Simultaneous Acoustic Localization of Multiple Smartphones with Euclidean Distance Matrices

Seyed-Mohsen Moosavi-Dezfooli[1], Yvonne-Anne Pignolet[2], Dacfey Dzung[2]
[1]École Polytechnique Fédérale de Lausanne, Switzerland, [2] ABB Corporate Research, Switzerland
seyed.moosavi@epfl.ch, yvonne-anne.pignolet@ch.abb.com, dacfey.dzung@ch.abb.com

Abstract

In this paper, we present an acoustic localization system for multiple devices. In contrast to systems which localize a device relative to one or several anchor points, we focus on the joint localization of several devices relative to each other. We present a prototype of our system on off-the-shelf smartphones. No user interaction is required, the phones emit acoustic pulses according to a precomputed schedule. Using the elapsed time between two times of arrivals (ETOA) method with sample counting, distances between the devices are estimated. These, possibly incomplete, distances are the input to an efficient and robust multi-dimensional scaling algorithm returning a position for each phone. We evaluated our system in real-world scenarios, achieving error margins of 15 cm in an office environment.

Categories and Subject Descriptors

C.2.1 [**Computer-communication networks**]: Network Architecture and Design—*Wireless Communication*

General Terms

Algorithms, Measurements

Keywords: indoor localization, acoustics, EDM, ETOA

1 Introduction

Smartphones and other mobile devices have become ubiquitous in our lives. As a consequence, many location-dependent applications emerge to support users at work, in shopping malls, airports, and exhibitions. While GPS provides localization outdoors, it is often not useable inside. Thus, a multitude of localization approaches with Wi-Fi and sensors have been devised.

In this paper we address localization using acoustic signals, as every commercial-off-the-shelf (COTS) smartphone is equipped with a microphone and speaker. In particular, we focus on the localization of several devices relative to each other. Such a system can be used for e.g., asset tracking, to ensure safety around (unmanned) vehicles and machines in industrial settings, for augmented reality applications, either alone or complementing other localization systems.

Starting from a simple acoustic ranging application, we propose methods and algorithms to enable the calculation of the position of several phones simultaneously. To this end we (i) design a pulse shaping and detection scheme which has not been used for acoustic ranging to the best of our knowledge and (ii) we propose an algorithm to schedule the actions of recording, emitting a pulse and stopping on the phones and (iii) solve the resulting multi-dimensional scaling problem with Euclidean distance matrices (EDMs).

Using COTS smartphones entails a set of disadvantages and constraints. Both the speaker and microphone systems are optimized for voice, i.e., they are designed for signals in the range of 20 Hz to 22 kHz, higher frequencies are affected by lowpass filters in the audio chain of smartphones [**?**]. Moreover, the API to access the audio system of a phone is limited, so using it for ranging and positioning algorithms is not straight-forward. We describe how we cope with these constraints when building a prototype on Samsung S4 mini phones and we validate our system in real-world scenarios reflecting the conditions of an office environment.

Contributions: We demonstrate in this paper how acoustics can be used to calculate positions of several devices relative to each other without anchor nodes. Compared with other systems, some of which rely on specialized hardware, our system features a low-cost deployment as well as accuracy. We use BeepBeep [**?**] ranging method as a basis, overcoming difficulties due to the multi-phone settings with a different pulse and detection scheme and proposing a scheduling scheme to deal with collisions to build a reliable system.We reduced the abstract problem behind to MDS and designed a novel weighting scheme dealing with possibly large (non-Gaussian) ranging errors. Hence, our system and validation work feature the following.

- Protocol and algorithm for relative positioning of several devices at the same time.
- Robustness. It is not necessary that each device determines the distance to all other devices. Incomplete distance matrices suffice for localization.
- No anchor points or synchronization. Even if clocks are not synchronized and no anchor positions are known, our system can localize devices.

International Conference on Embedded Wireless
Systems and Networks (EWSN) 2016
15–17 February, Graz, Austria

ISBN: 978-0-9949886-0-7

- Evaluation. We implement our system on Android smartphones and evaluate it in office environment scenarios. The mean location error is 5-15 centimeters depending on the environment and configuration, satisfying the requirement of many applications.

