

Research Article

Acoustic Self-Calibrating System for Indoor Smart Phone Tracking

Alexander Ens,¹ Fabian Höflinger,¹ Johannes Wendeberg,² Joachim Hoppe,¹ Rui Zhang,¹ Amir Bannoura,¹ Leonhard M. Reindl,¹ and Christian Schindelbauer²

¹Department of Microsystems Engineering, University of Freiburg, 79110 Freiburg, Germany

²Department of Computer Science, University of Freiburg, 79110 Freiburg, Germany

Correspondence should be addressed to Alexander Ens; alexander.ens@imtek.uni-freiburg.de

Received 1 August 2014; Revised 2 February 2015; Accepted 3 February 2015

Academic Editor: Aleksandar Dogandzic

Copyright © 2015 Alexander Ens et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents an acoustic indoor localization system for commercial smart phones that emit high pitched acoustic signals beyond the audible range. The acoustic signals with an identifier code modulated on the signal are detected by self-built receivers which are placed at the ceiling or on walls in a room. The receivers are connected in a Wi-Fi network, such that they synchronize their clocks and exchange the time differences of arrival (TDoA) of the received chirps. The location of the smart phone is calculated by TDoA multilateration. The precise time measuring of sound enables high precision localization in indoor areas. Our approach enables applications that require high accuracy, such as finding products in a supermarket or guiding blind people through complicated buildings. We have evaluated our system in real-world experiments using different algorithms for calibration-free localization and different types of sound signals. The adaptive GOGO-CFAR threshold enables a detection of 48% of the chirp pulses even at a distance of 30 m. In addition, we have compared the trajectory of a pedestrian carrying a smart phone to reference positions of an optic system. Consequently, the localization error is observed to be less than 30 cm.

1. Introduction

From the sustained rise and ubiquitous availability of mobile computers, smart phones, and handheld devices in everyday 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 device of the people, is already available.

The GPS-Module in commercial off-the-shelf (COTS) smart phones and hand-held devices makes navigation systems reliable to assist in outdoor areas [1]. The demand of localization systems begins to shift towards closed scenarios. For indoor environments, there is the need for new localization approaches, since the reliability of GPS vanishes in densely built-up urban areas and is completely void inside buildings. In addition, 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.

2. Related Work

Today several indoor localization systems are available, based on different methods and technologies. Some of these systems work with COTS smart phones. In addition, many participants have already COTS smart phones, which reduces the costs of the localization system. Figure 1 shows an overview of the different technologies and the achievable accuracy of indoor localization systems based on COTS smart phones which were developed by scientific research groups.

We use the principles of smart phone localization from our prior work [2] to apply our new developed algorithm

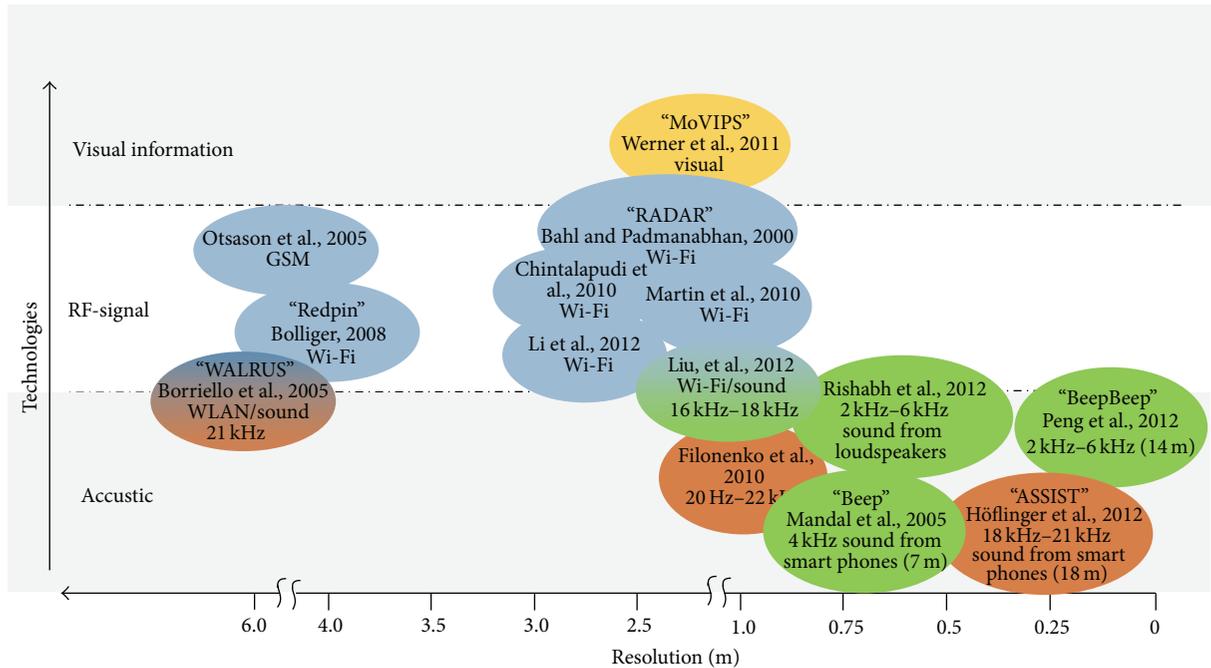


FIGURE 1: Overview of localization system based on smart phones.

(Cone Alignment) and particle filter. Further, we show localization with the integrated inertial measurement unit and compare the results with a reference motion tracking system. Furthermore, we showed in [3] an optimized receiver hardware to increase sensitivity and accuracy of the localization system.

2.1. Basic Indoor Localization

- (i) Many present localization systems use radio frequency (RF) signals for localization. The RF systems use the propagation of radio waves for position calculation. Therefore, the existing infrastructure can often be used. In the following, a brief description of indoor localization systems based on three different RF technologies is presented. Otsason et al. used the GSM communication with wide signal-strength fingerprints to locate the user in indoor environments [4]. For the localization, no infrastructure is required, but the accuracy strongly depends on the environment. Another possibility is using the Wi-Fi communication [5–7]. Current smart phones have a Wi-Fi module implemented to communicate with a network. RADAR [8] operates with the existing multiple Wi-Fi access points. Further, they use the received signal-strength indicator (RSSI) to calculate the distances between the Wi-Fi access points and the mobile phone. The accuracy depends on the number of Wi-Fi access points and the environment. The third technology is Bluetooth, which has the shortest range among the three technologies. However, the technology has some flaws for accurate positioning application. First

of all, Bluetooth adjusts the signal-strength when the signal becomes too strong or too weak. Moreover, Bluetooth takes a lot of time to discover new devices. As a result, these restrictions make Bluetooth positioning impractical and not feasible for high precision localization.

To sum up, the RF systems are susceptible to errors in dynamic environments. For example, the RSSI value depends on the environment and the smart phone. The RSSI value is distorted by objects in the direct path, in the vicinity and by environmental influences, like air humidity, and so forth. Additionally, the RSSI value also depends on the orientation of the antenna. The antenna directivity is influenced by specific smart phone types and the actual orientation to the anchor nodes. RF localization systems can localize people with low accuracy (1.5 m–3 m). Through combination with other technologies, this accuracy can be improved. The multimethod approach [9] uses a combination of built-in sensors of mobile devices and the capabilities of the end-users, which estimates positions with a scanner application. Redpin considered the signal-strength of GSM, Bluetooth, and Wi-Fi access points on a mobile phone to calculate the position [10].

- (ii) An alternative technology is pedestrian dead reckoning (PDR) with inertial sensors. By using the integrated MEMS sensors (accelerometers, gyroscopes), the current position can be calculated recursively based on the measured acceleration and angular rate of the movement. Inertial sensors based localization work without addition infrastructure. However, the

errors of the sensors are accumulated during the integration of the measurement values, which increases the localization error with the investigation time [11]. Therefore, position calculation based only on inertial sensors is usually fused with an absolute location method. Thus Kim et al. presented a smart phone localization system based on Wi-Fi access points and inertial sensors. Zhang et al. presented a smart phone localization system based only on inertial sensors [12]. Different methods were introduced to provide adaptive step lengths detection by analyzing vertical acceleration data. The experimental results showed that the obtained trajectory was able to follow the true path with an error margin of a meter in a walking distance of 45 m. Mautz compared different approaches based on inertial sensors which are integrated in smart phones or in external cases [13]. The localization accuracy varies greatly between 0.1% and 20% of the travelled distance and depends on the used methodology (algorithm) and sensors.

