Acoustic Localization in Mixed Environments with Line-of-Sight and Non-Line-of-Sight

Institut für Informatik
Technische Fakultät
Albert-Ludwigs-Universität Freiburg

Joan Bordoy

Juli 2020

Acoustic Localization in Mixed Environments with Line-of-Sight and Non-Line-of-Sight

Joan Bordoy

Zusammenfassung

Diese Arbeit behandelt die Lokalisierung von Zielen in Gebäuden mittels Schallwellen. Mit statischen Mikrofonen an bekannten Positionen kann ein Lautsprecher durch Messen der Ankunftszeitdifferenz (time difference of arrival, TDOA) der vom Lautsprecher ausgesendeten Schallwellen lokalisiert werden. Die größte Herausforderung besteht darin, zu identifizieren, welche Signale direkt vom Lautsprecher zu den Empfängern gelangen (*Sichtlinie*) und welche Signale von einer Wand oder einem Hindernis reflektieren (*Nicht-Sichtlinie*). Das Verwechseln eines Signals ohne Sichtverbindung mit einem Signal mit Sichtverbindung kann zu großen Lokalisierungsfehlern führen. Dies ist der schwerwiegendste Faktor, der die Robustheit eines Ortungssystems beeinträchtigt.

Wir stellen mehrere Ansätze zur Lokalisierung unter gemischten Sichtlinien- und Nicht-Sichtlinienbedingungen vor. Zunächst zeigen wir, wie man bei einer großen Anzahl von Messungen Messkombinationen verwenden und den Restfehler der Schätzungen analysieren kann, um festzustellen, welche der gemessenen Signale solche mit Sichtverbindung sind. Unsere experimentellen Ergebnisse legen nahe, dass dieser Ansatz einen ähnlichen Fehler erzielt wie andere Ansätze, die ein definiertes probabilistisches Bewegungsmodell für das Ziel erfordern.

In bestimmten Fällen ist die verfügbare Anzahl von Sichtliniensignalen reduziert. In diesen Fällen muss eine zusätzliche Informationsquelle genutzt werden. Wir zeigen experimentell und mit Simulationen, wie die von einer Inertial-Messeinheit gesammelten Informationen verwendet werden können, um den Effekt von Signalen ohne Sichtverbindung bei der letztendlichen Schätzung abzuschwächen. Zu diesem Zweck wird ein auf M-Schätzung basierendes Kalman-Filter verwendet.

Nicht-Sichtlinienmessungen als ungültige Messungen, die eliminiert werden sollten, zu betrachten, erfordert ein bestimmtes Maß an Überbestimmung in Form zusätzlicher Sensoren oder zusätzlicher Sichtliniensignale. Nicht-Sichtliniensignale enthalten jedoch wertvolle Informationen über den Ort des Ziels, die zur Schätzung seiner Position verwendet werden können. Eine Reflexion an einer Wand wirkt wie ein zusätzlicher *virtueller Empfänger*. Die größte Herausforderung besteht dabei darin, zu wissen, welche Messungen von derselben Wand reflektiert wurden (und somit zu demselben virtuellen Empfänger gehören). Um dies abzuschätzen, verwenden wir ein Joint Probability Data Association Filter. Wir nutzen die Tatsache, dass sich der Abstand zwischen Sender und Empfänger bei zwei aufeinander folgenden Messungen nur geringfügig ändert. Somit ist diese Abstandsänderung deutlich kleiner als die Abstandsdifferenz zwischen den reflektierten Signalen.

Nach dem Schritt der Datenassoziation kann die Positionen des Senders und der Empfänger geschätzt werden. Um die Variablen zu initialisieren, verwenden wir eine Fernfeld-Approximation. Wir gehen davon aus, dass die Entfernung, die ein Sender während einer bestimmten Zeit zurücklegt, deutlich kleiner ist als die Entfernung zu den Empfängern. Wir verbessern existierende Fernfeldalgorithmen für unser spezifisches Szenario, um die Anzahl der erforderlichen Empfänger zu reduzieren und eine robustere Schätzung zu erhalten. Anschließend zeigen wir mit simulierten und experimentellen Daten die Überlegenheit unserer Fernfeldalgorithmen, wenn sich der Sender in einer quasi-linearen Trajektorie bewegt oder das Messrauschen hoch ist.