2 Background

In this section, we discuss how to use audio technology, i.e. microphones and speakers, in order to measures the distance between two phones. We use this as a building block for our multiple-node localization system.

We apply a method proposed in [?] based on elapsed time between the two time-of-arrivals (ETOA) with sample counting. This method does not require a tight time synchronization which is a very difficult task in the Android environment and resolves the problem of timestamping which is not accurate enough due to the Android OS delay and its impossibility to acquire the exact instance of an event with the desired precision.

2.1 Pulse Shaping and Detection

Any time measuring method needs an accurate pulse detection scheme. The optimal detector w.r.t. Signal-to-Noise Ratio (SNR) is a matched filter [?]. To facilitate detection we need a pulse shape with a narrow autocorrelation function, i.e., high bandwidth. However, the available bandwidth is limited on smartphones because of the frequency response of their microphone and speaker.

A very simple option for a pulse shape is a finite duration sinusoidal signal, although it has a low bandwidth and a low detectability. In the presence of noise and interference, the accuracy of detection with pure sinusoids drops. Another candidate is a chirp signal, a frequency variant sinusoid, used in [?] and [?] for acoustic ranging. Pseudo-random sequences have been widely used in wireless communications contexts [?] due to their narrow autocorrelation. PN sequences are almost white noise but they differ in the distribution. As we will explain in more detail later, we use pseudo-random sequences for our setting because they enable an easier implementation for multi-user detection and have a narrower autocorrelation than the other variants.

2.2 Multi-Dimensional Scaling

Assume that we have several phones and we know the pairwise distances between them. We can collect all distances in a matrix called Euclidean Distance Matrix (EDM). Consider a list of points $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ in the Euclidean space $\mathbb{R}^\eta$ of dimension η. An Euclidean Distance Matrix (EDM) is a matrix $\mathbf{D}$ such that $\mathbf{D}[i,j] = d_{i,j}^2 = \|\mathbf{x}_i - \mathbf{x}_j\|^2$. In other words, each entry of $\mathbf{D}$ is an Euclidean distance-square between pairs of $\mathbf{x}_i$ and $\mathbf{x}_j$. The problem is how to find $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ based on the corresponding EDM. Many methods to solve this problem are proposed in the literature, e.g., the classical approach to solve this problem called classical Multi-Dimensional Scaling (cMDS) [?]. In an error-free setup where the all the pairwise distances are measured without error, cMDS exactly recovers the configuration of the points [?]. This method is simple and efficient. However, in a noisy situation it does not guarantee the optimality of the solution. Furthermore, it can only be used if all distances are known.

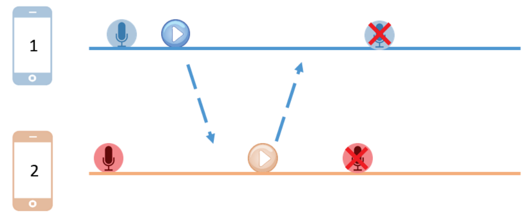


Figure 1. Elapsed time between two times of arrivals (ETOA) method: both phones start recording (indicated by microphone icon), emit a pulse each (play icon), and finally stop recording.

3 Pulse Shape and Detection

In our early experiments, we used a finite duration sinusoid pulse. The frequency of the pulse can be different for each phone to make detection easier. Because we would like to have a non-audible pulse, we choose 18 kHz and 17 kHz for phone 1 and 2 respectively. These frequencies are high enough to be hardly audible and are not too high to be distorted too much because of the frequency responses of microphones and speakers. The length of the pulse is set to 4000 samples at a sampling rate of 48 kHz, thus keeping the duration of the pulse below 0.1 second.

With finite duration sinusoid pulse shaping, we can choose different frequencies to make them more distinguishable. As they have finite duration, they are not completely orthogonal and can only be distinguished if the frequency difference is large enough. Therefore, we have used pseudo-noise as described in Section 4.2 in later experiments.

3.1 Assumptions and Parameter Selection

The sampling rate f_s, recording length and speed of sound ν_s influence the performance of acoustic ranging.

