The Telepathic Phone: Frictionless Activity Recognition from WiFi-RSSI

Stephan Sigg
Georg-August University Goettingen
Germany
Email: stephan.sigg@udo.edu

Ulf Blanke
Wearable Computing Lab
ETH Zurich, Switzerland
Email: blanke@eth.ife.ee.ethz.ch

Gerhard Tröster
Wearable Computing Lab
ETH Zurich, Switzerland
Email: troester@ife.ee.ethz.ch

***Abstract*—We investigate the use of WiFi Received Signal Strength Information (RSSI) at a mobile phone for the recognition of situations, activities and gestures. In particular, we propose a device-free and passive activity recognition system that does not require any device carried by the user and uses ambient signals. We discuss challenges and lessons learned for the design of such a system on a mobile phone and propose appropriate features to extract activity characteristics from RSSI. We demonstrate the feasibility of recognising activities, gestures and environmental situations from RSSI obtained by a mobile phone. The case studies were conducted over a period of about two months in which about 12 hours of continuous RSSI data was sampled, in two countries and with 11 participants in total. Results demonstrate the potential to utilise RSSI for the extension of the environmental perception of a mobile device as well as for the interaction with touch-free gestures. The system achieves an accuracy of 0.51 while distinguishing as many as 11 gestures and can reach 0.72 on average for four more disparate ones.**

I. Introduction

Mobile phones are a popular sensing platform for the multitude of sensors they incorporate and for their status as personal device kept close to or on the body [1], [2]. However, these mobile sensing platforms focus on inertial motion to recognize physical activity. When a device is no longer worn on the body, its sensing capabilities are greatly reduced. Indeed, although people are in the same room with their mobile device almost 90% of the time, their device is within arms reach less than 55% of a day [3], [4]. Therefore, the mobile phone can hardly serve as a continuous sensing platform with sensors such as accelerometers or gyroscopes.

To still obtain information about situations or activities, we need to exploit sensors that react on ambient stimuli. Possible choices are video [5], or audio for the classification of device-locations based on audio signatures [6] as well as localisation via audio-based fingerprinting [7]. However, video is restricted by the sensor's field of vision while audio is limited to general locations or situations [8].

We propose the use of another environmental sensor: the wireless interface to the Radio Frequency (RF) channel. By monitoring the fluctuation in the received signal strength indicator (RSSI) that is calculated at a receiver for each incoming packet, we attempt to classify the situation (e.g. crowd size), activities or gestures performed in proximity of a mobile phone (See figure 1). This approach allows operation even when the device is not carried by the user but near to her – a scenario where most activity recognition systems fail.

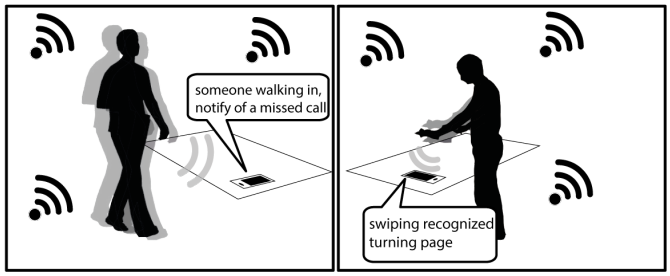


Fig. 1: Activity obtained from RSSI-signatures. Two example use-cases: user walking in with the smartphone implicitly reacting (left) and a no-touch explicit interaction (right).

We can utilise RSSI also in dark or quiet environments when audio or video might not provide sufficient information. In urban spaces, WiFi connectivity can be presumed (cf. section V-A). In addition, RF might be perceived as less privacy intrusive when compared to audio or video.

While there is some work on the device-free recognition of activities from RF-channel fluctuation [9], [10], [11], these systems require sophisticated Software Defined Radio (SDR) devices in order to obtain frequency domain features. In contrast, we attempt to utilise signal strength fluctuation on off-the shelf mobile phone hardware and from ambient WiFi traffic. On such devices, already the capturing of RSSI data in sufficient frequency is challenging. In addition, the data captured is less accurate and bursty. We discuss necessary pre-processing as well as the design of features suitable for highly bursty and low-resolution environmental RSSI data together with the final recognition step. In case studies we demonstrate the potential and limitations of using RSSI for recognition. The contributions of this work are:

1) System design and definition of feature space for RSSI-based 'frictionless' recognition
2) Analysis of RSSI-influencing factors in a controlled setting (e.g. direction, distance)
3) Feasibility study of situation, activity, and gesture recognition with off-the-shelf mobile phones.

Our results indicate that RF-based sensing of environmental *situations*, *crowd* and *individual activity* provides additional information for activity or context classification tools.

978-1-4799-3445-4/14/$31.00 ©2014 IEEE

II. Related Work

Device-free RF-based recognition was introduced by Youssef and others [12] as the localisation of an entity not equipped with any transmitter or receiver. In recent years, some groups work in this direction using hardware that ranges from SDR devices [13], laptop-class computers [14] over sensor nodes [15] or RFID tags [16] and achieve high accuracies of about 1 meter. This work is also related to a considerable body of practical and theoretical results on passive radar (cf. [17], [18] and references therein) where vehicles and individuals are detected and tracked from signals such as HF radio, UHF television broadcasts or DAB, DVB and GSM.

Recognition utilising signals on the wireless channel has been generalised in [11] to activities and we can further imagine also situations [19], gestures [10] or attention [20] to be identified by RF-based device-free implementations. These systems can be grouped into active and passive approaches conditioned on the presence of an active transmitter. Most previous work in this direction uses SDR devices.

Kassem et al. sense traffic situations by tracking frequency and speed of passing cars that intercept the direct line of sight between a pair of nodes [21]. The authors of [11] classify simple activities in an SDR-based active device-free system by extracting and interpreting features from a continuous signal between two nodes. Their approach explores also the multipath effects induced by persons that are not intercepting the direct path between nodes. It was later demonstrated that also simultaneously conducted activities from multiple persons can be distinguished by leveraging purely signal-strength based features [22]. Furthermore, it was shown by Pu and others that simultaneous detection of gestures from multiple individuals is possible utilising multi-antenna nodes and micro Doppler fluctuations [10], [23]. In a related system, Adib and Katabi employ MIMO interference nulling and combine samples taken over time to achieve the same result while compensating for the missing spatial diversity in a single-antenna system [9].