Unter den Nicht-Sichtlinien-Signalen sind die am Boden reflektierten Signale von besonderem Interesse. Diese Reflexionen enthalten wertvolle Informationen über die Höhe des Ziels. In dieser Arbeit geben wir eine Untergrenze für den quadratischen Mittelwertfehler an und überprüfen experimentell, wie diese Reflexionen die Lokalisierungsgenauigkeit erheblich verbessern können, wenn die Empfänger in ähnlichen Höhen platziert werden.

Bei den zuvor erwähnten Ansätzen wird ein Lautsprecher jedes Mal geortet, wenn er ein Signal aussendet. Dies bedeutet, dass die Anzahl der Lautsprecher, die gleichzeitig lokalisiert werden können, begrenzt ist, da sie sich ab einer gewissen Grenze gegenseitig stören oder ihre Position erst nach einem langen Zeitraum aktualisiert wird. Durch Lokalisieren eines sich bewegenden Empfängers mit statischen Lautsprechern als Ankerknoten können unbegrenzt viele Ziele gleichzeitig lokalisiert werden. Wir untersuchen diese Möglichkeit, indem wir einen Algorithmus vorstellen, der die in den akustischen Signalen codierten Informationen mit der geschätzten Position und Geschwindigkeit des Ziels verbindet, um einen Empfänger mit Geschwindigkeiten von bis zu 1,8 m/s zu lokalisieren. Die experimentellen Ergebnisse zeigen, wie der Empfänger mit einem Medianfehler von nur 5 cm lokalisiert werden kann, was bisherige Ansätze übertrifft.

Abstract

This thesis discusses the localization of targets inside buildings by means of acoustic waves. Having static microphones at known positions, a speaker can be located by measuring the time difference of arrival of the sound waves the speaker emitted. The main challenge is to identify which signals travel directly from the speaker to the receivers (*line-of-sight*) and which signals do it after bouncing to a wall or an obstacle (*non-line-of-sight*). Mistaking a non-line-of-sight signal by a line-of-sight signal can lead to large localization errors. This is the most remarkable factor which compromises the robustness of a location system.

We provide multiple approaches for localization in mixed line-of-sight and non-line-of-sight conditions. First we show how, when a high number of measurements are available, one can use combinations of measurements and analyze the residual error of the estimations in order to identify which signals are in line-of-sight. Our experimental results suggest that this approach achieves a similar error than other approaches which require a certain probabilistic motion model for the target.

In certain occasions, the available number of line-of-sight signals is reduced. Then, one must make use of an additional source of information. We show experimentally and with simulations how the information gathered by an inertial measurement unit can be used for mitigating the effect of non-line-of-sight signals in the final estimation. An M-estimation based Kalman filter is used for this purpose.

Considering non-line-of-sight measurements as invalid measurements which should be eliminated requires a certain degree of overdetermination in the form of additional sensors or additional line-of-sight signals. However, non-line-of-sight signals carry valuable information about the location of the target that can be used for estimating its position. A reflection from a wall acts like an additional *virtual receiver*. The main challenge is to know which measurements were reflected by the same wall (i.e. they belong to the same virtual receiver). In order to estimate this, we use a joint probabilistic data association filter. We exploit the fact that the distance between a sender and a receiver changes only slightly during two consecutive measurements. Then, this distance change is much smaller than the distance difference between the reflected signals.

