Acoustic Self-calibrating System for Indoor Smartphone Tracking (ASSIST)

Fabian Höflinger, Rui Zhang, Joachim Hoppe, Amir Bannoura, Leonhard M. Reindl
Department of Microsystems Engineering,
University of Freiburg, Germany
{fabian.hoeflinger, rui.zhang, reindl}@imtek.uni-freiburg.de

Johannes Wendeberg, Manuel Bührer, Christian Schindelhauer
Department of Computer Science,
University of Freiburg, Germany
{wendeber, schindel}@informatik.uni-freiburg.de

***Abstract*—In this paper, we present a novel smartphone indoor localization system. The smartphone user is localized with small effort, affordable equipment and with high accuracy in indoor areas. The system uses commercially available smartphones generating high pitched acoustic chirp signals beyond the audible range. The chirp signals are received by sound receivers which identify the specific sound produced from each smartphone. The receivers are connected to a WiFi network, such that they synchronize their clocks and exchange the time differences of arrival (TDoA) of the received chirps. In this way, using an iterative multilateration algorithm, the location of the smartphones can be calculated and the receiver positions are calibrated automatically. For generating the specific sound signals from the smartphone and for user navigation an Android software application was developed. The user interface is simple and is invoked by starting the software application, which automatically connects to a server and receives an ID using the internet connection of the smartphone. Furthermore, the user is assigned specific parameters, such that several devices can be distinguished by the appearance of the chirps. The position of the user is displayed on the smartphone in context of the environment, with a map and surrounding items.**

In the presented work we have verified our system in a real-world scenario. We compared our trajectory of a pedestrian carrying smartphone to the reference positions. We could locate the smartphones with error margin of 30 cm.

a centimeter margin of error.

Keywords: indoor localization, smartphone tracking, sound localization, TDoA, anchor-free

I. Motivation

From the sustained rise and ubiquitous availability of mobile computers, smartphones and handheld devices in every-day life, a multitude of exciting new location-dependent applications have emerged. Context sensitive applications support the user in everyday life. One of the most important contexts is user location for navigation. The demand for navigation in large structures as railway stations, airports, trade fair halls, or department stores is obvious, since the equipment – the mobile devices of the people – are already available.

The GPS-Module in commercial off-the-shelf (COTS) smartphones, and handhelds make navigation systems reliable to assist in outdoor areas [1]. The demands of localization systems begin to shift towards closed scenarios. For indoor environments, there is the need for new localization approaches since reliability of GPS vanishes in densely built-up urban areas and is completely void inside buildings. In additional, to effectively navigate people in their environments, for example to specific products in a supermarket or to particular exhibition booths on trade fairs, a more accurate localization system as GPS is needed. Hence, for indoor applications alternative technologies are required to provide the signal inside buildings with a low cost infrastructure.

978-1-4673-1954-6/12/$31.00 ©2012 IEEE

II. Related Work

A. Basic Indoor-Localization

Today many different indoor localization systems are available based on different methods, some of these systems work with COTS smartphones. Requirements for additional hardware are reduced when these localizations systems are employed for localization. In case of using a COTS smartphone for participating, which is already available for the user, cost of the localization system is reduced, especially for the user side.

Many present localization system use radio frequency (RF) signals for localization. The RF-Systems uses the propagation of radio waves which can utilize the available infrastructure. A brief description of indoor localization systems based on RF is presented here.

Current smartphones have implemented a WiFi-Modul to communicate with an infrastructure. RADAR [2] operates with existing multiple WiFi-Access Points. Via Received Signal Strength Indicator (RSSI) the distances between the WiFi-Access Point and the mobile phone can be calculated. Otsason *et al.* used the GSM communication with wide signal-strength fingerprints to locate the user in indoor environments [3]. Redpin considered the signal-strength of GSM, Bluetooth and WiFi-Access Points on a mobile phone to calculate the position [4].

Using a RF-System for localization people can navigate with low accuracy (1.5 m - 3 m). Through combination with other technologies the accuracy can be improved. With a combination of an infrastructure based localization system (e.g. RSSI localization with WiFi-Access Points) the position can be updated as reported in [5]. The multi-method approach [6] uses a combination of built-in sensors of mobile devices and the capabilities of the end-users, which estimates positions with a scanner application. Kim *et al.* presented a smartphone localization system based on WiFi-Access Points and inertial sensors. For localization with inertial sensors, no infrastructure is required, but the drift of the sensors and the errors are accumulated during the integration of the measurement values, which increases the localization error.

However, to sum up the RF-Systems are susceptible to errors in dynamic environments. For example, the RSSI value depends on the distance and on the environment. The RSSI value is distorted by objects in the direct path, in the vicinity and by environmental influences, like air humidity etc.. Additionally, the RSSI value also depends on the orientation of the antenna. The antenna directivity depends on the orientation and is not isotropic.

RF-Systems using Time of Flight (ToF) of electromagnetic waves exhibit measurement errors from multipath propagation in indoor environments, which is due to reflections on walls and objects, which can lead to false distance measurements.

Other smartphone localization systems use only the visual information of the surrounding [7], [8]. The integrated camera of the smartphones is used to create images and compare the images with a database. With a Simultaneous Localization and Mapping (SLAM) algorithm the position can be estimated. No additional infrastructure is needed and these systems are characterized with a high computational performance. Problems with shaking of the camera during walk and motion blur leads to failures [9], [10]. Similar or dynamic environments are mostly encountered in densely populated areas, e.g. shopping malls, where the localization errors is high.

B. *Sound Indoor-localization*

Another possibility is to use the propagation of sound for localization of the smartphones. Smartphones can generate sound from their built in speaker or smartphones can detect sound with the integrated microphone.

In comparison to other methods the position accuracy can be increased. The sound propagation is slow, compared to the speed of light; thereby the time stamp of the received signals are easier to determine. Furthermore, the received sound signals can be analyzed in detail and the suppression of multipath signals is straightforward.