While the above are active approaches that require a dedicated transmitter, Ding and others have presented a passive system leveraging RF noise from engines of vehicles [24]. In addition, Shi et al. recognised activities and locations from fluctuation in the signal strength of broadcast FM radio [25].

Also, active systems utilising non-SDR nodes have been studied. Most notably, Patwari and others estimated the breathing frequency of an individual surrounded by nodes from the RSSI of exchanged packets [26]. Following other directions, Xu et al. have counted crowd [27] from RSSI within a field of sensor nodes. Their unsupervised learning approach is able to predict the count of up to 10 stationary or moving individuals. Recently, the recognition of general activities from RSSI in a sensor network has been considered [28]. In particular, the activities standing, sitting, lying, walking and empty have been distinguished with an accuracy of 0.8-0.9.

For these studies, either a sophisticated SDR device or transmit-receive pairs of nodes were required. Both cases are hard to establish with end-user equipment in spontaneous use. We propose a usable RF-based device-free recognition approach on phones by leveraging received RSSI from packets of WiFi access points (APs). We are not aware of previous work on such RSSI-based passive device-free recognition system.

III. Capturing RSSI on Phones

In IEEE 802.11, data is exchanged in packets on 11 partly overlapping frequency channels. In normal communication, a WiFi receiver discards all packets not addressed to itself. However, we can force the interface into monitor mode to log all traffic. For each packet, the receiver calculates the signal strength from the 8 bit preamble. Due to the lower data rates, control packets differ in their estimated RSSI significantly.

While the APIs of contemporary mobile phone operating systems (OSs) provide means to access the RSSI, this information is averaged and refreshed at about 1 Hz only. Another access to the RSSI is possible via the interface directly with tools such as `airodump-ng` or `tcpdump`. This requires root permissions to access the interface in monitor mode.[1] WiFi-firmware with sufficient access to relevant parameters is sparse. More severe even for mobile phones, most handsets implement a similar chipset family (e.g. Broadcom bcm4329, bcm4330(B1/B2), bcm4334, bcm4335) for which the default firmware does not provide access to the desired information (even as root). The only solution to avoid root access and which abstracts from this chipset family is via an external antenna[2]. However, this considerably extends the dimensions and complexity of the hardware, so that we decided against it. Instead, we used a modified firmware for the above mentioned chipset family [29] on a Nexus One phone running Cyanogen mod 7.2 and executed tcpdump on the interface in monitor mode to capture RSSI of packets. In monitor mode, no data can be transmitted and consequently no impact can be taken on the frequency in which packets are received. We can, however, adjust the channel we listen on and might utilise data from multiple APs transmitting on the same channel. In summary, while it is practically possible to monitor RSSI, the support of manufacturers for the operating systems to perform this out of the box is limited. However, we can track RSSI fluctuation with a modified OS, but without hardware modifications.

Figure 2 shows an exemplary snippet of sampled RSSI. In the experiments conducted, the RSSI usually ranged from -98dBm to -47dBm. Since the RSSI calculated for control packets differs, we disregarded them for the generation of this data. At the time of this recording, the phone was lying on a table within approximately 0.5 meters distance of a person sitting at that table. We observe that the data is very bursty. While there might be only one packet within 0.1 seconds at times, we can also observe five or more packets in the same interval. Clearly, when compared to SDR-based recognition systems that have direct access to the physical channel, the amount of information available from RSSI is severely reduced. Even compared to active RSSI-based systems that contain a transmitter omitting packets at high rate, our passive approach has to deal with more bursty traffic and a lower packet arrival rate. In addition, the granularity of RSSI is low. In our case, the 1dB granularity observed in the figure could not be improved for the WiFi interface.

We conclude that it would be hard to apply any curve fitting that could successfully predict the RSSI evolution at a higher sample rate.

[1] Monitor mode is obligatory in our case since otherwise the tool is executed in Ethernet emulation which does not provide RSSI information

[2] github.com/brycethomas/liber80211/blob/master/README.md

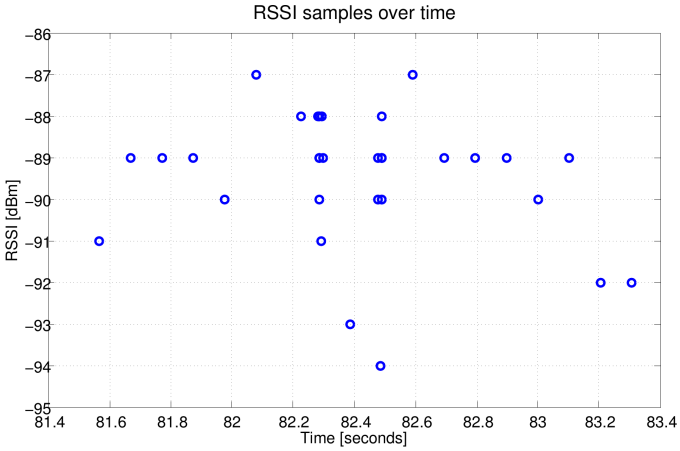


Fig. 2: RSSI from packets of a single AP

IV. Features for RSSI-based Recognition

Considering this structure of the data, we used simple features that express general properties such as the overall weight or mass as well as their spread.[3] As a tribute to the bursty traffic, the low granularity (cf. figure 2) and a fluctuating packet arrival rate, we simply fixed non-overlapping windows of two seconds and then utilised all RSSI values that would arrive during this period for feature calculation. The window length was set to 2s since we aim to design a system that would be practically usable with a good response time. A higher accuracy can be achieved with increased window size or via majority votes over successively calculated features (cf. section V-A). In total, 18 different features have been considered. On a data set with the three basic cases

1) A phone lying on a table in an empty room
2) A phone lying on a table with a person moving
3) A person holding and handling the phone

we applied a feature selection from the orange data mining toolkit [4]. From the remaining 9 features, we manually tweaked a combination that achieves good accuracy. Several combinations of *mean, median, variance, maximum* and the *difference between minimum and maximum* could achieve best and comparable classification results. For the case studies (section V), we decided for a combination of *mean, variance, maximum* and *difference between maximum and minimum*. For the gesture recognition, also the slope was considered.

