An Introduction To Acoustic-based Localization Techniques Using Off-The-Shelf Hardware

Matthias Linder, *University of Duisburg-Essen*
matthias.linder@uni-due.de

***Abstract*—In this paper we present the basics of acoustics, different kinds of localizations, and then explain and compare a set of acoustic-based localization techniques that make use of off-the-shelf hardware like mobile phones and personal computers. The basics of each technique are visually explained and advantages and disadvantages of each approach are shown.**

***Index Terms*—Acoustic Localization, Sound Context Recognition, Localization Techniques**

I. Introduction

KNOWING the location of a mobile device in an indoor environment can be very useful – Examples for this are indoor maps, home-automation, or e.g. observing the position of patients in medical treatment – the list of possible applications seems almost unlimited; Context-awareness is becoming more important every day [1].

While outdoor localization techniques like GPS are already widely in use, there currently is a lack of accurate and robust indoor localization techniques. There have been proposals to use RF-based localization [2] for indoor environments. While these can provide accurate results, they also suffer form a high setup cost, as the entire localization area has to have reliable WiFi-reception from multiple access points. Studies have shown that this is not the case in all countries [3]. Because of this limitation there is a concrete need for a low cost alternative that provides approximated localizations while focusing on a low setup cost: Acoustic-based localization.

Why is context-awareness such an important factor, and what are the requirements to localization techniques? To answer this, one should first have possible use-cases in mind. These are some example use-cases for localizations that come to mind:

1) Shopping Enhancement: One can imagine an application running on your smart-phone that would display offers on the fly as you are getting close to a specific store. Using localization, these offers can be fitted to the concrete, relevant environment.

2) Medical Patients: Imagine a person that is ill, and needs medication on an hourly basis. And imagine that person using his or her smartphone to remind him or here of the need for medication. It makes sense to ensure that this person carries around the smartphone at all time. Localization techniques could be used to determine whether the patient is carrying the phone in his/her pocket.

3) Indoor Maps: Did you ever get lost in a huge mall? Since outdoor localization techniques like GPS do not work well within indoor-environment, it makes sense to search for alternatives here.

4) Door Badges: Doors could automatically detect whether authorized persons are in its direct environment, and open on-demand. Instead of carrying around extra key fobs or badges, the users smartphone could accomplish this task.

5) Printing documents: In a huge company buildings with a lot of printers, you might not know where the next printer in the area is. Localization can help you find it, and could automatically cause the printer to print the document for you when you are near – without looking up its name or place before-hand.

All these examples are merely scratching the surface of what is possible, but they give a good insight into what is expected from the localization techniques used. Some of these examples can live with a high margin of errors; for others this could be fatal. Because of these different requirements, distinct localization techniques each serve to a different subset of these goals. [3]

In this paper we will explain the relevant basics of acoustics (Section III), then dive into the area of context localization (Section IV) by explaining the different kinds of localizations. The main part of this paper will then focus on explaining four different localization techniques: (1) Distance-based Localization (Section V), (2) Localization via Trilateration (Section VI), (3) Background Spectrum Localization using Matching Pursuit (Section VII), and (4) Material Spectrum Localization (Section VIII). For each technique we will first describe the technique, and later on show advantages and potential limitations.

II. Off-the-shelf Hardware

Seen from a theoretical perspective, it would most likely be easy to develop a new indoor positioning system that uses a 2D array of sensors (e.g. attached to a wall) to detect GSM beacons from smartphones, and uses them to localize the user within a room. And given enough sensors, it would also be rather easy to get this system accurate and noise-resistant enough for its intended application. So why is this not the solution to indoor localization? What is the down-side?

Deploying such an system in one room for an unique kind of application might be reasonable, but deploying it to every room in the whole world for a generic set of applications? The setup cost for such a system would be immense, and just not feasible. Although indoor localization is something very much preferred to be accomplished, it is not something that is substantial to our presence as human beings; nothing substantial to the survival of human kind; nothing that would solve the problem

of world hunger. This is why such an technique - if desired - has to have a reasonable setup cost in order to be successful.

So why shouldn't one use what already exists? Most persons today are running around with a smartphone in their hands – a smartphone very well capable of recording and playing sounds. In addition to this many publicly accessible rooms have a computer hidden away somewhere. If one can combine the existing resources, and create an accurate and reliable indoor localization system that can also be set up at a reasonable cost, then this system has the potential of getting widely deployed around the globe in a matter of months.

For this reason it makes sense to focus on existing, or cheap off-the-shelf hardware for these techniques (cf. [4]). Acoustic localization fits well into this perspective as the hardware required for localization is already distributed in most places, and processing is simpler when compared to other techniques like e.g. RF localization.

[1], [3], [5]

A. Smartphones

In this paper, we consider "off-the-shelf hardware" to especially also include smartphones. This makes sense as smartphones have become an integral part of our lives, and most people using application in need of context-awareness will therefore also carry around a smartphone. As they do not put any additional load on the setup cost, they are therefore considered to be part of the existing infrastructure.

Also, since modern smartphones contain several different kinds of sensors (e.g. accelerometers, microphones, GPS, multiple cameras, ...), they present a great target for deploying enhanced localization applications without the need of any additional hardware.

