the CSI patterns remain similar within repetitions of the same gesture and the repetitive property is independent of user numbers and locations. Furthermore, the superposition of synchronous gesture repetitions results in notable repetitive CSI patterns as if only one user were performing gestures. Conversely, the superposition of asynchronous gesture repetitions leads to dramatic irregularity in

CSI amplitudes due to random phase superposition of propagation paths. Therefore, we propose to harness the quality of periodicity in CSI patterns to distinguish synchronous and asynchronous gesture repetitions from multiple users in a non-invasive manner without body-worn inertial sensors.

III. Design and Implementation

To codify the above observations into a functional prototype, multiple challenges have to be addressed:

- **How to deal with the noises in CSI data?** Raw CSIs can be noisy: (1) CSIs are measurements estimated by preambles. The estimation can be affected by protocol parameter changes and noise levels. (2) CSIs contain irrelevant variations due to background dynamics. It is thus important to filter the raw CSI data to mitigate signal fluctuations not caused by gestures in the monitored area.
- **How to select the best subcarriers?** CSI depicts the frequency-selective fading in wideband WiFi channels. Some subcarriers may be more sensitive to user gestures than the others. It can improve robustness and reliability by selecting the proper subcarrier subset to track gestures.
- **How to quantify the quality of periodicity?** While CSIs exhibit repetitive patterns during gesture repetitions, two factors need to be considered when assessing their periodicity. (1) Users may not repeat the same gestures exactly in terms of gesture amplitude and timing due to *e.g.* fatigue. Thus the metric should tolerate variations in each gesture repetition. (2) The metric should require relatively few gesture repetitions to assess periodicity to ensure realtime performance.

This section elaborates on the design of *WiSync* to tackle the above challenges.

A. CSI Processing

1) Outlier Filtering: The first step is to remove outliers in the raw CSI data. As shown in Fig. 5, we observe two categories of outliers: (1) wideband abrupt variations (upper figure in Fig. 5a) and (2) narrowband jitters (upper figure in Fig. 5b). According to [13], occasional changes in Modulation and Coding Scheme (MCS) of WiFi transmissions may lead to diverse CSIs across the whole spectrum even for the same wireless link. To eliminate the wideband abrupt CSI variations due to such rate adaptation, we adopt the scheme in [13] and consider the received packets with small MCS indices as outliers. Since user gestures tend to alter certain propagation paths, and path changes tend to induce variations in the whole frequency band [14], it is unlikely that sudden jitters on a single subcarrier is caused by gestures. Hence we also take narrowband jitters as outliers, yet retain abrupt CSI changes caused by potential asynchronous gestures. Concretely, we utilize the Hampel identifier [17] to robustly estimate the locations of narrowband jitters. We identify x as an outlier in $X_N = \{x_i\}_{i=1}^{N}$ if

$$|x - med(X_N)| \geq g(N, \alpha_N) MAD(X_N) \quad (1)$$

where $med(X_N)$ and $MAD(X_N)$ are the median and Median Absolute Deviation (MAD) of X_N and $g(N, \alpha_N)$ is a determined by certain error levels [17]. We consider a packet as an outlier if fewer than three adjacent subcarriers are outliers

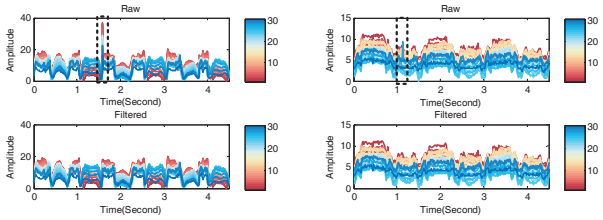


Fig. 5. CSI measurements before and after outlier filtering for (a) wideband abrupt changes and (b) narrowband jitters.

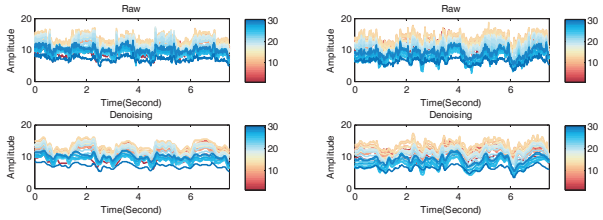


Fig. 6. Raw and denoised CSIs using wavelet shrinkage denoising for (a) synchronous gestures and (b) asynchronous gestures.

