WiEat: Fine-grained Device-free Eating Monitoring Leveraging Wi-Fi Signals

Zhenzhe Lin*, Yucheng Xie†, Xiaonan Guo†, Yanzhi Ren‡, Yingying Chen*, Chen Wang*§
*WINLAB, Rutgers University, USA
†Indiana University-Purdue University Indianapolis, USA
‡University of Electronic Science and Technology of China, P.R. China
§Louisiana State University, USA
zhenzhe.lin@rutgers.edu, yx11@iupui.edu, xg6@iupui.edu
renyanzhi05@uestc.edu.cn, yingche@scarletmail.rutgers.edu, chenwang1@lsu.edu

***Abstract*—Eating well plays a key role in people's overall health and wellbeing. Studies have shown that many health-related problems such as obesity, diabetes and anemia are closely associated with people's unhealthy eating habits (e.g., skipping meals, eating irregularly and overeating). Thus, keeping track of diet is becoming more important. Traditional eating monitoring solutions relying on self-report remain an onerous task, while the recent trends requiring users to wear dedicated yet expensive hardware are cumbersome. To overcome these limitations, in this paper, we develop a device-free eating monitoring system using WiFi-enabled devices (e.g., smartphone or laptop). Our system aims to automatically monitor users' eating activities by identifying the fine-grained eating motions and detecting the minute movements during chewing and swallowing. In particular, our system distinguishes eating from non-eating activities by using K-means clustering with principal component analysis on the extracted Channel State Information (CSI) from WiFi signals. It further adopts a soft decision-based eating motion classification through identifying the utensils (e.g., using a folk, knife, spoon or bare hands) in use. Moreover, we propose a minute motion reconstruction method to identify chewing and swallowing through detecting users' minute facial muscle movements. The derived fine-grained eating monitoring results are beneficial to the understanding of users' eating behaviors and estimation of food intake types and amounts. Extensive experiments with 20 users over 1600-minute eating show that the proposed system can recognize the user's eating motions with up to 95% accuracy and estimate the chewing and swallowing amount within 10% percentage error.**

***Index Terms*—WiFi sensing, CSI, Eating monitoring**

I. Introduction

Eating, as an essential activity for energy intake and nutrition supply, has been known to be closely related to people's health. A surfeit of food could lead to the excess of calorie intake, gaining body weight and various health-related problems such as cardiovascular diseases, diabetes, stomach cancers [1]. Whereas the imbalanced or insufficient food intake could not fulfill the daily body needs and further result in nutritional deficiency problems such as anemia, osteoporosis and scurvy, which impedes the cell recovery and growth, especially for the patients, teenagers, and seniors. The recent U.S. reports show that 70.2% of American people suffer from overweight or obesity [2] and 90% of the U.S. population have a nutrient deficiency [3]. It is thus important to keep tracking of diet and maintain a good dietary habit.

*Chen Wang's contribution to this work was when he was a graduate student at Rutgers University.

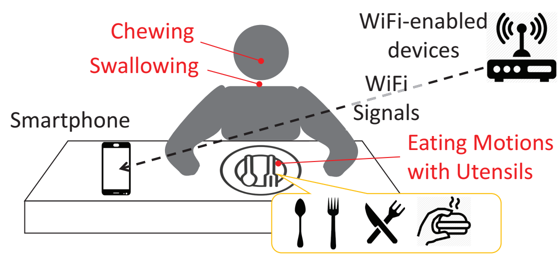


Figure 1. An illustration of the proposed eating monitoring system.

Traditional eating monitoring methods mainly rely on self-reports (e.g., food logs [4] or food journals [5]). Smartphone Apps such as YouEat [6] and Cara Care [7] allow users to record the food diaries via text logs or pictures manually. However, these self-report methods require user's participation to memorize or record the details in eating behaviors, and they all suffer from the subjective biases and the memory recall imprecision [8]. The food scanner or the calorie calculator advances the self-monitoring methods by tracking the type of food and calculating calorie consumption [9]. Nevertheless, these methods usually require expensive dedicated hardware and the users' active participation is still inevitable.

In recent years, the emerging mobile sensing technologies have enabled several automatic eating monitoring systems such as smart utensil-based method [10] and wearable-based methods [11], [12]. In particular, Smart-U, made of a special spoon, can recognize the types of food by its reflected light spectra [10]. Thomaz et al. [11] utilize the accelerometer in a smartwatch to capture the eating moments by recognizing the user's hand motions during eating. Amft et al. [12] propose to detect the air-conducted vibrations of food chewing by a condenser microphone embedded in an ear pad to understand the food textures. However, the smart utensil-based method is limited to a single utensil, while the wearable-based approaches can only detect partial body motions where the devices are worn (e.g., hand or jaw). They have limited capability to provide comprehensive eating monitoring. Moreover, these approaches all require additional dedicated devices.

Different from existing works, our goal is to develop a device-free eating monitoring system leveraging WiFi signals without the user's active participation. In particular, we propose to recognize food intake types (e.g., plant-based food, meat-based food, starch-based food and soup-based food)

978-1-7281-6607-0/20/$31.00 ©2020 IEEE

based on the combination of three-dimensional dietary information: utensil usage, chewing time and swallow activity. The key insight of this idea is that people would lean toward using different utensils depending on food intake types. According to a recent survey [13], fork and knife usage indicates veggie-based and meat-based food, whereas hand usage indicates starch-based food such as bread and pizza. Furthermore, chewing times and swallow times could provide a more comprehensive dietary information. The number of chews is highly related to the food texture and density [14]. The high-density food (e.g., steak and nuts) may require multiple chews, whereas it takes fewer chews to break down soft and water-filled food such as fruit and vegetables. Additionally, one-time chewing and direct swallowing together might indicate soup-based food, whereas chewing and swallowing for multiple times together indicates meat-based food. Based on the combination of these three-dimensional dietary information, our system takes one step further to provide an automatic dietary monitoring that enables users to track their daily meal composition and provides a potential solution to assist people on their dietary. For example, dietary information could help users to determine whether to reduce the food intake for bodyweight management or to increase certain types of food intake for obtaining sufficient minerals or vitamins.