III. Basics of Acoustics

Before diving into the different sound-based localization techniques, it makes sense to talk about the basics of acoustic waves first.

So what is an acoustic wave? An acoustic wave is a pulsating wave traveling through a medium (e.g. air) causing periodical increases and decreases in pressure. An important variable in this context is the *speed of sound*, which describes the distance sound waves can travel in a medium in a specific time.

$$c_{air} = 343.2m \cdot s^{-1}$$

Fig. 1. Speed of sound in norm-atmospheric conditions at sea-level

It is noteworthy that this speed is very much lower than the speed of electromagnetic waves, which is close to the speed of light (roughly 299 792 $km \cdot s^{-1}$ in air). As this variable will be used by some of the localization techniques described in the next sections, it is also important to know that the speed of sound greatly depends on atmospheric conditions like temperature and humidity, and therefore can change greatly over the course of a day [6], [7].

	min f [Hz]	max f [Hz]
audible for humans	20 Hz	20 000 Hz
human speech	1 024 Hz	8 192 Hz

Fig. 2. Common audio frequencies [7]

Fig. 3. A simple sinusoidal wave resulting in a sound with a fixed-frequency.

A. Describing sound waves

Whereas simple sinusoidal waves can be described by the two parameters *frequency* (number of repetitions in a fixed time, e.g. 5 Hz means 5 repetitions of the same cycle in one second) and *amplitude* (range from maximum to minimum value), most sound waves are of more complex nature: Several distinct, simpler waves are superpositioned on each other at different time ranges, and thus can create waves similar to the one seen in figure 4.

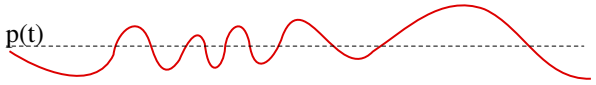

Fig. 4. A random sound resulting from several superpositioned waves. Sound can described as relative pressure over time.

When looking at the digital world, there are several different techniques available for describing these compound waves:

1) ADC: The first technique is called "analog-to-digital conversion". In this technique the sound wave is sampled at different moments in time at a fixed frequency, and the current value (meaning the current/relative pressure present) is converted into digital form by comparing the existing voltage at the recording device with different reference voltages. Using this principle it is possible to create a stair-like function for the acoustic wave. An important restricting factor here is the rate by which the sound is sampled: The *Nyquist-Shannon sampling theorem* dictates that in order to correctly analyze a sound wave with the frequency f, the sampling device needs to sample this wave with more than double that frequency $(2 \cdot f)$.

As we are talking about off-the-shelf hardware in this paper, and the usual sound controller in a mobile phone or PC has a maximum sample rate of 44.1 KHz, this limits the spectrum that we can use for localization to the band from 0 Hz to 22.05 KHz [1], [3].

The obvious advantage of this technique is that – given a high enough sample rate – the signal can be perfectly reconstructed. On the other hand one has to consider that the amount of data required for storing this data increases linearly with time – even for very simple sinusoidal sounds.

2) FFT: Another analytical technique which takes up less space at the cost of the time-component is the *Fast-Fourier*

Transformation. The basic principle behind this technique is that each sound wave can be seen as a collection of super-positioned sinusoidal sound waves of different frequencies, and thus every repetitive sound can be represented as a weighted-combination of those waves [7]. This obviously only works lossless as long as the sound wave only consists out of a repetitive pattern, as simple sinusoidal waves cannot encompass the nature of a sound changing to a different sound over time. On the other hand this technique allows one to directly see the energy of every frequency directly, which is useful when classifying different sounds.

B. RF versus Acoustics

Since there are known implementations for *WiFi/RF-based* localization algorithms (e.g. RADAR [2]), it makes sense to discuss the differences between *RF* and *acoustic* waves:

1) Wave type: Radio waves are high-frequency (3 KHz - 300 GHz) electromagnetic waves which do not require a medium for transportation, and thus they also work in a vacuum/space. Acoustic waves on the other hand are directly depended on the properties of medium they are travelling through [6].

2) Obstacles: Since radio waves do not rely on mechanical vibrations, they can easily pass through most materials. Acoustic waves are usually limited to the room where they originate. This is even hardened by the fact that most buildings are constructed in a way that should prevent acoustic cluttering, and thus the sound-absorption of these walls is strengthened by design, whereas the reverse is true for RF waves. As acoustic waves are limited to a room, they can be more reliable for in-room-based localization than RF localization [8]–[10].

3) Measurability: RF operates at very high frequencies. This means that its pure waveform cannot be recorded using off-the-shelf hardware, but instead only some commonly available factors like *Signal Strength* or the data contained within the RF signal can be used for localization. Acoustic waves, on the other hand, operate at much lower frequencies, and their distribution speed is lower. Taking this into account, standard off-the-shelf hardware can be used to analyze acoustic waves, and additional data like *time-of-flight* and *frequency distortions* can be used as features for localization [10]–[12].