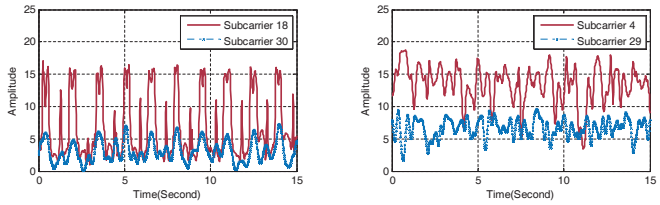
by the Hampel identifier. To remain uniform sampling, we replace the outliers by linear interpolation. Fig. 5 shows the CSIs before and after outlier filtering.

2) Denoising: The second step is to any suppress noise. CSIs are sensitive to change in wireless channels. We observe high-frequency fluctuations in raw CSIs. The noisy data may hide the periodic patterns of synchronous gesture repetitions. To improve the robustness of *WiSync*, we adopt wavelet shrinkage denoising [18]. We choose wavelet-based denoising rather than linear low-pass filters to preserve the local structures of different gestures and the sharp edges potentially caused by signal superposition of asynchronous gestures.

In wavelet shrinkage denoising, each subcarrier is decomposed via Discrete Wavelet Transform (DWT) using Symlet basis into approximations and details. Thresholding is then applied to the details to suppress noise. We choose the threshold that minimizes the Stein Unbiased Estimate of Risk (SURE) since it is smoothness-adaptive [19]. Such a thresholding scheme is suitable for denoising signals ranging from relatively smooth signals *e.g.* CSIs of mild and synchronous gesture repetitions to signals with random jumps *e.g.* CSIs of fierce and asynchronous gestures. Fig. 6 shows the CSIs on the 30 subcarriers before and after denoising. As is shown, most high-frequency noise is suppressed while the gesture-induced sharp edges still remain. The repetitive patterns of synchronous gestures (Fig. 6a) and the relatively random patterns of asynchronous gestures (Fig. 6b) are also preserved.

B. Subcarrier Selection

In the presence of multipath propagation, human movements are not the only factor affecting wireless signals. The constructive and destructive superposition of multipath components contributes to frequency-selective fading in wideband WiFi signals. Consequently, some subcarriers may be more sensitive



(a) (b)

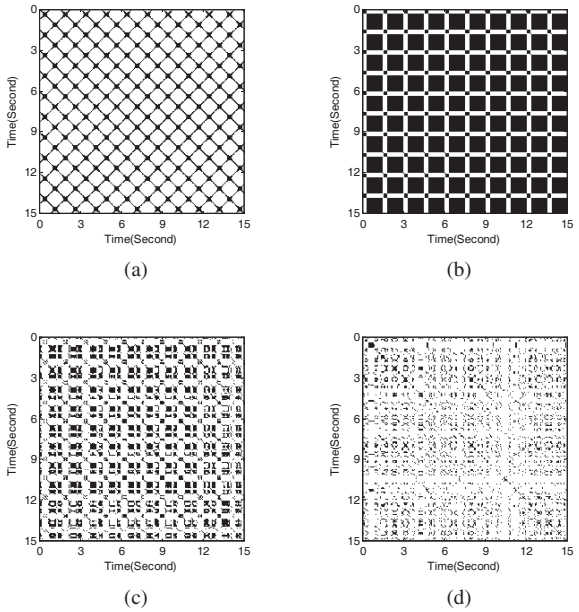
Fig. 7. Subcarriers with the highest and the lowest ranks for ten repetitions of (a) synchronous gestures and (b) asynchronous gestures.

to human movements and more reliable than others [20]. Hence it potentially improves the sensitivity and the accuracy by selecting a subset of subcarriers to track gesture repetitions.

Our subcarrier selection scheme is based on the following intuitions. (1) Since gestures induce fluctuations in CSI amplitudes, and more notable variations tend to indicate higher sensitivity to human movements, thus it is natural to select subcarriers with large variance. (2) Subcarriers with higher average power seem more reliable since their signal-to-noise ratios (SNRs) are higher. Hence selecting high-power subcarriers may improve reliability. In our work, we empirically rank each subcarrier by the product of its average power and its standard deviation. Fig. 7 plots the subcarriers with the highest and lowest ranks for 10 gesture repetitions. As shown, subcarriers with the highest ranks demonstrate more notable repetitive patterns for synchronous gesture repetitions while exhibiting random patterns in case of asynchronous ones. We make two comments here. (1) Note that selecting the most periodic (chaotic) subcarriers may favor synchronous (asynchronous) gestures yet induce high false asynchronous (synchronous) gesture detection rates. Thus the criterion of our subcarrier selection scheme is not to select the most periodic or chaotic subcarriers, but rather those with more notable changes and the more reliable ones. This is because we are unaware beforehand whether the CSI measurements are from synchronous or asynchronous gestures. (2) To further improve the robustness of *WiSync*, we heuristically select the top five ranked subcarriers and average their amplitudes as the input for periodicity analysis. We select multiple subcarriers instead of only one because we find that the top ranked subcarriers are sometimes distant from one another, with distances in central frequency greater than the typical coherent bandwidth indoors. This indicates these subcarriers are largely independent, and combining multiple independent and informative (subcarriers with notable fluctuations) observations may contribute to a more robust decision. We average the selected subcarriers to suppress residual narrowband noise after outlier filtering.