Filonenko *et al.* presented the practical limitations of sound generation with a speaker of a COTS smartphone [11]. Sound signals from 20 Hz to 22 kHz can be created with the implemented speaker of COTS smartphones. Higher frequencies of the sound signal are less audible and the transmitted sound is noise-affected. In *BeepBeep* [12] mobile phones transmits and receives audible sound impulses between 2 kHz and 6 kHz. The system needs no additional infrastructure and uses the Round Trip Time (RTT) between the smartphones to measure the distance between different smartphones. Borriello *et al.* presented the WALRUS [13] localization system where acoustic sound from PDAs/Laptops at a frequency of 21 kHz where transmitted to identify the location of the user within a room. Hennecke *et al.* presented a method for the acoustic self-localization of nodes in an ad-hoc array of COTS smartphones. The smartphones worked in the audible range with a short chirp impulse between 5 kHz and 16 kHz. The sound signals were received by the smartphone microphone [14].

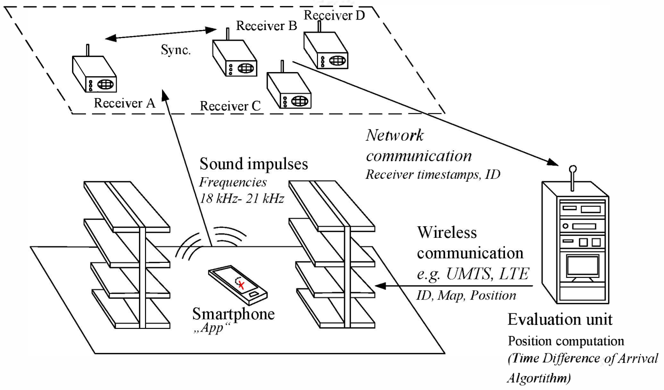

Figure 1. Overview of the ASSIST system with smartphone, the network of receivers and an evaluation unit.

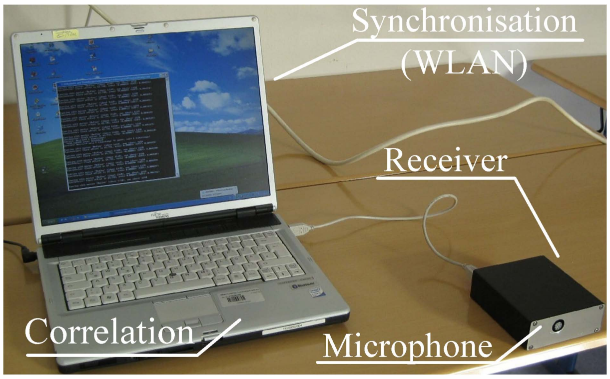

Figure 2. Special microphone connected to a notebook for detecting the incoming sound signals and calculation of the sound timestamps.

III. System Overview

For developing the concept of the Acoustic Self-calibrating System for Indoor Smartphone Tracking (ASSIST), the practical implementation and the data security of the user location was considered. The proposed indoor localization system ASSIST is schematically shown in Fig. 1. The system works with COTS smartphones and no additional equipment for the user. Cost of an indoor localization system can be significantly reduced when the users avail their own smartphones, further handling of the localization system becomes convenient and simplified.

For user localization the smartphones generate sound impulses beyond the human audible range. The sound impulses were received by self-built receivers which were placed at the ceiling or walls in a room. A minimum of three receivers are required to localize a mobile phone in one localization cell in two dimensions. Fig. 2 shows a receiver which filters and digitizes the received signals.

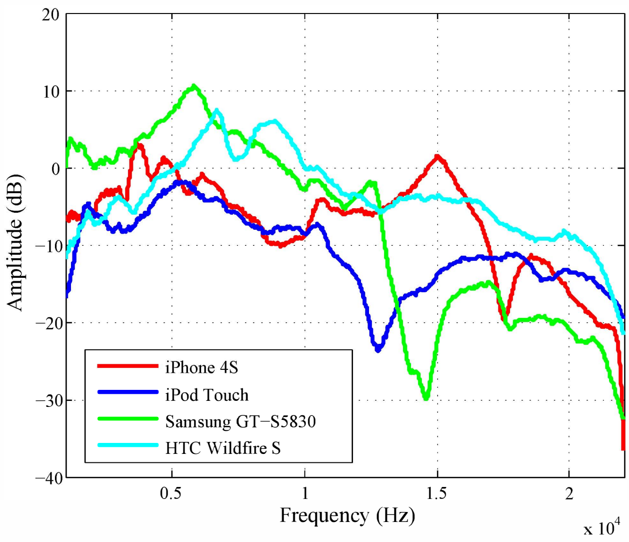

Figure 3. Frequency response of several COTS smartphones.

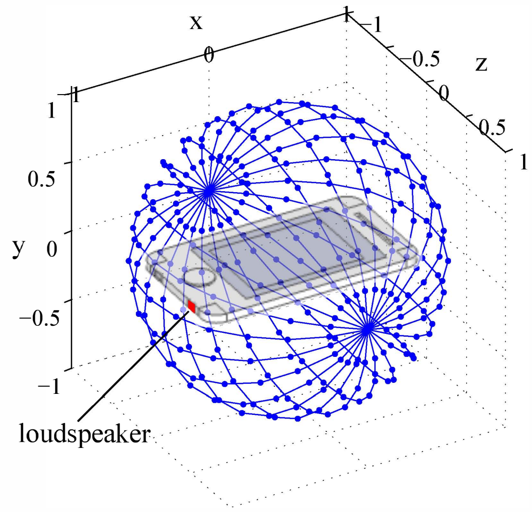

Figure 4. Measurement points for radiation characteristics of a sound speaker from a smartphone.

