Acoustic Receivers for Indoor Smartphone Localization

Joachim Hoppe, Fabian Höflinger, Leonhard M. Reindl
Department of Microsystems Engineering
University of Freiburg, Germany
hoppe.joachim@gmail.com
{hoeflinger, reindl}@imtek.uni-freiburg.de

Abstract—**This paper proposes solutions for designing acoustic receivers for detecting high frequency sound signals, emitted by commercial smartphones. The receivers can build the foundation of an indoor localization system. Detection takes place over a distance of up to 16 m, with an accuracy of 25 cm. In order to be competitive, the cost of the receivers is kept low and the installation effort is minimal. We compared three different approaches for sound signal detection: envelope detection, a single tone detection IC and chirp correlation. Furthermore, we estimated the audibility of the sound signals, based on measurements of the sound pressure level of commercial smartphones at different frequencies and distances.**

Keywords: smartphone localization, ultrasound, sound, tracking

I. Introduction

The indoor localization of smartphones has been becoming more and more important as smartphones have been spreading widely and play an important role in everyday life for many people. Up to this day, some methods for localizing smartphones in an indoor environment already exist. Due to the high availability of wireless networks, WLAN-based approaches, as [1] exemplarily, have been strongly under examination. Other methods make use of the internal sensors of smartphones like the built-in camera [2] or accelerometer and compass [3].

Some research groups have also started to investigate the built-in loudspeakers and microphones in smartphones for different purposes. Peng et al. implemented the ranging system BeepBeep [4], which is an essential step towards a localization system. BeepBeep allows distance measurements between two smartphones over a distance of up to 10 m, achieving a resolution of about 1-2 cm. The used sound signals are in the range of 2-6 kHz and therefore hearable for human beings.

Using sound signals in crowded rooms implies the use of frequencies that cannot be heard by human beings, to avoid annoyance. Filonenko et al. have shown, amongst others, that smartphones are able to emit sound beyond the hearing range of humans [5]. A method for transmitting data between two smartphones over a range of up to 80 cm using sound signals with frequencies between 20-23 kHz, therefore not hearable for most humans, was realized by Arentz and Bandara [6].

Tarzia et al. use smartphones as receivers to create a localization system working with room level resolution [7]. The localization is based on the acoustic background spectrum of each room. In their approach they exclusively use the built-in microphone of smartphones to take an acoustic fingerprint of a room. No additional hardware is required.

To realize a smartphone localization system, which does not need a costly infrastructure and allows highly accurate localization, we find it promising to use sound. We use smartphones to emit short sound pulses outside the human range of hearing. In this work we show the realization of acoustic receivers, which can be used in indoor localization systems. Our approach can compete with other approaches that use little or no additional infrastructure, because of the low price of our receivers combined with the fact that the installation effort can be minimized. The receivers can calculate their own positions automatically, as shown in [8].

II. Smartphones as Sound Emitters

A. Geometrical Spreading

We are using smartphones to emit sound signals. Assuming a smartphone behaves like a spherical sound source, the sound pressure p varies inversely proportional to the distance of the source r [9].

$$p \sim 1/r \tag{1}$$

The sound pressure level L_p, which is a logarithmic measure of the sound pressure, can be calculated with the rms value of the sound pressure.

$$L_p = 20 \cdot \log(p_{rms} / p_{ref}), \tag{2}$$

where p_{ref} is a reference sound pressure of 20 μPa.

Using (1) and (2) one can calculate the difference in sound pressure level, when changing the distance to the source from r_1 to r_2.

$$\Delta L_p = L_{p,r2} - L_{p,r1} = 20 \cdot \log(r_1/r_2) \tag{3}$$

For example, when doubling the distance to the source, which means that $r_2 = 2r_1$, the sound pressure level drops by -6 dB.

In a real world scenario, our receivers will, for example, be placed at the ceiling or walls of a supermarket, so we expect no smartphone to get closer to a receiver than 10 cm. We defined the size of one localization cell to be $10 \cdot 10$ m². The decrease of sound pressure level in this range, due to geometrical spreading, is around -40 dB.