Although using WiFi signal to recognize human activities has previously shown its initial success, such as location-oriented activity identification [15], fitness assistance [16] and vital sign monitoring (e.g., breathing rate) [17], it cannot be directly used for eating monitoring and the following challenges need to be addressed: 1) It is not easy to differentiate eating activities from many other human activities; 2) Various eating motions with different utensils (e.g., fork, spoon, knife and bare hand) all involve similar hand movements (i.e., delivering foods from a plate to mouth) and thus it is a challenging task to distinguish these eating motions based on the noisy WiFi signals. 3) The chewing and swallowing only exhibit minute facial muscle movements, which are relatively hard to be captured by the WiFi signal; 4) Smartphones are usually equipped with relatively small internal WiFi antennas, making the quality of the received WiFi signals be much lower than using the devices with external antennas. Thus, how the smartphone could provide WiFi sensing is still unexplored.

Toward this end, we develop WiEat, a system that leverages the channel state information (CSI) extracted from WiFi-enabled IoT devices (e.g., smartphone, laptop) to provide fine-grained eating monitoring. In particular, the proposed system adopts a cluster-based method to differentiate the eating motions from the many other non-eating activities by capturing the unique physiological characteristics of eating motions. We then propose to extract the unique spectrogram features of eating motions and develop a soft decision-based algorithm to further recognize how a user eats (i.e., type of utensils). Moreover, we utilize a Minute Motion Reconstruction method to capture the minute facial muscle movements of chewing and swallowing and develop an accumulated power spectral density method to detect the periods of these minute motions for deriving the statistics of chewing and swallowing. In addition, we use wifi-enabled devices (i.e., smartphone, laptop) to build the first generation system that has been extensively tested for both single person case and two people case. Figure 1 illustrates a target scenario where a user put his/her smartphone on a dining table while eating. The smartphone will continuously collect WiFi signals from a WiFi-enabled device (e.g., laptop or IoT devices). The collected data can be used to provide automatic eating monitoring. We validate these two cases because they can be achieved with only a pair of transceiver, and they covers the majority of daily eating scenarios (nearly 60% of Americans regularly ate on their own according to the American Time Use Survey [18]).
Our contributions are summarized as follows:

- We demonstrate that the CSI extracted from WiFi signal can be used to provide fine-grained eating monitoring, which not only recognizes the eating motions but also capture the minute muscle movements of chewing and swallowing.
- We develop a device-free eating monitoring system based on CSI to automatically track people's eating activity, which can be easily deployed on smartphones or WiFi-enabled IoT devices without incurring additional costs.
- We develop a soft decision-based approach grounded on the analysis of CSI spectrogram to identify various eating motions associated with different utensils. Moreover, we propose a minute motion reconstruction method to capture the minute facial muscle movements and develop an accumulated power spectral density method to derive the chewing and swallowing statistics.
- Extensive experiments with 20 people over 1600-minute eating show that our system can recognize the user's eating motions and estimate the fine-grained chewing and swallowing statistics with high accuracy.

II. Related Work

Traditional eating monitoring methods are mainly based on questionnaires or self reports [4], [5], [19]. Fallaize *et al.* [19] design Food4Me, an online Food Frequency Questionnaires (FFQ) system, to collect a user's nutrient intake information. The recent smartphones Apps [4], [5] enable the user to conduct self reports with more flexibility and convenience. However, these methods require the user's active participation and suffer from the subjective bias and memory recall imprecision.

To reduce the user's efforts, vision-based methods such as cameras [20], [21] are designed for automatically dietary monitoring. DietCam [20] performs automatic dietary assessment using the photo strings or short videos taken by the user's mobile device, while eButton [21] relies on a camera attached to chest location to capture and evaluate the diet. However, the vision-based approaches may raise privacy concerns, because images often capture the user's sensitive information (e.g., eating with whom and where). Instead, there are some studies focusing on developing smart utensils to analyze the food intake automatically. Smart-U [10] uses a dedicated spoon equipped with a LED light & sensor of recognizing different foods based on their reflected light spectra. But these methods limited to a dedicated spoon or knife are hard to provide the comprehensive eating monitoring, where people could eat flexibly with other utensils or bare hands.

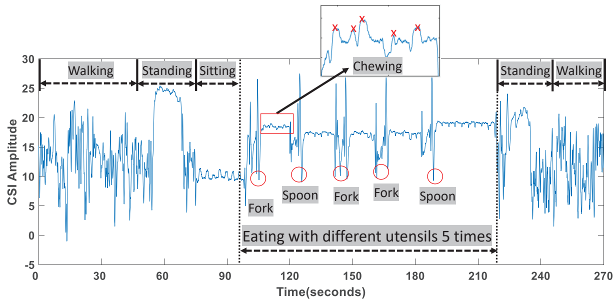


Figure 2. CSI amplitudes of one subcarrier under different human activities.

There are also active studies using the wearable devices (e.g., head-worn and wrist-worn devices) to provide automatic eating monitoring [11], [12]. Thomaz *et al.* [11] utilize the accelerometer on a smartwatch to capture the eating moments by recognizing the user's hand motions during eating. Similarly, Amft *et al.* [12] detect the air-conducted vibrations of food chewing by a condenser microphone embedded in an ear pad to understand the food textures. However, all the above methods are intrusive to the user by requiring the user to wear one or multiple dedicated devices.

Recent years, WiFi sensing has shown the initial success to provide the non-invasive human activity recognition [15]–[17], [22]. E-eyes [15] utilizes the WiFi signals to provide device-free location-oriented human activity identification. Guo *et al.* [16] use the WiFi signals from IoT devices to provide device-free fitness assistance. Liu *et al.* [17] develops a vital sign tracking system, which can detect minute human body motions like breathing via WiFi signals. Thus, in this paper, we propose to utilize the WiFi signals from a user's smartphone to provide fine-grained eating monitoring.

III. System

A. *Feasibility Study*

To design and implement an RF-based eating monitoring system, the basic idea is to explore the hidden relationship between human motion and the extracted Channel State Information (CSI). CSI is a fine-grained measurement of the wireless channel with 30 subcarriers, and each subcarrier measures the state of a subchannel with the amplitude and phase information. Compared with traditional received signal strength (RSS) measurement, CSI provides fine-grained channel state information that describes the propagation of wireless signals, including its fading, scattering, multipath, and wireless interference. Thus, when a user is present in the signal propagation paths, his/her body motions will affect the WiFi signals in the form of reflection, absorption, and refraction, which can be captured and revealed by the CSI pattern. However, eating is a complicated activity, which includes significant hand motions that deliver food to mouth as well as the minute muscle level jaw movements and pharynx movements that break down and ingest the food.