The sampling rate is very important because it determines the maximum frequency and bandwidth that can be used. Because the audio hardware of the smartphones are designed to play in audible frequency range, the highest possible sampling frequency for most of the phones are 48 kHz [?]. Since according to Nyquist's theorem a higher sampling rate implies a higher bandwidth, we thus choose $f_s = 48$ kHz.

The recording length is very important since a short duration can cause phones to miss pulses. On the other hand, longer recordings need more memory and the detection requires more computation. So there is a trade-off between the length of the recording and the chance of missing pulses.

To circumvent this issue, we use a simple communication protocol, illustrated in Figure 1. This protocols works over the existing Wi-Fi network. The phones let each other know via a WiFi connection that they started recording. After the reception of this message phone 1 emits its acoustic pulse. To ensure that phone 1 does not record indefinitely, phone 2 passes a message to phone 1 after it played the pulse. As soon as phone 1 receives it, it stops recording. This way we are sure that both phones have recorded both pulses and no one misses anything. It means that the recording length is not a constant and it varies according to the OS delays and network delays.

Acoustic distance measurement depends on the speed of sound, which is temperature dependent. Some recent smartphones have temperature sensors. Using this they can calculate the speed of sound according to the temperature sensor. As the phones we used, Samsung Galaxy S4 Mini, are not equipped with such a sensor, we assign the speed of sound according to the average room temperature of around $25°C$, i.e., $\nu_s = 340 m/s$.

In brief, we built an Android Phone App for acoustic ranging using ETOA measurement with sample counting and self-recording to calculate the distance between two phones using the above. Figures 1 illustrate the basic mechanisms.

4 Multiple-Node Localization

Above, we discussed how to measure distances between two devices using acoustics. Furthermore we described how Euclidean Distance Matrices (EDMs) can be used to infer positions under ideal conditions in Section 2.2. We now explain how to use these as building blocks to localize several phones simultaneously under noisy conditions.

4.1 Central Distance Collection

We cannot use the application described in Section 2 to measure the pairwise distances between several phones as is. With an increasing number of phones, several issues arise.

Measuring one pairwise distance at a time is unfeasible as the total required time is proportional to $N(N-1)/2$. Thus, we devise a scheme to do all measurements in one interval.

Clearly, the number of calculations to be executed by each phone increases as the number of phones grows. Also there is an extra calculation that we do not have in acoustic ranging, namely solving the MDS problem. Even though this can be done in a distributed way efficiently, it requires that each phone knows the pairwise distances of all nodes, which requires the exchange of $O(N^2)$ messages. Thus we decided to carry out all computations on a server, which incurs a linear message complexity of $O(N)$ and also reduces battery power consumption in the phones.

In addition, the server is not only used for collecting data and doing the calculations, it can also schedule the localization related activities and minimize the probability of missed pulses or of two pulses of two phone colliding.

The main responsibilities of the **server** include to (i) schedule the pulse emitting procedure, (ii) collect recordings from the phones, (iii) calculate distances and MDS, (iv) run algorithms to solve the MDS problem and localize phones.

For the measurements, each **phone** carries out the following steps: (i) start recording when receiving LISTEN(t_1,t_2) from server via a WiFi link, (ii) emit pulse after time t_1, (iii) stop recording after time t_2 has elapsed, (iv) send recording to the server via WiFi link.

Ideally, in each phone's recording we have N different recorded pulses. If we can detect these pulses in each recording, we can find the Round Trip Time (RTT) for each pair. The recorded signal of phone i can be written as $r_i[n] = \sum_j s_j[n - T_{i,j}]$, where $s_j(.)$ is the received signal from phone j and n is the index of the n^{th} sample of a signal. $T_{i,j}$ is the index of the sample which corresponds to the time phone i receives the pulse from phone j. For simplicity and because we are only interested in differences, not in absolute time, we can assume here that all the phones started the recording at the same time. The distance between phone i and j can thus be computed as $d_{i,j} = |(T_{i,j} - T_{i,i}) - (T_{j,j} - T_{j,i})|/(2f_s)\nu_s$,

