RoomSense: An Indoor Positioning System for Smartphones using Active Sound Probing

Mirco Rossi*
mrossi@ife.ee.ethz.ch

Julia Seiter*
jseiter@ife.ee.ethz.ch

Oliver Amft† *
amft@tue.nl

Seraina Buchmeier*
buserain@ethz.ch

Gerhard Tröster*
troester@ife.ee.ethz.ch

ABSTRACT
We present RoomSense, a new method for indoor positioning using smartphones on two resolution levels: rooms and within-rooms positions. Our technique is based on active sound fingerprinting and needs no infrastructure. Rooms and within-rooms positions are characterized by impulse response measurements. Using acoustic features of the impulse response and pattern classification, an estimation of the position is performed. An evaluation study was conducted to analyse the localization performance of RoomSense. Impulse responses of 67 within-rooms positions from 20 rooms were recorded with the hardware of a smartphone. In total 5360 impulse response measurements were collected. Our evaluation study showed that RoomSense achieves a room-level accuracy of $> 98\%$ and a within-rooms positions accuracy of $> 96\%$. Additionally, the implementation of RoomSense as an Android App is presented in detail. The RoomSense App enables to identify an indoor location within one second.

Categories and Subject Descriptors
H.3.4 [**Information storage and retrieval**]: Systems and software

Keywords
indoor positioning, room impulse response, pattern recognition, algorithms

1. INTRODUCTION
Indoor positioning is an essential part of context information and useful for various location-based services that can augment human capabilities, including indoor way-finding in buildings, patient localization, and tour guides [11, 19, 2]. While outdoors, mostly GPS, WLAN, and cellular information can be used to robustly estimate location [18], adequate network coverage is often lacking for locating within buildings [7]. Numerous approaches have already been proposed for indoor positioning using smartphones and other devices (see related work in Section 2 for details). Most notably, methods that used additional ambient infrastructure, such as beacons and wireless transmitters were successfully deployed. Among the indoor positioning methods that do not require additional infrastructure, positioning using acoustic patterns in rooms have large application potential.

Passive sound fingerprinting is performed by measuring the acoustic background spectrum of rooms. Passive fingerprinting depends on low variance in the acoustic spectrum and requires measurement times of multiple seconds to perform position estimations [18]. In contrast, an active acoustic fingerprint can be found by emitting a sound chirp and measuring the impulse responses. Different methods have been studied to derive an impulse response within a room [17]. However, the acoustic patterns were not analysed using pattern recognition methods. Moreover, the active sound fingerprinting has yet not been deployed and tested on smartphones to determine practicality for location-based services. Hence, it is not clear whether rooms and different positions within a room could be recognised using previously trained sound patterns with phones. Moreover, a short measurement duration is vital to determine the current position in many real-life applications. Due to the active probing, we expect that position estimation can be performed in far shorter times than required for the passive fingerprinting.

In this work we present RoomSense, a smartphone-based system to quickly determine indoor location. Our approach considers standard phones, thus the default smartphone microphone and speaker was used. Using the acoustic impulse response, we recognise room location within a building floor, similar rooms at different building levels, and different positions within a room. We selected 20 rooms and a total of 67 positions according to locations visited in the typical daily life of a university student.

In particular, this paper provides the following contributions:

1. We present the system architecture of RoomSense, which is designed to provide instantaneous indoor position estimates on two resolution levels: rooms and within-rooms positions. We use the Maximum Length Sequence (MLS) impulse response and a Support Vector Machine (SVM) based position recognition techniques to realise RoomSense. We identify the best-performing audio feature sets and further parameters to obtain robust estimates.
2. We evaluate RoomSense in a study comprising of recordings of impulse response measurements from 20 rooms and

*Wearable Computing Lab., ETH Zurich
†ACTLab, Signal Processing Systems, TU Eindhoven

totally 67 positions using a standard smartphone. Besides performance of room and within-rooms positioning, we vary the number of trained positions per room area. Finally, we evaluate accuracy when the signal-to-noise-ratio (SNR) was reduced.

3. We describe the implementation of RoomSense as an Android App. The App was designed to recognise a room or position within rooms in less than one second. The App can be used to learn new rooms or positions within rooms instantly.

The remainder of the paper is structured as follows: Section 2 describes related work on indoor positioning and impulse response methods. In Section 3, the system architecture is presented and Section 4 describes the evaluation study performed. Our evaluation results are presented in Section 5. Section 6 details the implementation of RoomSense as an Android App. Finally, Section 7 concludes the paper.