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

B. Air Absorption and Other Effects

The sound pressure also decreases as a result of air absorption. Damping, due to air absorption, increases with increasing distance and frequency. As reported in [9] the sound pressure of a sound wave with a frequency of 20 kHz drops by -1 dB over a distance of 12.5 m. Compared to the decrease by geometrical spreading of -40 dB over 10 m, the influence of air absorption can be neglected. In a real world scenario the expected dynamic range of the sound pressure variations are higher than 40 dB. One reason for this is the directivity of the smartphone speaker, as shown in [10] for the iPhone 4S. Additional damping occurs, for example when no line of sight exists between smartphone and receiver. In this case the sound has to travel a longer way and looses parts of its energy at reflecting boundaries.

III. Human Hearing at High Frequencies

Human hearing is best at frequencies where most of speech takes place, which is around 0.5-6 kHz. When moving away from this frequency range, the ability to hear worsens. 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. Usually it is plotted as a function of frequency. The hearing threshold of humans increases at high frequencies [11].

Sakamoto et al. have conducted measurements of the absolute hearing threshold in the frequency range from 8 kHz to 20 kHz, for different age groups [12]. 65 persons with normal hearing were tested. In the range from 18-20 kHz they report average hearing thresholds between 112-148 dB, which is highest for the oldest age group. The youngest age group shows the lowest average hearing threshold and the highest standard deviation. At 18 kHz the standard deviation has a maximum value of 22 dB. The standard deviation decreases with increasing frequency to a value of 7.5 dB at 20 kHz.

In their paper, the hearing threshold was measured under laboratory conditions. In the presence of background noise, which will appear in a crowded building, the hearing threshold will be raised through the psychoacoustic effect of masking.

There is no fixed frequency above that humans cannot hear. The audibility of a sound signal depends on its frequency and on its sound pressure. The sound pressure of a smartphone depends on the distance to the smartphone, see (3), and on the frequency response of the loudspeaker.

To evaluate the audibility of high frequency sound signals emitted by smartphones, we measured the sound pressure level of different commercial smartphones at 18 kHz and 20 kHz for the distances 1 cm, 10 cm and 5 m.

The smartphones were chosen arbitrarily from popular models. The difference between smartphone sound pressure and average hearing threshold for a specific frequency is given in Table I, in units of the corresponding standard deviation σ. A high number corresponds to a high percentage of people not being able to hear a sound. For example a difference of 1σ corresponds to 50 % + 68.3/2 % = 84.15 % of the test persons not being able to hear that specific sound. A difference of 3σ corresponds to 99.85 % and a difference of 4σ relates to 99.975 % not being able to recognize a sound. For the calculations the values of the average hearing threshold of the youngest age group were taken.

As expected, the audibility of the sound signals decreases with increasing frequency and distance to the smartphones.

One could draw the conclusion to use as high frequencies as possible. As we show in [10], frequencies up to 22 kHz can be reproduced by smartphone speakers. However the frequency response drops with increasing frequency. A compromise between audibility and emitted signal strength has to be made. In this work we use the frequency range from 18-22 kHz.

TABLE I. Difference between average hearing threshold and sound pressure level emitted by smartphones in units of the standard deviation.

Distance to smartphone	*Smartphone 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	7.2σ
	iPod Touch	1.7σ	6,5σ
5 m	iPhone 4S	3.5σ	11.7σ
	Samsung GT-S5830	3.4σ	12.0σ
	iPod Touch	3.5σ	11.3σ

IV. Approaches for Detecting High Frequency Sound Signals Emitted by Smartphones

A. Choosing a Microphone

The first part in the signal chain of our receivers is a transducer, which converts acoustical signals into electrical signals. For our purposes the transducer should be small and cheap. The microphone types we are left with are electret- and MEMS-microphones. We have measured the sensitivity as a function of frequency, i.e. the frequency response of several electrets-microphones, from the manufacturers Kingstate and Ekulit. We compared these measurements to the frequency response of a MEMS-microphone from Knowles Acoustics (see Fig. 1). The frequency response of the electret-microphones drops with increasing frequency. The MEMS-microphone shows a peak around 20 kHz. Thus, for detecting sound signals in the range of 18-22 kHz the use of the MEMS-microphone is preferred.