V. Case Studies

We conducted case studies in indoor environments at ETH Zurich and TU Braunschweig (cf. figure 3). Occasionally the phone was connected to a computer via `adb shell` as an alternative to the slow on-screen keyboard which made no difference for the recorded data. All recordings were conducted multiple times and over several days. We intentionally altered the environments between recordings (e.g. moving furniture, placing the device slightly different).

[3]No frequency domain features could be used; Features as zero crossings or direction changes were not meaningful on the undersampled signal.

[4]http://orange.biolab.si/

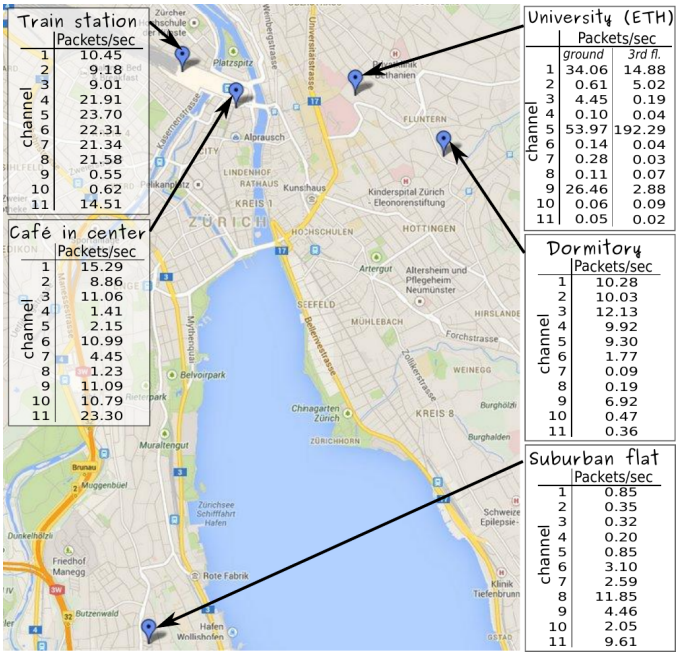


Fig. 4: Packets per second from the most active AP at various locations and over all 11 WiFi channels

Data was processed offline. However, we have developed a toolchain for the processing and classification that is sufficiently lightweight to be executed on the phone in realtime.[5] The tool groups packets for their source address (since the mean RSSI differs among senders) and disregards control packets (since also their RSSI level differs).

We now consider general RSSI properties and then investigate limits of RSSI-based recognition. The studies were conducted over two months in Braunschweig, Germany and Zurich, Switzerland. A total of 11 persons (9 male and 2 female; 26 to 37 years) have participated and overall about twelve hours of continuously sampled data has been produced.

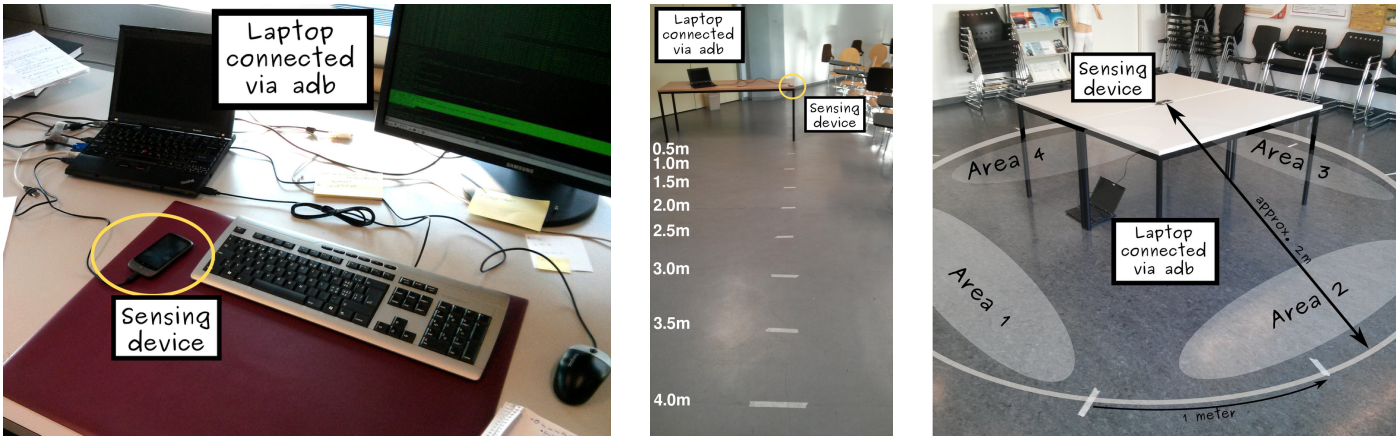
First, we investigate properties of urban WiFi with respect to traffic and sampling rate (section V-A). Then, we study coarse characteristics with respect to the presence of a user (sections V-B, V-C, V-D and V-E). Finally, we provide experiments on fine-grained gesture recognition.

A. WiFi Traffic in Urban Spaces

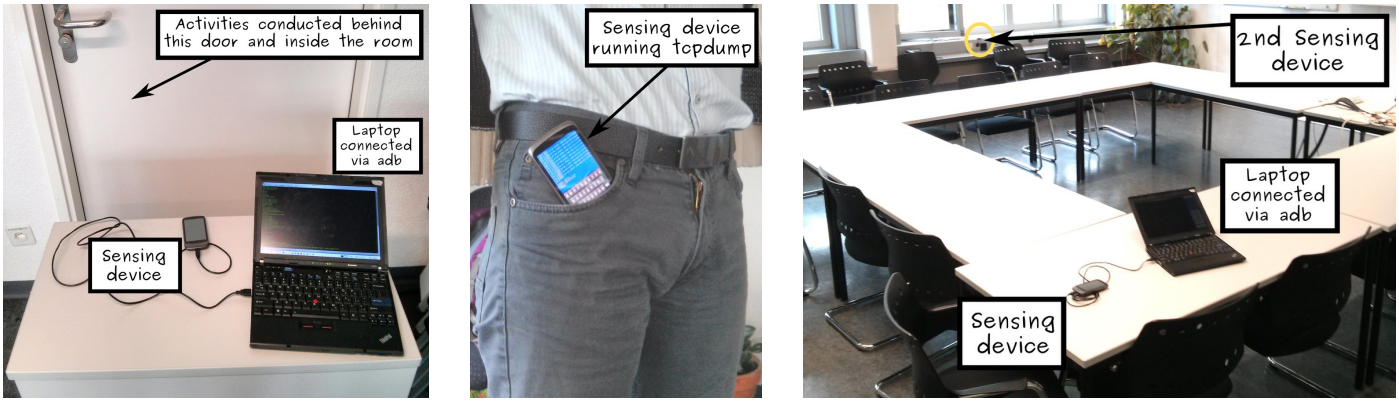
For the recognition of activities and gestures from RSSI, the rate of incoming packets is essential since this is the rate of fluctuation induced by environmental stimuli. We sampled packets over some days at various locations in Zurich to estimate a typical rate of packets in urban places. Figure 4 shows the number of packets per second from the most active AP at various locations on all channels. Short packets, such as acknowledgements, were removed (cf. section IV).