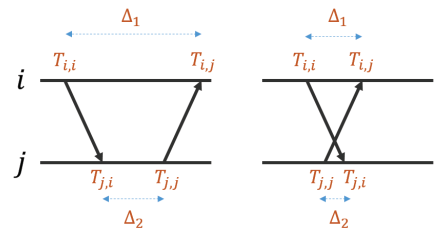


Figure 2. (left) RTT $= \frac{\Delta_1 - \Delta_2}{2}$. (right) RTT $= \frac{\Delta_1 + \Delta_2}{2}$.

where ν_s is the speed of sound and f_s is the sampling frequency. Figure 2 illustrates why this holds, even if the ordering of pulses leads to negative Δ_2.

4.1.1 *Scheduling: Increasing Reliability*

There are several factors that can cause errors in the measurements, falling in one of the two categories. (a) Android OS and networking delay (missed pulse error) and (b) acoustics errors (NLOS components, reverberations, obstruction).

Consider for example the case where a phone starts recording too late because of delays introduced by the Android OS and thus misses the pulses of other phones. Analogously, a phone can stop recording too early and miss pulses. A good schedule can minimize the probability of such errors.

Let $S = \{(t_1^{(1)}, t_2), (t_1^{(2)}, t_2), \ldots, (t_1^{(N)}, t_2)\}$ denote a schedule, where phone i emits its pulse $t_1^{(i)}$ ms after the reception of the message and to stop recording after t_2 ms. The server determines the schedule S and broadcasts it to the N involved phones over the WiFi network. Let $delay_{OS}$ be a bound on the maximum delay cause by the operation system and networking. To avoid errors, the schedule computed by the server should satisfy some conditions.

1. $\forall i: \ t_1^{(i)} > delay_{OS}$ (to avoid late recording errors)
2. $\min_i(t_2 - t_1^{(i)}) > delay_{OS}$ (to avoid early stopping errors)
3. $\forall i,j: \ |t_1^{(i)} - t_1^{(j)}| > \delta$ (to avoid colliding pulses)

We choose $t_j^{(i)}$s for N phones in the following way

$$t_1^{(i)} = D_{delay} + i \cdot D_0. \quad (1)$$

$$t_2 = 2 \cdot D_{delay} + N \cdot D_0. \quad (2)$$

The reason why we separate $t_1^{(i)}$ into two terms is the fact that there are two different types of error. The first type is to miss pulses and the second one is the collision of pulses. To prevent the former, we force the phones to wait for an amount of time, i.e. D_{delay}, before the first one sends a pulse to decrease the probability of missing any pulses because not all phones are in the recording state yet. We determined experimentally that $D_{delay} = 100$ ms is a good choice taking OS and networking delay into account. Collision errors are avoided by an additional amount of delay that varies from phone to phone, i.e. phone i waits iD_0 time before playing its pulse (assuming a pulse duration below D_0). To minimize the collision probability under i.i.d. OS and networking delay, given measurement time t_2, we set $D_0 = (t_2 - 2D_{delay})/N$. Thus, by increasing t_2 the recording phase is extended while the error probability is reduced. However, the probability that the phones have changed their positions in the mean-time increases and higher storage and computation costs are induced.

Another option to increase the probability of success is to repeat the measurements and to combine the results (weighted averages). To this end, we discuss in Section 4.3 how to fuse several (potentially incomplete) EDMs to get a better accuracy result, i.e. optimum weightings. Given a set of assumptions one can thus optimize along the trade-off between the required time for the measurements and the accuracy. However, this is out of the scope of this article. In our evaluation section we show that 5 repetitions provide an error margin of around 15cm in a noisy office environment.

Consequently, instead of using $N(N-1)$ pairwise recordings with up to two pulses each, we use one recording interval at each device with containing up to N pulses. This minimizes time and coordination, enabling evaluation in less than $1s$.

4.2 Pulse Shape and Detection Scheme

The detection scheme described in Section 3 for acoustic ranging does not satisfy all requirements for a multi-phone setting. Since we want to carry out all pairwise measurements together, two important issues arise:

(i) For N phones, we need N easily distinguishable pulse shapes because we would like to find their positions in each recording and compute their corresponding $T_{i,j}$s.