C. Periodicity Analysis

Autocorrelation is an intuitive approach to assess periodic signals and has shown to be effective in analyzing repetitive postural patterns such as step counts [21] and free-weight gym exercise counts [7]. However, we argue that autocorrelation is insufficient to discern the temporal patterns of synchronous and asynchronous gesture repetitions. On the one hand, unlike body-worn inertial sensors that directly measure the acceleration and orientation of body movements, WiFi signals



(a) (b) (c) (d)

Fig. 8. An illustration of structural characteristics of recurrence plots and the corresponding DET descriptors. (a) A sine wave of 10 periods ($DET \approx 1.00$); (b) A periodic square wave with a duty cycle of 20%, repeating 10 periods $DET \approx 1.00$); (c) Synchronous side lateral raises, 10 repetitions, two users ($DET = 0.96$); (d) Asynchronous side lateral raises, 10 repetitions, two users ($DET = 0.64$).

perceive gestures indirectly. Thus CSI amplitudes are noisier and contain irrelevant information *e.g.* wireless channels about the background environment. Therefore more noise-resilient periodicity analysis is required. On the other hand, the CSI amplitudes reflect the gesture superposition from multiple users. Even for asynchronous gestures, usually at least one user is still repeating gestures according to the tempo. Hence finer-grained periodicity quantification is preferable for capturing the irregular CSI patterns introduced by users not following the tempo.

In this work, we utilize Recurrence Quantification Analysis (RQA), a time-series analysis approach commonly dealing with non-linear dynamics [16], to quantify the periodic characteristics of CSI amplitudes. RQA has been successfully adopted in characterizing subtle dynamics of various rhythmic signals including respiratory patterns [22], heart rate variability [23], postural fluctuations [24] and long-term climate dynamics [16]. In essence, RQA provides a set of descriptors to quantify the Recurrence Plot (RP) of a time series.

An RP is a graphical view of the times at which a certain state recurs in the time series. Mathematically, for a time sequence $X_N = \{x_i\}_{i=1}^N$, its RP is the visualization of the squared matrix $\mathbf{R}_{N\times N}$, where

$$R_{i,j}^{\epsilon} = \Theta\left(\epsilon - \|x_i - x_j\|\right), i, j = 1, \ldots, N \qquad (2)$$

where ϵ, $\|\cdot\|$, $\Theta(\cdot)$ denote a threshold, a norm and the Heaviside step function [16]. In RP, a recurrence is defined as state x_j being sufficiently close to state x_i to allow for chaotic dynamics. We use the Euclidean distance as the norm metric and an adaptive threshold determined by background dynamics. That is, we set the threshold according to the

median variations of the corresponding subcarriers measured when there is no user within the monitored area. Although a static threshold *e.g.* the precision of CSI amplitudes also works, an adaptive threshold filters CSI variations irrelevant to gesture repetitions. Users can set the threshold by simply collecting CSIs without human movements in the monitored area within seconds, thus incurring little calibration overhead.

Fig. 8a and Fig. 8b depict the RPs of two periodic signals *i.e.* a sine wave and a periodic square wave. Their RPs exhibit more diagonal oriented and periodic structures such as diagonal lines (Fig. 8a) and checkerboard structures (Fig. 8b). This is because $R_{i+k,j+k} = 1, k = 1, \ldots, l$ (where l is the length of the diagonal line) indicates that the sequence visits the same region at different times while vertical and horizontal lines suggest staying in the same state [16]. Fig. 8c and Fig. 8d further present the RPs of 10 synchronous and asynchronous gesture repetitions. As shown, the RP of synchronous gesture repetitions exhibit more notable periodic patterns similar to the square wave. This may be because our denoising process keeps sharp edges (Fig. 7a).