4) Annoyance: It would be very annoying if all sound-based localization techniques would require the room to be flooded with a mass of sounds. While this is a good option for developing and testing a localization technique in the first place, this would be a show-stopper for all sound-based localization techniques when deployed into real use-cases. To solve this issue, there are two simple solutions: (a) only use passive techniques which don't actively send out sounds, or (b) use a frequency spectrum which cannot be heard by humans, but which can still be detected by off-the-shelf hardware. Studies have shown that the major amount of casually available sound-hardware is able to emit sounds at a frequency base around 21 KHz, which is in the area of inaudible sound waves [10]. Also, there have been experiments to well-sounding sounds and melodies for localization [10]. This could e.g. be used in a mall where music is an accepted "background noise".

IV. Localization basics

Since the basics of sound propagation are now clear, the remaining part of this paper will deal with the different sound-based localization techniques. The *process of localization* deals with detecting an approximative physical or logical location for an object by making use of features available in the environment.

A. Types of Localizations

Localization is not just localization – One can distinguish between different types of localizations; each with its own purpose. Different localization techniques give different results. While e.g. a position in the $(longitude, latitude)$ form seems like a reasonable approach first, absolute coordinates are not necessary in most cases, and can be disadvantageous in certain situations.

localization type	**localization**
absolute[1]	$(51.42716\ N, 6.800703\ O)$
relative	$(5, -3)$
discrete symbolical	noisy, large hall
concrete symbolical	LB103, Duisburg

Fig. 5. Examples for different localizations for the same object

1) Coordinate-based localization: As previously mentioned, the most obvious approach of describing the location of an object is some form of coordinate – be it a simple two-dimensional (x, y) tuple, a three-dimensional (x, y, z) position relative to some point of origin, or a latitude-longitude tuple describing the objects position on our planet earth using a well known standard.

The advantages of such an approach are obvious: It is very easy to determine the distance between two objects, and one can also easily transform the coordinates into other coordinate-based system by a simple transformation matrix. Also, as most existing localization techniques (GPS, GSM-based localization, ...) use similar notions, existing techniques can be used. Thus it makes sense to stick with this system further.

However, from this numeric approach there are also clear disadvantages: Imagine two rooms right next to each other, and a person standing exactly at the border between those two rooms. Since localization is not exact, but prone to errors instead, the person might end up being detected in the room right next-door (cf. figure 6) instead of where they are really standing. All techniques relying on coordinates are therefore prone to classification errors resulting from environments changing rapidly over small amount of space (e.g. a wall separating two spaces).

For applications relying on concrete classifications, this error can be severe. Just imagine a phone which is supposed to automatically mute itself in meeting rooms. If too close to a wall, the phone would constantly (and randomly) switch between the muted and unmuted state. In the worst case that would cause the application to permit what it was actually supposed to prevent from doing.

[1], [11]

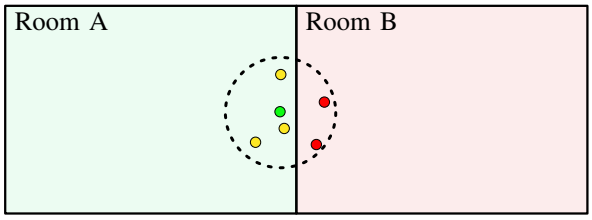

Fig. 6. Error margin for coordinate-based localization (green: real location, dashed: error margin, yellow: localizations without side effects, red: wrong classifications)

2) Logical/symbolical localization: Another approach to localization are logical, or symbolical locations. Instead of determining the objects physical location in the world, one can also determine the type of the environment the object is in. These can be abstract classifications like e.g. "Object A is on a table", "in a meeting room", "in a pocket", which can be determined by using analyzing the sound spectrum (cf. section VII), or concrete classifications like "Object B is in room LB103", or "B is in the Starbucks on 42nd Street crossing 9th Ave".

Whether one uses an concrete or abstract classification depends on the kind of information available: Abstract classifications can mostly be determined by using pre-defined rules/example sets for those classifications; concrete classifications require some person to set up the system for the environment before first using the system, which can result in a significant work load.

As hinted above, the big advantage of abstract symbolical localizations is that they can be established with a lower setup cost for the third party using the system. For most obvious context-aware use-cases (see section I) a symbolic location is enough, and thus one does not need to go through the effort of calculating a exact coordinate-based position. Since most symbolic-localization techniques are more focused on the acoustic features of an environment instead of positional features, the risk of the previously mentioned "room-hopping" is lower here.

[9], [12], [13]

B. Goals

To be able to compare different localization techniques, we first have to establish different criteria for evaluation. Although we focus on different use-cases (see section I), we can establish some basic properties which are – in general – preferable if present or absent. Thus we can distinguish between these major goals:

1) Accuracy: Obviously we do not want the technique in question to just spill out random results – they should have some accuracy with where the object really is. We are trying to get the highest degree of accuracy while keeping the setup cost as low as possible, and while remaining impervious to noise.

2) Noise-resistance: Unlike usual test scenarios where all environment conditions are static, there are a lot of things that can change in a real-world scenario. The weather might change, and thus cause a drop in humidity, which might affect the distribution of sound waves. Or there might be a random noise source (lets call it "$troll$") in the scenario, which emits random sounds and thus might try to alter/disturb our localization. The technique should be resistant against those noises occurring in the scenarios for which the localization technique is intended. We can distinguish between static noise (noise always present in an environment), random noise (e.g. people chatting) and noise caused by slow changes in the environment (e.g. weather or air-conditioning systems) [3].