To explore the relationship between human motion and the extracted CSI, we ask a participant to perform a series of daily activities, including walking, standing, sitting, eating with a fork and eating with a spoon. In the meanwhile, the participant's smartphone is placed on the table as illustrated in Figure 1. The CSI is extracted from the smartphone's side for further analysis. Figure 2 shows the CSI amplitude of one subcarrier, where we mark the ground-truth of the participant's activities. We observe repetitive patterns of CSI amplitude that associated with eating motions. This is because food intake process that contains repetitive motions of delivering food to mouth. Moreover, after each eating motion (i.e., food delivery), we also observer slight fluctuations of CSI amplitudes (e.g., marked by the red rectangular), which correspond to the minute jaw movements of chewing. However, it's hard to further differentiate between using a fork and using a spoon from the CSI amplitude. In addition, the minute movements caused by chewing and swallowing are easy to be submerged by noises.

B. *System Overview*

The basic idea of our system is to detect the fine-grained food intake activities and minute facial muscle movements (i.e., chews and swallows) leveraging WiFi signals. As shown in Figure 3, our system takes the CSI measurements from WiFi-enabled devices as input and extracts the relative phase and amplitude information. To mitigate the interference of environment, we first perform data calibration and noise removal to filter the outliers caused by the diffraction and reflection of the stationary objects (e.g., dining table and walls). To derive the user activity information associated with the calibrated data, we then apply relative short time energy (STE) to calculate the corresponding spectrogram pattern. Based on that, the daily activities performed by users are captured and segmented in terms of different frequency ranges. After this step, we propose a PCA-based method to extract unique behavioral characteristics of eating motions and utilize K-means clustering approach to further differentiate the food intake activities from daily activities.

The core of the proposed eating monitoring system consists of two components, soft decision-based eating motion classification and accumulated power spectral density (PSD) based chew and swallow estimation. Given the differentiated eating activities, we further categorize them into different food intake motions based on utensil usages. In particular, we extract the statistic features from each orthogonal frequency-division multiplexing (OFDM) subcarrier to capture the inherent behavioral characteristics of different eating motions. The learning-based classifier recognizes utensil types based on each CSI subcarrier separately and calculates the decision probability of classifying the eating motion into every pre-trained category (i.e., fork, spoon, knife + fork and bare hand). Probability-based soft decision integrates the decision results from different CSI subcarriers by adding each category probabilities of different subcarriers. After comparing the value of four utensil-category, we select the highest one as our final decision output.

The eating motion identification aims to quantify the chewing and swallowing motions by deriving the chewing period and measuring the statistics of chews and swallows. Specifically, during the interval of each identified eating motion (i.e., food delivery), minute motion reconstruction is performed to magnify the small facial muscle movements of chewing and swallowing by taking advantage of all the CSI subcarriers to reconstruct the minute motion information. To estimate

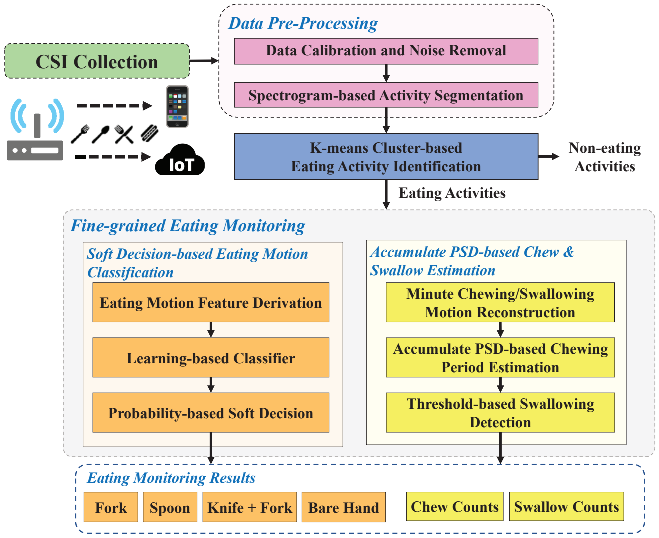


Figure 3. System Flow of WiEat.

the chewing period, we develop accumulate PSD-based chew detection to analyze the repetitive patterns of the chew motions from the CSI power spectral density accumulated over all the CSI subcarriers. Moreover, to detect the swallowing, the threshold-based swallowing detection recognizes the swallowing motions by capturing the inherent muscle movement differences between chewing and swallowing based on amplitude range and peak-to-valley time interval.

IV. Fine-grained Eating Monitoring

A. Data Pre-processing

1) Data Calibration and Noise Removal: WiFi signals suffer from RF interference and ambient noises. To eliminate the impact of such noises, we adopt the outlier removal approach as a first step. We discard the outlier-elements that are more than an interquartile range above the upper quartile or below the lower quartile. After removing the outliers, there still are some irregular impulses and fluctuations in CSI amplitude. Our key observation is that the ambient noises usually present on a fixed frequency range. Inspired by this, we apply a band-pass filter to remove interference caused by the ambient noises.

2) Spectrogram-based Activity Segmentation: After data calibration, our system utilizes a spectrogram-based method to segment the activities. We use the cumulative power spectral density (CPSD) to calculate the integrated frequency domain affected by human activities. According to large body movements always lasting for a short period and perform an ephemeral impulse in the frequency domain, we reconstruct the CSI complex value to enlarge the trifling variances into a distinct pattern. Inspired by this, we adopt Cumulative Short Time Energy (CSTE) to capture each eating activity. Specifically, we calculate the CPSD by accumulating all the power spectral density along the frequency dimension in the corresponding spectrogram. The CSTE is then calculated based on the CPSD by the following equation:

$$STE = \sum_{i=-\infty}^{\infty} [CPSD(i)W(n-i)]^2, \tag{1}$$

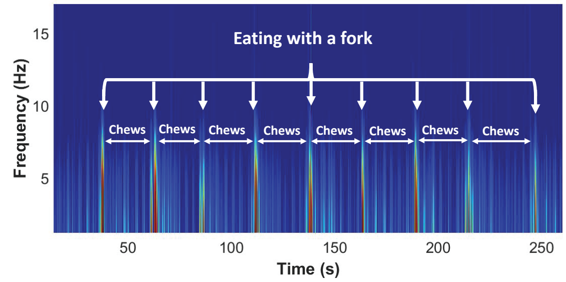


Figure 4. Illustration of spectrogram-based activity segmentation.

where $CPSD(i)$ is the cumulative power spectral density, $W(n)$ denotes the window function and n represents the frame shift of samples. In the relative spectrogram in frequency domain shown in Figure 4 we can observe clearer repetitive patterns, where the color degree of the wave represents the strength of power amplitude. The corresponding points of the zero points marked in the figure, determine the starting and ending time of the activities caused by users, which could be used for activity segmentation.