The locations span a University building at two distinct floors, a dormitory, a café in the city center, the main train

[5]The python tools to extract and process RSSI information from pcap files and to classify situations are available at http://www.stephansigg.de/DeviceFree/pcapTools.tar.gz



(a) Office environment at ETH (b) Lecture room at TU-BS (c) Scenario for the distinction of walking speed



(d) Activities conducted behind a closed door (e) Sensing device inside pocket (f) Meeting room at ETH

Fig. 3: Environments for our case studies. Surrounding furniture and objects were intentionally altered in all cases.

station and a flat in a suburb of Zurich. Only the University locations share APs. All other locations are well separated over the city. All locations have characteristic properties. While the café has the most equally distributed traffic over all channels, in the dormitory, traffic is clustered in few channels. University locations feature few, heavily trafficked channels while at the suburban flat only few channels are frequented. In all cases, we find at least one channel with 10 or more packets per second from a single AP. While this most frequented channel might differ spatially, a brief scan easily reveals most suited channels.

Since the receiver has no impact on the packet arrival rate, it relies on traffic from other devices. We considered the impact of the RSSI samples per second on the classification accuracy. In the case study (cf. figure 3a), we distinguish an empty office with the mobile phone lying on a table, the same room with a person walking next to the table and a person holding and handling the phone. Recordings were taken over four days at different times of day. Each activity is sampled for five minutes in a row. This was repeated on each day twice for all activities.

Table Ia shows the classification accuracy (CA), information score (IS), Brier score and area under the ROC[6] curve (AUC) [30], [31]. The IS measures how well a classifier learned a data set. It is higher when the correct class is predicted more often. Brier score measures the mean squared difference between a predicted probability for an outcome and the actual class. AUC is the probability that a classifier ranks a random positive instance higher than a random negative one.

	CA	IS	Brier	AUC
5 samples/s	.593	.594	.512	.813
7 samples/s	.607	.622	.502	.814
10 samples/s	.652	.703	.446	.831
15 samples/s	.671	.806	.408	.856
20 samples/s	.836	1.127	.229	.957

(a) Performance of a k-NN classifier with distinct sample rates

		Classification			
		activity	empty	holding	recall
Gr. truth	activity	**.829**	.014	.157	.829
	empty	.021	**.921**	.057	.921
	holding	.207	.036	**.757**	.757
	precision	.784	.949	.779	

(b) Confusion matrix for the k-NN classifier with 20 RSSI samples/sec

TABLE I: Impact of the sample rate on the classification

For these results, we used a k-NN with $k = 20$ (best results reached with $k \in [10..20]$), and a 10-fold cross validation. While higher sample rates improve accuracy, also 10 to 15 samples per second allow an indication about a class. The Confusion matrix for 20 samples per second is depicted in table Ib. Observe that activity and holding suffer from slight confusion. In the empty room almost no confusion is seen. Then the signal is stable and not influenced by movement.

[6]Receiver Operating Characteristic

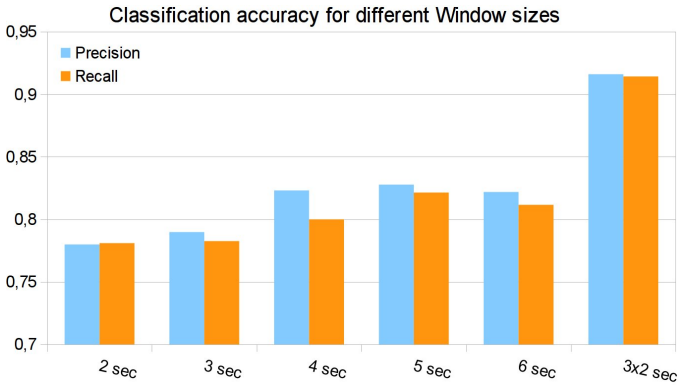


Fig. 5: Accuracy for the distinction between three basic cases with varying feature window size. A majority vote over three windows of 2 seconds outperforms greater windows

	Distance [meters] .5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	em	CA	IS	Brier	AUC
1 AP	x							x	x	.809	1.115	.258	.939
				x				x	x	.730	.796	.434	.866
							x	x	x	.528	.472	.599	.743
		x		x		x		x	x	.483	.933	.644	.831
	x	x	x	x	x	x	x	x	x	.379	1.19	.762	.823
2AP	O	O	O	O	O	O	O	O	O	.427	1.329	.722	.857

(a) Performance using 1 (x) and 2 (O) APs

		Classification .5m	4.0m	empty	recall
Gr. truth	.5m	**.981**	.019		.981
	4.0m	.026	**.768**	.206	.768
	empty	.013	.310	**.677**	.677
	precision	.962	.700	.766	

(b) Classification accuracy with fairly separated locations

	0.5m	1.0m	1.5m	2.0m	2.5m	3.0m	3.5m	4.0m	empty
0.5m	0.736	0.171	0.057	0.021	0.014	0	0	0	0
1.0m	0.1	0.471	0.114	0.236	0.029	0.014	0.014	0.014	0.007
1.5m	0.021	0.15	0.457	0.207	0.093	0.05	0.021	0	0
2.0m	0.007	0.164	0.2	0.336	0.071	0.093	0.057	0.071	0
2.5m	0	0.079	0.086	0.157	0.214	0.171	0.093	0.157	0.043
3.0m	0	0.079	0.064	0.093	0.171	0.121	0.243	0.136	0.093
3.5m	0	0.036	0.029	0.114	0.093	0.164	0.264	0.2	0.1
4.0m	0	0.036	0.021	0.114	0.129	0.136	0.171	0.221	0.171
empty	0	0.021	0.014	0.007	0.057	0.05	0.114	0.15	0.586

(c) Confusion matrix over all distances

		Classification 0.5m	4m	empty	recall
Gr. truth	0.5m	**1.0**			1.0
	4m	.027	**.846**	.128	0.846
	empty	.007	.134	**.859**	0.859
	precision	.968	.863	.871	

(d) Accuracy with two APs

TABLE II: Classification of activity in various distances.