3) Setup cost: Since we want to use off-the-shelf hardware, any extra setup cost – be it through extra hardware, or through extra man-hours – should be kept to an minimum so that the techniques can be used by a wide range of people without high additional effort.

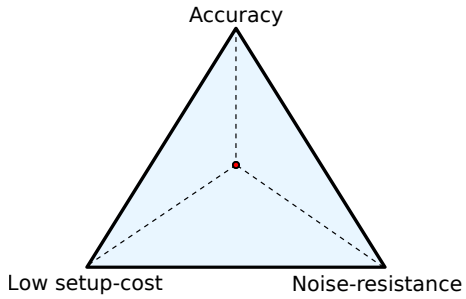

Fig. 7. Each criteria inversely effects the other two criteria

It is obvious that all these criteria influence each other. The more hardware you throw at something, the higher accuracy you can achieve; an higher demand in accuracy results in a smaller margin in error, and thus a higher demanded noise-resistance, and a higher noise-resistance can again be achieved by (again) "throwing more hardware" at a problem, which increases the setup cost.

C. Testing architecture

When designing techniques for localization, testing these algorithm is an important aspect. It is also one that can easily be done wrong. Testing the algorithm with only a limited set of data is bound to produce unrealistic evaluations. Think going around a campus and collecting different sound samples at different locations is enough? It is certainly necessary, but when e.g. taking into account the differences between day and night, or weekday and weekend at an university campus, it becomes obvious that there is a lot more to consider. The set of samples has to be diverse enough to encompass differences in location, time, hardware and weather, as all of these can influence noise-levels and important factors like speed of sound.

Once a set of enough data has been collected, the algorithm has to be tested against this set of data. One possible way of doing so is the *leave-one-out classification* [3]: The set of available samples is divided into a set of known information which is used for classification/localization, and an unknown test set given to the system to be tested. As these two sets can be arbitrarily chosen out of the set of data available, a large number of different test cases can be constructed from

a relatively small data set, without requiring any physical work like e.g. running around different rooms collecting test values. The same data can also be used to compare different localization techniques against each other so that one can find the optimal localization technique for a certain scenario.

As the basis for localization have now been satisfiably conveyed, we can now focus on the different localization technique implementations.

V. Distance-based Localization

The simplest form of position-based localization is to determine the distance between two objects. A system capable of doing so could potentially determine whether two objects are in the same room (e.g. by checking the condition "sound can be heard and target is closer than 20 meters") by using a single sound source and a single receiver [8].

In this section the Robust Range Estimation technique by Lewis Girod et al. will be presented as an example, which is based upon a localization-sound which is being emitted from the speaker (e.g. a smartphone), and is received by fixed-position microphone (e.g. a computer workstation). Speaker and microphone are both inter-exchangeable since direction is of no importance.

A. Making yourself heard

As sound is something which present in nearly all situations, one cannot assume that every sound wave that gets recorded by the receiver is actually from the sound source which is to be localized. In order to distinguish the localization-sound from random environment noise, the sound has to be unique and distinct to its environment. The easiest way to achieve this is to create a frequency-function/pseudo-noise sequence which is played by the speaker, and later on matched against by the receiver by using a matching/correlation function $f(received, expected)$ which peaks when the received and expected sounds are most similar [8]. This, however, puts a small limitation on this technique: The receiver always has to know the kind/characteristics of the sound being played by the emitter.

B. Time of flight

In order to determine the distance between sender and receiver, the *time of flight* of the sound waves is considered. If a sound is emitted at t_0, and received at the microphone at t_r, the distance between both can be calculated with:

$$d = \frac{(t_r - t_0)}{speed\ of\ sound} \tag{1}$$

As the speed of sound is dependent on atmospheric conditions, and changes up to ten percent during the same day, assuming it to be a constant would cause major inaccuracies. To help against this the speed of sound can be calibrated by calculation d along a fixed, known distance, e.g. the distance between speaker and microphone in a smartphone [8].

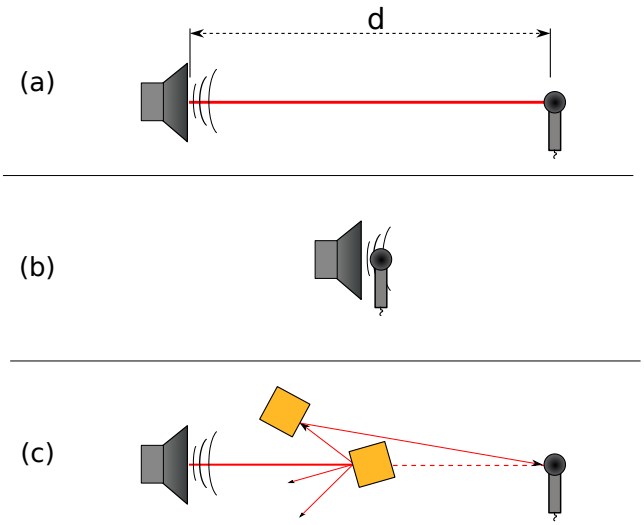

Fig. 8. *
(a) Distance-estimation by using the *time of flight*
(b) Calibration of the *speed of sound* using a known distance
(c) Obstacles in the path cause a longer signal-way, and thus causes over-estimation

If the path between sender and receiver is blocked (see figure 8c), and there therefore is no line-of-sight (short: LOS) between both objects, the distance estimated by the algorithm will cause an over-estimation: The detected distance will be higher than the actual distance. While is this one potential error source, it is noteworthy that this technique can never under-estimate any distance since this would require a sound-wave to warp in space. Thus this error is bounded to one direction, and classification algorithms could consider this in their calculations [8].