B. Differentiate Eating Activities from Non-eating Activities

In this section, we focus on differentiating the eating activities from non-eating indoor activities with a K-means cluster-based eating activities identification method.

Eating activities are defined as the movement of delivering food from cutlery to mouth. The basic idea is to capture the wireless channel variances of all activities and then use the cluster-based method to differentiate them. We first show how different activities influence the CSI estimate. As shown in Figure 5, the eating activities and non-eating activities can be categorized into two clusters based on two principal components through the principal component analysis (PCA). We observe that eating activities associated with different utensils (i.e., using forks, using knives, using spoons, and using bare hands) are gathered in the middle position, whereas other non-eating activities such as reading, chatting and typing surround externally. This is because the eating activities exhibit repetitive movements from hand to mouth, which have high similarity to each other. Although there are a few activities also exhibit repetitive movements, especially, smoking, we exclude them from our consideration because smoking is not allowed in indoor environments in most of the places [23].

Inspired by the above observations, we propose to use a cluster-based method to differentiate eating from non-eating activities. Specifically, a K-means clustering method is used to partition all the activities into two clusters based on two geometric centroids denoted as μ_1 and μ_2. The threshold is selected by the distance between the centroid of eating activities and testing activities. The distance can be calculated as $D_c = \|\mu_1 - \mu_2\|$. Through the analysis above, we show that the cluster-based method has the capability of identifying the eating activities by applying the threshold ξ to the D_c as follows:

$$\begin{cases} D_c \leqslant \xi, \ eating\ motions \\ D_c > \xi, \ non-eating\ motions \end{cases} \tag{2}$$

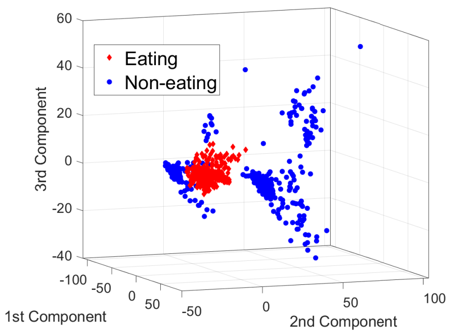


Figure 5. Clusters of eating activities and non-eating activities.

C. Soft Decision-based Eating Motion Classification

After detecting the eating activities, our system further recognizes the detailed eating motions based on utensils recognition, which provides the information about what a user eats and how much he/she eats.

1) Eating Motion Feature Derivation: To further classify the eating motion with utensils, we need to derive a series of reliable features extracted from the CSI readings. Additionally, unique features of eating motions could eliminate the environmental noises in terms of WiFi signals that suffer from ambient interference. Based on our preliminary experimental investigations and detailed analysis of extracted CSI, we particularly choose 14 features extracted from both time domain (e.g., *mode*, *average rectified value*, *interquartile range*, etc.) and frequency domain (e.g., *root mean square frequency*, *power*, etc.) of each subcarrier.

2) Learning-based Classifier: We adopt the learning-based classifier to further identify the eating motions with utensils. We use Support Vector Machine (SVM) implemented by LIBSVM [24] with linear kernel to build the classifier. Specifically, for each segmented CSI raw reading, we extract a set of fourteen features from all thirty subcarriers and then derive a two-dimensional with 30×14 vectors as the input for learning-based classifier. We denoted the two-dimensional vectors from selected thirty subcarriers as $\upsilon = [\upsilon_1, ..., \upsilon_i, ..., \upsilon_{30}]$, where υ_i includes the fourteen features mentioned above. Four prediction probabilities regarding different eating motions with utensils are defined as $\rho_f, \rho_k, \rho_s, \rho_h$, corresponding to forks, knives, spoons, hands, respectively. Then the estimated prediction probabilities for four different eating motions with utensils from the i^{th} subcarrier can be obtained by using the following equation 3:

$$\begin{cases} P^i = \max\{\rho_f^i, \rho_k^i, \rho_s^i, \rho_h^i\} \\ P_{total} = \sum_{i=1}^{30}[\rho_f^i + \rho_k^i + \rho_s^i + \rho_h^i] \end{cases} \tag{3}$$

3) Probability-based Soft Decision Strategy: Even though some carriers show lower sensitivity to users' eating motions, they still contribute useful information to the classification decision. The traditional methods by majority vote over the hard decision results (i.e., one of the categories) or subcarrier selection are hard to utilize the useful information from all the CSI subcarriers. Different from these methods, we develop a new probability-based soft decision strategy to leverage all the subcarriers and infer the eating motions with various utensils. In particular, the probability of classifying the eating motion to a utensil category based on each CSI subcarrier can be integrated with an assigned weight. The integrated probabilities of the all utensil category are compared, and the utensil category with the largest integrated probability is the final decision. The soft decision-based eating motion classification can be described as:

$$argmax\left(\sum_{i=1}^{30}[\rho_f^i \cdot w_f^i], \sum_{i=1}^{30}[\rho_k^i \cdot w_k^i], \sum_{i=1}^{30}[\rho_s^i \cdot w_s^i], \sum_{i=1}^{30}[\rho_h^i \cdot w_h^i]\right), \tag{4}$$

where the assigned weight value w is determined based on the variance of the CSI at each subcarrier, with the larger variance showing higher sensitivity to the eating motions.

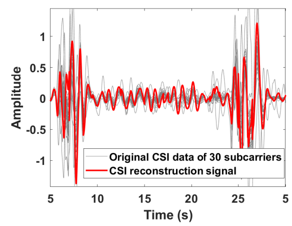


Figure 6. Reconstructing chewing/swallowing on CSI.

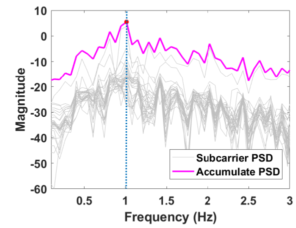


Figure 7. Accumulated PSD of CSI.

D. Chewing and swallowing Detection

Chewing is a physical degradation or digestion of food, and swallowing is the phase following chewing. To distinguish the food intake type and to calculate the amount of food a person eats, it is essential to detect chewing and swallowing activity.