The classification accuracy is impacted by the sampling window size (cf. figure 5). A majority vote over three successive windows of two seconds can reach higher accuracy than a greater window size. However, since the system is more responsive with shorter windows, we choose 2s windows.

B. Distance to the phone

How does the distance to the sensing hardware impact the capability to detect an activity. The case studies depicted in figure 3b were conducted at TU-Braunschweig over two consecutive days with repetitions of experiments on both days. On the floor, locations were marked in increasing distance of 0.5m up to 4.0m. At these locations, an individual walked around or move for at least 5min for each distance and day.

We investigated the distinction between an empty environment, a person moving in 4 meters distance and a person moving closer to the mobile phone (cf. table IIb). The classification accuracy deteriorates when the locations are closer together (cf. table IIa). However, when we tolerate an error of about 0.5-1m, reasonable accuracy can be achieved (cf. table IIc). Furthermore, distance to an activity can be estimated

		Classification s 1	s 2	s 3	s 4	recall
Gr. truth	side 1	**.486**	.193	.121	.2	.486
	side 2	.3	**.321**	.086	.293	.321
	side 3	.286	.136	**.179**	.4	.179
	side 4	.221	.121	.214	**.443**	.443
	precision	.376	.417	.298	.332	

(a) Naive Bayes

		Classification s 1	s 2	s 3	s 4	recall
Gr. truth	side 1	**.471**	.243	.15	.136	.471
	side 2	.314	**.279**	.236	.171	.279
	side 3	.3	.314	**.214**	.171	.214
	side 4	.293	.257	.171	**.279**	.279
	precision	.342	.255	.278	.368	

(b) Classification tree

		Classification s 1	s 2	s 3	s 4	recall
Gr. truth	side 1	**.421**	.207	.221	.15	.421
	side 2	.271	**.336**	.214	.179	.335
	side 3	.271	.186	**.279**	.264	.279
	side 4	.221	.143	.264	**.371**	.371
	precision	.355	.385	.285	.385	

(c) k-NN classifier

TABLE III: Confusion matrices for the distinction of the direction in which a person was performing activities

from RSSI. In conclusion, there is good potential to classify activities also in this distance so that for indoor environments a mobile phone can cover a typical room sufficiently.

In addition, we employed another equally active AP operating at the same frequency. Although the signal strength between both differed by about 10 dB, classification accuracy was comparable using packets from either AP. In addition, when features are created from RSSI information of both APs, the accuracy can be further improved (cf. table IId). We used the same features for both access points, effectively doubling the number of features for one time window.

C. Direction of Movement or Activity

To identify locations of performed activities, in addition to distance also relative direction must be distinguished. We conducted a study in the environment depicted in figure 3c in which the mobile was placed in the center of a 2m×2m table. In parallel to the four borders of the table a subject conducted activity (walking up and down) in approximately 1m distance. In figure 3c, the regions are marked as 'Area 1–4'. The experiment was repeated multiple times for each side and each time for at least five minutes continuously.

We then attempted to distinguish at which side the activity was performed. However, it turned out that it is hardly possible to tell this from the RSSI. We were not able to find a subset of features that would achieve reasonable accuracy with three distinct classifiers[7] (cf. table III for exemplary results).

D. Detection of activity behind a door/wall

WiFi signals can traverse obstacles such as walls or doors but the signal will be damped at this occasion so that the recognition of activity based on this data might be more challenging. We distinguished activity inside or outside a room. As depicted in figure 3d, we placed the phone inside a room next to the door. Then, a person was present and moving either inside or on the other side of the door. In the third case, nobody was present in the room or outside on the corridor. For each case, RSSI samples have been recorded for at least five minutes. Table IV depicts the results. While all three cases can be distinguished, the activity conducted outside the room is indeed most confused. This is, because although there is increased fluctuation, signals are weak so that classes are more likely confused for one of the other classes which represent either stronger activity or weakly fluctuating signals.

[7]For the results depicted in this table, we utilise a Naive Bayes classifier with 100 sample points and a Loess window of .5, a classification tree with two or more instances at its leaves and a k-NN classifier with $k = 20$

	CA	IS	Brier	AUC
Naive Bayes	.710	.784	.423	.880
Classification tree	.669	.855	.663	.795
k-NN	.724	.843	.393	.903

(a) Performance of various classifiers

Gr. truth \ Classification	empty	inside	outside	recall
empty	**.814**	.036	.150	.814
inside	.064	**.743**	.193	.743
outside	.2	.186	**.614**	.614
precision	.755	.770	.642	

(b) Confusion matrix

TABLE IV: Classification of activity inside and outside a room

	CA	IS	Brier	AUC
5 samples/s	.681	.777	.409	.881
10 samples/s	.717	.823	.388	.894
15 samples/s	.767	.910	.355	.905

(a) Performance for different sampling rates

Gr. truth \ Classification	$.5\frac{m}{sec}$	$1.0\frac{m}{sec}$	$2.0\frac{m}{sec}$	recall
$.5\frac{m}{sec}$	**.864**	.071	.064	.864
$1\frac{m}{sec}$	.121	**.657**	.221	.657
$2\frac{m}{sec}$	.050	.171	**.779**	.779
precision	.834	.730	.732	

(b) Confusion of walking speeds

TABLE V: Classification of walking speed ($k = 18$; $10\frac{\text{samp.}}{\text{sec}}$)

	CA	IS	Brier	AUC
Phone 1	.759	1.309	.354	.946
Phone 2	.805	1.397	.304	.937

(a) Performance of a k-NN classifier with data from various phones

Gr. truth \ Classification	0P	1P	5P	10P	recall
0 Persons	**1.0**				1.0
1 Persons		**.857**	.129	.014	.857
5 Persons		.129	**.671**	.2	.671
10 Persons		.114	.193	**.693**	.693
precision	1.0	.779	.676	.764	