2. RELATED WORK

Indoor positioning is an actively researched field. Various approaches have been proposed using additional ambient infrastructure such as sensors or transmitters that were installed in buildings to localize a wearable device [5, 19, 3]. Infrastructure-based methods have a typical location error of less than one meter. However, a dedicated technical infrastructure is needed for the localization, which is not always practical or affordable.

Alternative approaches are using already existing wireless infrastructure such as cellular network and Wi-Fi information for the localization task [11, 21, 12]. Here, the signal strength of cellular or Wi-Fi station is used to determine location of a mobile device and the position of cells and network stations is known in advance. When using wireless infrastructure, localization accuracy depends on the density of cellular/Wi-Fi stations in the environment. E.g., Haeberlen et. al. reported an accuracy of $95\,\%$ over a set of 510 rooms [11]. However, at least five Wi-Fi stations were in range at all measurements. The positioning approach is less suitable where station coverage is unknown or sparse.

Recently sound-based positioning approaches have been proposed that require no additional infrastructure to perform indoor positioning. Passive sound fingerprinting uses ambient sound to generate position estimates, whereas active fingerprinting approaches emit and then record a specific sound pattern for the positioning. Wirz et al. [20] proposed an approach to estimate the relative distance between two devices by comparing ambient sound fingerprints passively recorded from the devices' positions. The distance was classified in one of the three distance regions ($0\,m$, $0\,m{-}12\,m$, $12\,m{-}48\,m$) with an accuracy of 80%. However, no absolute position information was obtained by this method. Tarzia et al. [18] proposed a method based on passive sound fingerprinting by analysing the acoustic background spectrum of rooms to distinguish different locations. The location was determined by comparing the measured sound fingerprint for a room with fingerprints from a database. A room's fingerprint was created by recording continuous ambient sound of $10\,s$ length. The system was implemented as an iPhone App to localise between different rooms. The localization performance was high for quiet rooms, but dropped when people were chatting or when the background spectrum had large variations. Azizyan et al. [4] used a combination of smartphone sensor data (WiFi, microphone, accelerometer, color and light) to distinguish between logical locations (e.g. McDonalds, Starbucks). Their passive acoustic fingerprints are generated by recording continuous ambient sound of $1\,min$ and extracting loudness features.

We expect an active sound fingerprinting approach to reduce recognition time compared to the passive approach and to be robust against noise. Zhang et al. [22] proposed an approach to estimate the relative distance between two devices with active sound fingerprinting. Their method was tested in a measurement range of $2\,m$ and had an median distance error of $2\,cm$. So far, only Kunze and Lukowicz presented an absolute positioning approach where active sound fingerprinting was considered [13]. Their system could recognize specific as well as more abstract locations where a phone was placed (e.g. table, floor) when combining information from acceleration and sound sensors. However, no localization on a room resolution level was considered in their work.

Different methods have been proposed to measure room impulse responses however these techniques were not applied for indoor position estimation and pattern recognition. Stan et al. [17] compared different impulse response measurement techniques, including maximum length and inverse repeated sequences, time-stretched pulses and sine sweeps. They considered the room impulse response to be one of the most important acoustical characteristics of a room. Furthermore, the MLS measurement technique showed several advantages compared to the other methods: MLS is perfectly reproducible and immune to various noise types. Furthermore, MLS is deterministic and hence allow summing and averaging of multiple repetitions to improve the signal-to-noise ratio.

In this work, we propose to use room impulse response based on MLS and pattern recognition for indoor positioning on a smartphone at room and within-room position resolutions. Instead of relying on the acoustic background spectrum as in passive fingerprinting, we characterize room acoustics using impulse response measurements.

3. ROOMSENSE ARCHITECTURE

The RoomSense system emits a short acoustic wave and measures the impulse response. This response is further processed as sound fingerprint of a within-room position and eventually the extracted sound features are classified to estimate room and within-room position. The sound pattern models of room positions are derived in a training phase based on annotated acoustic impulse response data. Figure 1 illustrates the RoomSense system architecture comprising impulse response measurement, front-end processing, and classification components. This section presents the RoomSense system architecture in detail.

3.1 Impulse Response Measurement

The first main component of the system is the impulse response measurement for an indoor position. The impulse response is a response of a dynamical system to a Dirac input impulse. It is a time-dependent function. The behaviour of a linear and time invariant system can be obtained by a convolution of the input signal with the impulse response [15]. Assuming that loudspeaker and microphone setup are motionless, the sound propagation and reflections within a room can be regarded as a close approximation to a linear and time-invariant system [10]. Room impulse responses can therefore be used to completely describe the acoustic characteristics of a position in a room. Common measurement techniques for acoustic room impulse responses are maximum length, time-stretched pulses, and sine sweeps [17].