1) Minute Chewing/Swallowing Motion Reconstruction: Although chewing and swallowing motions can be detected to have an impact (i.e., slight vibrations) on CSI, the minute motion caused by chewing and swallowing still can not be identified accurately from the CSI. There are two main challenges we need to address. First, not all subcarriers are sensitive to tiny motions. Some subcarriers are less susceptible to chewing and swallowing activities because different subcarriers have different central frequencies. Besides, the performance of some subcarriers are not consistent, and they might present sensitive amplitude for one period time but have dull amplitude for another period time. Second, environmental-related uncertainty and noise (i.e., wireless interference and multipath reflection) will also cause CSI fluctuation, which makes it hard to differentiate the CSI patterns caused by real chewing and swallowing motions in CSI time series.

To provide a robust chewing and swallowing detection model, we first utilize a Butter-worth band-pass filter with a cutoff frequency of 0.8 Hz$\sim$$3$ Hz to remove unwanted noise, because chewing and swallowing activities are usually fixed around a low-frequency range [25]. Furthermore, a moving average filter is used to remove outliers in the signal.

After we get the filtered signal, based on Mouth Motion Profile [26], we propose to reconstruct the CSI from all subcarriers to get a single representative CSI sequence, which amplifies the CSI fluctuation caused by chewing and swallowing without losing the original information. In particular, a sliding window is applied on all subcarriers parallelly, and a series of CSI segments are selected from various subcarriers and then assembled one by one into a single new CSI according to time series. For each sliding window (the sliding window size is 250 ms), the mean amplitude values

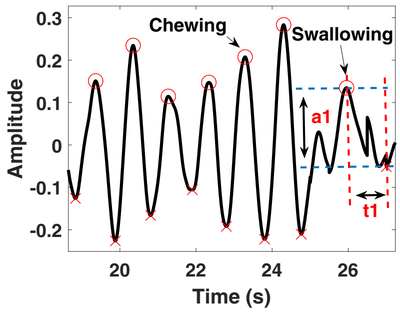
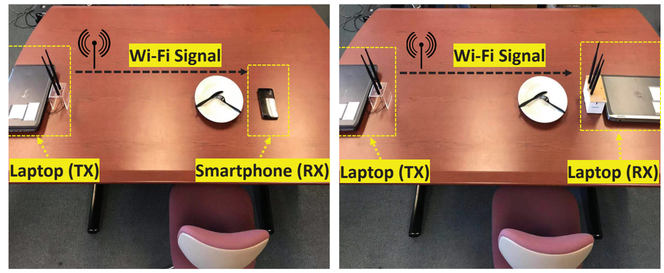


Figure 8. Chewing and swallowing differentiation.



(a) Smartphone-Laptop setup (b) Laptop-Laptop setup

Figure 9. Illustrations of two experimental settings.

of all subcarriers are calculated, respectively. Then the ten subcarriers which have the nearest mean values to the average mean value of all subcarriers are picked. This step will filter out those abnormal and less-sensitive subcarriers. Furthermore, among these selected subcarriers, the segment which has the maximum peak to peak value within one sliding window is chosen as a representative segment for this time slot. By moving the sliding window along time sequence, we could select a series of CSI segments and produce a single new CSI sequence. As shown in Figure 6, the fluctuation of the new signal is more significant than that from any of the original single subcarrier, which proves the effectiveness of the reconstruction.

2) Accumulate PSD-based Chewing Period Estimation: To further reduce the impact of environmental-related noise, and eliminate those fake fluctuation caused by noise, it is necessary to estimate the chewing period. Based on the estimated period, it is convenient to remove those fake chewing and swallowing motions, which are too close to adjacent chewing and swallowing motions. In our system, instead of calculating the chewing period from time domain, we find it is more effective to acquire a stable period of chewing from frequency domain. Specifically, we first transfer time domain CSI amplitude to frequency domain by using power spectral density (PSD). Since PSD would produce one strong peak at the frequency that corresponds to the periodicity of dominant repetitive fluctuations, we observe that the CSI measurements in our frequency window (i.e., 0.8 Hz$\sim$$3$ Hz) also present one strong peak that corresponds to the dominant chewing rate. Because the target CSI measurement is only derived between two adjacent identified eating motions, and we have filtered the interferential frequency components by our frequency window, we believe that the estimated period is dominated by mouth motion.

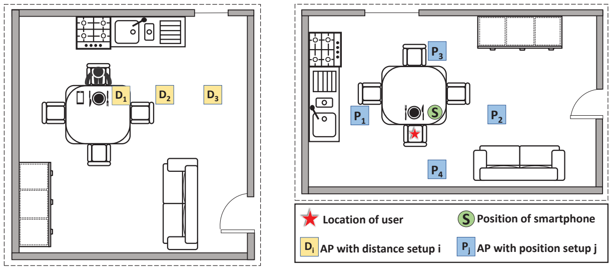
Moreover, to further utilize different subcarriers which present conform results in frequency domain and estimate a more robust chewing period, accumulated power spectral density (APSD) method is proposed in our system to accumulate the results collected from all subcarriers. The accumulated power spectral density of 30 subcarriers with N CSI amplitude measurements can be presented as:

$$APSD = 10log_{10}\sum_{i=1}^{30}\frac{(abs(FFT(c_i)))^2}{N}, \tag{5}$$

where c is the CSI measurements of subcarrier i. Since the swallowing times are far less than chewing times in real life, the highest peak is identified as the chewing rate of the specific period between two eating motions. Figure 7 shows an example to infer the chewing rate of one period time between two eating motions. The ground truth of the chewing rate is 1 Hz, which is measured and verified by camera-based method. Figure 7 depicts that there is one strong peak around 1 Hz in the APSD, which implies that our APSD could effectively estimate the chewing rate of a user.

Based on the average period of chewing, we propose a threshold-based peak detection approach to identify the candidate CSI patterns caused by chewing and swallowing on CSI data. In original peak detection algorithm, a peak could be found if the value of data is larger than its two neighboring data. However, this algorithm has two limitations. First of all, some peaks that are actually caused by one chewing and swallowing motion can all be marked out. Second, environmental noises could also produce some tiny peaks. In our system, two thresholds are set empirically to remove these fake peaks. Firstly, based on the average chewing period, a minimal distance β between two neighbor peaks is used to restrict the peak-to-peak separation, and only peaks that recur at regular intervals are preserved. Second, a minimum amplitude of the peaks γ is used to filter out those fake peaks caused by environmental noises. After filtering these fake peaks, we get the number of all CSI patterns caused by mouth motions, and in the next part, we can distinguish chewing and swallowing motions from those preserved real peaks.