After the data association step, one can estimate the positions of the sender and the receivers. In order to initialize the variables, we use a far-field approximation. We assume the distance moved by a sender during a certain time is much smaller than the distance to the receivers. We improve existing far-field algorithms for our specific scenario in order to reduce the number of required receivers and provide a more robust estimation. Then, with simulated and experimental data we show the superiority of our far-field algorithms when the sender moves in a quasi-linear trajectory or the magnitude of the measurement noise is large.

Among the non-line-of-sight signals, the signals which are reflected at the ground are of special interest. These reflections carry valuable information about the height of the target. In this thesis we provide a lower bound for the root mean square error and verify experimentally how these reflections can considerably increase the localization precision when the receivers are placed at similar heights.

In the previously mentioned approaches a speaker is located every time it emits a signal. This means, the number of speakers that can be located simultaneously is limited, as at a certain point they would interfere with each other or their position would be updated only after a large period of time. By locating a moving receiver and having static speakers as anchor nodes one can locate unlimited targets simultaneously. We explore this possibility by presenting an algorithm which fuses the information encoded in the acoustic signals with the estimated position and velocity of the target in order to locate a receiver with velocities up to 1.8 m/s. The experimental results show how the receiver can be located with a median error of only 5 cm, outperforming previous approaches.

Acknowledgments

First and foremost, I would like to thank my advisor Prof. Christian Schindelhauer for letting me be a PhD student at his chair. His invaluable guidance let me grow as a scientific researcher. He always gave me the right amount of freedom, support and constructive criticism in a way that I always felt motivated to do better research. In our discussions he always provided me with new ideas and suggestions. I would like to thank also Prof. Leonhard Reindl for his guidance and discussions every time I presented him my work.

I wish to thank also Johannes Wendeberg for his guidance since I first did my master thesis and then started as a PhD student. We had many discussions during these years and his excellent work on algorithms for indoor localization was a great inspiration. I would also like to thank Fabian Höflinger for his recommendations and in general all the members of the company Telocate for being such great work mates. Working part-time in the company let me understand the necessities of real-life applications and have access to magnificent resources which eased the process of doing experiments and recording data for later analysis. Moreover, I always was encouraged to publish the algorithms which came up from the necessities and discussions inside the company.

I thank also my coauthors Alexander Traub-Ens, Rui Zhang and Dominik Jan Schott for their remarkable contribution and scientific discussions. It was also an honor to be part of the chair of Computer Networks and Telematics and share thoughts with their members, especially with Sven Köhler and Fabian Schillinger. I would like to express my gratitude also to Kerstin Pfeiffer and Beate Botschek for their help in administrative matters.

My sincere thanks to Amir Bannoura, Jean-Michel Boccard and the members of the post-graduate scholar traineeship who were fellow PhD students and shared their thoughts about the process with me.

Thanks also to Prof. Álvaro Hernández Alonso for agreeing to be my second supervisor and for his suggestions during my first international conference.

Last, thanks to my family and to Johanna Florez for their patience and unconditional support during these years.

Contents

1 Introduction

Nowadays, GPS (Global Positioning System) is a well known solution for localization. The number of applications which require it has been growing over the past years. While it is a widely used system, it cannot be used inside buildings. Therefore, indoor localization has been a field of research which has gained increasing attention. The number of applications for which such systems have been envisioned is immense. Among others, guidance for visually impaired persons, localization of persons and assets or navigation inside airports.

A large number of people use a smartphone to navigate outdoors. Then, there exists a demand for an indoor location system which is capable of locating out-of-the-box smartphones, allowing their users to navigate also inside buildings.

An indoor localization system can also play a decisive role in intralogistics. With such a system one can track a great magnitude of assets and even optimize the efficiency of the employees. In such a scenario the localization system is required to track multiple targets simultaneously. In addition, the system must be affordable and require a limited energy consumption.

In applications such as navigation for visually impaired people, it is crucial to have a robust system which is capable of locating a target with high precision and accuracy.