(ii) Each phone receives not only different pulses but also different power levels. This means that a recorded signal contains N different pulses where the received power depends on the distance between the phones. The detection scheme should thus not be sensitive to the power level.

One possible solution is to detect the pulses iteratively. We can detect one pulse at a time and then cancel its contribution from the recorded signal. We repeat this procedure on the canceled recording in the previous step for another pulse recursively. This method would work perfectly if the pulses were not distorted and noise free. In practice, we have still a residual of larger pulse after cancellation.

Pseudo-Random Binary Sequences: In principle any pulse shape with a narrow autocorrelation function can be used in such a localization system. Due to the constraints posed by the built-in microphone and speaker, we select Pseudo-noise (PN) sequences in the frequency range 15-20 kHz and durations of 1000 samples. Though 15 kHz is still audible by humans, it is noticed only as a very short pulse. PN sequences have a large bandwidth with a narrow autocorrelation function. These characteristics depend on the length of the sequence and facilitate detection. The longer the sequence, the better the detectability.

As the phones receive signals from other phones as well as the one emitted by themselves, the signals vary in their power levels. To avoid the problem of different power levels if we use the traditional matched filter approach, we propose a CDMA-like detection scheme that correlates a binary signal to detect pulses. A PN binary pulse shape of length L is defined as $s[n] = b_n$ for $n = 0, 2, \ldots, L-1$, where b_ns are realizations of i.i.d. binary random variables with $P(b_n = 1) = P(b_n = -1) = \frac{1}{2}$. These sequences are suitable as they do not convey any information in their amplitude. Hence, additive noise with reasonable variances can be canceled easily by a sign filter. Thus, we can ignore the amplitude and apply the matched filter detection on a binary sequence.

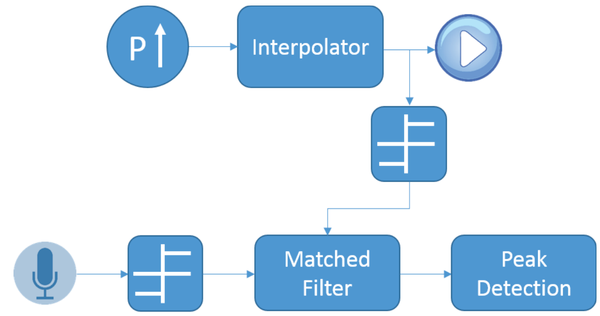


Figure 3. Transmitter (top) and receiver (bottom) of detection scheme

The proposed detection scheme is illustrated in Figure 3. On the transmitter part, we first upsample the generated PN binary sequence by a factor of P. For the inserted zeros by upsamplers, we interpolate the values. The resulting pulse is our new pulse shape. We do the interpolation and upsampling to decrease the required bandwidth and make it low-pass. Therefore, we used $P = 4$ to reduce the bandwidth and be able to modulate the signal to higher frequencies. However, for very high frequencies, greater than 20 kHz, audio components are more affected by distortions caused by the microphone and loud speaker.

On the receiver side, instead of directly applying a matched filter that corresponds to the transmitted pulse shape, we pass it through a sign filter. Then we apply a matched filter that corresponds to the signed version of the pulse shape. The output of the matched filter will be fed into a peak detector in order to find $T_{i,j}$s.

The proposed detection scheme shows a better performance compared to using a matched filter directly. Though it may be surprising at the first glance, this is due to the lack of the optimality condition for the matched filter. The matched filter receiver is the optimum linear filter in the sense of SNR. However, in this case we do not know the distortion by the acoustic propagation channel, therefore a matched filter based on only the pulse shape does not necessarily work better in all circumstances.