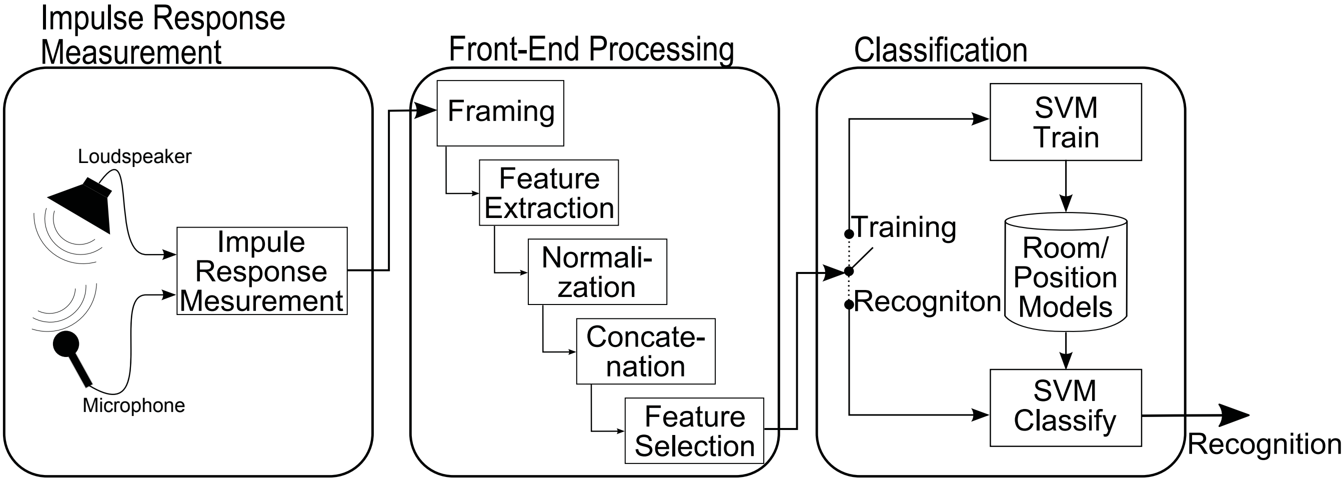


Figure 1: RoomSense architecture illustrating the main components of the system.

MLS Measurement Technique

The Maximum Length Sequence (MLS) measurement technique is based upon the excitation of the acoustical space by a periodic pseudo-random signal having almost the same stochastic properties as a pure white noise [17]. Maximum length sequences are binary, periodic signals. They are characterised by their order M. The period length of the MLS is $L = 2^M - 1$. A possible method to generate a MLS signal is to use maximal feedback shift register. The shift register can be represented by the following recursive function:

$$a_m[n+1] = \begin{cases} a_0[n] \oplus a_1[n], & m = 3 \\ a_{m+1}[n], & \text{otherwise} \end{cases}$$

where $\oplus$ denotes the XOR operation. Let the MLS signal with order M be $x[n] = a_M[n]$ and the impulse response of the LTI system be $h[n]$. The output $y[n]$ of the system stimulated by $x[n]$ can be denoted by: $y[n] = x[n] * h[n]$. Since the auto correlation of pseudo-random maximum length sequence ϕ_{xx} has approximately the shape of a delta pulse, the room impulse response can be obtained by circular cross-correlation between the determined output signal and the measured input signal. Or in other words: Taking the cross-correlation of $y[n]$ and $x[n]$, we can write: $\phi_{yx} = h[n] * \phi_{xx} = h[n]$, with the assumption ϕ_{xx} is a Dirac impulse.

System Parameters

For our system, we chose the MLS measurement technique with a common parameter set [10]. The order $M = 15$ was set and the sampling frequency was configured to $f_s = 48\,kHz$. An MLS sequence with a length of $0.68\,s$ was played by the loudspeaker and recorded by the microphone with the same sampling frequency f_s. The played MLS sequence is hearable as a short noisy sound. With this parameter set a impulse response of the time interval $t = [0, 0.68]\,s$ and frequency interval $f = [0, 24]\,kHz$ is generated. Since time-synchronisation between loudspeaker and microphone is not supported by a common smartphone's hardware, the first arriving impulse - assumed to be the largest peak in the impulse response - is considered to a fixed time $t_{fa} = 450\,ms$ within the response. An illustration of a measured impulse response is shown in Figure 3.