Further, we compared three different types of signal processing: envelope detection, single tone decoder and chirp correlation.

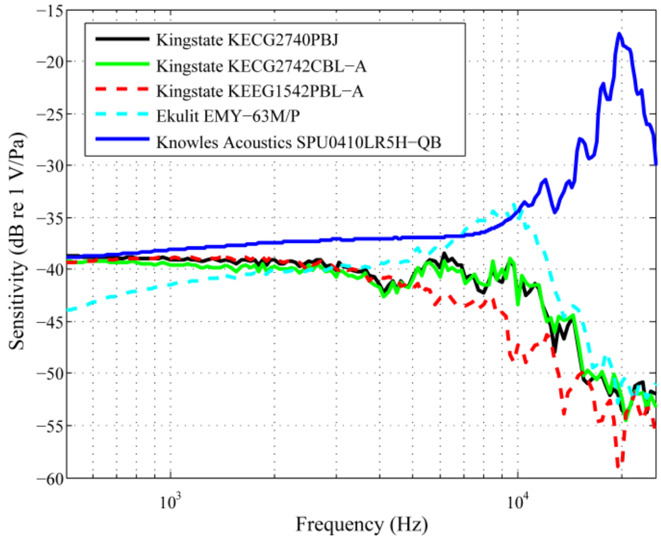

Figure 1: Frequency response of electret- and MEMS-microphones in the range from 500 Hz – 25 kHz.

B. *Envelope Detection*

Our first approach uses only the amplitude of an incoming signal to detect its presence. The signal processing is completely analog. A tone with a fixed frequency and duration is emitted by the smartphones. The signal is converted by a microphone, pre-amplified, filtered and amplified again. Afterwards an incoherent demodulator detects the envelope of the incoming signals (see Fig. 2).

The filter is an active 8^{th} order highpass filter, made of four cascaded Sallen-Key structures. Each filter stage is tuned to the same cutoff-frequency of 17.5 kHz. The amplifier stages are decoupled with passive bandpass filters.

The output signal from the envelope detector is sent to an Atmel ATxemga 128 for analog to digital (A/D) conversion. The sampling rate is 25.6 kHz with a resolution of 8 bit. The digitalized values are then sent to a computer via the universal serial bus (USB). Here threshold detection takes place.

We carried out an experiment to verify the accuracy of this approach. A smartphone was used to emit pulses with a frequency of 19 kHz and duration 50 ms. A measurement of the time difference of arrival (TDOA) was done repeatedly between two receivers and the positioning error was calculated for each of the 300 measurements. The distance between the smartphone and the receivers was 2 m. All positions were kept constant during the measurements. The receivers were synchronized via WLAN. Synchronization between the receivers and the smartphone is not needed. The experiments were conducted in an acoustically untreated hangar. Using the relation

$$s = c \cdot t, \qquad (4)$$

the time differences were converted into distances. In (4) s is the distance, which sound travels during time t. The speed of sound c is 343 m/s, at a temperature of 20 °C. The standard deviation we measured was 16 cm. In another experiment we measured the percentage of received data points in relation to the number of sent pulses.

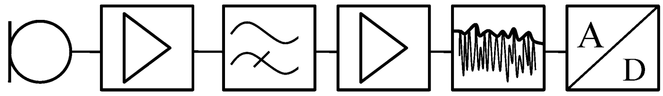

Figure 2: Diagram of the signal processing chain, used for envelope detection. The building blocks are: Microphone, pre-amplifier, filter, amplifier, envelope detector and A/D converter.