- (iii) Other existing smart phone localization systems use information of the surrounding. Further, the magnetic field fluctuations and anomalies inside buildings [14] can be used to create landmarks for localization. Another possibility is using the fluorescent light as a medium to transmit position information by using a pulse-frequency modulation technique [15]. Hence, a smart phone can receive the encoded light information through the integrated camera and can calculate the position. It is also possible to use only the visual information of the surrounding [16, 17] for localization. The integrated camera of the smart phones is used to create images and compare the images with a database. Moreover, with a simultaneous localization and mapping (SLAM) algorithm the position can be estimated. Thus, no additional infrastructure is needed, but these systems are characterized with a high computational performance. Problems with shaking of the camera during walk and motion blur lead to failures [18, 19]. Similar or dynamic environments are mostly encountered in densely populated areas, for example, shopping malls, where the localization errors are high.

2.2. Sound Indoor Localization. Sound is feasible for high accuracy indoor localization. Smart phones can generate sounds from their built-in speaker or they can detect sounds with the integrated microphone. In comparison to other technologies, 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 is easier to determine. Precise measurement of the time of arrival (ToA) is very important for exact position determination. Errors exist from the clock skew and drift between devices and differences of the propagation speed. In contrast, through the high propagation speed of light, small errors lead to high position deviations. Furthermore, the received sound signals can be analyzed in detail and the suppression of multipath

signals is straight forward. A brief description of current indoor localization systems based on sound is presented in this section.

Most of the research groups uses the time of flight (ToF) or round trip time (RTT) measurement for smart phone positioning. However, there are several intrinsic uncertainty factors of a ToF measurement which lead to the ranging inaccuracy. For COTS smart phones, there exists a variable latency, a changeable misalignment between the timestamps of the command from the transmitted signal and the transmitted signal from the loudspeaker. Another problem is the synchronization of the smart phones and receivers. These delays can easily add up to several milliseconds, which imply a ranging error of several cm.

Borriello et al. presented the WALRUS [20] localization system, where acoustic sound for PDAs/Laptops at a frequency of 21 kHz was received. The wireless network provides a synchronizing pulse along with the information about the room to determine the location in a room-level accuracy. Liu et al. improved the Wi-Fi localization accuracy with an acoustic ranging [21]. The phones are using nearby peer phones as reference points and calculate the relative distances with the acoustic RTT. It means smart phone *A* transmits the impulses and smart phone *B* receives the impulses and transmits a new impulse to the smart phone *A*. For the distance measurement, no synchronization is necessary except for time delay. Liu et al. use this additional distance measurement to increase the accuracy to 1-2 m of the Wi-Fi localizations system. A pure sound localization system is Beep [22] and BeepBeep [23]. Peng et al. showed that a localization system can use mobile phones which transmit and receive audible sound impulses between 2 kHz and 6 kHz [23, 24]. Further, the system needs no additional infrastructure and uses the RTT between the smart phones to measure the distance between different smart phones in a resolution of about 1-2 cm. For the system, the latency is measured and transmitted to the other smart phone. In this case, a very precise position measurement is possible. Rishabh et al. use the loudspeakers which are installed in shopping malls and consumer stores to play music for public entertainment [25]. The use of barely audible (low energy) pseudorandom sequences in their approach poses very different challenges to other approaches which use high-energy ultrasound waves. The approach was tested in a meeting room and reported a promising initial result with localization accuracy of 50 cm.

In most of the state-of-the-art systems, the anchor nodes are used as transmitters. Those receivers detect the sound signal emitted by the anchor nodes. However, this method suffers from certain disadvantages.

- (i) The sound signals are received at different positions during a movement (see Figure 2). Thus, the mobile device needs the information of the environment, especially the positions of the beacons to calculate the own position.
- (ii) The microphone of COTS smart phones can only detect relatively low frequencies (i.e., frequencies in the audible range) due the limitation of its built-in microphone (made for normal speaking which uses

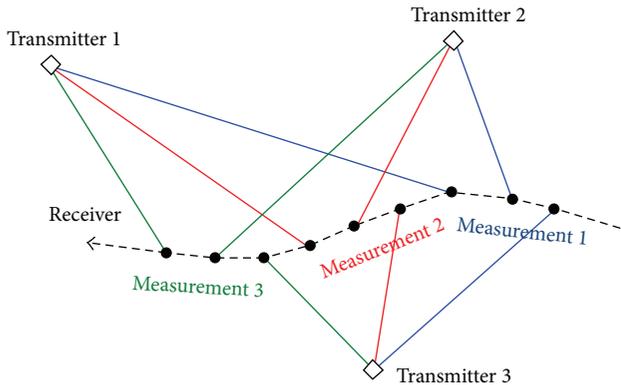


FIGURE 2: Principle of moving device position estimation with stationary transmitters [44].

the band between 80 Hz and 12 kHz). Outside this frequency range, the microphone has low sensitivity to receive sound from larger distances. Additionally, there exists a maximum sampling rate of the analog to digital converter of COTS smart phones. The corresponding sampling frequency needs to be greater than twice of the maximum signal frequency. As a result, the sound emitted by the handheld device lies in the audible range, detectable by the user. Furthermore, this frequency band is crowded with natural sounds, making it more difficult to distinguish the localization signal from noise.

- (iii) Due to permanently receiving the sound signals by the mobile device, an increased power consumption on the mobile side is necessary for signal identification and calculation [26, 27].

3. System Overview

In the presented work, the practical implementation of the concept acoustic self-calibrating system for indoor smart phone tracking (ASSIST) as discussed in [2] is considered. Using this concept, the above-mentioned disadvantages (Chapter II-B) were avoided. The proposed indoor localization is schematically shown in Figure 3. The system works with COTS smart phones and requires no additional equipment from the user. The following is a brief description of the system.

In ASSIST, the smart phones generate sound impulses beyond the human audible range. The sound impulses were received by self-built receivers which can be placed at the ceiling or on the walls of a room. A minimum of three receivers is required to localize a mobile phone in one localization cell in two dimensions. The receivers were connected to a Wi-Fi network to synchronize the timestamps of the incoming signal. Additionally, the receivers were connected with a wireless network to an evaluation unit. The evaluation unit is connected to the smart phones via cellular communication (GPRS/UMTS/LTE), which serves the ID of the specific sound and provides the map with the actual position of the

TABLE 1: Difference between average hearing threshold and sound pressure level emitted by smart phones in units of the standard deviation.

Distance to smart phone	Smart phone type	18 kHz	20 kHz
1 cm	iPhone 4S	1.6σ	5.0σ
	Samsung GT-S5830	0.9σ	4.1σ
	iPod Touch	1.0σ	4.4σ
10 cm	iPhone 4S	1.9σ	6.4σ
	Samsung GT-S5830	1.8σ	4.1σ
	iPod Touch	1.7σ	4.1σ
5 m	iPhone 4S	3.5σ	11.7σ
	Samsung GT-S5830	3.4σ	12.0σ
	iPod Touch	3.5σ	11.3σ

user. In ASSIST, the absolute acoustic localization system is supported by the integrated inertial sensors. In areas where no receivers are available, the integrated inertial sensors can be used to localize the user for short periods.

3.1. Human Sense of Hearing at High Frequencies. In an applicable localization system based on sound signals, the frequency range of the used signals should be outside of the audible range. Choosing the correct frequency range is therefore essential. The following section elaborates different frequency ranges of human cognition and various hearing thresholds. Human hearing capability is the best at frequencies where most of the speech takes place, which is around 0.5–6 kHz. The absolute hearing threshold defines the minimum sound pressure level, which a pure tone needs to have in order to be recognizable for a human being. Sakamoto et al. have conducted measurements of the absolute hearing threshold in the frequency range from 8 to 20 kHz, for different age groups [28]. In the range from 18 to 20 kHz they reported average hearing thresholds between 112 and 148 dB SPL (sound pressure level). One should note that the hearing threshold was measured under laboratory conditions. For the case of background noise (typical environment of a crowded building), the hearing threshold will be raised through masking.

To evaluate the audibility of high frequency sound signals emitted by smart phones, we measured the sound pressure level of different commercial smart phones for different frequencies and distances. The sound pressure level values of the smart phones were then compared to the lowest values of average hearing threshold and corresponding standard deviation σ . The difference between smart phone sound pressure and average hearing threshold for a specific frequency was calculated in units of σ (Table 1).