(b) Confusion matrix (Phone 2)

TABLE VI: Classification of crowd (k-NN; 20 samples/s)

Gr. truth \ Classification	Stat.–Empty	Stat.–pres.	Walking–Empty	Walking–pres.	recall
Stationary–Empty	**.921**	.079			.921
Stationary–presence	.15	**.693**	.079	.079	.693
Walking–Empty		.171	**.457**	.371	.457
Walking–presence		.107	.364	**.529**	.529
precision	.860	.660	.508	.540	

TABLE VII: Classification of presence when device is carried

E. Detection of Walking Speed

Walking speed can be derived from signal strength with an SDR-based active device-free system [20]. We investigate the performance of a passive RSSI-based system. In the setting shown in figure 3c, a person walked around the table with the mobile phone in its center in a distance of about 2m. The phone sampled the RSSI while the person was moving at $0.5\frac{m}{sec}$, $1\frac{m}{sec}$ and $2\frac{m}{sec}$. This experiment was conducted for at least 5 minutes at each recording and repeated for each velocity twice and also clockwise and counter-clockwise. The speed was controlled autonomously by the subject. For this purpose we marked the circle with an interleaving of 1m and equipped the subject with a stopwatch so that she could adjust her speed. Best accuracy was achieved considering median, mean, minimum and standard deviation. Results are depicted in table V. All velocities can be well distinguished. The confusion is greater for velocity pairs that are closer to each other.

F. Sensing Crowd

An important ingredient for context-recognition is the size of the surrounding crowd. Different sizes can indicate different situations. For instance, having a conversation between few people or listening to or giving a talk in a meeting with multiple people. We attempted to distinguish between the empty room depicted in figure 3f and the same room occupied by 1, 5 or 10 persons. In the room, two phones where placed to record the RSSI. Phone 1 is located near the entrance on a table and the second one is placed beside a window across the room. The latter was farther away from the nearest AP which is located right next to the door outside the room. For the case study, 10, 5, 1 or no person would be present for at least five minutes. Participants were instructed not to stand still for longer periods of time but otherwise should move or act freely. They have then, for instance, moved around, stood in front of a poster and discussed it or leaned over a map to plan a weekend trip. Table VI shows that this broad distinction of the number of persons present is possible with reasonable accuracy. Empty room is perfectly recognised with 100% of accuracy. While different crowd sizes are confused in particular for the 5 and 10 persons, the performance is still far above random guess.

Observe that with the data captured by the phone placed near the window (Phone 2) the recognition accuracy is higher. We account this to the fact that individuals in the room continuously have resided in the area between the AP outside the room and the window. Therefore, the impact on the WiFi packets due to blocking or damping was greater for this phone.

G. Detect Activity while the Device is Carried

When the phone is carried, we expect significant noise for the recognition of situations from packets blocked by the user carrying the phone. We investigated whether RSSI can still be utilised to classify simple situations. For instance, it might be possible to derive whether a person is alone or in company. For this study, the phone was carried in the pocket of a person (cf. figure 3e). Then, the person was standing or walking alone and while another person was walking in proximity. For each class, data has been recorded for at least five minutes. Table VII depicts the results. We are well able to detect whether the person wearing the device is stationary and alone or if there is movement either by the device holder or by someone else. However, when the device holder is herself moving, the distinction of other activity is more confused.

H. Recognition of gestures

Finally, we considered the recognition of gestures. For this study, we placed the phone on the table as depicted in figure 3a and performed 10 single-handed gestures (figure 6). Each gesture lasted for approximately 0.4 to 0.7 seconds and was performed in a distance of about 1cm to 20cm. Only for two gestures, *Take up* and *Wipe*, the hand was in actual physical contact with it. Each gesture was repeated 100 times for a total of 1100 recordings of the distinct cases.

Best results have been achieved with the features *mean, variance, signal peaks within 10% of the maximum* and the *fraction between the mean of the first and second half of a feature window*. Table VIII shows the classification results. We observe that, while some gestures have a reasonable accuracy and average results are far above guess, a high confusion for other classes inhibits correct classification. In particular, the gesture *Hold over* can hardly be be distinguished. Furthermore, some of the swiping gestures are confused. Therefore, we merged cases into a single gesture. Table IX shows two levels