3.2 Front-End-Processing

Front-end processing steps aim at extracting position- and room-dependent audio features from the impulse response. Initially, the impulse response is processed in frames with a sliding window with a window size of $31ms$ and 50% overlap. Each window is smoothed with a Hamming filter. In a pre-evaluation, we found that this framing parameter setting resulted in the largest recognition performance. Similar settings could be found in other audio recognition systems, e.g. in [8]. Subsequently, audio features were extracted for each frame. Common audio features as well as specific room acoustic features have been evaluated (see Table 1). The performance results of the feature sets are presented in the evaluation (Section 5). Feature vectors f_i were extracted from each frame i. In a next step the feature vectors were normalised with $F_i = \frac{f_i - m_i}{\sigma_i}$, where m_i are the mean values and σ_i are the standard deviation values of all feature vectors of the training set. After this step, all feature vectors F_i with $i = \{1, 2, ..., n\}$ were concatenated to one feature vector $F_{All} = \{F_1, F_2, ..., F_n\}$. Finally, the Minimum-Redundancy-Maximum-Relevance (MRMR) [16] feature selection was used to select the M_{sel} most relevant features F_{SEL}. In our evaluation M_{sel} was tuned to maximise recognition performance.

type	feature names	coef
room acoustic [10]	Reverberation Time (T)	3
	Early Decay Time (EDT)	1
	Clarity (C)	2
	Definition (D)	2
	Center Time (CT)	1
common audio [14]	Auto Correlation Function (ACF)	12
	Linear Bands (LINBANDS)	10
	Logarithmic Bands (LOGBANDS)	10
	Linear Predictive Coding (LPC)	12
	Mel-Freq. Cepstral Coefficients (MFCC)	12

Table 1: Common audio features and specific room acoustic features considered in the evaluation with their number of coefficients *coef.*

3.3 Position Classification

The classification aims to generate an estimation of room and within-rooms position based on the generated feature vector F_{SEL}. We used the Support Vector Machine (SVM) classifier with a Gaussian kernel [6], which include the cost parameter C and the kernel parameter γ. These parameters were optimized with a parameter sweep as described later in the evaluation section 5. The one-against-one strategy was used, which is provided by the LibSVM library [6].

In a training phase, the training set feature vectors including the position and room labels were derived and *SVMTrain* was used to create pattern models for all rooms and within-rooms positions. In the testing phase *SVMClassify* used the stored models to classify a new feature vector F_{SEL} regarding room and within-room position.

4. EVALUATION STUDY

An evaluation study has been conducted to analyse the recognition performance of RoomSense. An impulse response dataset of 67 positions within 20 rooms was collected. The impulse response measurements of our dataset were collected with a Samsung Galaxy SII Android smartphone. Figure 2 illustrates the locations of loudspeaker and microphone at the smartphone. Distance between loudspeaker and microphone was $2.4\,cm$.

Figure 2: Samsung Galaxy SII smartphone used during the evaluation study. The distance between loudspeaker and microphone was $2.4\,cm$.

4.1 Recording Procedure and Dataset

Table 2 lists the rooms of the compiled impulse response dataset. The rooms were chosen to cover regularly visited rooms of a university student during one working day. Rooms from two different buildings were selected, marked as 'Work' (denoting an office building) and 'Home' (denoting the participant's home). Some rooms in the dataset are very similar: *work corridor 1* and *work corridor 2* have the same floor plan and furniture arrangement, whereas *work office 1* and *work office 2* have the same floor plan but a different set of furniture.

id	rooms	size $[m^2]$	positions
1	Work coffee room	20	2
2	Work corridor 1	65	7
3	Work corridor 2	65	7
4	Work entrance 1	20	2
5	Work entrance 2	20	2
6	Work lab	20	2
7	Work lecture room	50	6
8	Work meeting room 1	50	6
9	Work meeting room 2	25	3
10	Work office 1	25	3
11	Work office 2	25	3
12	Work office 3	30	3
13	Work toilet	15	2
14	Home bathroom	15	2
15	Home bedroom	25	3
16	Home corridor	15	2
17	Home entrance	15	2
18	Home kitchen	25	3
19	Home living room	30	4
20	Home office	25	3

Table 2: Overview on rooms and within-rooms positions included for the impulse response dataset. Rooms were selected according to the frequently visited places of a university student, including *Work* and *Home* buildings.

To investigate within-rooms position estimation, we selected a position for approximately every $9\,m^2$. It is conceivable that in larger rooms, the localisation service need to give more detail than in smaller ones. Depending on the size of the room, 2 to 6 recording positions were selected (see Table 2). For each within-room position, two orientations were chosen. One orientation was determined by pointing loudspeaker and microphone to the middle of the room. The second orientation was chosen in the opposite direction, thus rotated by $180°$.