As expected, the audibility of the sound signals is worse when frequency increases and the distance to the measured smart phone as well. The measurements and calculations show that with a chance of 0.13% (3.5σ) a sound signal with 18 kHz can be heard in a distance of 5 m in a quiet room. Therefore, for the least upper bound of the auditory threshold,

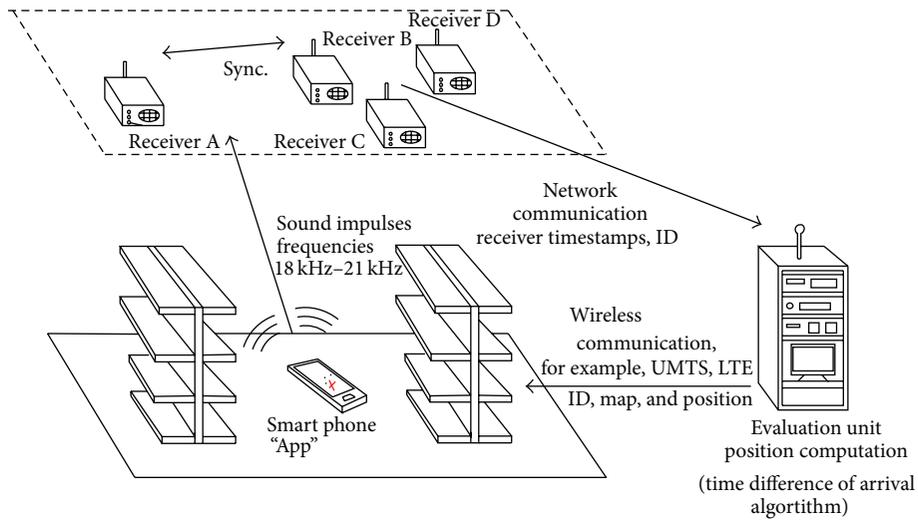


FIGURE 3: Overview of the ASSIST system with smart phone, the network of receivers and an evaluation unit.

a frequency of 18 kHz is chosen to guarantee that the signal is outside of the audible range.

3.2. *Transmitter.* In our system, the smart phone speakers transmit the sound signals for the localization. To analyze the maximum frequency limitation and the maximum acoustic bandwidth of a smart phone speaker, several COTS-smart phones were tested. Therefore, the frequency response and the radiation characteristic were measured.

For the measurement of the frequency response, sound with white noise was transmitted from several smart phones and recorded with a broadband measurement microphone Earthworks M50.

The frequency response is depicted in Figure 4, which shows a damping factor of 20 dB in a range of 1kHz to 22.5 kHz. The sound amplitude of frequencies with more than 21 kHz decreases rapidly with higher sound frequency. In addition, Filonenko et al. presented in a study the practical limitations of sound generation with a speaker of a COTS smart phone [29]. Frequencies above 22 kHz are significantly affected by noise. Also, our results show that the maximum frequency of localization system based on smart phones is 21 kHz. Up to this limit, the sound signals from the speaker have a high amplitude which enables them to transmit sound over long distances.

For the measurement of the smart phone radiation characteristics, the sound signals were measured within a distance of 25 cm from a microphone at different positions. Therefore, a smart phone holder is designed to allow a manual rotation of the smart phone and inclination angle around the holders axes. The smart phone is placed along the horizontal axis. The speaker is located on the opposite side of the measuring microphone. The measurements start at an inclination angle of 0° and the smart phone rotates around the holders axes with an angle of 15°. This corresponds to the movement of the microphone along a circle around the smart phone. The advantage of this rotation around the holders axes is its

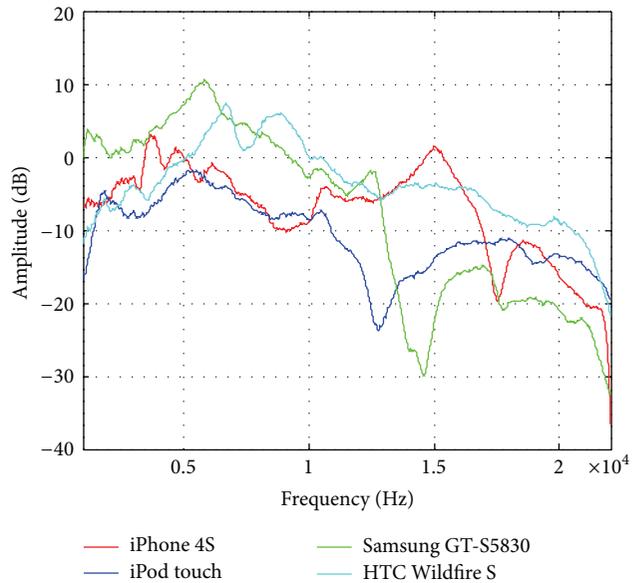


FIGURE 4: Frequency response of commercial available smart phones.

simple implementation. Eventually, the inclination angle of the smart phone is increased to reach 180°. The 3D measured radiation characteristics are shown in Figure 5 and an axis-plot is depicted in Figure 6. As expected, the radiation is anisotropic and has a small directivity into the direction of the ear.

The sound pressure is plotted logarithmically. As a reference, the sound pressure is located within a direct orientation of the speaker to the microphone. The reference sound pressure level of 0 dB is assigned to a distance of 35 dB as the origin of the coordinate system.

Generating audio signals with a smart phone requires approximately 33 mW [26]. Moreover, the signal length is

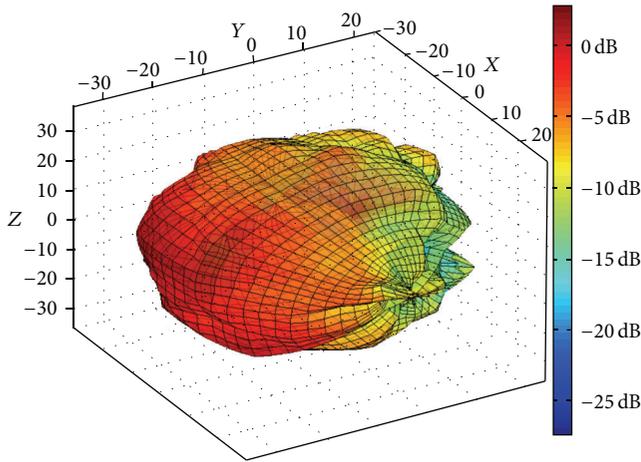


FIGURE 5: 3D radiation pattern of iPhone 4S sound speaker.

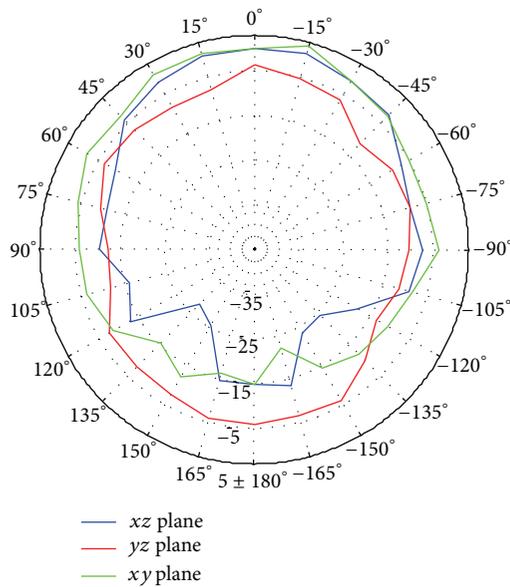


FIGURE 6: Radiation characteristic in dB of iPhone 4S with implemented loudspeaker in the front side.

2 ms and hence the smart phone requires $66 \mu\text{W}$ s for every transmitted burst. However, 58% of the power is consumed by decoder. Hence, calculating chirps requires less power than decoding audio from an MP3 file. In addition, the smart phone requires 55 mW for transmission of the calculated position to the smart phone. Compared to localization systems, where the position is calculated on the smart phone, the power for listening of the signals and calculation takes approximately 150 mW [26, 27] and the CPU load is about 80%. Moreover, the calculation is limited, due to the low power CPU, to simple localization algorithms (not enough computation power for a particle filter). As a result, our localization system benefit from longer battery life of the smart phone and better position estimation (complex algorithms can be run on the server).

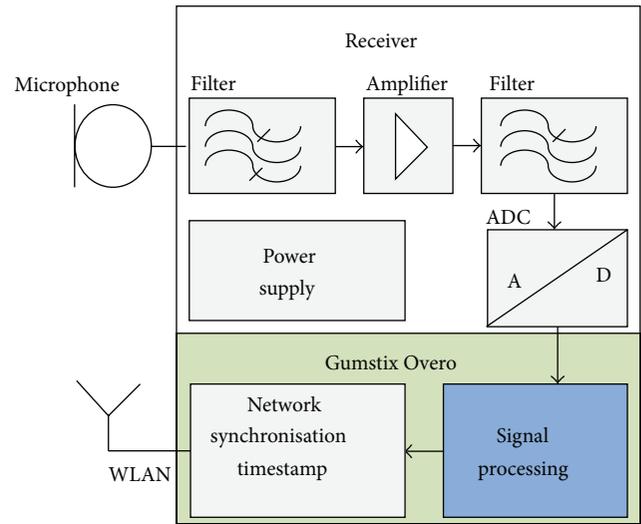


FIGURE 7: Block diagram of the receiver with signal processing (described in Section 4) and Wi-Fi communication.