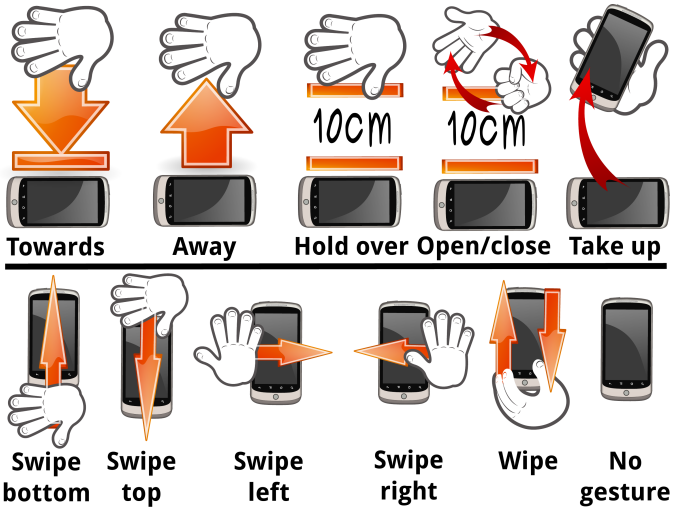


Fig. 6: Gestures performed

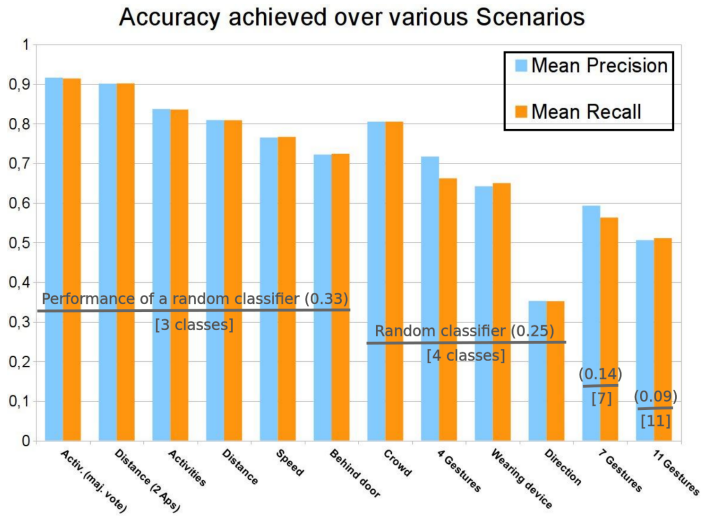


Fig. 7: Accuracies achieved for various scenarios considered

Ground truth \ Classification	Away	Hold over	Towards	No gesture	Open/close	Take up	S. bottom	S. left	Wipe	S. right	S. top	recall
Away	**.54**				.06	.03	.15	.12		.05	.05	.540
Hold over		**.26**	.16	.05	.16	.08	.03	.06	.04	.14	.02	.260
Towards		.09	**.71**	.07	.04	.01	.01		.06		.01	.710
No gesture		.04	.06	**.67**	.05	.01		.01	.15		.01	.670
Open/close		.1	.07	.09	**.47**	.07	.01		.14	.03	.02	.470
Take up	.01	.08	.03	.02	.09	**.46**	.06	.03	.09	.06	.07	.460
S. bottom	.13	.06		.01	.04	.09	**.36**	.06		.2	.05	.360
S. left	.12	.01	.01	.01		.08	.07	**.49**		.07	.14	.490
Wipe		.04	.1	.08	.16	.09	.01		**.51**		.01	.510
S. right	.03	.03		.01	.03	.01	.1	.01	.01	**.68**	.09	.680
S. top	.07	.02	.01		.03	.08		.21		.11	**.47**	.470
precision	.600	.356	.617	.663	.416	.455	.450	.495	.510	.507	.500	

TABLE VIII: Confusion matrices for the distinction of gestures

of merging gestures. When merging to 7 distinct gestures[8] we achieve a mean accuracy of about 0.56. In the table, labels are shortened to the first two letters for space limitations. While most gestures are well recognized, especially the swipe gestures still achieve mediocre performance. When further reducing to the four gestures *away, towards, no gesture* and *swipe* by merging all swipe gestures, an average accuracy of 0.66 is achieved. This suggests that some gestures can indeed

[8] *Hold over, Open/close, Take up* and *Wipe* were labelled as *No gesture*

Classification	Aw	No	To	Sb	Sl	Sr	St	recall
Aw	**.58**	.09		.13	.11	.05	.04	.58
No		**.872**	.05	.014	.012	.034	.018	.872
To		.4	**.59**				.01	.59
Sb	.15	.22		**.32**	.04	.22	.05	.32
Sl	.12	.11	.01	.06	**.48**	.08	.14	.48
Sr	.04	.15		.06	.01	**.67**	.07	.67
St	.03	.18	.01	.01	.24	.1	**.43**	.43
prec	.630	.791	.686	.492	.511	.519	.518	

(a) Confusion of 7 distinct gestures – all remaining gestures shifted to 'no gesture'

Classification	Away	Towards	No gesture	Swipe	recall
Away	**.45**	.06		.49	.45
Towards		**.834**	.052	.114	.834
No gesture		.41	**.56**	.03	.56
Swipe	.063	.128	.005	**.805**	.805
precision	.643	.810	.667	.747	

(b) Confusion of 4 gestures

TABLE IX: Performance with fewer gestures

be utilised to interact with phones or other WiFi capable devices. Possible applications cover a touch-free, frictionless interface to control mobile devices also through clothes, an extended interface for wearable devices or interface-free devices in an Internet of Things.

I. Discussion

We have investigated a passive, device-free RSSI-based activity recognition system considering several situations captured by a mobile phone. Figure 7 summarises shows the accuracies achieved in our case studies relative to a random classifier[9] as a baseline. In general, the overall accuracy falls with increasing number of classes to distinguish. However, short of the recognition of direction, the results are far above random guess in all cases. The simple distinction of distance and three well separated situations reached best results and could be further improved by considering multiple APs or majority votes over several windows of features.

The failure in the distinction in which direction activity was performed indicates limitations of passive device-free RSSI-based recognition. Since the system has to rely on data transmitted from an AP which can be located in an arbitrary direction and the device might be in arbitrary orientation, it is hardly possible to obtain fine grained information on environmental situation. The sequence of received RSSI samples is highly bursty and of low granularity and rate. Consequently, the classes that can be distinguished are limited too. Additional studies we conducted on the recognition of further activities (sitting, standing, walking, reading, typing on a computer) could not yield a useful recognition accuracy.

However, our results show that an RSSI-based passive device-free recognition system can provide basic environmental awareness when classical phone-based recognition systems fail (e.g. when the phone is not carried on the body). In addition, for special cases such as the distinction of gestures where movement is conducted in close proximity to the device,

[9] The random Classifier takes each possible choice with equal probability

RSSI-based passive recognition might provide an innovative ad-hoc alternative to more complex solutions.

Unfortunately, our solution requires a modified WiFi firmware, root access and is currently limited to a small set of phones. Much work is still required in order to allow operating-system supported non-root access to RSSI information in sufficient frequency.

VI. Conclusion

We have proposed and discussed the utilisation of RSSI information from mobile phones for the characterisation of situations, activities and gestures. We reported problems to be solved for the acquisition of RSSI from received packets on mobile phones and discussed the structure of the data as well as features suited for the recognition of activities and gestures.

In case studies we investigated the feasibility of RSSI-based recognition on mobile phones for multiple scenarios. Summarising, these results show that it is possible to distinguish simple activities and to some extent also gestures from RSSI fluctuation captured by a mobile phone. However, it also shows the limitations of this device-free recognition approach for instance, regarding a localisation of activities. Furthermore, the accuracies achieved stay below what would be possible with classical sensors such as accelerometers.

However, we could demonstrate, that there is a good potential to extend the perception of a phone beyond its boundaries into the environment. A recognition in distances of 4 meters is still feasible. RSSI-based recognition can cover cases where classical sensors can not provide meaningful results.

Regarding the recognition of gestures, we see a good potential to extend the interface of body-worn devices with RSSI-based gesture recognition.

References

[1] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing," in *Mobicom 2012*, 2012, pp. 173–184.