3) Threshold-based Swallowing Detection: Swallowing counting is challenging because the CSI measurements between chewing and swallowing are both sinusoidal-like patterns. Nevertheless, we observe that the CSI measurements of chewing and swallowing are still distinguishable on the reconstructed CSI representative. Compared with chewing motions, swallowing motions are more slowly and the movement range of the throat when swallowing is more moderate than jaw activity. Therefore we propose a threshold-based mechanism to classify the CSI patterns into two classes based on two measurements, including CSI amplitude range and corresponding peak-to-valley time interval. Specifically, as shown in Figure 8, a swallowing motion occurs following several times of chewing. The peak-to-valley time interval of CSI sinusoidal patterns is calculated as $t_j = l_{j+1} - l_j$, while the range of CSI amplitude between the peak and corresponding valley is derived from $a_j = p_j - v_j$. p_j and v_j are corresponding to the peak and valley of the j^{th} CSI sinusoidal pattern respectively, and l_{j+1} and l_j are the time stamp of the peak and valley of the same sinusoidal pattern on time serials. After



(a) Three different transcevier distances (i.e., 1 m, 2 m and 3 m) (b) Four different transceiver positions

Figure 10. Illustrations of two experimental settings.

we obtain the measurements of each CSI sinusoidal pattern, a threshold-based method is proposed to classify the candidate CSI sinusoidal pattern into two classes. The average values of the amplitude $\overline{a}$ and period $\overline{t}$ of all CSI sinusoidal patterns within a specific period time (e.g., between two adjacent eating motions) are identified as the reference threshold of these two classes. Those CSI sinusoidal patterns whose amplitude a_j is less than $\overline{a}$ and also the period t_j is larger than $\overline{t}$ are counted as sw_i, which represents one swallowing time. Once we recognize swallowing from chewing, it is very convenient to derive the statistic of chewing and swallowing respectively.

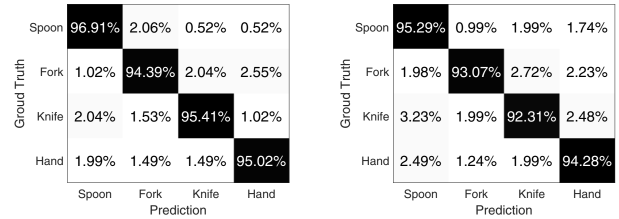
V. Performance Evaluation

A. Experimental Methodology

Experimental Setup: To evaluate WiEat's performance in detecting eating activity, we build a prototype with laptops and smartphones. Specifically, we conduct experiments in two setups including *Smartphone-Laptop Setup* and *Laptop-Laptop Setup*. The *Smartphone-Laptop Setup* describes the practical scenario when a user places her/his personal smartphone on a dinning table during eating. In this setup, the smartphone is connected to a laptop to receive WiFi signals and sense the user's eating activities. *Laptop-Laptop Setup* describes a different scenario, where an IoT device (e.g., smart TVs and restaurant table tablets) can use the WiFi signals received from a WiFi-enabled device to capture the user's eating. Both setups are deployed in three representative indoor environments to evaluate our system, including a office and two dining rooms.

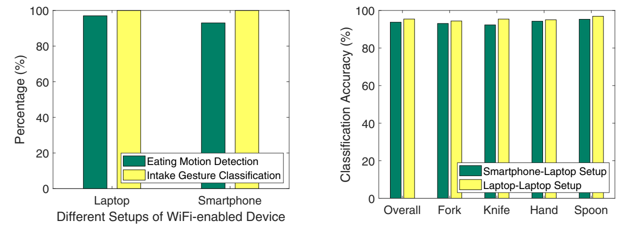
Devices: We conduct experiments with two different smartphone models, Nexus 6 and Huawei Mate 10 smartphones. The Nexus 6 has 3 GB RAM and a 2.7 GHz Snapdragon 805 processor while the Huawei Mate 10 is equipped with a 4GB RAM and a 2.36 GHz Kirin 970 processor. We also utilize two Dell E6430 laptops equipped with 802.11n WiFi wireless card IWL 5300 NICs [27] and 6dBi rubber ducky external omni-directional antennas for extracting CSI data. One laptop is used to imitate the IoT device and the other serves as the WiFi-enabled device. Both laptops are running Ubuntu 14.04.4 LTS with the kernel 4.2. And the WiFi cards work at 5GHz frequency band with $1000 pkt/sec$ transmitting rate.

Data Collection: We recruit 20 participants to perform eating activities with the two experimental setups at three representative indoor environments. In total, 1600 minute eating period data is collected and the ground truths are measured and verified by camera-based method during the experiments. To collect CSI in the smartphone-laptop setup, we configure the laptop (serving as a WiFi-enabled device) to run in the net-link mode, which sends Internet Control Message Protocol (ICMP) echo and gets the reply from the smartphone to collect the CSI data from the smartphone [28]. For the laptop-laptop setup, we configure the both laptops (i.e., an IoT device and a WiFi-enabled device) to work under the injection mode. Moreover, for both setups, we test three distances (1 m, 2 m, and 3 m) and four different placements to deploy the devices. Unless mentioned otherwise, half of the data we collect are used to build the users' profiles, and half are used to evaluate the performance.



(a) Laptop-Laptop Setup (b) Smartphone-Laptop Setup

Figure 11. Utensils Identification: confusion matrix for different setups.



(a) Eating moments recognition (b) Eating motion detection and intake gesture classfication

Figure 12. Impact of the receiver selection on system performance.

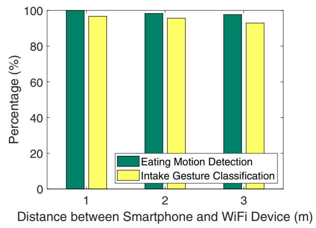
Evaluation Metrics: $Detection\ Rate$ is the ratio of the number of correctly detected eating activities over the total number of eating activities. $Accuracy$ is the ratio of the number of correctly detected activities over the total number of activities. $Percentage\ \ Error$ of estimating the chewing/swallowing count is defined as:

$$percentage\ error = \frac{|estimated\ number - ground\ truth|}{ground\ truth}$$

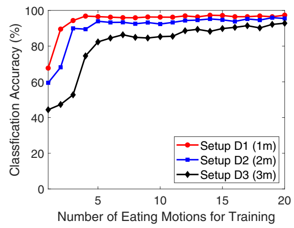
B. Overall Performance

We evaluate the overall performance of food intake gesture recognition, chewing and swallowing estimation under different real three indoor environments including two dining rooms and a laboratory. Considering the fact that users usually put his/her smartphone on a dining table, we perform a detailed study of dietary information under various factors including: the impact of receiver selection, the impact of transceiver distance, and the impact of transceiver position.