Using TDOA as the localization principle, the system is independent of the exact transmission time of the pulse. Hence, the operation system requires no modification or patch to ensure deterministic behaviour. On the contrary, localization systems based on TOF or round trip measurements rely on precise transmission and receiving time of the signal. Thus, the operation system is patched to ensure deterministic real-time behaviour. Consequently, the user does not need root rights to modify the operating system.

3.3. Receiver. Ten prototype receivers for receiving the sound signals from the smart phones were built. Figure 7 shows the block diagram of the receiver. The receivers calculate the low level signal processing (correlation, threshold) and the localization is calculated on a central server.

The first part in the signal chain of our receivers is a transducer, which converts acoustical signals into electrical signals. The designed system uses a small, low cost transducer, powered by a maximum voltage of 5 V. Further, MEMS-microphones from Knowles Acoustics were used and the sensitivity as a function of frequency was calculated and compared for different measurements as depicted in Figure 8. The MEMS-microphone shows a peak around 20 kHz. For detecting sound signals in the range of 18–22 kHz, the use of this MEMS-microphone is preferred.

An 8th order Butterworth low-pass filter with a cut-off frequency of 17.5 kHz was used to eliminate ambient noise. Before digitizing the data, the signal is analog amplified by a factor of $v = 414$, which is a trade off between sensitivity and false detections. Subsequently, the sound signals were digitized using an analog digital converter (ADC) having a resolution of 15 bits per sample and a sampling rate of 88.15 kHz. The digitized signals are correlated and the threshold is applied to the result. Further, the peaks and the IDs (identification numbers) of the sound signals are estimated and the

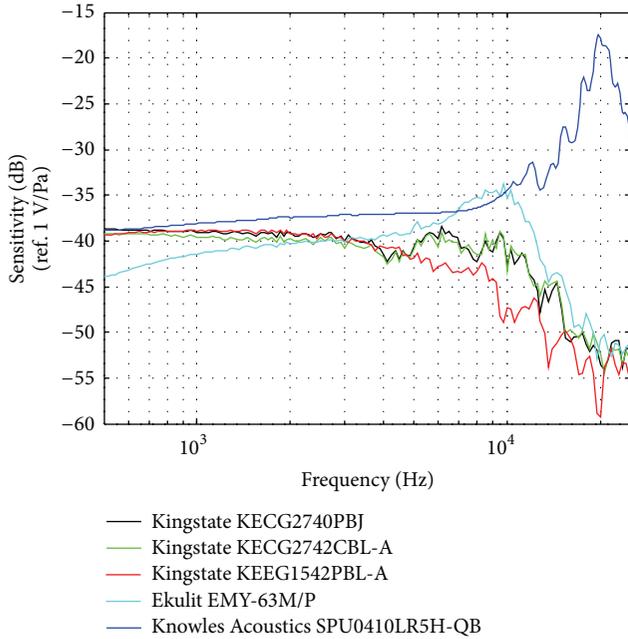


FIGURE 8: Frequency response of electret- and MEMS-microphones in the range from 500 Hz to 25 kHz.

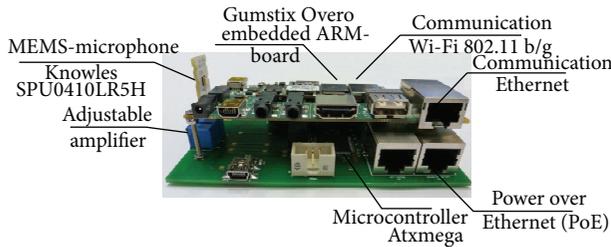


FIGURE 9: Photo of the receiver with analog module (bottom) and Gumstix Overo Board (top) for the signal processing.

timestamps are transmitted from the receiver (Figure 9) via Ethernet-Interface to a central server (e.g., notebook).

To determine the time of arrival (ToA) of the received sound impulses, a precise time synchronization is needed, as the accuracy of the localization system relies on synchronization precision between the receivers. The receivers are connected to a Fast-Ethernet network to synchronize their clocks. The connected receivers (slaves) negotiate a master receiver which acts as a time reference. Subsequently, the other clients (receivers) adjust their clocks to the master considering time offset and time drifts. The slaves ping to the master to get the current time of the master via UDP-protocol. This time is corrected by round trip time from the slave. Time offset and the time drift are both considered by an adaption of the Network Time Protocol algorithm. Both time offset and clock drift between slave and master are obtained by linear regression from the set of the time stamps. The implementation of synchronization can be found in [30]. With a 802.11 b/g Wi-Fi connection, a synchronization precision of greater than 0.1 ms can be achieved [31]. As a result, the

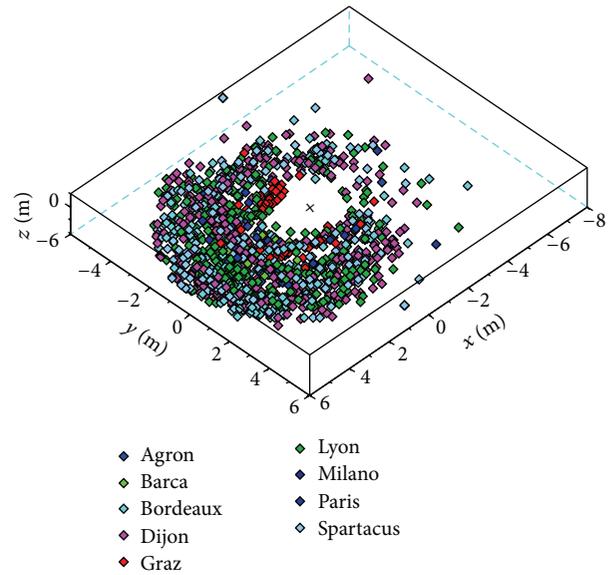


FIGURE 10: Opening angle of the receivers.

theoretical localization synchronization error for the speed of sound (340 m/s) is 3.4 cm.

Figure 10 shows the opening angle of the receivers. Therefore, measurement data from a localization experiment with 10 receivers was used. The positions of the smart phones emitting the received signals are plotted relatively to the receiver. All receivers are located in the same position in the center of Figure 10 and aligned in the same direction, marked with a cross. The opening of the microphone is in the positive x -direction. Thus, the figure shows the positions of the smart phone, where the corresponding receiver was able to detect the signal that was sent out by the smart phone. This can be seen as the opening angle of the receivers. The opening angle depends mainly on the directivity of the microphone and on the detection threshold of the receiver. As expected, the microphone receives the highest number of signals in the direction of the microphone. Going towards the back of the receiver, the number of received signals decreases and has a minimum at 180° from the front.

3.4. Software Application. We have developed an Android software application (app), which transforms a standard COTS device into a transmitter for ASSIST. Fundamentally our designed application has three functionalities: (I) communication with the evaluation unit (server), (II) sound control, (III) and visualization of the current position on the map.

- (i) The system works when the user downloads and starts the app in an area which supports the ASSIST infrastructure. The user interface is simple as one starts the app, which connects to an evaluation unit and receives an ID using its internet connection. Every registered hand-held device in a localization cell is assigned a unique ID. The smart phone is connected to the internet without a special infrastructure, only



FIGURE 11: The developed Android software application on the screen of a smart phone.

a mobile network is mandatory. In this work long term evolution (LTE) is used for wireless data communication which is the latest standard technology of mobile data transmission. The smart phones and the server communicate 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 building map. Based on these data, the smart phone regularly sends out the chirp signal to guarantee localizing the user.

- (ii) The app controls the loudspeaker of the smart phone and generates the specific sound signals (which is described in chapter 4) inside the smart phones.
- (iii) The current position and the map are transmitted from the evaluation unit to the smart phone. The position of the user is displayed on the screen of the smart phone in context to the environment, with a map and surrounding items. Figure 11 shows an example of the software application on a smart phone 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 real time from the evaluation unit. Displaying the current position on the smart phone has a time latency of approximately 12 ms to 410 ms. This is due to the window size of the signal processing ($N_{\text{wind}} \in (2 \text{ ms}, 400 \text{ ms})$), the calculation (1-2 ms), and the transmission of the data to the smart phone by Wi-Fi (2-10 ms).

4. Localization with TDoA

In our approach, we use TDoA-Algorithms to calculate the position of the smart phones. When using TDoA-Algorithms for localization, the processing time inside the smart phones is not relevant. For using other localization algorithms, the position accuracy would be affected in a negative sense if the processing time is not measured. Only by knowing the propagation speed of sound and the precise arrival times at the receivers, the position of the smart phone device can be calculated. The receivers are connected to a Wi-Fi network, such that they synchronize their clocks and exchange the time differences of arrival of the received sound impulses. A smart phone transmits acoustic signals at a position S_0 relative to the receivers with the positions M_i ($i = 1, \dots, n$). Further, the receivers detect the signals at different timestamps T_i , which depend on the distance R_i between the receiver and transmitter. Moreover, the distance from the smart phone to the receiver R_i can be described by the coordinates as follows:

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

The speed of sound C_{air} can be calculated in the air according to the following equation:

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

The speed of sound depends on the temperature ϑ of the environment. At a temperature of 25°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. Hence, the theoretical maximum localization error, caused by synchronization error, is 3.4 cm.