C. Synchronization

In order to determine the times t_0 and t_r correctly, sender and receiver both have to be synchronized. One easy way of to achieve this is to use an RF link for transferring the *begin localization* message, as electro-magnetic waves travel at speeds close to light as they are not bound to a specific medium [7].

D. Hardware-delay

Most off-the-shelf hardware has one issue when it comes to calculating the time of flight: *Hardware and software delays*. Most sound-cards introduce a fixed delay of up to $50ms$ while sampling [8], and – since most computers and smartphones run non-real-time kernels - additional software delays are introduced by kernel schedulers and sound drivers.

Several solutions are possible to help against this delay: (a) custom sound card drivers can minimize the error introduced by schedulers since assigning a timestamp to a sample can occur at a much earlier stage. (b) the calibration method described in figure 8c can also be used to determine the fixed delay if the speed of sound is known. and (c) averaging multiple localizations can help against dynamic delays caused by schedulers.

E. Echo

As indoor rooms are prone to echo, the same sound waves might be received multiple times, causing the matcher function to output multiple matches. If an echo from a localization-sound is used, it will cause even more over-estimation. To help against this the strongest and earliest peak in the matching/correlation function has to be used, and different localization attempts should use distinct probing sounds.

F. Evaluation

This simple, yet effective algorithm allows for simple distance/presence based localization, and can achieve accuracies in the *cm* level [8] if calibrated to a specific set of hardware. In the worst case (blocked sight, random delays) this algorithm will cause slight over-estimation, which has to be considered by the application using the information.

VI. Localization via Trilateration

Since knowing only the distance between two objects is not always enough, it is desirable to know the real two-dimensional (x, y) or three-dimensional (x, y, z) position of an object in the room. In order to achieve this using the time of flight technique, *trilateration* can be used.

The technique presented here is a result of the works of James Scott et al. in their paper "Audio Location: Accurate Low-Cost Location Sensing" [11].

A. Trilateration

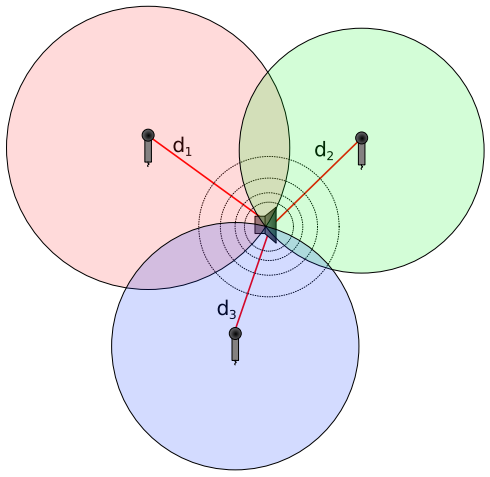

Fig. 9. To localize an object with a 2D/3D-position at least three receivers are required

The setup for this technique consists out of at least three receivers, which are – in optimum – placed around the sound source while having an equidistance between each receiver [11]. It is assumed that each receiver has a fixed location which is known to the system. These receivers could e.g. be the personal workstations of employees in a company building. Similar synchronization and matching techniques as described in section V are used to calculate the time of flight t_i between the sender and each receiver r_i.

The distance between sender and receiver i shall be d_i:

$$d = \frac{(t_i - t_0)}{speed\ of\ sound_i} \tag{2}$$

As the distance is not bound to a specific direction, one can imagine it being a circle around every receiver. The localization/intersection point (p_x, p_y) we are looking for can then be found at the intersection point of all three circles (see figure 9). It can be calculated with a linear system of equations of circle-equations:

$$d_1^2 = (p_x - x_1)^2 + (p_y - y_1)^2 \tag{3a}$$
$$d_2^2 = (p_x - x_2)^2 + (p_y - y_2)^2 \tag{3b}$$
$$d_3^2 = (p_x - x_3)^2 + (p_y - y_3)^2 \tag{3c}$$

As the quadratic equations will yield two solutions, the two unknowns p_x and p_y need to be solved with at least three equations to correctly determine the point of origin. At this point it also becomes obvious that any more receivers would result in an apparent overfull equation system. But this is only the case because the system of equations above represent an optimistic version without any margin for errors – in reality one would not consider the border of each circle to be a line, but to be a ring instead. More receivers can be used to detect and reduce this error.

[7], [11]

B. Advantages

The advantages of this technique over *Distance-based Localization* are apparent:

1) 3D Positions: This technique does not limit localization to a mere distance between sender and receiver, but allows for relative and absolute positioning within a room.