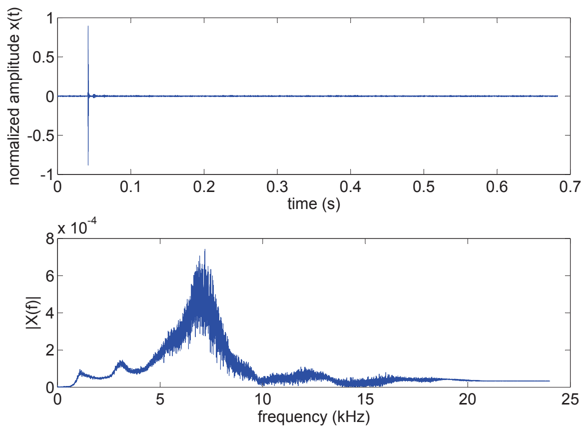


Figure 3: Illustration of an impulse response measured with a Samsung Galaxy SII smartphone. Time and frequency domain is shown.

All impulse response measurements were carried out with the Samsung Galaxy SII smartphone. During the measurement the smartphone was held with one hand at approximately $1.20\,m$ over ground in an ergonomic posture (see Figure 9). Neither hands, other body parts nor objects covered the loudspeaker and microphone (see Figure 2). An example of a measured impulse response is shown in Figure 3. Additionally, during all measurements the state of the room has been kept unchanged: Windows and doors were closed and no furniture has been moved.

For each orientation, 40 measurements were carried out. Since every position has two orientations, 80 measurements per position were gathered. Overall, 67 position within 20 rooms were defined which corresponds to 5360 impulse response measurements.

5. EVALUATION

We evaluated the recognition performance of the RoomSense system using the impulse response dataset introduced in Section 4. This section presents the results of the evaluation. In Section 5.1 the recognition performance of different feature sets is compared. The recognition accuracy for the room localization and for the within-rooms position estimation is presented in Section 5.2 and 5.3, respectively. Furthermore, the effect of noise is analysed in Section 5.4. Note that for all evaluation results the SVM parameters C and γ, and the feature selection parameter M_{SEL} (see Section 3) were swept to reach the best recognition performance.

5.1 Feature Comparison

Figure 4 depicts the room localization accuracy of the system for different feature sets (as introduced in Table 1). Positions are classified as one of the 20 rooms (see Table 2). The accuracy was computed with a leave-one-sample-out cross validation, where the tested sample is left out from the training set. Using all features the highest recognition accuracy was reached (ALL, $98.7\,\%$). A similar result was achieved by the MFCC features ($98.2\,\%$), followed by the other common audio features LINBANDS ($94.4\,\%$), ACF ($93.5\,\%$), LOGBANDS ($92.5\,\%$), and LPC ($82.4\,\%$). With the acoustic room features the lowest accuracy was reached (ACOUSTIC_ROOM, $60.3\,\%$).
Since the MFCC is the best performing feature set, this set was used for the following evaluations and for the App implementation of RoomSense.

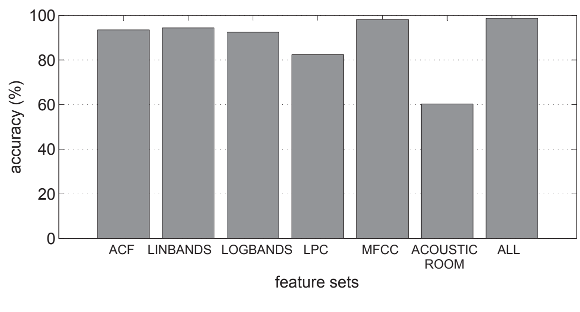


Figure 4: Performance comparison of the different feature sets.

5.2 Room Localization

As already presented in the Section 5.1 the system's room localization performance using MFCC features was $98.2\,\%$. In this evaluation it is assumed that the system is trained with impulse responses of all positions and orientations. Thus, for a room characterization impulse responses of each tested position and orientation would be needed. For the assumption that the system is trained by an orientation-independent training set, we carried out a leave-one-orientation-out cross validation, where the orientation of the tested sample is not trained. In this case the recognition performance dropped to $85.1\,\%$. This performance drop shows that impulse responses depend not only on the measured room but also on the measurement's position and orientation within the room. Nevertheless, the similarity of room impulse responses within a room compared to other rooms still enables to perform room localization. Figure 5 depicts the confusion matrix of this evaluation.