At a distance of 10 m we received 74 % of the points. For greater distances the percentage of received points was decreasing. At a distance of 12 m the number dropped to 9 %. Only points within a range of $\pm 2\sigma$ were included.

One disadvantage of using envelope detection is that it does not allow distinguishing between different smartphones at the same time. This problem can be solved by using a time-division multiplexing (TDM) scheme. This approach also misses robustness against background noise. Especially short sounds, with a broad spectrum, cannot always be filtered out adequately.

C. *Single Tone Detector*

With the use of a single tone decoder, which can normally be found in dual tone multy frequency (DTMF) circuits, the detection of the signal's presence can be made more robust. The single tone decoder can be tuned to a certain frequency. The output is a digital signal, being high or low, indicating a valid input signal or not. We used the integrated circuit (IC) LM567CN. It uses the in-phase and quadrature-phase components of an incoming signal to detect the presence of a tone. We modified the circuit for envelope detection by exchanging the actual envelope detector for the DTMF chip.

Compared to envelope detection the robustness of signal detection could be improved. Another advantage is that when using several single tone decoders, tuned to different frequencies, several smartphones can be used at the same time.

Through experiments we found out that the standard deviation for a TDOA measurement was 8.74 m. We considered this value too high, to follow this approach further, as we are interested in performing localization in the centimeter range.

D. *Chirp Correlation*

In this approach we use linear chirp signals. Some of their characteristics make them applicable for localization. Equation (5) shows a formula to describe such a function.

$$chirp(t) = \sin(2\pi t(f_0 + k\, t/2)), \qquad (5)$$

with time t, start frequency f_0, $k = \Delta f/\tau$, Δf = stop frequency – start frequency and chirp duration τ. The frequency of the sinusoid is swept linearly from a start frequency to a stop frequency over the duration of the chirp. The auto-correlation function of a linear chirp signal 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. Numerous authors reported promising results by using this type of signal for detecting sound and ultrasound signals; see [4] and [13]. Also they have been used in radar and sonar applications since a long time. The available frequencies for chirps range from 18–22 kHz. Besides A/D conversion, the ATxmega 128 is now used to mix the incoming signals with a square wave and for lowpass filtering. The square wave has a frequency slightly above 22 kHz. The spectrum of an incoming chirp is shifted down to the frequency, which is the difference between the highest frequency of the chirp and the frequency of the square wave. After downmixing, groups of eight samples are added together to increase the resolution by three bit. At the same time this process performs lowpass filtering, due to the averaging of eight samples. The data rate is reduced to about 11 kHz, with a resolution of 15 bit per sample. Reduction of data rate was necessary because of its limitation by the USB transmission to a maximum of 25.6 kHz for 8 bit samples. Fig. 3 shows a diagram of the signal processing in the ATxmega 128. The signal is then sent to a computer, where it is cross- correlated to a reference chirp with the same properties. The cross-correlation is carried out as a convolution, which equals a multiplication of the two signals in frequency domain. The spectra of the input chirp and reference chirp are calculated with the Fast Fourier Transform (FFT). The inverse FFT is used to convert the signal back to time 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 of arrival of the chirp. Comparing the times of arrival of multiple receivers, one can realize TDOA based localization. The TDOA measurement, done as for the other two approaches, showed a standard deviation of 25 cm. For distances greater than 16 m the number of received pulses is decreasing. For the measurements, chirps with frequencies from 18-19 kHz and duration of 50 ms were used. The highest errors were caused by multipath propagation. Multipath propagation happens in indoor environments when sound gets reflected by the walls, floor or ceiling of a room.