To quantify the structural patterns in RPs, standard RQA defines five variables including recurrence rate (RR), percent determinism (DET), maximal line length (D_{max}), entropy (ENT) and trend (TND). Among them, we find DET the most suitable for our application. Mathematically, DET calculates the fraction of points forming diagonal lines:

$$DET = \frac{\sum_{l=d_{min}}^{N} l H_D(l)}{\sum_{i,j=1}^{N} R_{i,j}} \tag{3}$$

where

$$H_D(l) = \sum_{i,j=1}^{N} (1 - R_{i-1,j-1})(1 - R_{i+l,j+l}) \prod_{k=0}^{l-1} R_{i+k,j+k} \tag{4}$$

denotes the histogram of the diagonal line lengths. DET depicts the predictability of a system with more periodic behaviors than chaotic processes [16]. We set $d_{min} = 2$ to exclude short diagonal lines since periodic signals often inscribe long diagonal segments and are more deterministic. As shown, the RP of the sine wave has a DET of almost 1. Note that the solid rectangular structures in the RP of the square wave can also be considered to be diagonal lines. Thus its RP also approaches 1. As expected, the DET of synchronous gesture repetitions is larger than that of asynchronous ones, and we thus use DET as the metric to distinguish synchronous and asynchronous gestures.

D. Putting It Together

Fig. 9 illustrates the processing flow of *WiSync*. It consists of two stages: calibration and monitoring. During the calibration stage, CSI amplitudes when no users move in the monitored area are collected to estimate the background dynamics. A threshold θ_b is derived as the standard deviation of the CSI measurements representing environment-induced variations. Another threshold θ is estimated with one user performing gestures since CSI measurements are blind to one user and synchronous users in theory. We select θ as $mean(DET) - std(DET)$, where $mean(DET)$ and $std(DET)$ are the mean and standard deviation of the DETs of one user performing gestures, respectively. A more accurate yet labor-intensive way is to derive θ based on two users performing synchronous and asynchronous gesture repetitions. Although such calibration is indispensable for deriving the optimal threshold, our evaluation demonstrates that a general threshold can fit various gestures, user diversity, user orientation and different locations. Hence no per-location or per-user calibration is required. During the monitoring stage, the receiver collects CSI measurements, which are sequentially processed to filter outliers and suppress noise, and then selects the subcarriers with notable changes and inputs for periodicity analysis as previously discussed, where the corresponding θ_b for the selected subcarriers are averaged as the threshold for their RPs. The decision module then determines the CSI measurements are from synchronous (asynchronous) gestures if the calculated DET is larger (smaller) than θ. For synchronous gestures, we can also provide further analysis *e.g.* tempo estimation, repetition segmentation and counting.

IV. Performance

In this section, we interpret the experiment methodology, followed by a detailed performance evaluation of *WiSync* in terms of (1) accuracy, (2) robustness and (3) scalability.

A. Methodology

Hardware. We implement a prototype of *WiSync* with commodity WiFi infrastructure. A single-antenna Tenda wireless router is used as the transmitter operating in IEEE 802.11n AP mode at 2.4GHz Channel 11. A mini PC with Intel WiFi Link 5300 NIC and three external omnidirectional antennas works as the receiver. The receiver runs Ubuntu 10.04 LTS with the 2.6.36 kernel and pings packets from the transmitter at a rate of 60 packets per second. CSI information is extracted from the received packets by the Linux CSI tool [15], and is further processed using MATLAB scripts.

Scenarios. We evaluate the performance of *WiSync* in two typical indoor environments: (a) an office of $12 \times 8m^2$ and a two-bedroom apartment of $8 \times 6m^2$. Fig. 10 shows the layouts. The office is made of concrete walls with wooden doors. The bedrooms are separated by hollow walls with wooden doors of a thickness of 8cm. The transmitter and the receiver are placed 0.8m above the floor at different locations in the test rooms. We experiment with four users in total. Two exercise regularly, and the others seldom do. To evaluate the robustness of *WiSync*, we extensively test it in different scenarios including LOS, NLOS and through-wall propagation (Fig. 11). We also test the performance with different relative user headings including $0°$, $90°$ and $180°$. During the measurement, one to three other students are also working at their desks, occasionally moving around in the same room, or standing near the wireless link.

Data Collection. We test eight fitness and gaming gestures as in Fig. 2. For each gesture, we collect 100 sets of measurements. During each measurement set, users perform ten repetitions of the gesture. For synchronous actions, we broadcast tempo using a smartphone metronome application, and each user performs the gestures according to the tempo. For asynchronous actions, one user performs gestures following the same metronome tempo using earphones, while the other

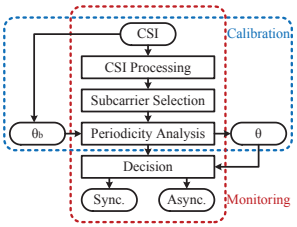


Fig. 9. Processing flow of *WiSync*.