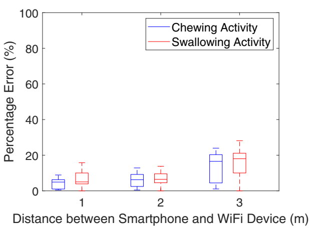
Impact of Receiver Selection: We first evaluate WiEat's performance in smartphone-laptop setup and then compare to the laptop-laptop setup. Figure 11 presents the identification results for four different eating motions based on utensils held by users. As shown in Figure 11(a) and Figure 11(b), WiEat achieves average accuracy in around 95% and 94%, respectively. We observe that the laptop-laptop setup gives



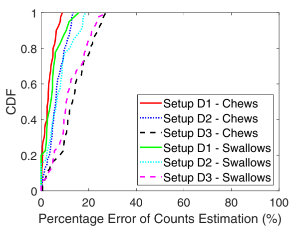
(a) Eating motion detection and intake gesture classfication



(b) Different training size on intake gesture recognition

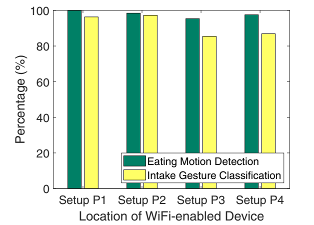


(c) Chews and swallows estimation

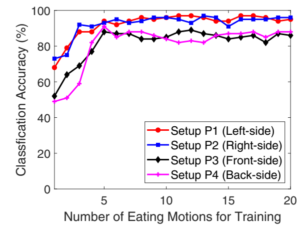


(d) CDF of estimation error

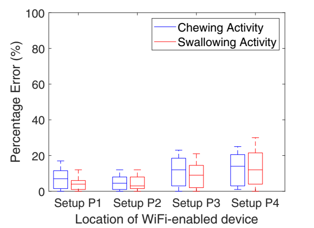
Figure 13. Impact of the transceiver distance on system performance.



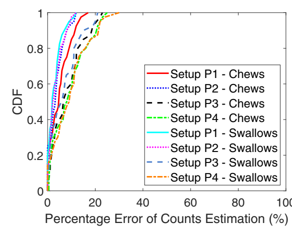
(a) Eating motion detection and intake gesture classfication



(b) Different training size on intake gesture recognition



(c) Chews and swallows estimation



(d) CDF of estimation error

Figure 14. Impact of the transceiver position on system performance.

a slightly better performance than smartphone-laptop setup. The reason might be the receiving laptop installed with the external antenna, which could reduce the interference of multi-path when transmitting. Moreover, Figure 12(a) depicts that our system could achieve a detection rate of 100% on both smartphone-laptop setup and laptop-laptop setup, confirming that the effectiveness and reliability of our system both on different transmitting WiFi-enabled device. Figure 12(b) presents the performance of our soft decision-based classification algorithm on differentiating eating utensils including spoon, fork, fork & knife and hands. We observe that our system can achieve over 95% average accuracy across different eating utensils for both smartphone-laptop and laptop-laptop setup. Additionally, it is encouraging to find that all the four types of utensils can all be recognized well with the lowest accuracy as 92% on smartphones and 93% on laptops. This demonstrates that our system is robust for utensils identification even if the different WiFi-enabled transmitters are used.

Impact of Transceiver Distance: Different distances between the transceivers would affect the accuracy of eating monitoring due to the increments of multi-path and diffraction. Figure 10(a) shows that the transceivers are placed at two sides of the dining table with different distances from 1 to 3 meters. Figure 13(a) depicts that our system could achieve over 97% eating motion detection and 93% intake gesture classification accuracy even for 3 m distance case. Figure 13(b) shows the overall accuracy for intake gesture recognition under different training set size. We observe that WiEat could achieve sufficient accuracy to recognize fine-grained eating motions based on the utensils held by users with only several training sets. For chews and swallows estimation, Figure 13(c) depicts that the average percentage error slightly increases when transceiver pair is separated by 3 m. This is because the such tiny jaw open-close movements and throat vibrations are hard to be captured under weaker WiFi signals in a longer distance. Figure 13(d) illustrates the Cumulative Density Function (CDF) of the chews and swallows percentage error for three transceiver distance settings. The above results show that more than 80% percentage error of the number of chews and swallows estimation are below 20%, indicating that our system is robust and effective under a longer distance.

Impact of Transceiver Position: We further evaluate our system with different location of the transceivers as shown in Figure 10(b). Figure 14(a) shows that left-side and right-side settings give better performance than front-side and back-side settings for eating motion detection and intake gesture classification. This is because human body dominate and partially block the WiFi links, increasing the amount of interference and diffraction of WiFi signals when transmitting. Figure 14(b) shows the classification accuracy of food intake gesture under different number of training eating motions. Consistent with the previous observations, left-side and right-side settings obtain better classification accuracy, and the classification accuracy increases with the growing number of training eating motions. Figure 14(c) and Figure 14(d) present the mean percentage error and the CDF of chewing and swallowing estimation. We observe that the average percentage error are all below than 12% even for the worst case *Setup P4*. The above results show that our system is effective under different relative positions of the WiFi-enabled device.

VI. Discussion

In this section, we discuss the practicality of WiEat and the effectiveness of the measure experimental data under various practical positions. According to survey statistics [18], there are around 40% people usually eat with others. In that case, we evaluate two people case and the results show WiEat can achieve promising accuracy. We will further explore the potential of WiEat in a multi-person scenario. As shown in Figure 15, we conduct the experiments of two people eating together in three common daily-life scenarios. The relative position of two people includes face to face, side by side, and one sitting in the right angle of another. In addition, we use directional antennas to boost the reception of the WiFi signal. WiEat can achieve a promising accuracy for intake gesture classification for three deployments, confirming

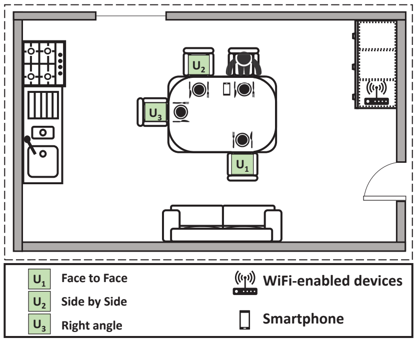


Figure 15. Illustration of two people cases for three deployments.

the feasibility of monitoring a user under the interference of surrounding people. We also notice that among three scenarios, the accuracy of *Setup U3* is highest. This is because if the distance between a user and surrounding people is longer than the corresponding radius of the Fresnel Zone, the impact of people nearby can be negligible [16]. The results show that our system has the practicality to work with two people cases. A more comprehensive study of the system performance in multiple people scenario with various environments will be explored in our future work.