[2] P. Lukowicz, A. Pentland, and A. Ferscha, "From context awareness to socially aware computing," *IEEE Pervasive Computing*, vol. 11, no. 1, pp. 32–41, 2012.

[3] A. K. Dey, K. Wac, D. Ferreira, K. Tassini, J.-H. Hong, and J. Ramos, "Getting closer: An empirical investigation of the proximity of user to their smart phones," in *Ubicomp 2011*, 2011.

[4] S. N. Patel, J. A. Kientz, G. R. Hayes, S. Bhat, and G. D. Abowd, "Farther than you may think: An empirical investigation of the proximity of users to their mobile phones," in *Ubicomp 2006*, 2006, pp. 123–140.

[5] J. Aggarwal and M. Ryoo, "Human activity analysis: A review," *ACM Computing Surveys*, vol. 43, no. 3, pp. 16:1–16:43, Apr. 2011.

[6] K. Kunze and P. Lukowicz, "Symbolic object localization through active sampling of acceleration and sound signatures," in *Ubicomp 2007*, 2007.

[7] D. Schuermann and S. Sigg, "Secure communication based on ambient audio," *IEEE Transactions on mobile computing*, vol. 12, no. 2, 2013.

[8] J. M. Chaquet, E. J. Carmona, and A. FernáNdez-Caballero, "A survey of video datasets for human action and activity recognition," *Comput. Vis. Image Underst.*, vol. 117, no. 6, pp. 633–659, Jun. 2013.

[9] F. Adib and D. Katabi, "See through walls with wi-fi," in *ACM SIGCOMM'13*, 2013.

[10] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Mobicom 2013*, 2013.

[11] S. Sigg, M. Scholz, S. Shi, Y. Ji, and M. Beigl, "Rf-sensing of activities from non-cooperative subjects in device-free recognition systems using ambient and local signals," *IEEE Transactions on Mobile Computing*, vol. 99, no. PrePrints, 2013.

[12] M. Youssef, M. Mah, and A. Agrawala, "Challenges: Device-free passive localisation for wireless environments," in *MobiCom 2007*, 2007, pp. 222–229.

[13] A. Popleteev, "Device-free indoor localization using ambient radio systems," in *Adjunct Proceedings of Ubicomp 2013*, 2013.

[14] M. Seifeldin, A. Saeed, A. Kosba, A. El-Keyi, and M. Youssef, "Nuzzer: A large-scale device-free passive localization system for wireless environments," *IEEE Transactions on Mobile Computing*, vol. 12, no. 7, 2013.

[15] D. Zhang, Y. Liu, X. Guo, M. Gao, and L. M. Ni, "On distinguishing the multiple radio paths in rss-based ranging," in *Proceedings of the 31st IEEE International Conference on Computer Communications*, 2012.

[16] B. Wagner and D. Timmermann, "Adaptive clustering for device-free user positioning utilizing passive rfid," in *Adjunct Proceedings of UbiComp 2013*, ser. UbiComp '13, 2013.

[17] D. Tan, H. Sun, Y. Lu, M. Lesturgie, and H. Chan, "Passive radar using global system for mobile communication signal: theory, implementation and measurements," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 152, no. 3, pp. 116–123, 2005.

[18] F. Colone, P. Falcone, C. Bongioanni, and P. Lombardo, "Wifi-based passive bistatic radar: Data processing schemes and experimental results," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1061–1079, 2012.

[19] M. Scholz, S. Sigg, D. Shihskova, G. von Zengen, G. Bagshik, T. Guenther, M. Beigl, and Y. Ji, "Sensewaves: Radiowaves for context recognition," in *Video Proceedings of the 9th International Conference on Pervasive Computing (Pervasive 2011)*, 2011.

[20] S. Shi, S. Sigg, W. Zhao, and Y. Ji, "Monitoring of attention from ambient fm-radio signals," *IEEE Pervasive Computing, Special Issue - Managing Attention in Pervasive Environments*, accepted for publ.

[21] N. Kassem, A. Kosba, and M. Youssef, "Rf-based vehicle detection and speed estimation," in *75th IEEE Vehicular Technology Conference (VTC Spring)*, 2012, pp. 1–5.

[22] S. Sigg, S. Shi, and Y. Ji, "Rf-based device-free recognition of simultaneously conducted activities," in *Adjunct Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2013)*, ser. UbiComp '13, 2013.

[23] Y. Kim and H. Ling, "Human activity classification based on micro-doppler signatures using a support vector machine," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1328–1337, 2009.

[24] Y. Ding, B. Banitalebi, T. Miyaki, and M. Beigl, "Rftraffic: Passive traffic awareness based on emitted rf noise from vehicles," in *International Conference on ITS Telecommunications*, 2011, pp. 393 –398.

[25] S. Shi, S. Sigg, and Y. Ji, "Activity recognition from radio frequency data: Multi-stage recognition and features," in *IEEE Vehicular Technology Conference (VTC Fall)*, 2012.

[26] N. Patwari, J. Wilson, S. Ananthanarayanan, S. K. Kasera, and D. Westenskow, "Monitoring breathing via signal strength in wireless networks," 2011, submitted to IEEE Transactions on Mobile Computing, 18 Sept., 2011, available: arXiv:1109.3898v1.

[27] C. Xu, B. Firner, R. S. Moore, Y. Zhang, W. Trappe, R. Howard, and N. An, "Scpl: Indoor device-free multi-subject counting and localization using radio signal strength," in *ACM/IEEE IPSN*, 2013.

[28] M. Scholz, T. Riedel, M. Hock, and M. Beigl, "Device-free and device-bound activity recognition using radio signal strength full paper," in *Augmented Human 2013*, 2013.

[29] O. Ildis, Y. Ofir, and R. Feinstein, "Wardriving from your pocket," 2013. [Online]. Available: http://www.recon.cx/2013/slides/Recon2013-Omri%20Ildis%2c%20Yuval%20Ofir%20and%20Ruby%20Feinstein-Wardriving%20from%20your%20pocket.pdf

[30] I. Kononenko and I. Bratko, "Information-based evaluation criterion for classifier's performance," *Machine Learning*, vol. 6, no. 1, pp. 67–80, 1991.

[31] K. A. Spackman, "Signal detection theory: valuable tools for evaluating inductive learning," in *6th international workshop on Machine learning*, 1989, pp. 160–163.