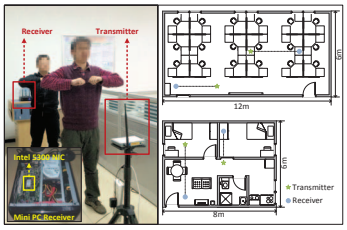


Fig. 10. Deployment of *WiSync*.

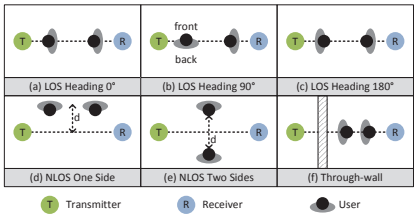


Fig. 11. Illustration of testing scenarios.

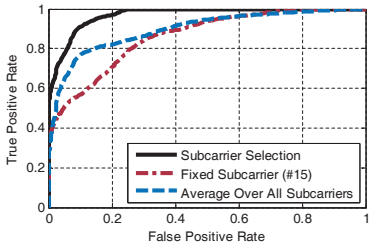


Fig. 12. ROC curves of *WiSync*.

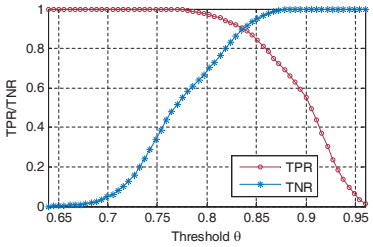


Fig. 13. TPR/TNR vs. thresholds.

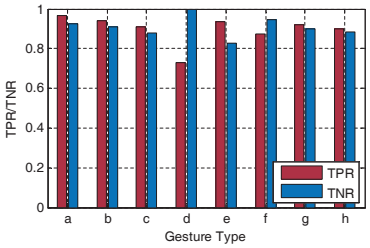


Fig. 14. Accuracy for each gesture.

user is free to perform the gestures at his/her own pace or stops in the middle as if he/she felt tired or unable to keep up. The tempo is fixed at 40bpm. We collect 1000 synchronous and 1000 asynchronous sets of measurements in total. According to both video recordings and inertial sensors, the users are able to perform gestures following the tempo at an average time difference of fewer than 0.2s.

Metrics. We mainly focus on the following metrics to evaluate the performance of *WiSync*. (1) True Positive Rate (TPR): The fraction of cases where *WiSync* correctly identifies synchronous gesture repetitions among all synchronous data. (2) False Positive Rate (FPR): The fraction of cases where *WiSync* mistakes asynchronous gesture repetitions for synchronous gesture repetitions among all asynchronous data. Since our scheme leverages repetitive patterns without segmentation, we only label each set of gestures rather than each repetition of a gesture as "synchronous" or "asynchronous". However, after correctly identifying repetitive synchronous gestures, it is possible to automatically segment the sets of gestures and estimate the average tempo.

B. Overall Performance

We plot the Receiver Operating Characteristic (ROC) curves in Fig. 12 to depict the overall tradeoff between TPRs and FNRs over a wide range of thresholds. As is shown, *WiSync* achieves a balanced synchronous gesture identification accuracy of 80% with an FPR of 20% on a fixed subcarrier (subcarrier 15 in our case). The balanced identification accuracy rises to 90% with an FPR of 10% with subcarrier selection and drops by 10% if it is averaged over all subcarriers. This indicates subcarrier selection moderately improves identification reliability by selecting the more sensitive subcarriers. In contrast, averaging over all subcarriers completely ignores frequency-selective fading, thus leading to modest performance degradation. Fig. 13 plots the TPRs and True Negative Rates (TNRs) of synchronous gestures under a range of thresholds with subcarrier selection. *WiSync* achieves balanced TPR and TNR of 90.43% using a threshold of 0.837. We use this threshold in the subsequent evaluations unless otherwise specified.

Fig. 14 shows the performances for the 8 testing gestures separately using the threshold for balanced overall detection performance. While the threshold fits the majority of the gestures, we find a forward lunge has a TPR of only 70% yet a TNR of 100%. This indicates the CSI measurements for synchronous forward lunges contain relatively more dynamics and thus less repetitive than other gesture types. One reason may be that forward lunges mostly involve lower limb movement and the torso may induce more fluctuations to keep balance. Therefore, while a general threshold can fit a range of gesture types, re-calibration would help to further optimize the gesture-specific performance.