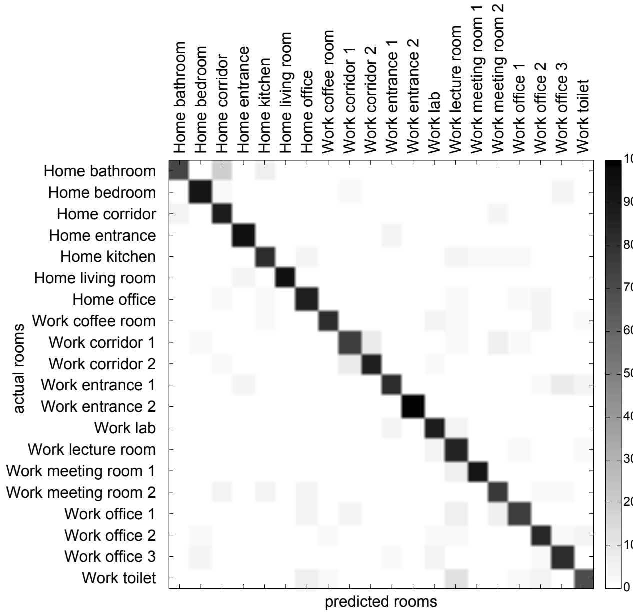


Figure 5: Room localization performance computed with a leave-one-orientation-out cross validation. The accuracy is $85.1\,\%$.

We further analysed the system's performance for different densities of training positions per room. We varied the number of training positions per room-area from one training position per $9\,m^2$ to one training position per $63\,m^2$. Additionally, positions of the test samples were never trained (expect for $9\,m^2$ where all positions were used for training). Figure 6 shows the result of this evaluation. The results show the dependency of recognition accuracy and density of the training positions per room: Performance dropped from $98.2\,\%$ for one training position per $9\,m^2$ to $49.8\,\%$ for one training position per $63\,m^2$. We conclude that impulse responses are dependent of the position of the measurement equipment. Variations of impulse responses in larger rooms are higher than in smaller rooms. Thus, to reach a room localization accuracy of $\sim 80\,\%$ at least one position every $\sim 18\,m^2$ should be trained.

5.3 Within-Rooms Position Estimation

In this section the system's within-rooms position estimation is analysed. Figure 7 shows the result in comparison to the room localization performance. For the assumption that all tested positions are exactly known by the system, we trained the system with all 67 positions on both orientations. Tested samples were then classified as one of the 67 positions. We performed a leave-one-sample-out

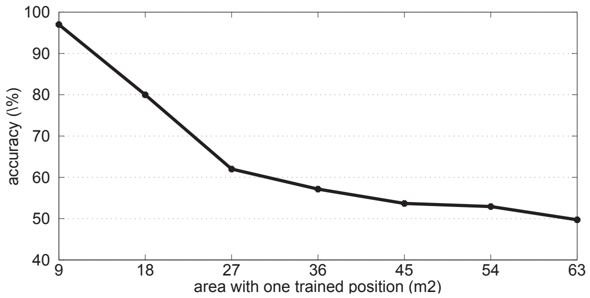


Figure 6: Room localization performance with different densities of training positions per room. The training position per room-area was varied from one position/$9m^2$ up to one position/$63m^2$.

cross validation, where the tested sample is left out from the training set. This evaluation resulted in an accuracy of $96.4\,\%$, which is similar to the room localization performance. In a second evaluation, we analysed the orientation-dependency of the within-rooms position estimation. A leave-one-orientation-out cross validation leaving out the test sample's orientation from the training set was carried out. In this case the accuracy dropped to $51.3\,\%$. We conclude that the orientation-dependency of the impulse response is high. Within-rooms position estimation is possible, however, the orientation of the tested measurement has to be trained in advance.

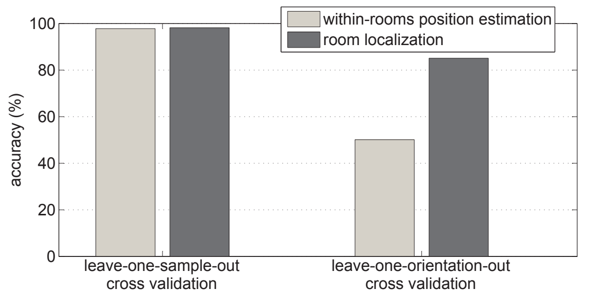


Figure 7: Accuracy of within-rooms position estimation compared to room localization. Leave-one-out cross validation was performed, where either the test sample or the test sample's orientation was left out from the training set.

5.4 Noise Robustness

Figure 8 shows the noise robustness of the room localization and within-rooms position estimation. Additive white Gaussian noise was added to the recorded maximum length sequence of the tested measurement samples. The recorded maximum length sequence is assumed to be noiseless. The SNR was varied between 10 and 50 dB. For both localization levels, a leave-one-sample-out cross validation was performed, were test sample was left out from the training set. Noise robustness is similar for both localization levels. The recognition performance constantly drops while decreasing the SNR from $98.2\,\%$ and $96.4\,\%$ for an SNR of 50dB to $66.6\,\%$ and $65.9\,\%$ for an SNR of 10dB. We conclude that localization is possible (accuracy of $> 80\,\%$) in environments with an SNR of $> 30\,dB$.