A solution which is capable of fulfilling the above mentioned requirements is localization with acoustic waves. Using frequencies which are above the audible range, one can use inexpensive devices for localization. A speaker can be located by having multiple static receivers attached to the ceiling. If the receivers are synchronized, one can use the time at which those signals are received by the microphones to locate the target with time difference of arrival (TDOA). The targets can be devices like smartphones or laptops which emit inaudible sound signals. This allows localization with an error in the order of centimeters [1].

One can also have static speakers as anchor nodes in order to locate moving receivers. Then, the targets receive signals emitted by the static senders and are capable of locating themselves. By doing this, the number of targets that can be located is unlimited.

As with acoustic localization, most of the existing solutions for indoor localization rely on static anchor nodes placed over the localization area. Then, by multiple means, one can know how far or in which angle is the target to some of the nodes and estimate its position.

Measuring the signal strength of an electromagnetic signal (RSSI) is a simple and cheap solution to know whether a target is close to an anchor node. However, one cannot locate a target with high precision and accuracy.

Certain localization systems measure the angle of arrival of the electromagnetic or acoustic waves emitted by a portable device. This can lead to an error in the order of centimeters when the target is close to the anchor node. While this is a promising approach, it is not optimal for

1

places with high ceilings, as the angle variation is too small to provide a precise estimation.

Measuring the time at which electromagnetic signals are received is a challenging task, as the light velocity is much higher than the sound velocity. Ultrawideband (UWB) systems manage to do so by using a high bandwidth and low spectral density (e.g. [2]). The price is usually higher than acoustic systems.

All the above mentioned systems have the issue that sometimes a target must be located in mixed line-of-sight/non-line-of-sight conditions. This means that the target emits a signal that might be obstructed and not received by an anchor node (see Fig. 1.1). In addition, the anchor node might receive signals which have bounced from this target to a reflector (e.g. a wall) and then to the anchor node (non-line-of-sight signals). Then, one has to decide which signals were really emitted by the target and did not bounce to any reflector (line-of-sight signals). This is a challenging task and can lead to high errors if it is not correctly done. The robustness of an indoor localization system depends mostly on doing a correct data association.

This issue is not present in other localization solutions which use an inertial measurement unit (IMU) carried by the target. The IMU provides information about the acceleration of the target, which can be integrated in order to estimate its position. This has been proved to be effective for localization, although its accumulative error makes it unusable after a certain period of time. However, this technology can be fused with others in order to achieve a more robust estimation. For instance, one can use an acoustic system to correct the accumulative error of the IMU and the inertial measurement unit to improve the identification of the acoustic line-of-sight signals.

This thesis focuses on acoustic and ultrasonic localization systems. The reason behind that is that it is a cost effective technology which can be used in multiple scenarios. In addition, it can be used with out-of-the-box smartphones. We make use of the flexibility which this technology offers. One can have inexpensive unsynchronized speakers as anchor nodes or as targets. In addition, one can have a more expensive system where the speakers and microphones are synchronized. The emitters can send signals regularly at specific intervals or at random time instants. All of these scenarios have a different application case and offer different challenges when locating a target in presence of non-line-of-sight measurements. They will be further explained in this thesis.

1.1 Non-Line-of-Sight Measurements

As mentioned above, the task of deciding which measurements are line-of-sight and non-line-of-sight is crucial for the robustness of a localization system. Especially in systems which rely on acoustic measurements, as acoustic signals are considerably attenuated when passing through a material.

A high number of researchers have proposed methods to properly decide which signals are line-of-sight, either with data processing and channel estimation or by selecting which position is more likely according to the received measurements and the movement of the target. Both categories of algorithms are often used together in order to improve the robustness of the sys-
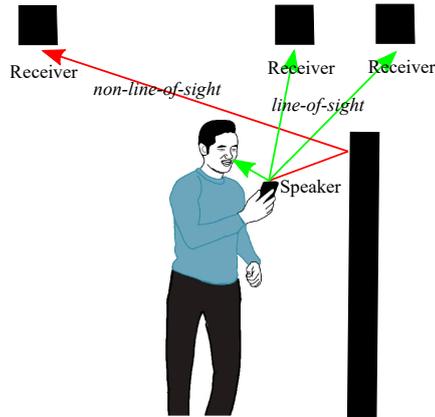
Figure 1.1: In certain occasions the signal emitted by a speaker is blocked and only non-line-of-sight signals are received. If these signals are mistaken for line-of-sight signals, the localization error can increase dramatically.

tem. The algorithms presented in this thesis fall into the second category. We present multiple approaches for discarding non-line-of-sight measurements in different scenarios but also for using these measurements as additional information.