Smart phones generate specific sound signals at time t_0 . Thus, the distance R_i from (1) can be calculated by multiplying the speed of sound C_{air} with the transmitted time $T_i - t_0$ as given in the following equation:

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

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

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

Equation (4) is the hyperboloid description for 2 receivers. As a result, iterative TDoA-Algorithm with a minimum of 3 receivers can calculate the location of the smart phones in 2D.

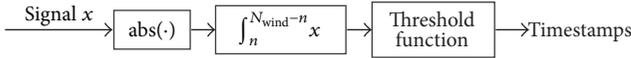


FIGURE 12: Description of the signal processing with envelope detection (blue block in Figure 7). The absolute value of the signal is integrated over the floating window with the size N_{wind} .

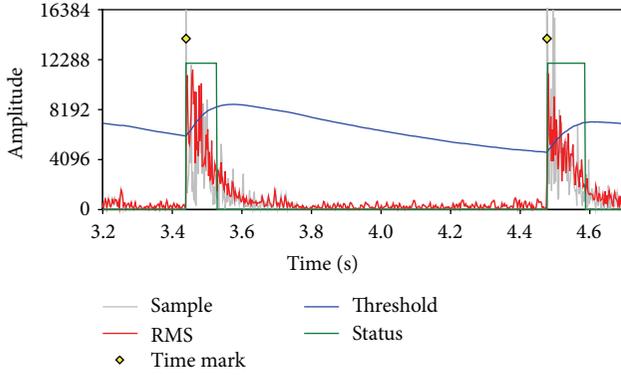


FIGURE 13: Threshold detection of the incoming sound signals.

4.1. Envelope Detection and Particle Filter. Our first approach uses only the amplitude of an incoming sound signal to detect its presence. Therefore, the smart phone generates short sound impulses with 18 kHz.

The approach of using envelope detection of sound signals is relatively easy but suffers from different drawbacks. The amplitude of sound decreases rapidly with distance. In the presence of background noise, one cannot distinguish between wanted and unwanted signals. Figure 12 shows the functional diagram of the signal processing and Figure 13 shows the threshold detection.

To increase the robustness against measurement outliers and incorrect initialization, we implemented a particle filter for localization of the smart phone. The algorithm is described in [32]. Our method is robust against measurement outliers and incorrect initialization. This is achieved through a probabilistic sensor model for TDOA data which explicitly considers the measurement uncertainty and takes into account disproportional errors caused by measurement outliers.

4.2. Chirp Impulse and Self-Calibration. In a second approach, we use a chirp impulse to increase the performance of the system by using pulse compression.

4.2.1. Chirp Impulses. We use linear chirp signals to transmit the sound signal. A linear chirp is a signal in which the frequency increases or decreases linearly with time (up- and down-chirps). Some of their characteristics make them applicable for localization. Signals with maximum energy are essential for receiving short signals over 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

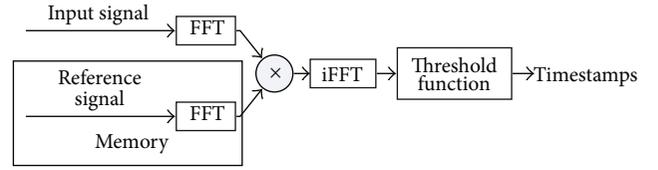


FIGURE 14: Description of the signal processing with the chirp (blue block in Figure 7). The signal is multiplied in the Fourier domain to estimate the peaks.

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 increase the SNR for a given bandwidth.

When autocorrelating a linear chirp signal, the resulting function shows a high and narrow peak. This characteristic allows high temporal accuracy for detecting signals. Cross correlating chirps in different frequency bands or up and down chirps, the resulting function does not show a distinct peak. This characteristic can be used to have multiple emitters operating at the same time. References [23, 28] show this for the detection of sound and ultrasound signals.

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:

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

The received signal is cross correlated with a stored up and down reference chirps. The mathematical formula for cross-correlation of two signals x and y is

$$z(t) = \int x(\tau) y(t + \tau) d\tau, \quad (6)$$

where $x(t)$ is the received signal and $y(t)$ is the saved reference signal. Further, the maximum of the cross-correlation function $z(t)$ is achieved at the perfect matched time. Hence, we use a matched filter to maximize the SNR. To detect different smart phones, up and down chirps are used to transmit the ID of the specific smart phone as a binary data stream. The cross-correlation is carried out as a convolution, which in turn equals a multiplication of the two signals in the frequency domain. The spectra of the input chirp and reference chirp are calculated with the fast Fourier transform (FFT). After multiplying the spectra of input chirp and reference chirp, the inverse FFT is used to convert the signal back to the time domain. Figure 14 shows the principle of the threshold function in the frequency domain. When a chirp, equal to the reference chirp, is present in the FFT-window, a peak occurs in the output signal. The position of the peak can be related to the time, when the input chirp was arriving at the receiver. Comparing the times of arrival of multiple receivers, one can realize time difference of arrival (TDoA) based localization.

Using a constant static threshold limits the transmission range by a high value to reduce false detections. However, an adaptive threshold, which detects the presence of the signal

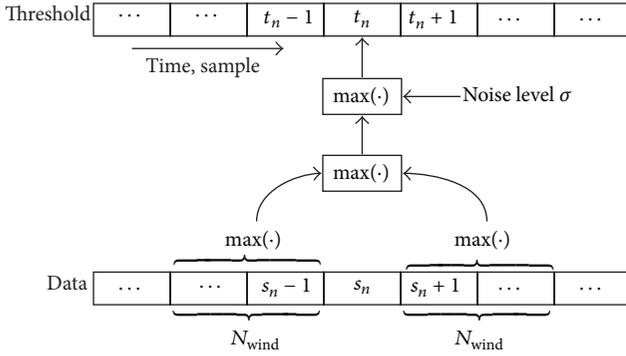


FIGURE 15: Adaptive threshold calculation by GOGO-CFAR. The greatest value is taken from both windows.

and increases the threshold and decreases the threshold for lower signal values can improve the sensitivity of the system. Moreover, an adaptive threshold can also reduce false detection by echoes due to increasing the threshold after the receipt of the signal. Therefore, we modified the constant failure alarm ratio (CFAR) algorithm to calculate the adaptive threshold [33]. Furthermore, we used only the maximum values of the windows to take from both windows the greatest value. Figure 15 shows the principle function of the cfar algorithm. The algorithm takes two windows, one before the point and one after the threshold point. Then, the greatest value is taken in each window (greatest of, GO) and is compared with the minimum noise level σ .

Figure 16 shows the GOGO-CFAR threshold for low correlation amplitudes for a distance of 20 m. Hence, the peaks are detected and the threshold is above the noise level. Furthermore, false detections by echoes are reduced by increasing the window size. This is shown in Figure 16 at the time 0.91 s to 0.94 s. Where some echoes causes a high correlation value; however, the threshold is above this disturbance.

4.2.2. Calibration Phase. Absolute localization systems uses fixed and installed anchor nodes as infrastructure. Further, the localization system has to know the position (x_i, y_i) of every receiver (anchor node) a priori. A multilateralists TDoA-Algorithm requires position information of the receivers to calculate the relative position (x, y) of the mobile object (e.g., smart phone). 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. The localization system ASSIST uses an Anchor-free localization algorithm to calibrate the system. During the calibration phase, the positions of the receivers in the indoor scenario are calculated automatically. Moreover at least three measured receiver positions are required for the orientation of the system on a map. The sending time of signals by a smart phone are not known to the localization system, as this would require synchronization of the smart phone using a very unreliable network, or bidirectional exchange of sound signals. Only the times of reception of

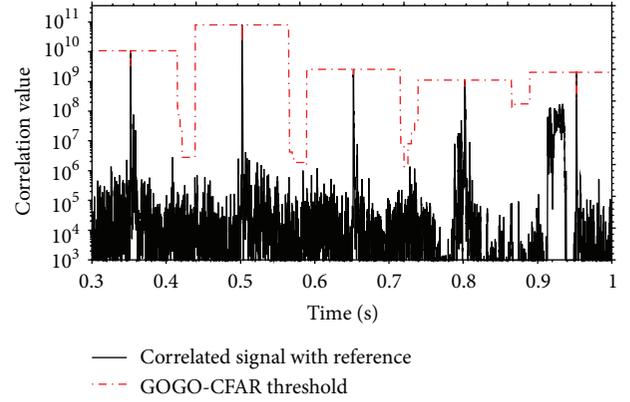


FIGURE 16: Adaptive threshold GOGO-CFAR for real signal with small correlation amplitude (20 m distance). Echoes and noise is suppressed.

the signals at the receivers can be measured. However, the receivers are synchronized and the TDoA values can be calculated. This forms 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 [34–36].