6. ROOMSENSE IMPLEMENTATION

The RoomSense system was implemented in an Android smartphone setting. The main components (see Section 3) were implemented in Java SE 7 and are running on an Android smartphone

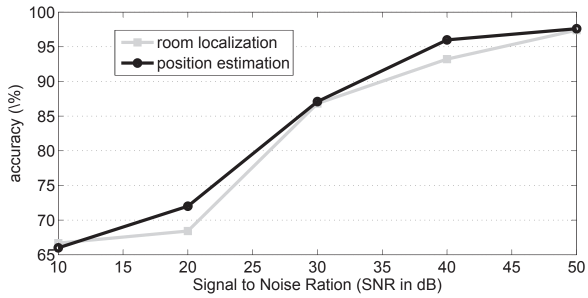
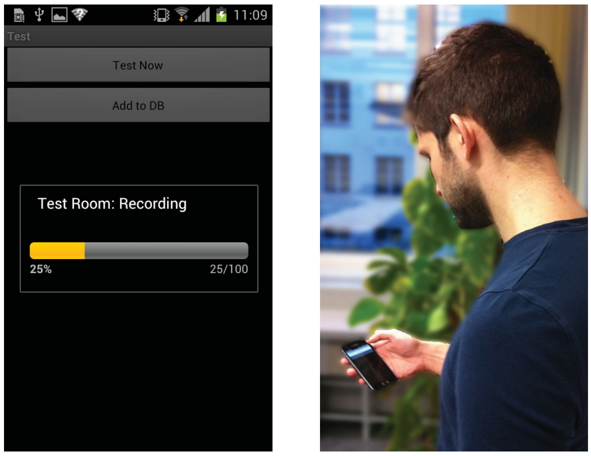


Figure 8: Noise robustness of room localization and within-rooms position estimation. Measurement samples were corrupted with white Gaussian noise. Noise level was varied between 10 and 50 dB.



(a) The UI. Recognize or train new locations.

(b) App usage during impulse response measure.

Figure 9: RoomSense user interface (UI) on an Android smartphone and its usage. Both recognizing a location and training new location is possible.

or PC environment. For the implementation we referred to [9] to implement the MLS impulse response measurement, we used the MFCC implementation of FUNF open sensing framework [1] to derive the features and the LibSVM Library [6] for the SVM modeling and prediction.

Figure 9 shows an illustration of the user interface (UI). The App can be used to recognize a location by pressing 'Test Now': The impulse response measurement is immediately started, the generated signal is processed (*front-end-processing*), and a location prediction is generated (*classification*). On the Samsung Galaxy SII the duration of the overall process is about one second whereas IR measurement requires most of the time ($0.68\,s$). Additional the App enables to extend the training set by pressing 'Add to DB'. Training data for new or existing room positions can be recorded and integrated in the room/position models.

7. DISCUSSION AND CONCLUSION

In this work we presented a new method for indoor positioning using an active sound fingerprinting approach. After characterising rooms according to the impulse response using acoustic features, pattern classification was used to estimate positions on a room and within-rooms position level.

Our evaluation study showed that our system achieved excellent recognition performance (accuracy $>$ 98 %) for localize a position in a set of 20 rooms. Our evaluation of different feature sets revealed that MFCC features outperform any other feature group, including the specific room acoustics features in the classification task. Even in a more challenging setting in which a position is localized in a set of 67 within-rooms positions an excellent accuracy ($>$ 96 %) can be achieved. For room localization the positioning performance depends on the density of positions trained. Larger rooms could still be identified (accuracy $\sim 80\,\%$), if at least one position is trained for every $\sim 18\,m^2$ of room area. Additionally, the orientation of a measurement also effects the performance of room localization. Nevertheless, the similarity of room impulse responses within a room compared to other rooms still enables to perform accurate room localization. For within-rooms positioning the orientation-dependency is higher. If the tested orientation is not trained, within-room positioning performance drops to 51.3%.

Overall, the sound localization approach presented in this work has large application potential for indoor location-based services as it requires very short measurement times until a robust position estimate can be derived. In our study, less than 1 s was required to obtain the presented estimation performances. Moreover, the impulse response measurement showed robustness against noise. We consider that the short estimation time and noise-robustness can be advantageous over passive fingerprinting approaches. Due to the use of different study conditions, our performance results are however not directly comparable with the passive approach presented by Tarzia et al. [18]. Our active sound localization method may however be unsuitable for continuous use or applications where the user is unaware of the position estimation, due to the hearable probing. Nevertheless, we expect that our method and results open opportunities for indoor location estimation applications using smartphones (e.g. using an ultrasonic frequency range, which is not hearable for humans). The smartphone implementation and system parameters proposed in this work could serve as reference.