In this thesis we also show how acoustic reflections can be used to locate a target even when the positions of the anchor nodes and the reflectors are unknown. The reflections act as virtual anchors, providing additional information about the movement of the target. By assuming the moved distance by the target during a certain period of time is smaller than the distance from the target to the anchor nodes (*far-field assumption*) one can simplify the equations and provide an initial estimation for other algorithms which require an initial guess in order to successfully estimate the positions of the target and anchor nodes.

Furthermore, in this thesis we analyze how much the localization precision can be increased by using acoustic reflections as additional information. More specifically, we focus on waves reflected by the ground. Multiple smartphones have their speaker on the rear side, on the opposite site of the screen, which means the ground reflection is often received. Moreover, these reflections contain information about the height of the target. A slight change in the height of the target leads to a noticeable time difference between the line-of-sight measurements and the reflected signals. This allows estimating the height of a smartphone with a much lower error than using only line-of-sight measurements. This is because often the anchor nodes are placed at similar heights, which impedes a proper target height estimation. This thesis shows analytically and experimentally how much improvement ground reflections provide when estimating the position of the target in three dimensions.

1.2 Thesis Goals and Objectives

This thesis aims to provide a robust position estimation in presence of mixed line-of-sight/non-line-of-sight conditions. This is achieved under different scenarios:

- In certain scenarios the position of the target is updated frequently and therefore one can assume it follows a smooth trajectory over the time. In other scenarios this is unfeasible.
- The assumed a priori knowledge varies depending on the scenario. Then, the sending time of the signals, the positions of the anchor nodes and the positions of the reflectors can be either known or unknown.
- One can locate a moving sender or a moving receiver.

Finally, this thesis proves that reflections can be used as additional information when locating a moving speaker in order to reduce the localization error. It shows:

- How can reflections be used without knowing the positions of reflectors and anchor nodes.
- How much improvement can they provide when estimating a target position in three dimensions.

1.3 Outline of the Thesis

This thesis is organized as follows. This chapter contains an introduction to the topic and the main contributions to the state of the art. Chapter 2 presents an overview of the related work. Chapter 3 presents an introduction to acoustics with emphasis on soundwave propagation. Moreover, it shows the simplified mathematical models which relate the reception times of such acoustic signals and the positions of speakers and microphones. Additionally, it provides a description of the localization systems that will be used in this thesis.

An introduction to probabilistic localization is provided in Chapter 4. We show how a target can be located by assuming certain motion and sensor models with Gaussian distributed error. This is of special importance when locating a moving receiver with unsynchronized speakers due to the reduced number of mathematical constraints. Parts of this chapter were published in [3, 4].

Chapter 5 already introduces the problematic of identifying NLOS measurements. This chapter defines the model that governs such measurements and shows how they can lead to high localization errors when a LOS model is assumed. Different algorithms are proposed to mitigate their effect on the final position estimation. We propose a novel algorithm and compare it with existing solutions. In addition, we show how data from an inertial measurement unit (IMU) can be fused with the acoustic data in order to provide a reliable estimate, as we already showed in [5].

In Chapter 6 we show how the movement of a sender leads to time differences in the receivers which can be used to locate a target in three-dimensions without knowing where the receivers are and using the reflections as virtual receivers. We compare this novel approach with the

state of the art, showing its superiority when tracking a moving target in three-dimensions and having imprecise measurements. The majority of this chapter has been published in [6].