C. Robustness Evaluation

1) Impact of Repetitive Counts: Since *WiSync* utilizes the repetitive patterns to identify synchronous gesture repetitions, the number of gesture repetition may affect the identification performance. Fig. 15 plots the balanced TPRs/TNRs ranging from 200 to 2000 packets. As the reference tempo is 40bpm and the sampling rate is 60 packets per second, one gesture repetition roughly corresponds to 100 packets. As is shown, the balanced TPRs/TNRs steadily increase from 200 to 1000 packets, and remain stable until they reach more than 1800 packets then they drop slightly. This is because periodicity analysis requires at least two complete repetitions to identify the correct periodic patterns. We note that the balanced accuracies remain above 85% with about six gesture repetitions. Thus *WiSync* relies on relatively few repetitions to identify synchronous gestures. With more than 18 gesture repetitions, *WiSync* suffers slight performance degradation. This may be because too many repetitions can cause fatigue and thus incur irregularity even in synchronous gesture repetitions.

2) Impact of NLOS Propagation: One advantage of WiFi-based gesture monitoring is that it can work in NLOS propagation. We evaluate the performance of *WiSync* in a two-bed

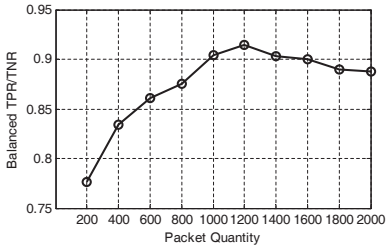


Fig. 15. Impact of gesture counts.

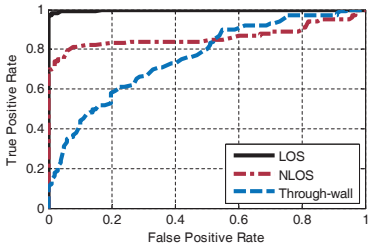


Fig. 16. Impact of NLOS propagation.

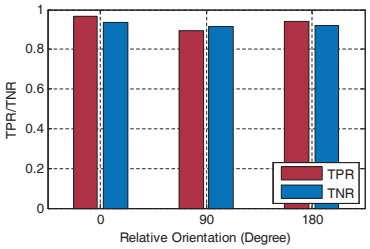
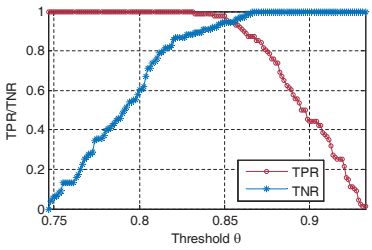
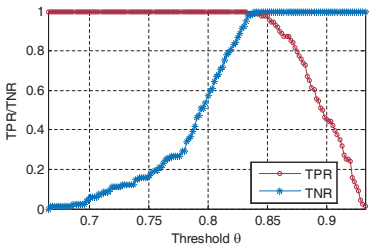


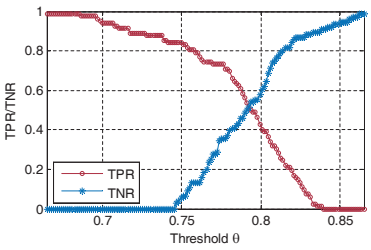
Fig. 17. Impact of orientation.



(a)



(b)



(c)

Fig. 18. TP/TN with 4 users. (a) 4 synchronous users vs. 2 synchronous users and 2 asynchronous users; (b) 4 synchronous users vs. 1 user repeating gestures with 3 asynchronous users; (c) 2 asynchronous users out of 4 users vs. 3 asynchronous users out of 4 users.

apartment under NLOS propagation. Fig. 16 illustrates the ROC curves for LOS, NLOS and through-wall propagation, respectively. As expected, the balanced identification accuracy degrades moderately for NLOS propagation (83%) and drops to below 70% for through-wall scenarios due to severe attenuation and more interferences. We expect future WiFi technology to bring radar-like processing [9] to commodity WiFi platforms to truly realize WiFi radars for smart homes.

3) Impact of User Orientation: We evaluate the identification performance at three relative headings between two users performing side lateral raises (Fig. 11). Fig. 17 plots the TPRs/TNRs using the general threshold. As is shown, *WiSync* performs slightly better for parallel relative orientations than an orthogonal one, yet achieves TPRs/TNRs of at least 89% for all the three relative orientations. This indicates that *WiSync* is robust to different user orientations. This is because the multipath propagation facilitates users of various orientations to affect part of the wireless channels and our subcarrier selection scheme also adaptively picks the more sensitive subcarriers with different user orientations.