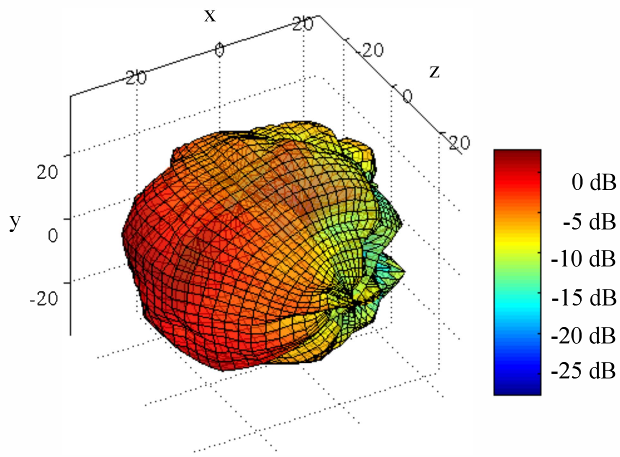

Figure 5. 3D radiation pattern of iPhone 4S sound speaker.

The receivers were connected to notebooks for signal processing and communicating with the network, to synchronize the timestamps of the incoming signal. Additionally, the notebooks with receivers were connected with a wireless network to an evaluation unit. The evaluation unit is connected to the smartphones via cellular communication (GPRS/UMTS/LTE), which serves the ID of the specific sound and provides the map with the actual position of the user.

In areas where no receivers are available, the integrated inertial sensors can be used to localize the user for short periods. In ASSIST the absolute acoustic localization system is supported by the integrated inertial sensors. The smartphone shows the actual position of the data from the inertial sensors, if the position information from the acoustic localization is not available.

A. *Transmitter*

The speaker of a commercial smartphone is controlled from a software application (app). It is used as a sound emitter of specific sound signals. In an applicable localization system based on sound signals, the frequency range of the used signals should be outside audible range. Study [15] shows that the maximum frequency of acoustic perception depends on the age and the volume of the sound source.

Hoppe *et al.* analyzed that using COTS smartphones with a frequency of 18 kHz limits the percentage of audible people. Higher sound frequencies generated from smartphones can not be heard by humans even within small distances [16]. Therefore the localization systems based on sound should use frequencies above 18 kHz as a frequency requirement.

In order to test the maximum sound frequency limitation of the speaker and to define the maximum acoustic bandwidth of COTS-smartphones, several smartphones were tested. The frequency response and the radiation characteristic were measured.

For the measurement of the frequency response, sound with white noise was transmitted from several smartphones. The frequency response of the commercially available smartphones is depicted in Fig. 3, which shows a damping factor of 20 dB in range of 1 kHz to 22.5 kHz. The sound amplitude of frequencies more than 21 kHz from a smartphone speaker, decreases rapidly with distance.

Results shows that the best frequency range of sound based localization system in urban environments is between 18-21 kHz [16]. In this area the sound signals from the speaker have high amplitude to transmit sound in a far-field and the signal is outside of the audible range. We decided to separate channels by placing them into different frequency bands (FDMA). Through the acoustic bandwidth of 3 kHz, different channel can be used to transmit at the same time sound impulses from different smartphones.

For the measurement of the smartphone radiation characteristics, the sound signals were measured within a distance

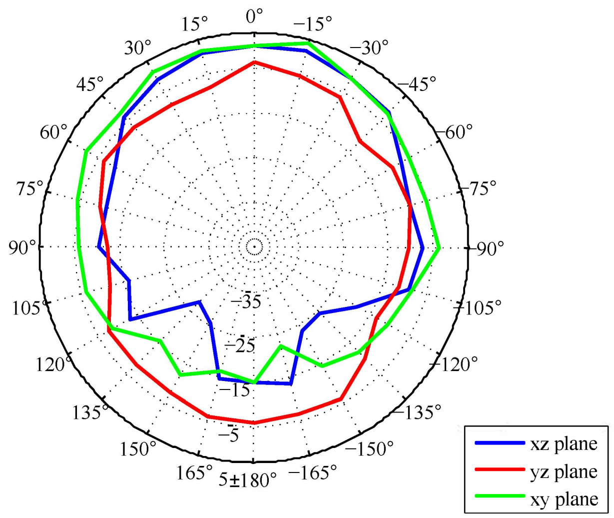

Figure 6. Radiation characteristic in dB of iPhone 4S with implemented loudspeaker in the front side.

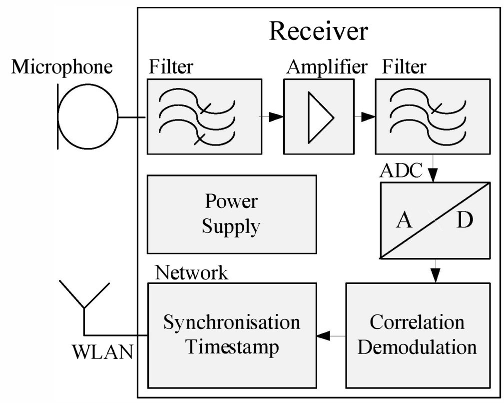

Figure 7. Blockdiagram of the receiver with correlation detection and WiFi communication

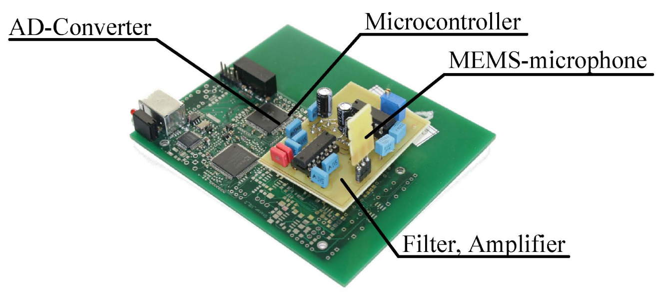

Figure 8. Receiver-PCB with filter module, analog digital converter and USB-Interface.

of 25 cm from a microphone at different positions. Fig. 4 shows the measurement points. The following measuring plots are normalized to the maximum radiation. The 3D measured radiation characteristics are shown in Fig. 5 and an axis-plot is depicted in Fig. 6. The radiation is concentrated near the position of the loudspeaker. Other radiations showed a decrease in sound levels. The loudspeaker should be placed in the direction of the receivers, for better radiation from the transmitted sound. COTS smartphones usually have a built in loudspeaker in the front side of the smartphone. In case of placing the receivers on a ceiling, the user should rotate the smartphone 180°. In our software application the rotation feature is provided as a built in feature. Not all COTS smartphones have a loudspeaker in the frontside but they are installed in the backside as well. In such cases additional rotation is not required.