8. ACKNOWLEDGEMENT

This work was supported by the EU FP7 Marie Curie Network iCareNet under grant number 264738.

9. REFERENCES

[1] funf|open sensing framework: http://funf.media.mit.edu/.

[2] G. Abowd, C. Atkeson, J. Hong, S. Long, R. Kooper, and M.Pinkerton. Cyberguide: A mobile context-aware tour guide. In *Wireless Networks 8*, pages 421–433, 1997.

[3] M. Addlesee, R. Curwen, S. Hodges, J. Newman, P. Steggles, A. Ward, and A. Hopper. Implementing a sentient computing system. *IEEE Computer 34,*, 34(8):50–56, 2001.

[4] M. Azizyan, I. Constandache, and R. R. Choudhury. SurroundSense: mobile phone localization via ambience fingerprinting. In *MobiCom '09 Proceedings of the 15th annual international conference on Mobile computing and networking*, pages 261–272, 2009.

[5] G. Borriello, A. Liu, T. Offer, C. Palistrant, and R. Sharp. Walrus: wireless acoustic location with room-level resolution using ultrasound. In *Proc. Intl. Conf. on Mobile Systems, Applications, and Services (MobiSys)*, pages 191–203, 2005.

[6] C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):188–205, 2011.

[7] G. Dedes and A. G. Dempster. Indoor GPS Positioning: Challenges and Opportunities. *62nd IEEE Vehicular Technology Conference*, pages 412–415, 2005.

[8] A. Eronen, V. Peltonen, J. Tuomi, A. Klapuri, S. Fagerlund, T. Sorsa, G. Lorho, and J. Huopaniemi. Audio-based context recognition. *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(1):321–329, 2006.

[9] A. Farina and F. Righini. Software implementation of an mls analyzer with tools for convolution, auralization and inverse filtering. In *Audio Engineering Society Convention 103*, 9 1997.

[10] I. O. for Standardization. Acoustics - Application of new measurement methods in building and room acoustics. ISO/DIS 18233), 2004.

[11] A. Haeberlen, E. Flannery, A. Ladd, A. Rudys, D. Wallach, and L. Kavraki. Practical robust localization over large-scale 802.11 wireless networks. In *Proc. Intl. Conf. on Mobile Computing and Networking (MobiCom)*, pages 70–84, 2004.

[12] J. Hightower, S. Consolvo, A. LaMarca, I. Smith, and J.Hughes. Learning and recognizing the places we go. In *Proc. Intl. Conf. on Ubiquitous Computing (UbiComp)*, pages 159–176, 2005.

[13] K. Kunze and P. Lukowicz. Symbolic Object Localization through Active Sampling of Acceleration and Sound Signatures. In *Proceedings of the 9th international conference on ubiquitous computing*, pages 163–180, 2007.

[14] D. Mitrovic, M. Zeppelzauer, and C. Breiteneder. Features for content-based audio retrieval. *Advances in Computers*, 78:71–150, 2010.

[15] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab. *Signals and systems*. University of Michigan. Prentice Hall, 1997.

[16] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.

[17] G. Stan, J. Embrechts, and D. Archambeau. Comparison of different impulse response measurement techniques. *Journal of the Audio Engineering Society*, 50(4):249–262, 2002.

[18] S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik. Indoor Localization without Infrastructure using the Acoustic Background Spectrum. In *Proceedings of the 9th international conference on mobile systems, applications and services*, pages 155–168, 2011.

[19] R. Ward, A. Hopper, V. Falcao, and J.Gibbons. The active badge location system. *ACM Trans. Information Systems*, 10(1):91–102, 1992.

[20] M. Wirz, D. Roggen, and G. Tröster. A wearable, ambient sound-based approach for inrastructureless fuzzy proximity estimation. In *Proceedings of the 14th IEEE International Symposium on Wearable Computers*, pages 1–4, 2010.

[21] M. Youssef and A. Agrawala. The horus wlan location determination system. In *Proc. Intl. Conf. on Mobile Systems, Applications, and Services (MobiSys)*, pages 2005–2018, 2005.

[22] Z. Zhang, D. Chu, X. Chen, and T. Moscibroda. Swordfight : Enabling a new class of phone-to-phone action games on commodity phones categories and subject descriptors. In *the International Conference on Mobile Systems, Applications, and Services (MobiSys'2012)*, pages 1–14, 2012.