D. Scalability Evaluation

We ask four users to perform side lateral raises in three cases. Case 1: four users perform gestures synchronously; Case 2: two users perform gestures synchronously while the other two perform arbitrarily; Case 3: three out of the four users perform randomly while only one user performs gestures to the tempo. Fig. 18 plots the TPRs/TNRs under a wide range of thresholds to distinguish Case 1 and 2; Case 1 and 3; as well as Case 2 and 3, respectively. Comparing Fig. 18a and Fig. 18b, we observe that the balanced TPR/TNR to distinguish Case 1 and Case 3 outperforms the one for Case 1 and Case 2. This is because more asynchronous users (three asynchronous users in Case 3) induce more randomness in the CSI measurements, thus are more distinguishable from Case 1, *i.e.*, four synchronous users. The thresholds for the balanced TPR/TNR only differ slightly from the general threshold derived for the measurements of two users. This indicates that the CSI patterns of four synchronous users are similar to those of two synchronous users, both of which look as if only one user were performing the gestures.

Fig. 18c shows the TPRs/TNRs with which differ the cases with two asynchronous users among four and those with three asynchronous users among four. Unfortunately, the balanced accuracy is close to random guesses. Therefore, while our scheme can easily scale from two users to four without recalibration, finer-grained time-series analysis is required to further distinguish different numbers of asynchronous users.

V. Related Work

Out work is related to previous research in fitness tracking, in-air gesture recognition, and wireless sensing systems.

Wearable Sensors. Wearable fitness tracking systems are both common and commercially available. Fitbit wristbands [8] and Apple HealthKit [25] can track steps, distances, and estimates calories burnt. A core module of these systems is a pedometer, which exploits the repetitive patterns of walking and running to count steps. In addition to tracking aerobic exercise such as running, researchers have also explored the idea of monitoring resistance training using wearable sensors. RecoFit [7] further improves the recognition and counting accuracy using an arm-worn accelerometer and a gyroscope. Since these systems rely on inertial sensors to capture body postures, they require per-user worn sensors or smartphones, thus are unable to support multiple users simultaneously. *WiSync* is motivated by the thriving growth in fitness tracking, yet targets multi-user monitoring for corporative gaming and sports training without body-mounted sensors.

In-air Gesture Recognition Systems. Human-computer interaction and interactive gaming are shifting from *on-device* to *around-device*. A key enabler is in-air gesture recognition. Leap Motion [6] emits multi-channel infrared signals and analyzes reflected signal patterns to recognize hand gestures. SoundWave [4] leverages microphones to extract the Doppler effects from sounds reflected by human gestures for gesture-based computer operation. The popular Xbox Kinect [5] utilizes advanced computer vision technology to achieve fine-grained gesture recognition, and brings with it a series of multi-player interactive games. While these systems relieve users from wearing sensors, they require a LOS path between users and the devices. In contrast, *WiSync* can operate in NLOS and through-wall scenarios.

Wireless Sensing Systems. Since wireless signals can penetrate walls, wireless-based systems can enable whole-home activity recognition and tracking. WiVi [9] utilizes inverse synthetic aperture radar technology for through-wall motion sensing. WiSee [10] extracts body-reflected Doppler features for gesture recognition. While these systems can obtain fine-grained motion information, they involve sophisticated radar-like processing and hardware modification, thus impeding immediate deployment. Alternatively, researchers adopt a pattern matching approach using ubiquitous WiFi networks without extra or customized infrastructure. Melgarejo *et al.* [12] leverage directional antennas to recognize American Sign Language gestures. E-eyes [13] utilizes location-aware WiFi fingerprints to identify daily activities. CARM [14] proposed a CSI-based speed model for accurate location-independent activity recognition. Wi-Sleep [26] extracts repetitive patterns from WiFi signals for breadth monitoring. WiGest [11] uses RSS to detect signal change primitives and further recognize hand gestures. However, these efforts mostly target at the activity recognition of a *single* user, yet *WiSync* deals with synchronized gestures of *multiple* users. *WiSync* can also be integrated with existing activity recognition systems [11], [12] to support multi-user activity quality assessment with commodity WiFi devices.