4.3 Robust Positioning with incomplete EDMs

We cannot use cMDS mentioned in Section 2.2, as pairwise distance measurements are noisy or might even be missing if two phones are too far from each other to detect each other's pulse. Instead, we use a convex and differentiable cost function called *s-stress* to solve the MDS problem despite incomplete EDMs as proposed by Takane et al. [?]:

$$\min_{\mathbf{X}\in\mathbb{R}^{N\times\eta}} \sum_{i,j} w_{i,j} \left(\mathbf{D}(\mathbf{X})[i,j] - d_{i,j}^2\right)^2 \tag{3}$$

We are interested in cases where the number of spatial dimensions η ($\eta = 2$ or 3) is significantly smaller than the number of nodes N. Parhizkar [?] proposed an algorithm to minimize this function in a (distributed) manner based on the alternating gradient descent optimization method. In each iteration it uses the coordinate descent method by optimizing along one of the variables cyclically. To the best of our knowledge, no other MDS methods combine 1) operation without parameter-tuning, 2) configuration independence, 3) fast convergence, and 4) cope with missing/noisy data.

This algorithm has the advantage that it lets us fuse several sets of measurements easily. Thus, we can increase the accuracy by repeating the experiment several times. When there are only two phones, we can simply take the average over the measured values. Now, suppose we have repeated the measurements for multiple-node localization and obtained several EDMs, one per measurement. The naive approach is to average over each component separately and form a new EDM. We can solve the MDS problem for this new EDM, to minimize the cost function in (3). In other words, this new EDM contains the mean value of the measured distances. We improve over the naive approach by weighting the measurements. By modeling the measurement noise as additive Gaussian noise, one can show that if we choose the weight $w_{i,j}$ inversely proportional to the squared variance of the measurements between node i and j, the error is minimized, i.e.

$$w_{i,j} = \frac{1/\sigma_{ij}^4}{\sum_{i',j'} 1/\sigma_{i',j'}^4}. \tag{4}$$

A lower weight reflects a higher variance which means more uncertainty in measuring the corresponding distance. Since we do not know the exact variances of the measurements, we estimate it by the sample variance.We evaluated both the naive and the optimum weighting strategy in Section 5.

5 Evaluation

Acoustic Ranging: We first compare the performance of the single tone method described in Section 3 to the CDMA-like approach of Section 4.2, depicted in Figure 4. In the single tone approach, we used two sinusoidal pulses at frequencies 18 and 19 kHz for each phone. The accuracy and confidence are much better for the binary PN sequence.

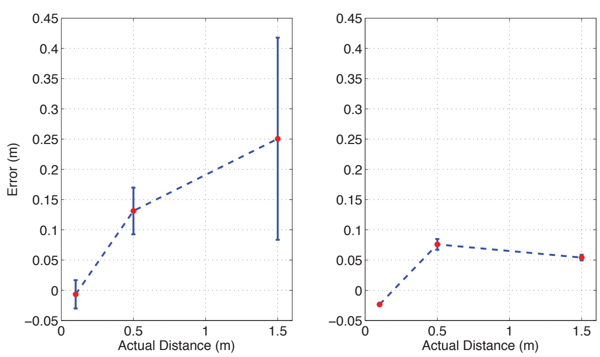


Figure 4. Error vs the actual distance for the finite duration sinusoid (left) and PN sequence (right) schemes: average and standard deviations of 10 measurements with phones at distance 0.1m, 0.5m and 1.5m from each other.

Due to this and the reasons explained in Section 4.2, we used PN sequences in the remainder of our evaluation. To evaluate the accuracy, we applied the proposed ranging method with binary PN sequences on distances up to 6.5m, the confidence intervals for such accuracy is around 4 cm in the worst case

Acoustic Multiple-Node Localization: Here, we show the results of applying the localization scheme explained in the previous section in four different settings. We localized the phones in two configurations, a cross-shaped configuration of 5 phones where all the distances are mid-range and a three by three configuration of 6 phones where the distances are short-range and long-range. We repeated the experiment in two indoor environments: an empty quiet room. and an office environment with several people, computers, desks and other obstacles and noise.