B. Receiver

A receiver was built, for receiving the sound signals from smartphones outside the human audible range. Hoppe *et al.* analyzed different approaches to detect specific sound impulses from smartphones till 16 m distance [16]. An envelope detection can be used to detect sound signals from smartphones only in a range of 10 m. The high attenuation of acoustic signals in air result in a large sensitivity to interference sources at larger distances. Ambient sounds can create measurement errors which cannot not be filtered with an envelope detection. In addition, using the envelope detection, the ID can be only modulated in the time domain which leads to identification errors. Furthermore, the flat edges of the envelope detection complicate the detection limit of the signal and so the resolution.

Hoppe *et al.* shows another possibility to correlate the incoming signal with the transmitted signal to detect the arrival time of the incoming signal in farther distance and to estimate the time of arrival [16]. ASSIST uses the proposed correlation concept.

Fig. 7 shows the block diagram of the receiver. The sound signals were received from the receivers with MEMS microphones, which have good properties in a bandwidth from 18 kHz to 23 kHz. To eliminate incoming sound signals from the environment, the incoming sound signals were filtered using a 8^{th} order Butterworth low pass filter with a frequency cut of 17.5 kHz. Before digitizing the data the signal is amplified by a factor of $V = 414$.

The sound signals were digitized with analog digital converter (ADC) having a resolution of 15 bit per sample and a sampling rate of 88.15 kHz. The digitized signals were transmitted from the receiver (Fig. 8) via USB-interface to a Notebook which correlates the incoming digital signals with the transmitted reference signal. The notebook analyzes the specific sound ID from different smartphones and estimates the receiving time of the incoming signal with the synchronized clocks.

The notebooks of the receivers are connected to a WiFi network, such that they synchronize their clocks. The connected notebook clients negotiate a master receiver which acts as a time reference. Then, the other clients adjust their clocks to the master considering time offset and time drifts. With a 802.11 b/g WiFi connection, a synchronization precision of greater than 0.1 ms can be achieved [17].

C. Evaluation Unit

Every registered handheld device in a localization cell is assigned an unique ID from the evaluation unit. The evaluation unit is connected to the receivers via network where receivers exchange their received timestamps trough network protocol UDP with the evaluation unit. The received timestamps were than used to calculate the positions of the smartphones with an iterative TDoA-Algorithm. An Apache web server with PHP 5 and MySQL processes the inquiries of the positions of the smartphones.

D. Software application

We have developed an Android software application "app" that we have developed for commercial smartphones, which transforms a standard COTS device to a transmitter for ASSIST. The "app" was deveolped with Android SDK for Android smartphones. Fundamentally our designed "app" has three functionalities: (1) Communication with the evaluation unit (server), (2) Sound control, (3) Visualization on the map for the current position.

The system works when the user download and start the 'app' in an area which supports the ASSIST infrastructure. The user interface is simple as starting the "app" , which connects to an evaluation unit and receives an ID using its internet connection. Every registered handheld device in a localization cell is assigned a unique ID.

The smartphone is connected to the internet without a special infrastructure, only a mobile network is mandatory. In this work Long Term Evolution (LTE) is used for wireless data communication which are the latest standard technology of mobile data transmission. The smartphones and the server communicates using the secure communications protocol HTTPS in JavaScript Object Notation (JSON) format. Specific parameters were assigned to each user, such that several devices can be distinguished by the appearance of the chirps. The necessary parameters conceived from the evaluation unit are frequency, impulse t_{impuls}, interval duration of the chirp signal and the map of the building. Based on these data, the smartphone regularly sends out the chirp signal to guarantee the localization of the user.

The "app" controls the loudspeaker of the smartphone and generates the specific chirp sound signals. The following algorithm is used to create the chirps inside of the smartphones.

Algorithm 1 Chirp generation of duration t_p in the frequency interval $[f_\text{initial}, f_\text{final}]$.

$k = (f_\text{final} - f_\text{initial})/t_\text{p}$
for all $i \in [0, t_\text{p} \cdot \mathcal{F})$ **do**
  $t = i/\mathcal{F}$
  $f_\text{current} = (f_\text{initial} + \frac{1}{2}kt)$
  $\text{S}[i] = \sin(2\pi t \cdot f_\text{current})$
end for
play(S)

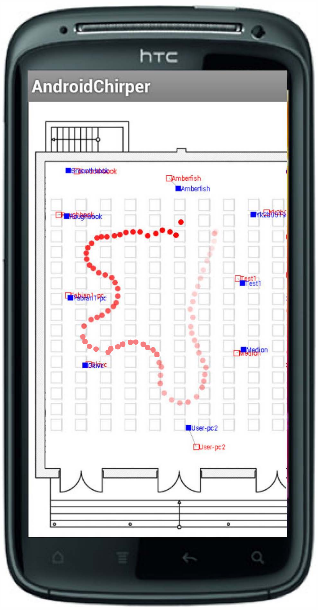

Figure 9. The developed Android software application on a screen of a smartphone.

Here, k corresponds to the slope of the frequency change, f_initial is the starting frequency, f_final is the final frequency and t_p is the duration of the chirp. $\mathcal{F}$ is the sampling rate of the audio chirps (44100 Hz) and f_current is the frequency at the current time t.