For the general case of arbitrarily distributed signal positions [37] 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 [38] 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 [39, 40] solves iteratively a nonlinear 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% (with only six received signals and four receivers). Through using the algorithm, a quick-setup system for smart phone localization is created. There is no need to measure the positions of the receivers.

5. Experimental Results

We show measurement results for localization with the acoustic system and a possibility of using an IMU for localization.

5.1. Constant Frequency Sound Pulse. In the first experiment, we use pulses with constant frequency and use the envelope to detect the presence of the signal. Figure 17 shows the real-world indoor scenario with ten receivers, which were placed

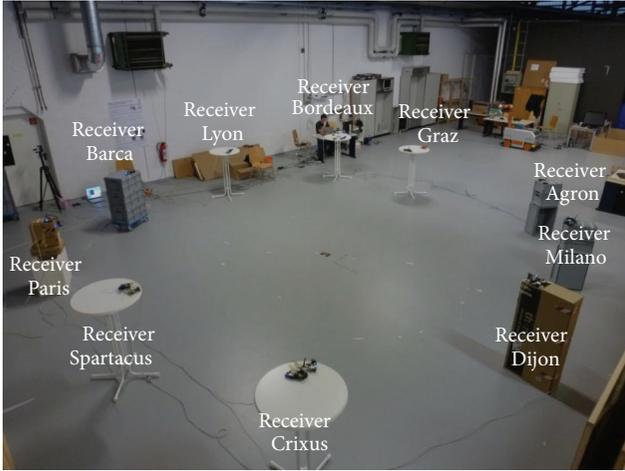


FIGURE 17: Photo of the real-world experiment environment.

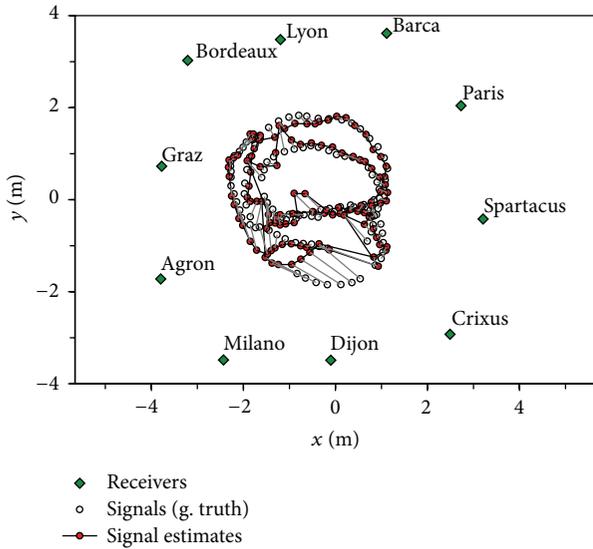


FIGURE 18: Trajectory of the real-world experiment with envelope detection and particle filter. The mean positioning error of the signal estimate is 0.25 m ($\sigma = 0.26$ m).

around the area of the optical motion capture system. The absolute accuracy of the motion capture system is in the range of about 3 mm [41]. Figure 18 shows the smart phone localization by a particle filter. Further, the particle filter localizes the smart phone with an error of $\sigma = 0.26$ m. Figure 19 shows the cumulative distribution of the distance errors.

We analyzed in an additional experiment the receiver range. Therefore, an iPhone 4S is positioned at a variable distance between 1 and 12 meters from two receivers. At each distance of 1 meter, the smart phone transmits 500 acoustic pulses. The length of each acoustic pulse is 50 ms with a frequency of 18 kHz. Figure 20 shows the correct received pulses in percentage over different distances. The accuracy that was achieved is within an interval of $\pm 2\sigma$.

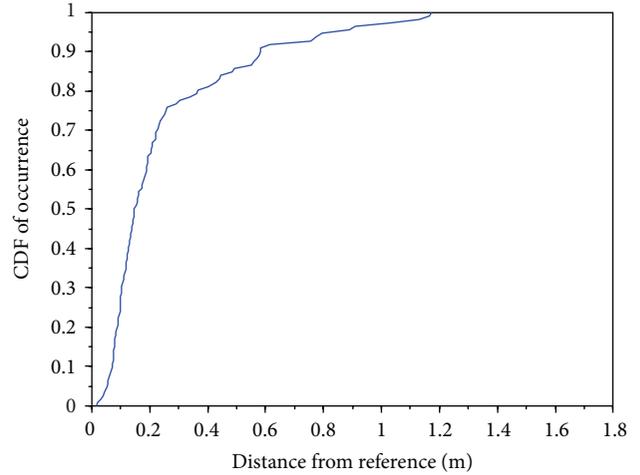


FIGURE 19: Cumulative distribution of distance errors for Figure 18.

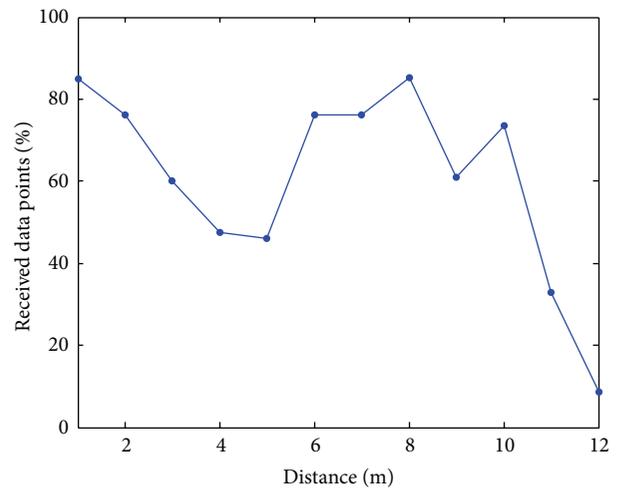


FIGURE 20: Measurement values of ASSIST with envelope detection depending on the distance.

5.2. Chirp Sound Pulse. The measurement deviation of the system was evaluated in static experiments. Figure 21 shows the distribution of measurement errors and the normal distribution. ASSIST shows a standard deviation of $\sigma = 25$ cm. Signals with multipath propagation lead to increased standard deviation.

Further, 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 smart phone which transmitted acoustic chirp impulses between 19 kHz and 20 kHz with a length of 50 ms. Figure 22 shows the calculated positions and the defined walking track. Thus, the data shows a well match compared to the track. The trajectory shows a systematic error which depends on the localization algorithm

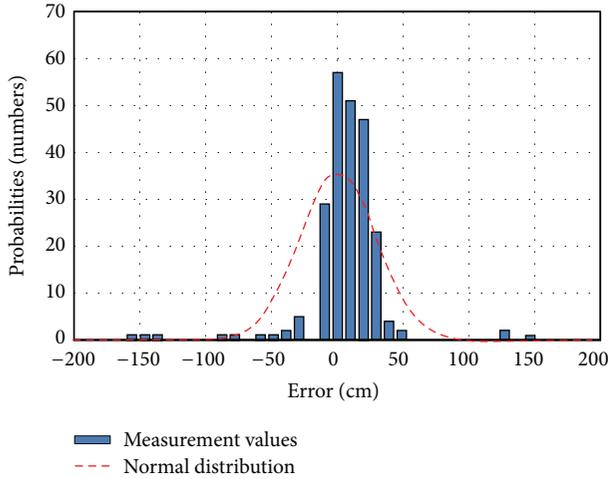


FIGURE 21: Measurement errors of ASSIST with correlation and TDoA-Algorithms in a static experiment.

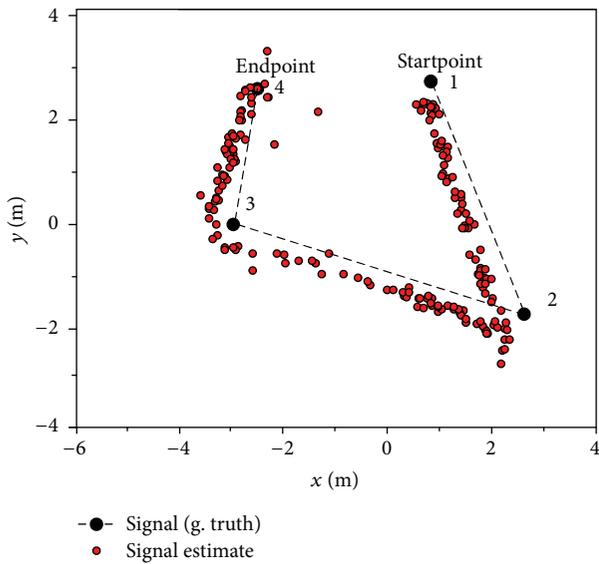


FIGURE 22: Real-world experiment of ASSIST. Data set *Exp. 2d*. Microphone error mean 1.81 m, ($\sigma = 1.32$ m). Signal position error mean: 0.34 m ($\sigma = 0.18$ m).

and some measurement errors from the multipath propagation.