In Figure 5, an example of the result obtained from one set of measurements for each configuration is depicted. As expected, the accuracy in an office is lower because of the obstacles and noise in the environment. It is impossible to keep all influencing factors the same in the two environments. For example the quality of the Wi-Fi network, which has a great effect on the delays with which the phones start recording, varies considerably. The error of the examples in Figure 5 is shown in Table 1 in centimeter. The error is the average deviation from the actual positions, i.e. $e = \frac{1}{N}\sum_{i=1}^{N} \|\hat{\mathbf{x}}_i - \mathbf{x}_i\|$, where $\hat{\mathbf{x}}_i$ and $\mathbf{x}_i$ are the estimated and actual positions respectively and rotation and translation have been applied for error minimization. The second setup, consisting of 6 phones, shows a better overall performance especially in the office. This might be due to the fact that the number of phones has a great impact on the performance of the MDS algorithm in [**?**]. However, more experiments are needed to verify this hypothesis. The overall accuracy is on the decimeter level.

Table 1. Error comparison (single measurement set).

	Cross-shaped	Three-by-Three
Empty room	1.30 cm	5.2 cm
Office	34.8 cm	8.2 cm

For multiple-node localization, the effect of repeating measurements depends on the weighting strategy. The average distance error is shown in Table 2. Compared to the results using one set of measurements we observe that error can be reduced up to 50%. The optimal weighting scheme outperforms the equal weighting scheme by up to 30%.

Table 2. Error comparison for different weightings in noisy office (5 sets of measurements).

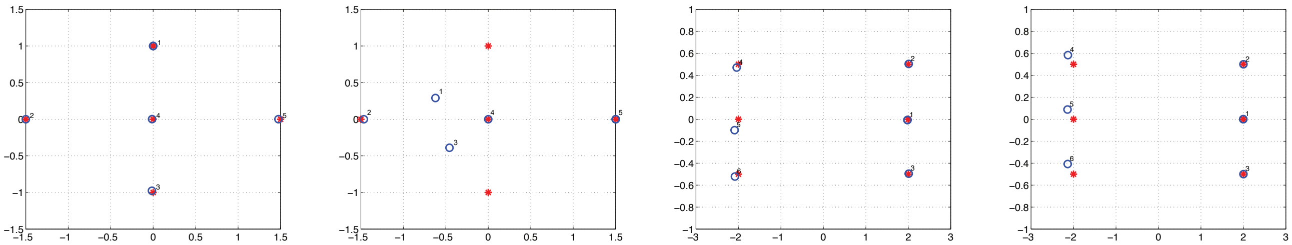
	Equal Weighting	Optimum Weighting
Cross-shaped	19.89 cm	13.6 cm
Three-by-Three	11.4 cm	9.7 cm

In summary, our localization scheme behaves as expected and is able to provide error bounds of around 30cm in noisy environments using one set of measurement or around 15cm when combining several measurements.

6 Related Work

Many different indoor localization systems are available today, based on pedestrian dead-reckoning, Wi-Fi or other radio signals, cameras, etc. or a combination thereof. We refer to [**?**] for an overview, since we focus here on acoustics.

The fact that many devices can generate sound from their built-in speaker and detect sound with the integrated micro-



(a) (b) (c) (d)

Figure 5. Examples with one measurement set. Red asterisks correspond to actual, blue circles to estimated locations. (a) and (c) 5 resp. 6 phones in a quiet empty room, (b) and (d) 5 resp. phones in an office environment (in meter).

phone has been used in a number of different approaches. There are distance-free localization methods which use audio devices to capture acoustic impulse response as an input to a pattern classification algorithm, e.g. [**?**]. Here, we focus on the distance-related methods.

While there is a multitude of acoustic ranging methods [**?**, **?**, **?**, **?**, **?**, **?**], using different pulse shapes and calculation methods, positioning has not received the same amount of attention. From a system design perspective, it is highly valuable to know how to schedule distance measurements between several nodes with unknown positions and how to process the results to derive positions. To our knowledge, the current literature does not address these issues.

Marziani et al. [**?**] use an RTT and CDMA-based method, to find pairwise distances. However, they do not determine the actual positions of the nodes. Chakraborty et al. [**?**] presented a TDOA-based localization scheme while Whistle [**?**], applies a ranging method similar to BeepBeep with 10-20 cm accuracy for localizing a sound source with TDOA The problem of finding the position of multiple nodes simultaneously using acoustics is not discussed in any of the work we are aware of.