The current position and the map are transmitted from the evaluation unit to the smartphone. The position of the user is displayed on the screen of the smartphone in context to the environment, with a map and surrounding items. Fig. 9 shows an example of the software application on a smartphone screen. The current position of the user is shown with a dark red point with minimal transparency. The trajectory of the user is shown with decreasing transparency of the red points. The previously calculated data points are more transparent than the actual points, this allows the user to visualize his walk in a chronological sequence.

Depending on the connection speed, the positions of the user are provided in a real time from the evaluation unit.

IV. LOCALIZATION

The system works in two phases: A.) Calibration for the installation of the system. B.) User localization.

A. Calibration phase

For installation of an absolute localization system, the anchor nodes should be installed in the infrastructure. For every receiver (anchor node) the position (x_i, y_i) is needed. A multilateration TDoA-Algorithm requires position information of the receivers to calculate the relative position (x, y) of the mobile object (e.g. smartphone).

Normally the system customer has to measure the exact positions of the receivers; which is required for installation.

This measurement increases for large buildings, since the number of receivers depends on the size of the building (To cover large buildings completely a large number of receivers are needed). The localization system ASSIST uses special algorithm to calibrate the system.

During the calibration phase, the positions of the receivers in the indoor scenario are calculated automatically. Then, only at least three measured receiver positions are required for the orientation of the system on a map.

In a self-calibrating TDOA localization system the positions of receivers and of the signal emitter, here the smartphone, are unknown a priori. The send times of signals by a smart phone are not known to the localization system, as this would require synchronization of the smartphone over a very unreliable network, or bi-directional exchange of sound signals. Only the times of reception of the signals at the receivers can be measured. As the receivers are synchronized, the time differences of arrival can be calculated, forming a system of hyperbolic equations for the signal position and any pair of receivers. The goal of the calibration phase is to approximate the relative positions of the receivers, with respect to the map.

There are several self-calibrating TDoA-Algorithms available to calculate the positions of the receivers (anchors). In the *far field case* the signals originate from the distance, such that the propagation front of the signals approximates a line, sweeping over the receivers. Then, the positions of receivers, and subsequently of the signal directions can be calculated directly [18], [19], [20].

For the general case of arbitrarily distributed signal positions Biswas and Thrun [21] proposed a solution, which maximizes the likelihood of receiver and signal positions, given a Gaussian distribution of measurement errors. For at least eight receivers in the plane or ten receivers in space Pollefeys and Nister [22] showed a direct solution using matrix factorization.

For the calibration phase of the localization system an iterative optimization algorithm is used. The "Iterative Cone Alignment" algorithm [23] solves iteratively a non-linear optimization problem of TDoA by a physical spring-mass simulation. The success rate of solving the calculation of the receiver positions was increased to 99.4% [23]. (With only six received signals and four receivers). Through using the algorithm a quick-setup system for smartphone localization is created. There is no need to measure the positions of the receivers. In real-world indoor scenarios the receiver positions of the smartphone localization system could be located with a Cone Alignment algorithm in the range of centimeters.

B. Localization phase

The evaluation unit calculates the position of the handheld device with a TDoA algorithm. The receivers are connected in a WiFi network, such that they synchronize their clocks and exchange the time differences of arrival of the received chirps. By knowing the propagation speed of sound and the arrival times at the receivers, the position of the handheld device can be calculated.

During the sound localization the position accuracy can be increased by repeated measurements. Sound propagation is slow, compared to the speed of light; thereby the time stamp of the received signal is more accurate. The sound signals can be analyzed in detail and the suppression of multipath signals is straightforward.

A smartphone transmits acoustic signals at a position S_0 relative to the receivers with the positions M_i (i=1...,n). The receivers receive the signals at different timestamps T_i which is depending on the distance R_i between the receiver and transmitter.

For position localization of the smartphone with the received timestamps from the receivers an iterative TDoA-Algorithm is used.

The distance from the smartphone and the receiver R_i can be described by the coordinates as follows:

$$R_i = \left|\vec{M}_i - \vec{S}_0\right| = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2}. \qquad (1)$$

The smartphones generates specific sound signals at the time t_0. The sound speed C_{air} can be calculated in air according to the following equation:

$$C_{\text{air}} = 331.5\,\text{m/s}\sqrt{1 + \frac{\vartheta}{273.15\,^{\circ}\text{C}}}. \qquad (2)$$

The speed of sound depends on the temperature ϑ of the environment. At a temperature of $25\,^{\circ}$C the speed of sound is 346 m/s. The receivers generate timestamps in the time of arrival T_i of the received signal.

In case of using sound waves instead of electromagnetic waves the influence of the position accuracy from the synchronization of the receiver is decreased. The synchronization of the receivers is necessary for generating the timestamps for the TDoA-Algorithms. The receivers are connected together via wireless network (WLAN) which provides a precise time synchronization up to an order of 0.1 ms. In this case the theoretical maximum localization error from synchronization is 3.4cm.

The distance R_i from equation 1 can be also calculated by multiplying the speed of sound C_{air} with the transit time $T_i - t_0$ as given below:

$$R_i = C_{\text{air}}(T_i - t_0). \qquad (3)$$

Time $\tau_{1,2}$ shows the time difference of the receiving signal between receiver 1 and receiver 2.

$$\tau_{1,2} = (T_1 - t_0) - (T_2 - t_0) = \frac{R_1 - R_2}{C_{\text{air}}} \qquad (4)$$

Equation 4 is the hyperboloid description for 2 receivers. In this way, using an iterative TDoA-Algorithm with minimum of 3 receivers the location of the smartphones in 2D can be calculated.