V. Conclusion and Outlook

We compared three different approaches for receiving high frequency sound signals, emitted by smartphones. Localization with envelope detection results in a maximum distance of 10 m with an accuracy of 16 cm. However, it suffers from immunity against background noises and the ability to distinguish between different smartphones simultaneously. Single tone decoders can increase the robustness of signal detection, but their temporal accuracy is low. The use of chirp correlation is preferred, as with 16 m it shows the highest distance over which localization can be performed. The accuracy is around 25 cm. Furthermore it is possible to distinguish between several smartphones at once. With the achieved results applications can be in supermarkets and airports navigating people to desired products or gates, respectively. Other uses are at exhibitions or in hospitals. In future work we want to investigate the number of chirps which can be used in the available frequency range from 18-22 kHz. We also want to increase the accuracy of the chirp correlation by filtering out multipath propagations and improving our signal processing algorithms.

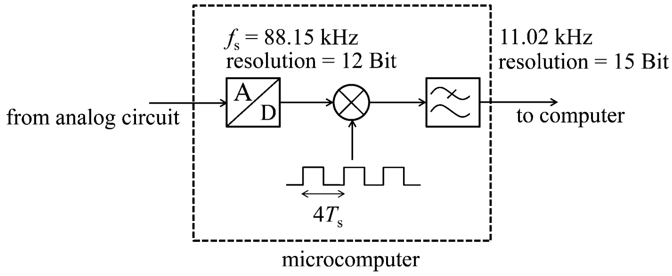

Figure 3: Diagram of the signal processing used in the ATxmega 128. The incoming signal is sampled with a rate of 88.15 kHz and a resolution of 12 bit. The signal is then multiplied with a rectangular function with a frequency of $f_s/4$. Groups of eight samples are added together to increase the resolution by 3 bit and lowpass filtering the signal.

References

[1] K. Chintalapudi, A. Padmanabha Iyer and V.N. Padmanabhan, “Indoor localization without the pain,” ACM, Proceedings of the 16th annual international conference on Mobile computing and networking, pp. 173–184, 2010.

[2] M. Werner, M. Kessel and C. Marouane, “Indoor positioning using smartphone camera,“ in Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1-6, 2011.

[3] J.A.B. Link, P. Smith, N. Viol and K. Wehrle, “FootPath: accurate map-based indoor navigation using smartphones,” in Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1-8, 2011.

[4] C. Peng, G. Shen, Y. Zhang, Y. Li and K. Tan, “BeepBeep: a high accuracy acoustic ranging system using COTS mobile devices,” ACM, Proceedings of the 5th international conference on Embedded networked sensor systems, pp. 1-14, 2007.

[5] V. Filonenko, C. Cullen and J. Carswell, “Investigating ultrasonic positioning on mobile phones,” in Proceedings of the 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1-8, 2010.

[6] W. A. Arentz and U. Bandara, “Near ultrasonic directional data transfer for modern smartphones,” ACM, Proceedings of the 13th international conference on Ubiquitous computing, pp. 481-482, 2011.

[7] S. P. Tarzia, P. A. Dinda, R. P. Dick and G. Memik, “Indoor localization without infrastructure using the acoustic background spectrum,” ACM, Proceedings of the 9th international conference on Mobile systems, applications, and services, pp. 155-168, 2011.

[8] J. Wendeberg, F. Höflinger, C. Schindelhauer and L. Reindl, “Anchor-free TDOA self-localization,” in Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1-10, 2011.

[9] R. Lerch, G. Sessler, D. Wolf, Technische Akustik: Grundlagen und Anwendungen. Springer Berlin Heidelberg, 2009.

[10] F. Höflinger, J. Wendeberg, R. Zhang, A. Bannoura, M. Bührer, J. Hoppe, L. M. Reindl, C. Schindelhauer, “Acoustic Self-calibrating System for Indoor Smartphone Tracking,“ in Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2012.

[11] C. J. Plack, The Sense of Hearing. Lawrence Erlbaum Associates, 2005.

[12] M. Sakamoto, M. Sugasawa, K. Kaga and T. Kamio, “Average thresholds in the 8 to 20 kHz range as a function of age,” Scand Audiol, Vol 27, pp. 189-192, 1998.

[13] B. P. Flanagan, “Self Localization Using A Modulated Acoustic Chirp,” Technical report, The MITRE Corporation, 2007.