The smart phone track shows an average deviation of 0.34 m ($\sigma = 0.18$ m).

In an additional experiment, the receiver range was analyzed. Figure 23 shows the signal detection rate for arrived chirp signals as a function of distance which are inside of $\pm 2\sigma$. The developed receivers (chapter III-C) with a constant threshold were able to receive more than 70% of the transmitted signals up to a distance of 16 m from a smart phone (red curve with circles). The percentage of received signals drops at distances above 16 m. Moreover, with the adaptive GOGO-CFAR threshold, the detection rate is approximately 88% at a distance of 20 m (black curve).

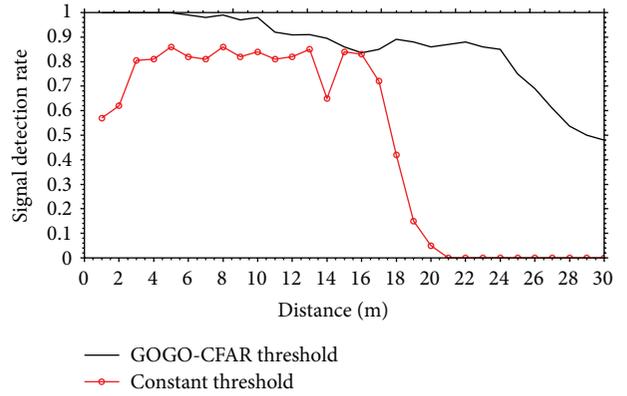


FIGURE 23: Measurement values of ASSIST with chirp correlation depending on the distance. Black line is with adaptive GOGO-CFAR threshold and red line with circles shows the results for constant threshold.

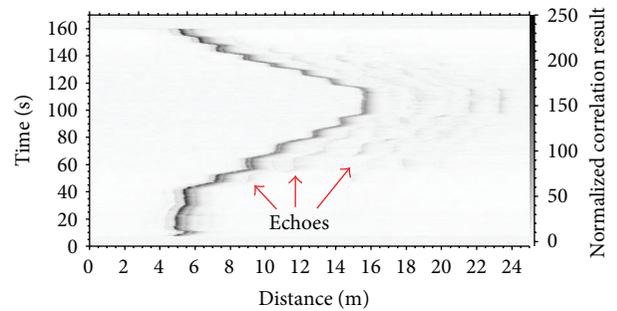


FIGURE 24: Analysis of the echoes. The abscissa represents the time delay and the ordinate represents the signal number. There are multiple echoes due to reflections.

Furthermore, at a distance of 30 m, we measured a detection rate of 48%.

5.3. Echo Analysis. Reflections at walls or hard surfaces (e.g., cabinet) induce echoes and disturb the line of sight signal. Furthermore, the echoes reduce the accuracy of the localization system, which is assumed to work on the line of sight signal. Figure 24 shows the echo analysis for different signals. The abscissa represents the distance of the signal and the ordinate time of the measurement. The brightness indicates the amplitude of the correlation. Every line is generated by using the modulo operation of 300 ms, which is the time interval of the transmitter. We started the measurement at approximately 5 m and moved to 16 m and back. In particular, at distances above 10 m, the echoes become strong and are visible as the shadows of the main signal in Figure 24. Thus, multiple (up to four at approximately 16 m) echoes can be distinguished. To achieve a robust localization, we perform the adaptive GOGO-CFAR threshold to remove the echoes and work on the line of sight signal.

5.4. 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.

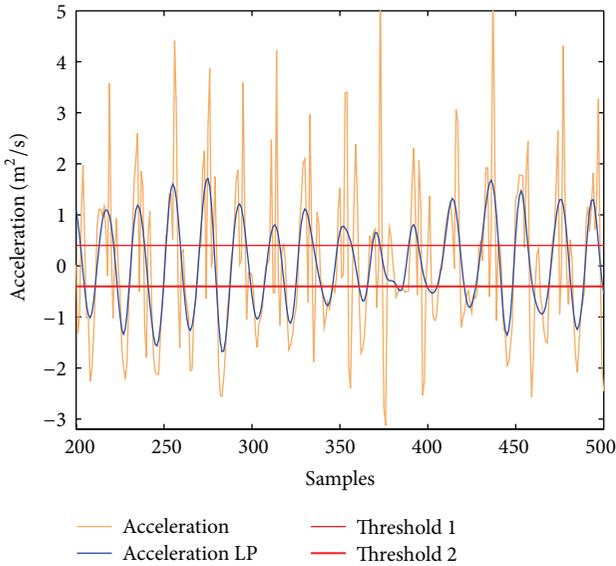


FIGURE 25: Measurement values of the acceleration sensor in z-axis during a walk with the two threshold values.

Currently, many different sensor types are integrated in the smart phones. For example, the commercial smart phone 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 observation time. The inertial sensor unit additionally supports a method to perform acoustic localization.

Since the smart phone is usually held by hands, methods as zero velocity update [42] 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. Figure 25 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 [43].

In an experiment, the data from inertial sensors of the smart phone 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 Figure 26. The red-dashed 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.

6. Conclusions

In this paper, we presented a smart phone indoor localization system based on sound. The user of the system needs no additional hardware except a COTS smart phone. Through

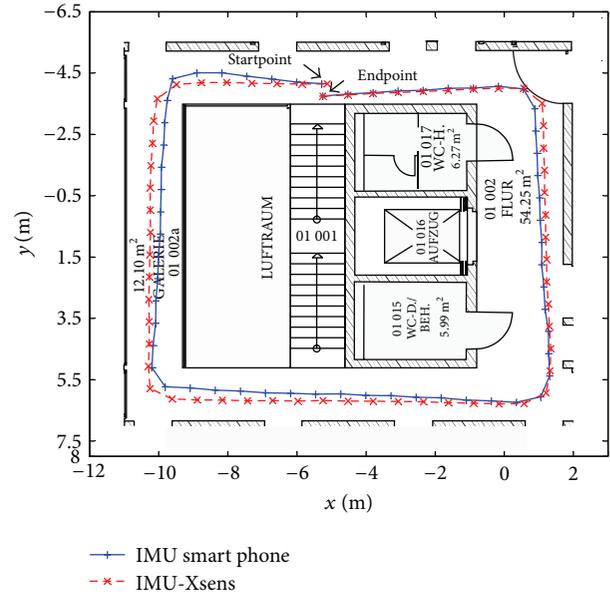


FIGURE 26: Trajectory of data from an experiment. Data from the inertial sensors of a smart phone (blue line) and a reference inertial measurement unit (red-dashed line).

our self-built receivers, which were synchronized with a Wi-Fi network, the arrived signal can be correlated and the position is calculated with a TDoA-Algorithm. 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.

In our paper, two different approaches were tested. The envelope detection shows an easy implementation but a limited transmission range of 11 m. The particle filter showed a robust localization of the smart phone with an error of $\sigma = 0.26$ m. In a second approach chirp, impulses were used and correlated in the receivers, which increased the range to 16 m. Moreover, the adaptive GOGO-CFAR threshold extend the detection rate to 48% at a distance of 30 m. Additionally, a self-calibration algorithm is used, which localizes the anchor nodes in a range of $\sigma = 1.32$ m and the smart phone in a range of $\sigma = 0.18$ m.

In areas with a poor receiver coverage, the localization with built-in inertial sensors in smart phone 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.

7. Outlook

In our future investigations, we will improve the acoustic localization in situations where there is no line of sight

between the smart phone and the receivers. Error minimization can be achieved, through fusion of the data from the inertial sensors and the data from the acoustic localization. Additionally, we will use the self-calibration algorithms in combination of the particle filter to increase the robustness.

We will modulate an identifier onto the signal to distinguish between different senders and enable multiuser applications. Another possibility is to use time-division multiplexing (TDM) to identify different smart phones. The time domain is divided into time slots which can be used from the respective of smart phones to generate the sound.

In some applications (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 reference system to measure the systematic error precisely.

Disclosure

This paper is an extended version of [2].

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work has partly been supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) within the Research Training Group 1103 (Embedded Microsystems). Special thanks are due to our technicians Hans Baumer, Christoph Bohnert and Uwe Burzlaff for the redesign of the hardware.