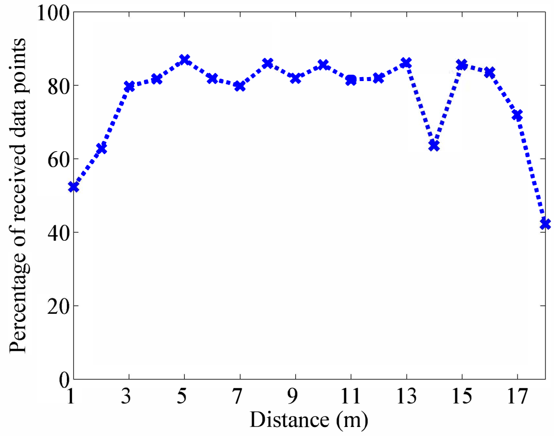

Figure 10. Measurement values of ASSIST depending on the distance.

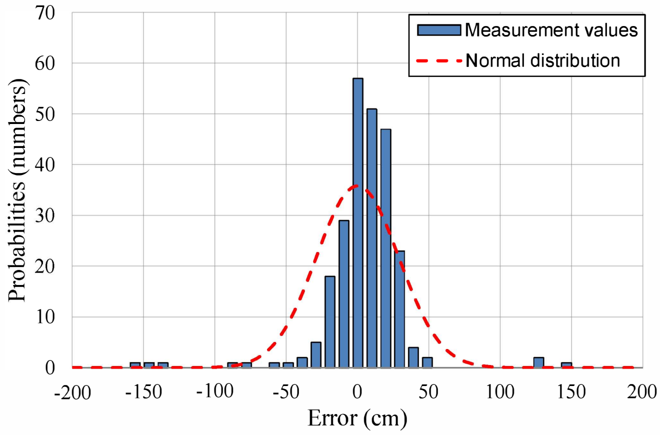

Figure 11. Measurement errors of ASSIST with correlation and TDoA-Algorithms in a static experiment.

V. Chirp Correlation

Signals with maximum energy are essential for receiving short signals in large ranges. The influence of interfering signals or white and Gaussian noise can be reduced by increasing the signal energy, where the signal-to-noise (SNR) ratio is increased. The increase of signal energy can be done either by increasing the signal amplitude or the signal length. In radar or sonar applications chirp signals are used to compress the original signal (pulse compression), while the signal time duration is reduced but the signal energy is retained.

A linear chirp is used to transmit the sound signal. It is a signal in which the frequency increases or decreases linearly with time (up- and down-chirps). The chirp impulse works between $0 \leq t \leq T$ with a start frequency of f_0 and an end frequency of f_1. It can be described according to the following equation:

$$s(t) = sin\left(2\pi\left(f_0 + \frac{f_1 - f_0}{2T}t\right)t\right) \quad . \tag{5}$$

Matched filtering is commonly used to detect the received signals. The cross-correlation of the received signal with the saved transmitted signal is computed. For detecting different smartphones diverse chirps with different frequency bands can be used to correlate the signals and to identify the smartphone.

The mathematical formula for cross-correlation of two time signals is:

$$z(t) = \int x(\tau) * y(t+\tau)\,\mathrm{d}\tau \ . \tag{6}$$

Where $x(t)$ is the received signal and $y(t)$ is the saved reference signal. The maximum of the cross-correlation function $z(t)$ is achieved at the perfect matched time.

For ASSIST the chirp signals are transmitted from smartphones and correlated with the specific ID signals for each smartphone inside of the notebooks. The notebooks calculate the exact time stamp the received signal. We used three different frequencies and a bandwidth of 1 kHz for the ID. Addiotionally in the same frequency range, up and down chirps can be differenced by using correalation (as given in equation 6), which further implies that it is possible to distinguish six different smartphones in one cell.

In an experiment the receiver range was analyzed. We used 95% of the received signals which was not deficient from mutipath propagation. Fig. 10 shows the arrived signals as a function of distance which are inside of $\pm 2\sigma$. The developed receivers (in section III B) were able to receive more than 70% of the transmitted signals up to a distance of 16 m from a smartphone. The percentage of received signals decreases at distances above 16 m,.

VI. Experimental Results

A. Sound localization

The measurement deviation of the system was evaluated in static experiments. Fig. 11 shows the measurement errors and the normal distributions. ASSIST shows a standard deviation of $\sigma = 25$ cm. Signals with mulitpath propagation leads to increased standard deviation.

Also, we verified our system in dynamic real-world scenario; 2D experiment. For a reference, we defined a walking track of 14 m which was exactly measured. In our experiment, we placed seven receiver devices in an oval of 10 m times 10 m around the walking track in a height of 1 m. A person walked along the defined track.

We calculated the positions of the smartphone which transmitted acoustic chirp impulses between 19 kHz to 20 kHz with a length of 50 ms. Fig. 12 shows the calculated positions and the defined walking track. The data shows a well matching compared to the track. The trajectory shows a systematically error which depends on the localization algorithm and some measurement errors from the multipath propagation.

The smartphone track shows an average deviation of 0.34 m (σ = 0.18 m).

B. Localization with Inertial Sensors

In areas where no infrastructure is available, the integrated inertial sensors can be used to localize the user for a short period.

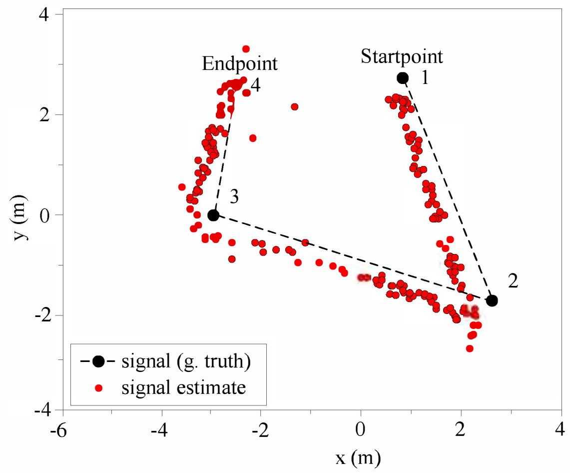

Figure 12. Real world experiment of ASSIST. Data set *Exp. 2d*. Microphone error mean 1.81 m, (σ = 1.32 m).Signal position error mean: 0.34 m (σ = 0.18 m).