2) Multiple receivers: When using multilateration (more than three receivers), the errors introduced by each *time of flight* measurement, and the errors caused by a blocked line of sight or similar can be lessened, and so the accuracy and robustness of this system scales with the numbers of receivers. This basically means that – given enough hardware – one can get the system as robust as one likes.

3) Less line-of-sight issues: With redundant receivers blocked line-of-sights become less of an issue, as an over-estimation spike can be detected and systematically ignored.

C. Disadvantages

There are, however, also some disadvantages to this technique:

1) Distribution of receivers: Receivers have to be evenly distributed across the localization area, as otherwise the margin of errors will increase as nodes right next to each other are subject to the same erroneous measurements. This is especially an issue when considering the z-axis and looking at regular office spaces: As all workstations are usually located on the floor, precision on this axis suffers.

2) Cost of setup: Whereas the previous technique could go with just one receiver, we require at least three to be able to estimate an position using this technique. This does not only require more hardware, but also a significant contribution to the setup (e.g. distance calibration) of each receiver.

3) Two-Room-Problem: Small errors in localization might result in huge mistakes in classification, as described in section IV-A1.

VII. Background Spectrum Localization

All the previous techniques have been using the principle of time of flight, and thus are prone to errors resulting from wrong time measurement. There are, however, also techniques that do not rely on the time a signal takes, but instead use *features* of the sound wave itself to create a logical localization.

This section focuses on the works of Martin Azizyan et al. in the paper "SurroundSense" [9], and on the MP-based feature detection as described by Selina Chu et al. [13].

A. Background Spectrum

SurroundSense assumes that each room has a different ambient fingerprint. Although this technique does not merely focus on sound, but instead also includes visual and other feedback, the sound part is still of importance.

The goal is to create an unique *fingerprint* (or, in other words, a hash-value) uniquely describing the current room/environment, which is then compared with existing fingerprints in a fingerprint database. The closest matches with existing fingerprints, which can e.g. be determined by a *k-nearest neighbor algorithm* [7], are then combined into a localization result.

B. Fingerprinting

To create a fingerprint, the recording device first captures the environment sound over a fixed period of time. There are many different ways of classifying the sound. We will focus on a technique called "Matching Pursuit" (short: MP) [13].

The principle of MP is that given a sound s and a dictionary of existing sound-pieces D, one can calculate a weighted sum that completely describes the sound s (eqn. 4).

$$s = \sum_{i=1}^{|D|} w_i \cdot d_i, \quad d_i \in D, \ (D \subset \text{Sounds}, s \in \text{Sounds}) \tag{4}$$

$$w = (w_1, w_2, \ldots, w_{|D|}) \in \mathbb{R}^{|D|} \tag{5}$$

An example: The environment sound (starbucks) and the dictionary d would produce the weight-vector:

$$d := (\text{coffeemaker}, \text{people}, \text{hamster}, \text{door}) \tag{6}$$

$$weights(\text{starbucks}) = (0.5, 0.3, 0.0, 0.2) \tag{7}$$

Although this already produces a fingerprint of some sort, is it not guaranteed that two similar sounds will produce the same fingerprint. If the dictionary also contains atomic-samples like e.g. a $1KHz$-sound, ..., one could either use the fully assembled $coffeemaker$ sound, or the atomic parts. In order to prevent this, the MP-algorithm does not look for an arbitrary weight vector, but the minimal one.

Since detecting the minimal weighted set is an *NP-complete* problem [13], the MP-algorithm instead uses an approximated sub-optimal minimal set:

```
matchingPursuit(sample, dictionary) {
 int[] weights = new[len(dictionary)]
 sound = sample
 do {
  weight, strongest =
       (FROM dictionary d
        SELECT (weight(sound, d), d)
        ORDER BY weight(sound, d) DESC)[0]
  weights[index(strongest)] = weight
  sound -= strongest * weight

 } while (strength(sound) > threshold)
 return
}
```

The algorithm always picks out the strongest-weighted fit, and then subtracts this fit from the sound. This is repeated until the sound does not contain anything else but irrelevant noise. An graphic example is provided in figure 10.

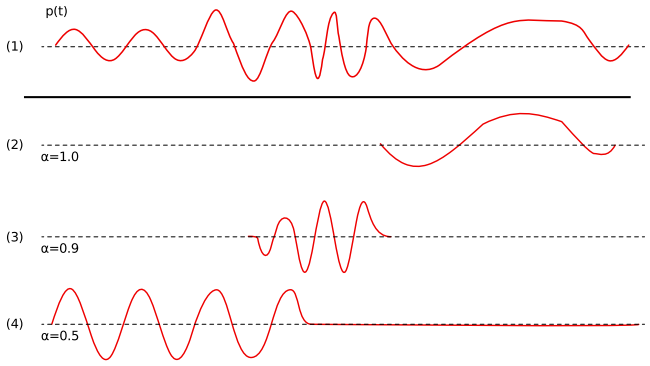

Fig. 10. *
(1) The environment sound, combined out of several distinct sections
(2) A 100% match which is found first when MP
(3) Partial match which is found in the remaining sound bits
(4) Sinusoidal component that makes up the remaining sound, and is weightened less strongly as it would otherwise have a stronger amplitude than the original sound

It is obvious that the dictionary should consist out of distinct sounds, as otherwise errors in detection may occur. To do so, one can e.g. use the *Gaborfunction* which provides distinct atoms [7], [13].