References

- [1] K. Schreiner, "Where we at? Mobile phones bring GPS to the masses," *IEEE Computer Graphics and Applications*, vol. 27, no. 3, pp. 6–11, 2007.
- [2] F. Höflinger, J. Wendeberg, R. Zhang et al., "Acoustic self-calibrating system for indoor smartphone tracking (ASSIST)," in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '12)*, pp. 1–9, Sydney, Australia, November 2012.
- [3] F. Höflinger, J. Hoppe, R. Zhang et al., "Acoustic indoor-localization system for smart phones," in *Proceedings of the IEEE 11th International Multi-Conference on Systems, Signals and Devices (SSD '14)*, pp. 1–4, Barcelona, Spain, February 2014.
- [4] V. Otsason, A. Varshavsky, A. LaMarca, and E. de Lara, "Accurate GSM indoor localization," in *UbiComp 2005: Ubiquitous Computing*, pp. 141–158, Springer, Berlin, Germany, 2005.
- [5] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan, "Indoor localization without the pain," in *Proceedings of the 16th annual International Conference on Mobile Computing and Networking (MobiCom '10)*, pp. 173–184, ACM, Chicago, Ill, USA, 2010.
- [6] E. Martin, O. Vinyals, G. Friedland, and R. Bajcsy, "Precise indoor localization using smart phones," in *Proceedings of the International Conference on Multimedia (MM '10)*, pp. 787–790, ACM, Firenze, Italy, 2010.
- [7] W. Li, Y. Zhu, and T. He, "WiBee: Building WiFi radio map with ZigBee sensor networks," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM '12)*, pp. 2926–2930, Orlando, Fla, USA, March 2012.
- [8] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *Proceedings of the IEEE 19th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '00)*, vol. 2, pp. 775–784, March 2000.
- [9] M. Bourimi, G. Mau, S. Steinmann et al., "A privacy-respecting indoor localization approach for identifying shopper paths by using end-users mobile devices," in *Proceedings of the 8th International Conference on Information Technology: New Generations (ITNG '11)*, pp. 139–144, Las Vegas, Nev, USA, April 2011.
- [10] 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 (MELT '08)*, pp. 55–60, September 2008.
- [11] F. Höflinger, J. Muller, R. Zhang, L. M. Reindl, and W. Burgard, "A wireless micro inertial measurement unit (IMU)," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 9, pp. 2583–2595, 2013.
- [12] R. Zhang, A. Bannoura, F. Höflinger, L. M. Reindl, and C. Schindelbauer, "Indoor localization using a smart phone," in *Proceedings of the 8th IEEE Sensors Applications Symposium (SAS '13)*, pp. 38–42, February 2013.
- [13] R. Mautz, *Indoor Positioning Technologies*, ETH Zurich, Zürich, Switzerland, 2012.
- [14] B. Gozick, K. P. Subbu, R. Dantu, and T. Maeshiro, "Magnetic maps for indoor navigation," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 12, pp. 3883–3891, 2011.
- [15] A. D. Cheok and L. Yue, "A novel light-sensor-based information transmission system for indoor positioning and navigation," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 1, pp. 290–299, 2011.
- [16] N. Ravi, P. Shankar, A. Frankel, A. Elgammal, and L. Iftode, "Indoor localization using camera phones," in *Proceedings of the 7th IEEE Workshop on Mobile Computing System and Applications (WMCSA '06)*, pp. 1–7, April 2006.
- [17] M. Werner, M. Kessel, and C. Marouane, "Indoor positioning using smartphone camera," in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '11)*, pp. 1–6, September 2011.
- [18] B. Williams, G. Klein, and I. Reid, "Real-time SLAM relocation," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV '07)*, pp. 1–8, Rio de Janeiro, Brazil, October 2007.
- [19] D. Wagner and D. Schmalstieg, "First steps towards handheld augmented reality," in *Proceedings of the 7th IEEE International Symposium on Wearable Computers (ISWC '03)*, pp. 127–135, October 2003.
- [20] 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 (MobiSys '05)*, pp. 191–203, ACM, June 2005.
- [21] H. Liu, Y. Gan, J. Yang et al., “Push the limit of WiFi based localization for smartphones,” in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking (MobiCom '12)*, pp. 305–316, ACM, New York, NY, USA, August 2012.
- [22] A. Mandal, C. V. Lopes, T. Givargis, A. Haghghat, R. Jurdak, and P. Baldi, “Beep: 3D indoor positioning using audible sound,” in *Proceedings of the 2nd IEEE Consumer Communications and Networking Conference (CCNC '05)*, pp. 348–353, January 2005.
- [23] C. Peng, G. Shen, and Y. Zhang, “BeepBeep: a high-accuracy acoustic-based system for ranging and localization using COTS devices,” *Transactions on Embedded Computing Systems*, vol. 11, no. 1, article 4, 2012.
- [24] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, “BeepBeep: a high accuracy acoustic ranging system using COTS mobile devices,” in *Proceedings of the 5th ACM International Conference on Embedded Networked Sensor Systems (SenSys'07)*, pp. 1–14, November 2007.
- [25] I. Rishabh, D. Kimber, and J. Adcock, “Indoor localization using controlled ambient sounds,” in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '12)*, pp. 1–10, November 2012.
- [26] A. Carroll and G. Heiser, “An analysis of power consumption in a smartphone,” in *Proceedings of the USENIX Conference on USENIX Annual Technical Conference (USENIXATC '10)*, p. 21, USENIX Association, Berkeley, Calif, USA, 2010.
- [27] K. Lin, A. Kansal, D. Lymberopoulos, and F. Zhao, “Energy-accuracy trade-off for continuous mobile device location,” in *Proceedings of the 8th Annual International Conference on Mobile Systems, Applications and Services (MobiSys '10)*, vol. ACM, pp. 285–297, ACM, New York, NY, USA, June 2010.
- [28] M. Sakamoto, M. Sugawara, K. Kaga, and T. Kamio, “Average thresholds in the 8 to 20 kHz range as a function of age,” *Scandinavian Audiology*, vol. 27, no. 3, pp. 189–192, 1998.
- [29] V. Filonenko, C. Cullen, and J. Carswell, “Investigating ultrasonic positioning on mobile phones,” in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '10)*, pp. 1–8, September 2010.
- [30] J. Wendeberg, *Laufzeitbasierte Schallortung mit unbekanntem Sender- und Empfängerpositionen [M.S. thesis]*, University of Freiburg, 2009.
- [31] B. Sundararaman, U. Buy, and A. D. Kshemkalyani, “Clock synchronization for wireless sensor networks: a survey,” *Ad Hoc Networks*, vol. 3, no. 3, pp. 281–323, 2005.
- [32] J. Wendeberg, J. Muller, C. Schindelhauer, and W. Burgard, “Robust tracking of a mobile beacon using time differences of arrival with simultaneous calibration of receiver positions,” in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '12)*, pp. 1–10, Sydney, Australia, November 2012.
- [33] H. Rohling, “Radar CFAR thresholding in clutter and multiple target situations,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 19, no. 4, pp. 608–621, 1983.
- [34] T. Janson, C. Schindelhauer, and J. Wendeberg, “Self-localization application for iPhone using only ambient sound signals,” in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '10)*, pp. 259–268, November 2010.
- [35] J. Wendeberg, T. Janson, and C. Schindelhauer, “Self-localization based on ambient signals,” *Theoretical Computer Science*, vol. 453, pp. 98–109, 2012.
- [36] S. Thrun, “Affine structure from sound,” in *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS '05)*, pp. 1353–1360, MIT Press, Cambridge, Mass, USA, December 2005.
- [37] R. Biswas and S. Thrun, “A passive approach to sensor network localization,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '04)*, vol. 2, pp. 1544–1549, October 2004.
- [38] M. Pollefeys and D. Nister, “Direct computation of sound and microphone locations from time-difference-of-arrival data,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 2445–2448, April 2008.
- [39] J. Wendeberg, F. Höflinger, C. Schindelhauer, and L. Reindl, “Anchor-free TDOA self-localization,” in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '11)*, pp. 1–10, September 2011.
- [40] J. Wendeberg, F. Höflinger, C. Schindelhauer, and L. Reindl, “Calibration-free TDOA self-localisation,” *Journal of Location Based Services*, vol. 7, no. 2, pp. 121–144, 2013.
- [41] J. Müller, N. Kohler, and W. Burgard, “Autonomous miniature blimp navigation with online motion planning and replanning,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '11)*, pp. 4941–4946, September 2011.
- [42] L. Ojeda and J. Borenstein, “Non-GPS navigation for security personnel and first responders,” *Journal of Navigation*, vol. 60, no. 3, pp. 391–407, 2007.
- [43] R. Zhang and L. M. Reindl, “Pedestrian motion based inertial sensor fusion by a modified complementary separate-bias Kalman filter,” in *Proceedings of the 6th IEEE Sensors Applications Symposium (SAS '11)*, pp. 209–213, February 2011.
- [44] J. Bordoy, P. Hornecker, F. Höflinger, J. Wendeberg, C. Schindelhauer, and L. M. Reindl, “Robust tracking of a mobile receiver using unsynchronized time differences of arrival,” in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN '13)*, Montbeliard, France, November 2013.