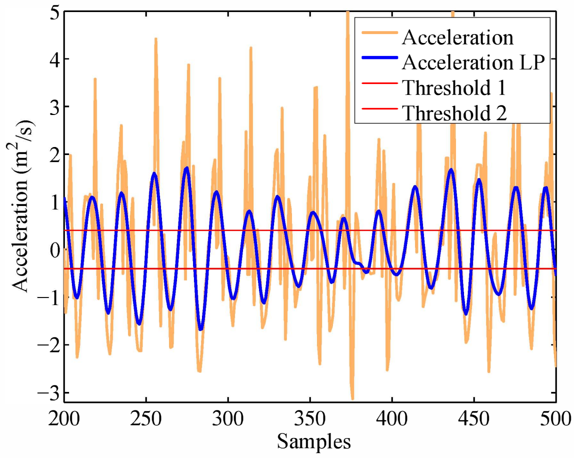

Figure 13. Measurement values of the acceleration sensor in z-axis during a walk with the two threshold values.

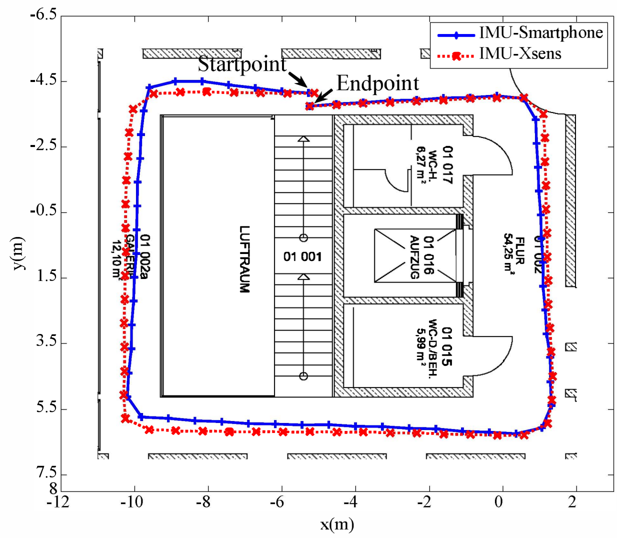

Figure 14. Trajectory of data from an experiment. Data from the inertial sensors of a smartphone (blue line) and a reference inertial measurement unit (red dash line).

Currently, many different sensor types are integrated in the smartphones. For example the commercial smartphone Samsung Galaxy S2 provides the data of the integrated inertial sensors like gyroscope, accelerometer and a magnetic field sensor.

User localization based on the inertial sensors leads to measurement errors with increased computational time. The inertial sensor unit additionally supports a method to perform acoustic localization.

Since the smart phone is normally held by hands, methods as zero velocity update [24] can not be implemented. In order to deliver correct position information, step length and orientation information must be determined. For normal walking, each step is set roughly as 0.70 m. The step detection is accomplished by analyzing the accelerations. New step is detected only when the acceleration signal crosses two predefined thresholds with a rising edge. Fig. 13 shows the acceleration in z-axis during a walk with the two threshold values. The orientation information is obtained by Kalman Filter based sensor data fusion, as discussed in [25].

In an experiment the data from inertial sensors of the smartphone without the ASSIST localization system was used for detecting a walk of 45 m distance in a building. The trajectory of the walk is shown in Fig. 14. The red dash line shows a reference path which was measured with an inertial measurement unit from Xsens. The blue line shows the calculated path with the data from inertial sensors of smartphone Samsung Galaxy S2. The calculated maximum deviation from the 45 m real track was 1 m.

VII. Conclusions

In this paper, we presented a smartphone indoor localization system based on sound. The user of the system needs no additional hardware except a COTS smartphone. Through our self-build receivers which were synchronized with a WiFi network the arrived signal can be correlated and with a TDoA-Algorithm the position can be calculated. The first experiments showed that it is possible to use the system in a real environment to localize a user in an indoor environment with less effort.

The system does not require a special knowledge to be installed by the provider, hence the installation effort is minimized. Through an Anchor-free-Algorithm, the receivers work as a Plug-and-Play-System and there is no need for additional information, since the positions of the receivers can be calculated.

The comparison of the different smartphones localization systems is given in Table I. For ASSIST, only receivers are needed which can detect the sound signals. Since the signals are correlated, the range of a receiver can be increased to 16 m. The accuracy of the ASSIST system is much better then RF based localization systems. A real world experiment showed

TABLE I
COMPARISON OF LOCALIZATION SYSTEMS FOR SMARTPHONES

System	Technology	Accuracy	Comment
RADAR	RF (WLAN)	2 m	Existing infrastructure can be use, cheap
MoVIPS	Visual	1 m	High data rate for communication
WALRUS	RF (WLAN), Sound	Room	Receivers are needed
ASSIST	Sound (outside of audible Range)	< 30 cm	Receivers are needed

that it is possible to detect a smartphone within an error margin of 30/,cm.

For areas with a poor receiver coverage, the localization with built-in inertial sensors in smartphone was tested. The integrated inertial sensors can be used as an additional localization method to support ASSIST for a short time. The maximum deviation from the reference track of 45 m was 1 m.

VIII. OUTLOOK

In our future investigations we will improve the acoustic localization in situations where there is no line of sight between the smartphone and the receivers. Error minimization can be achieved, through fusion of the data from the inertial sensors and the data from the acoustic localization.

We will modulate a pseudo noise code (PN code) on the transmitted sound to identify more users for multiuser applications,. Currently, we can distinguish only eight smartphones in the same localization cell. Another possibility is to use Time-division multiplexing (TDM) to identify different smartphones. The time domain is divided into time slots which can be used from the respective of smartphones to generate the sound.

In some application, e.g. in a supermarket the receivers should be installed in the ceiling. In this case, the position of the user should be localized in a 2D area with defined height. The algorithm must be modified for 2.5D applications.