As the resulting fingerprint is close to being minimal, comparison is simpler as similar sounds will end up with the same

classification, and confusion through too many classification features is kept to a minimum.

C. Advantages

1) Time is not an issue: Since this technique does not focus on time of flight, time delays, line of sights and the other issues mentioned previously cannot occur with this technique.

2) Visual aid: Instead of just relying on sound, the feature-base can also be enhanced with additional information available on standard smartphones in order to increase accuracy and robustness.

D. Problems

1) Processing cost: Calculating the minimal set of weighted sounds can require a lot of processing power, and thus might be inappropriate for small handheld devices – especially if localization is run periodically in the background.

2) Fingerprint Database: A fingerprint database has to be created and stored in an accessible way. Maintaining the database by keeping it up to date can cause severe cost, although automated update/merge algorithm can help against fingerprints [13] getting outdated.

3) Distinctness required: The algorithm assumes that different places are very different in their ambient sound levels. This might work in a mall, but can be an issue in an office environment where all rooms look and sound similar.

VIII. Material Spectrum Localization

Another technique, called "Symbolic Object Localization" by Kai Kunze et al. [12], is similar to the Background Spectrum analysis as it also analyzes the features of a sound. However, instead of relying on a collection of ambient sounds, this technique actively probes the environment on a set of different frequencies by playing different sounds, and records the feedback/echo of the probes.

The feedback is then analyzed to determine a set of characteristic properties for the environment - in this case a specific material on which the probing device, e.g. a smartphone, resides. By comparing the detected fingerprint with existing fingerprints, one can decide on the closest matching fingerprint in the database, and return this as the localization result.

A. Probing range

When sending out a sinusoidal probing sound at a certain frequency by the speaker of the smartphone, the response might look similar to what can be seen in figure 11.

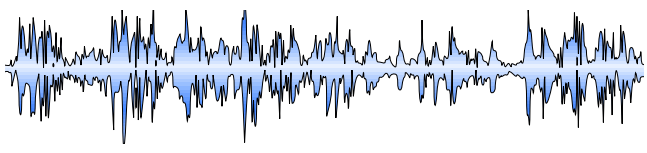

Fig. 11. An examplary sound wave (pressure over time) resulting from a sinusoidal probe with increasing frequency

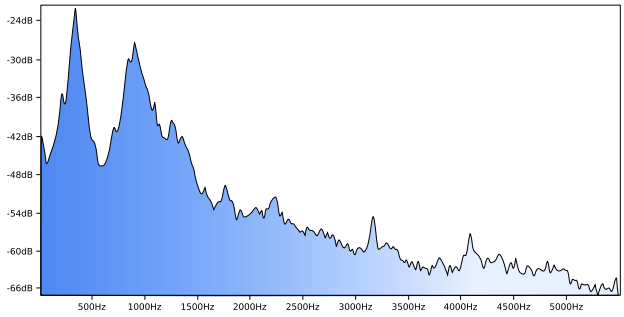

Fig. 12. Spectrum analysis via FFT (see section III-A2). The highest energy is present in the $0\,Hz - 1.5\,KHz$ range

When considering the spectrum analysis, and comparing different materials, it becomes obvious that most of the relevant energy is present in the band from $0\,Hz - 4096\,Hz$. Since the echo response greatly depends on the probing frequency, the material has to be probed at different steps in this band in order to create a characteristic fingerprint for an material (cf. figure 13).

material	response frequency	highest energy
concrete, unpainted	$1059\,Hz$	0.035
brick wall, painted	$1223\,Hz$	0.025
carpet on concrete	$1114\,Hz$	0.037

Fig. 13. Spectrum statistics for different materials when probed at the $0\,Hz - 4096\,Hz$ band (chart compiled from data presented in [12] using median, page 5)

It becomes obvious that – even though the materials in question are fairly similar – their median frequency and energy is rather distinct, and therefore provides a good basis for classification. Similar materials can also be distinguished by other objects in the area: A monitor put onto a desk will significantly change its material echo properties.

[7], [12]

B. Database

In order to retrieve localization results from the gathered finger prints, the probing device needs to have access to a database of existing fingerprints. We distinguish between two different kinds of reference fingerprints:

1) Specific/trained fingerprints: Similar to the approach presented in the "Background Spectrum" technique, one can compare the fingerprint with fingerprints representing specific logical locations, e.g. "kitchen desk in BC 203", or "my table at home". Since each of these locations is likely to have a different material characteristics, success rates of up to 70% have been reported when using a reasonable number of reference fingerprints (cmp. [12]).

The advantage is obvious: Users can set up the system for their own need. The less reference fingerprints are available, the lower the chance of polluting localization result since the *Hamming distance* between all fingerprints is higher.