VI. Conclusion

In this work, we present *WiSync*, a non-invasive synchronous gesture tracking scheme on a single link with commodity WiFi infrastructure. *WiSync* works by analyzing the differences in periodicity during gesture repetitions of the received signals to discern synchronous and asynchronous gesture repetitions without gesture recognition. We prototype *WiSync* on commodity WiFi devices and evaluate it in typical indoor environments. Experimental results show that *WiSync* achieves a balanced synchronous gesture identification accuracy of 90.43% for two users with eight gesture types. *WiSync* is also robust to relative user orientations and can scale to four users. We envision *WiSync* as an early step towards general multi-user supports for wireless gesture sensing, which can benefit a range of teamwork training, cooperative gaming and social interaction applications.

Acknowledgement

This work is supported in part by the NSFC under grant 61522110, 61332004, 61472098, 61303209, 61572366, and Beijing Nova Program under grant Z151100000315090.

References

[1] "Unisoft Just Dance," 2009, http://just-dance.ubi.com/.

[2] A. M. Piper, E. O'Brien, M. R. Morris, and T. Winograd, "SIDES: A Cooperative Tabletop Computer Game for Social Skills Development," in *Proc. of ACM CSCW*, 2006.

[3] M. R. Morris, A. Huang, A. Paepcke, and T. Winograd, "Cooperative Gestures: Multi-user Gestural Interactions for Co-located Groupware," in *Proc. of ACM CHI*, 2006.

[4] S. Gupta, D. Morris, S. Patel, and D. Tan, "Soundwave: Using the Doppler Effect to Sense Gestures," in *Proc. of ACM CHI*, 2012.

[5] "Kinect for Xbox," 2014, http://www.xbox.com/en-US/xbox-one/accessories/kinect-for-xbox-one.

[6] "Leap Motion," 20101, https://www.leapmotion.com/.

[7] D. Morris, T. S. Saponas, A. Guillory, and I. Kelner, "RecoFit: Using a Wearable Sensor to Find, Recognize, and Count Repetitive Exercises," in *Proc. of ACM CHI*, 2014.

[8] "Fitbit," 2007, https://www.fitbit.com/.

[9] F. Adib and D. Katabi, "See Through Walls with Wi-Fi!" in *Proc. of ACM SIGCOMM*, 2013.

[10] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-Home Gesture Recognition Using Wireless Signals," in *Proc. of ACM MobiCom*, 2013.

[11] H. Abdelnasser, M. Youssef, and K. A. Harras, "WiGest: A Ubiquitous WiFi-based Gesture Recognition System," in *Proc. of IEEE INFOCOM*, 2015.

[12] P. Melgarejo, X. Zhang, P. Ramanathan, and D. Chu, "Leveraging Directional Antenna Capabilities for Fine-Grained Gesture Recognition," in *Proc. of ACM UbiComp*, 2014.

[13] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free Location-oriented Activity Identification using Fine-grained WiFi," in *Proc. of ACM MobiCom*, 2014.

[14] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of wifi signal based human activity recognition," in *Proc. of ACM MobiCom*, 2015.

[15] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 Packet Delivery from Wireless Channel Measurements," in *Proc. of ACM SIGCOMM*, 2010.

[16] C. L. Webber Jr, *Recurrence Quantification Analysis*. Springer, 2007.

[17] L. Davies and U. Gather, "The Identification of Multiple Outliers," *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 782–792, 1993.

[18] C. Taswell, "The What, How, and Why of Wavelet Shrinkage Denoising," *Computing in Science and Engineering*, vol. 2, no. 3, pp. 12–19, 2000.

[19] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.

[20] O. Kaltiokallio, M. Bocca, and N. Patwari, "Enhancing the Accuracy of Radio Tomographic Imaging using Channel Diversity," in *Proc. of IEEE MASS*, 2012.

[21] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort Crowdsourcing for Indoor Localization," in *Proc. of ACM MobiCom*, 2012.

[22] S. Rathnayake, I. Wood, U. Abeyratne, and C. Hukins, "Nonlinear Features for Single-Channel Diagnosis of Sleep-Disordered Breathing Diseases," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 8, pp. 1973–1981, 2010.

[23] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths, "Recurrence-plot-based Measures of Complexity and Their Application to Heart-Rate-Variability Data," *Physical Review E*, vol. 66, p. 026702, 2002.

[24] M. Riley, R. Balasubramaniam, and M. Turvey, "Recurrence Quantification Analysis of Postural Fluctuations," *Gait & Posture*, vol. 9, no. 1, pp. 65–78, 1999.

[25] "iOS 8 HealthKit," 2014, http://www.apple.com/ios/whats-new/health/.

[26] X. Liu, J. Cao, S. Tang, and J. Wen, "Wi-Sleep: Contactless Sleep Monitoring via WiFi Signals," in *Proc. of IEEE RTSS*, 2014.