In addition, we will improve ASSIST through reducing measurement errors from multipath propagation. The experimental results should be evaluated with a reverence system to measure the systematical error precisely.

ACKNOWLEDGMENT

This work has partly been supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) within the Research Training Group 1103 (Embedded Microsystems).

REFERENCES

[1] K. Schreiner, "Where we at? mobile phones bring gps to the masses," *Computer Graphics and Applications, IEEE*, vol. 27, no. 3, pp. 6 –11, may-june 2007.

[2] P. Bahl and V. N. Padmanabhan, "Radar: an in-building rf-based user location and tracking system," in *Proc. IEEE Nineteenth Annual Joint Conf. of the IEEE Computer and Communications Societies INFOCOM 2000*, vol. 2, 2000, pp. 775–784.

[3] V. Otsason, A. Varshavsky, A. LaMarca, and E. de Lara, "Accurate GSM Indoor Localization," in *Ubicomp*, 2005, pp. 141–158.

[4] P. Bolliger, "Redpin-adaptive, zero-configuration indoor localization through user collaboration," in *Proceedings of the first ACM international workshop on Mobile entity localization and tracking in GPS-less environments*. ACM, 2008, pp. 55–60.

[5] Y. Kim, H. Shin, and H. Cha, "Smartphone-based wi-fi pedestrian-tracking system tolerating the rss variance problem," in *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*. IEEE, 2012, pp. 11–19.

[6] M. Bourimi, G. Mau, S. Steinmann, D. Klein, S. Templin, D. Kesdogan, and H. Schramm-Klein, "A privacy-respecting indoor localization approach for identifying shopper paths by using end-users mobile devices," in *Proc. Eighth Int Information Technology: New Generations (ITNG) Conf*, 2011, pp. 139–144.

[7] N. Ravi, P. Shankar, A. Frankel, A. Elgammal, and L. Iftode, "Indoor localization using camera phones," in *Proc. 7th IEEE Workshop Mobile Computing Systems and Applications WMCSA '06*, 2006.

[8] M. Werner, M. Kessel, and C. Marouane, "Indoor positioning using smartphone camera," in *Proc. Int Indoor Positioning and Indoor Navigation (IPIN) Conf*, 2011, pp. 1–6.

[9] B. Williams, G. Klein, and I. Reid, "Real-time slam relocalisation," in *Proc. IEEE 11th Int. Conf. Computer Vision ICCV 2007*, 2007, pp. 1–8.

[10] D. Wagner and D. Schmalstieg, "First steps towards handheld augmented reality," in *Proc. Seventh IEEE Int Wearable Computers Symp*, 2003, pp. 127–135.

[11] V. Filonenko, C. Cullen, and J. Carswell, "Investigating ultrasonic positioning on mobile phones," in *Proc. Int Indoor Positioning and Indoor Navigation (IPIN) Conf*, 2010, pp. 1–8.

[12] K. Driscoll, "Beepbeep: Embedded real-time encryption," in *FSE*, 2002, pp. 164–178.

[13] G. Borriello, A. Liu, T. Offer, C. Palistrant, and R. Sharp, "Walrus: wireless acoustic location with room-level resolution using ultrasound," in *Proceedings of the 3rd international conference on Mobile systems, applications, and services*. ACM, 2005, pp. 191–203.

[14] M. H. Hennecke and G. A. Fink, "Towards acoustic self-localization of ad hoc smartphone arrays," in *Proc. Joint Workshop Hands-free Speech Communication and Microphone Arrays (HSCMA)*, 2011, pp. 127–132.

[15] J. Lee, S. Dhar, R. Abel, R. Banakis, E. Grolley, J. Lee, S. Zecker, and J. Siegel, "Behavioral hearing thresholds between 0.125 and 20 khz using depth-compensated ear simulator calibration," *Ear and Hearing*, vol. 33, no. 3, p. 315, 2012.

[16] J. Hoppe, F. Höflinger, and L. Reindl, "Acoustic receivers for indoor smartphone localization," in *Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2012, pp. 1–10.

[17] B. Sundararaman, U. Buy, and A. Kshemkalyani, "Clock Synchronization for Wireless Sensor Networks: A Survey," *Ad Hoc Networks*, vol. 3, no. 3, pp. 281–323, 2005.

[18] T. Janson, C. Schindelhauer, and J. Wendeberg, "Self-Localization Application for iPhone using only Ambient Sound Signals," in *Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Nov. 2010, pp. 259–268.

[19] J. Wendeberg, T. Janson, and C. Schindelhauer, "Self-Localization based on Ambient Signals," *Theoretical Computer Science*, vol. 453, pp. 98 – 109, 2012.

[20] S. Thrun, "Affine Structure From Sound," in *Proceedings of Conference on Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 2005.

[21] R. Biswas and S. Thrun, "A Passive Approach to Sensor Network Localization," in *Proceedings of the International Conference on Intelligent Robots and Systems, 2004. (IROS 2004). 2004 IEEE/RSJ*, vol. 2, 2004, pp. 1544–1549.

[22] M. Pollefeys and D. Nister, "Direct computation of sound and microphone locations from time-difference-of-arrival data," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*. IEEE, 2008, pp. 2445–2448.

[23] J. Wendeberg, F. Höflinger, C. Schindelhauer, and L. Reindl, "Anchor-free tdoa self-localization," in *Proc. Int Indoor Positioning and Indoor Navigation (IPIN) Conf*, 2011, pp. 1–10.

[24] L. Ojeda and J. Borenstein, "Non-gps navigation for security personnel and first responders," *Journal of Navigation*, vol. 60, no. 3, p. 391, 2007.

[25] R. Zhang and L. M. Reindl, "Pedestrian motion based inertial sensor fusion by a modified complementary separate-bias kalman filter," in *Proc. IEEE Sensors Applications Symp. (SAS)*, 2011, pp. 209–213.