2) Symbolic fingerprints: As the setup cost for the previous approach might be too high, another approach is to stick with abstract localizations like "wooden table", "in pocket", "concrete floor", "bathtube full of water"[2]. Since the materials available in every day life are fairly common, the system can thus be pre-trained. This kind of localization can be enough for systems that do not need any logical localization, or where the logical context is provided by the user ("put phone on vibrate when in pocket").

3) Increasing accuracy: Since most smart phones are capable of vibrating, one can also use the vibrator instead of the speaker to probe lower frequencies. As the signal emitted by the vibrator is much stronger, the response can be more distinct.

C. Limitations

This technique falls with the number of reference fingerprints and their similarity. If, by any chance, your work desk has the exact same material characteristics as your desk at home, this technique will fail to detect the proper location. This can be an issue as it is not a predictable error, but one that depends on the context in which this localization technique is used.

Also, probing all different frequencies of a material can take up significant time (roughly 8 seconds for the band proposed here, [12]), and thus does also put a high power load onto the phones battery.

IX. Conclusion

The different techniques described in this paper all have advantages and disadvantages. One huge advantage of all sound-based localizations is that nearly every user carries around a mobile device capable of playing and recording sounds, and thus the need for extra hardware is minimal. As the need for context-awareness is all present for mobile applications, it makes sense to further invest into this area.

On the other hand one has to consider that all techniques have their limitations: Some are very susceptible to noise; others may perform well in some situations, but terrible on others, providing an unpredictable pattern. This is why it is important to not only focus on one single technique, but instead combine all techniques proposed here into one to provide an accurate and reliable basis for localization. A similar approach is already taken by modern smartphones when it comes to outdoor positioning: Positions are approximated and enhanced using GSM based locations, and nearby WiFi hotspots, and are made exact using GPS – when available. We propose that a similar approach should be taken when it comes to audio localization.

Also, instead of just focusing on acoustic localization, the accuracy can be improved further by taking into account all sensors available on a smartphone. The ambient sound image can e.g. be enhanced using visual aids [9].

[2]Using this kind of localization might require a water-proof smartphone. Test at own risk.

If all of these suggestions are implemented, there is nothing in the way stopping location-awareness for mobile devices from becoming present and useful in everyday life.

References

[1] A. Madhavapeddy, D. Scott, and R. Sharp, "Context-aware computing with sound," in *IN PROCEEDINGS OF THE 5TH INTERNATIONAL CONFERENCE ON UBIQUITOUS COMPUTING*, 2003, pp. 315–332.

[2] P. Bahl and V. N. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *INFOCOM'00*, 2000, pp. 775–784.

[3] S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik, "Indoor localization without infrastructure using the acoustic background spectrum," in *Proceedings of the 9th international conference on Mobile systems, applications, and services*, ser. MobiSys '11. New York, NY, USA: ACM, 2011, pp. 155–168. [Online]. Available: http://doi.acm.org/10.1145/1999995.2000011

[4] P. N. Television, "Macgyver," 1985 - 1992.

[5] M. Stager, P. Lukowicz, and G. Troster, "Implementation and evaluation of a low-power sound-based user activity recognition system," in *Proceedings of the Eighth International Symposium on Wearable Computers*, ser. ISWC '04. Washington, DC, USA: IEEE Computer Society, 2004, pp. 138–141. [Online]. Available: http://dx.doi.org/10.1109/ISWC.2004.25

[6] J. Chen, K. Yao, and R. E. Hudson, "Source localization and beamforming," 2002.

[7] Wikipedia, "Wikipedia," 2013. [Online]. Available: http://wikipedia.org

[8] L. Girod and D. Estrin, "Robust range estimation using acoustic and multimodal sensing," in *In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2001)*, Maui, Hawaii, October 2001. [Online]. Available: 0102

[9] M. Azizyan, I. Constandache, and R. Roy Choudhury, "Surroundsense: mobile phone localization via ambience fingerprinting," in *Proceedings of the 15th annual international conference on Mobile computing and networking*, ser. MobiCom '09. New York, NY, USA: ACM, 2009, pp. 261–272. [Online]. Available: http://doi.acm.org/10.1145/1614320.1614350

[10] C. V. Lopes, A. Haghighat, A. Mandal, T. Givargis, and P. Baldi, "Localization of off-the-shelf mobile devices using audible sound: architectures, protocols and performance assessment," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 10, no. 2, pp. 38–50, Apr. 2006. [Online]. Available: http://doi.acm.org/10.1145/1137975.1137980

[11] J. Scott and B. Dragovic, "Audio location: accurate low-cost location sensing," in *Proceedings of the Third international conference on Pervasive Computing*, ser. PERVASIVE'05. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 1–18.

[12] K. Kunze and P. Lukowicz, "Symbolic object localization through active sampling of acceleration and sound signatures," in *Proceedings of the 9th international conference on Ubiquitous computing*, ser. UbiComp '07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 163–180. [Online]. Available: http://dl.acm.org/citation.cfm?id=1771592.1771602

[13] S. Chu, S. Narayanan, and C.-C. J. Kuo, "Environmental sound recognition with time-frequency audio features," *Trans. Audio, Speech and Lang. Proc.*, vol. 17, no. 6, pp. 1142–1158, Aug. 2009. [Online]. Available: http://dx.doi.org/10.1109/TASL.2009.2017438