7 Conclusions

In this article we proposed a localization system to position several phones simultaneously. Our system uses ETOA measurements with sample counting to compute distances between phones. We used the s-stress cost function to formulate the problem of finding the positions from the distances as an optimization problem to which we applied an alternating gradient descent algorithm. Furthermore, we described a pseudo-noise-based pulse shaping and detection scheme and a method to schedule multi-node measurements reliably despite OS and networking delays, which has not been addressed in other work to the best of our knowledge. In addition to reliability, accuracy is an important performance measure of a localization system. To improve the accuracy, we take measurements several times and combine them using an optimal weighting strategy.

A more detailed version of this paper is available at arXiv.

8 Acknowledgments

The authors would like to thank Reza Parhizkar and Martin Vetterli for their feedback and advice on this work.

9 References

[1] T. Akiyama, M. Nakamura, M. Sugimoto, and H. Hashizume. Smart phone localization method using dual-carrier acoustic waves. In *Indoor Positioning and Indoor Navigation (IPIN)*, 2013.

[2] G. Borriello, A. Liu, T. Offer, C. Palistrant, and R. Sharp. Walrus: Wireless acoustic location with room-level resolution using ultrasound. In *Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2005.

[3] J. Chakraborty, G. Ottoy, M. Gelaude, J.-P. Goemaere, and L. De Strycker. In *Computer and Information Technology (ICCIT)*.

[4] C. De Marziani, J. Urena, A. Hernandez, J. Garcia, F. Alvarez, A. Jimenez, M. Perez, J. Carrizo, J. Aparicio, and R. Alcoleas. Simultaneous round-trip time-of-flight measurements with encoded acoustic signals. *Sensors Journal, IEEE*, 12(10):2931–2940, 2012.

[5] V. Filonenko, C. Cullen, and J. D. Carswell. Indoor positioning for smartphones using asynchronous ultrasound trilateration. *ISPRS International Journal of Geo-Information*, 2(3):598–620, 2013.

[6] Google. Android documentation. `http://developer.android.com/reference/android/media/package-summary.html`, 2014.

[7] S. Haykin. *Digital Communication Systems*. Wiley, 2013.

[8] F. Hoflinger, R. Zhang, J. Hoppe, A. Bannoura, L. Reindl, J. Wendeberg, M. Buhrer, and C. Schindelhauer. In *Indoor Positioning and Indoor Navigation (IPIN)*.

[9] J. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.

[10] K. Liu, X. Liu, L. Xie, and X. Li. Towards accurate acoustic localization on a smartphone. In *INFOCOM*, 2013.

[11] A. Mandal, C. Lopes, T. Givargis, A. Haghighat, R. Jurdak, and P. Baldi. Beep: 3d indoor positioning using audible sound. In *Consumer Communications and Networking Conference*, 2005.

[12] R. Mautz. Indoor positioning technologies. *ETH Zurich, Department of Civil, Environmental and Geomatic Engineering, Institute of Geodesy and Photogrammetry*, 2012.

[13] R. Parhizkar. *Euclidean distance matrices: Properties, algorithms and applications*. PhD thesis, EPFL, 2013.

[14] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan. Beepbeep: A high accuracy acoustic ranging system using cots mobile devices. In *Conference on Embedded Networked Sensor Systems (Sensys)*, 2007.

[15] M. Rossi, J. Seiter, O. Amft, S. Buchmeier, and G. Tröster. Roomsense: An indoor positioning system for smartphones using active sound probing. In *Proceedings of the 4th Augmented Human International Conference*, Mar. 2013.

[16] Y. Shang, W. Ruml, Y. Zhang, and M. P. J. Fromherz. Localization from mere connectivity. In *MobiHoc*, 2003.

[17] Y. Takane, F. Young, and J. Leeuw. Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, 42(1):7–67, March 1977.

[18] B. Vucetic and S. Glisic. *Spread Spectrum CDMA Systems for Wireless Communications*. 1997.

[19] R. Yu, B. Xu, G. Sun, and Z. Yang. Whistle: synchronization-free tdoa for localization. In J. Beutel, D. Ganesan, and J. A. Stankovic, editors, *SenSys*, pages 359–360. ACM, 2010.