VII. Conclusion

In this paper, we explore the feasibility of using the WiFi-enabled devices to provide users with automatic eating monitoring. We show that the channel state information extracted from a user's smartphone or IoT devices could be utilized to both recognize the user's fine-grained intake gesture based on utensils and detect the minute facial muscle movements of chewing and swallowing, which could further infer the food intake types. We develop a device-free system to distinguish eating activities from non-eating activities based on a K-means cluster methods and then adopt a soft decision-based learning approach to classify the eating motions according to the utensils used by the user. Moreover, we reconstruct the minute facial muscle movements based on the CSI and develop the accumulated power spectral density method to derive the chewing and swallowing statistics. Extensive experiments involving 20 people over 1600-minute eating period are conducted. The results show that the proposed system can achieve up to 95% accuracy for identifying users' eating motions and 10% percentage error for chewing and swallowing amount estimation.

VIII. Acknowledgment

This work is partially supported by the National Science Foundation Grants CNS-1826647, CNS-1814590, CNS-1815908 and CNS-1717356. This work is partially supported by the National Natural Science Foundation of China under Grant NSFC-61802051.

References

[1] W. H. Organization, "What are the health consequences of being overweight?" https://www.who.int/features/qa/49/en, 2019.
[2] "Overweight and Obesity Statistics," https://www.niddk.nih.gov/health-information/health-statistics/overweight-obesity, 2019.
[3] "Report Suggests That 90 Percent of Americans Have a Nutritional Deficiency," https://www.thejoint.com/minnesota/roseville/roseville-32003/report-suggests-that-90-percent-of-americans-have, 2019.
[4] J. Loewy, "My macros+ — diet and calories," https://itunes.apple.com/us/app/my-macros-diet-calories/id475249619/, accessed April 21, 2019.
[5] Y. GmbH, "Yazio — diet and food tracker," https://itunes.apple.com/us/app/calorie-counter-yazio/id946099227/, accessed April 21, 2019.
[6] "Feel great about what you ate," https://youate.com/, 2019.
[7] "Your personal food and symptom diary," https://cara-app.com/?_branch_match_id=651011669040444429, 2019.
[8] R. Hill and P. Davies, "The validity of self-reported energy intake as determined using the doubly labelled water technique," *British Journal of Nutrition*, vol. 85, no. 4, pp. 415–430, 2001.
[9] T. M. Futurist, "The Top Five Digital Health Innovations For Food Tracking and Eating," 2017, url=https://medicalfuturist.com/top-5-digital-innovations-for-food-tracking-and-healthy-eating.
[10] Q. Huang, Z. Yang, and Q. Zhang, "Smart-u: Smart utensils know what you eat," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 1439–1447.
[11] E. Thomaz, I. Essa, and G. D. Abowd, "A practical approach for recognizing eating moments with wrist-mounted inertial sensing," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015, pp. 1029–1040.
[12] O. Amft, "A wearable earpad sensor for chewing monitoring," in *SENSORS, 2010 IEEE*. IEEE, 2010, pp. 222–227.
[13] "How Do You Eat These Things?" https://www.buzzfeed.com/sarahaspler/how-do-you-eat-these-things, 2019.
[14] "Chewing Your Food: Is 32 Really the Magic Number?" https://www.healthline.com/health/how-many-times-should-you-chew-your-food, 2016.
[15] 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 signatures," in *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 2014, pp. 617–628.
[16] X. Guo, J. Liu, C. Shi, H. Liu, Y. Chen, and M. C. Chuah, "Device-free personalized fitness assistant using wifi," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 4, p. 165, 2018.
[17] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, "Tracking vital signs during sleep leveraging off-the-shelf wifi," in *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing (ACM MobiHoc)*, 2015, pp. 267–276.
[18] "The most American thing there is: eating alone," https://www.washingtonpost.com/news/wonk/wp/2015/08/18/eating-alone-is-a-fact-of-modern-american-life/, 2015.
[19] R. Fallaize, H. Forster, A. L. Macready, M. C. Walsh, J. C. Mathers, L. Brennan, E. R. Gibney, M. J. Gibney, and J. A. Lovegrove, "Online dietary intake estimation: reproducibility and validity of the food4me food frequency questionnaire against a 4-day weighed food record," *Journal of medical Internet research*, vol. 16, no. 8, p. e190, 2014.
[20] F. Kong and J. Tan, "Dietcam: Automatic dietary assessment with mobile camera phones," *Pervasive and Mobile Computing*, vol. 8, no. 1, pp. 147–163, 2012.
[21] M. Sun, L. E. Burke, Z.-H. Mao, Y. Chen, H.-C. Chen, Y. Bai, Y. Li, C. Li, and W. Jia, "ebutton: a wearable computer for health monitoring and personal assistance," in *Proceedings of the 51st annual design automation conference*. ACM, 2014, pp. 1–6.
[22] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Communications Surveys & Tutorials*, 2019.
[23] "List of smoking bans," https://en.wikipedia.org/wiki/List_of_smoking_bans, 2019.
[24] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at url-http://www.csie.ntu.edu.tw/ cjlin/libsvm.
[25] M. Farooq and E. Sazonov, "Automatic measurement of chew count and chewing rate during food intake," *Electronics*, vol. 5, no. 4, p. 62, 2016.
[26] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We can hear you with wi-fi!" *IEEE Transactions on Mobile Computing*, vol. 15, no. 11, pp. 2907–2920, 2016.
[27] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: gathering 802.11 n traces with channel state information," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, pp. 53–53, 2011.
[28] M. Li, Y. Meng, J. Liu, H. Zhu, X. Liang, Y. Liu, and N. Ruan, "When csi meets public wifi: Inferring your mobile phone password via wifi signals," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2016, pp. 1068–1079.