Chapter 7 proposes the inverse approach, using the relative movement between a receiver and multiple static speakers for localization. This approach uses the measured timestamps, the predicted position and velocity of the target and modulated symbols in order to predict the most likely line-of-sight measurements. Parts of this chapter were published in [3].

Acoustic reflections can be used to reduce the error when locating a target. In Chapter 8 we analyze how much the ground reflection can improve the height estimation. Lower bounds are found for the root mean square error (RMSE). The improvement is also shown with real-life experiments. In addition, an algorithm to identify the ground reflections is provided. Parts of this chapter were published in [7].

In the last chapter we provide the conclusions of this thesis and outline open problems.

1.4 Collaborations

This thesis focuses on algorithms for robust localization in mixed line-of-sight/non-line-of-sight conditions. The other parts required for a localization system to properly locate a target were developed and conceived either partly or completely by other researchers. More in detail:

- The localization system for tracking a smartphone using static microphones was initially conceived by Johannes Wendeberg and Fabian Höflinger. This was the initial idea of their spin-off Telocate GmbH. The receivers were designed and built by Fabian Höflinger.

- The localization system which aims to locate a moving receiver using unsynchronized emitters was initially conceived by Heinrich Hippenmeyer, Fabian Höflinger, Leonhard Reindl, Christian Schindelhauer, Alexander Traub-Ens and Johannes Wendeberg. It was part of the industrial project eCULTS. The algorithms presented in Section 4.2, which assume line-of-sight measurements, were developed in close collaboration with Johannes Wendeberg. The signal modulation and demodulation was done by Alexander Traub-Ens. Moreover, Chapter 7 shows improvements made for better line-of-sight identification over the existing methods presented by Alexander Traub-Ens.

- The data from the inertial measurement unit used for testing the data fusion algorithm presented in Section 5.2 was processed by Rui Zhang.

1.5 Co-authored Publications

Parts of this thesis have been already published in peer-reviewed conferences proceedings or journals. These publications are:

- Joan Bordoy, Christian Schindelhauer, Fabian Höflinger and Leonhard Reindl. Exploiting acoustic echoes for smartphone localization and microphone self-calibration. In *IEEE Transactions on Instrumentation and Measurement, Volume: 69 , Issue: 4, Pages: 1484 - 1492*, 2020

- Joan Bordoy, Rui Zhang, Fabian Höflinger, Christian Schindelhauer and Leonhard Reindl. Robust Extended Kalman filter for NLOS mitigation and sensor data fusion. In *IEEE International Symposium on Inertial Sensors and Systems (INERTIAL),* 2017
- Joan Bordoy, Johannes Wendeberg, Fabian Höfflinger, Christian Schindelhauer and Leonhard Reindl. Exploiting ground reflection for robust 3D smartphone localization. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2016
- Joan Bordoy, Alexander Traub-Ens, Ali Sadr, Johannes Wendeberg, Fabian Höflinger, Christian Schindelhauer and Leonhard Reindl. Bank of Kalman filters in closed-loop for robust localization using unsynchronized beacons. In *IEEE Sensors Journal, Volume: 16, Issue: 19, Pages: 7142 - 7149*, 2016
- Alexander Traub-Ens, Joan Bordoy, Johannes Wendeberg, Leonhard M Reindl and Christian Schindelhauer. Data fusion of time stamps and transmitted data for unsynchronized beacons. In *IEEE Sensors Journal,Volume: 15, Issue: 10, Pages: 5946 - 5953*,2015
- Alexander Traub-Ens, Joan Bordoy, Johannes Wendeberg, Christian Schindelhauer and Leonhard M. Reindl. Unsynchronized ultrasound system for TDOA localization. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2014

1.6 Notations

The following symbols and variables will be often used in this thesis:

$\|\cdot\|$	Norm operator
$\lvert\cdot\rvert$	Absolute value operator
$\mathcal{N}(\mu,\sigma^2)$	Normal distribution with mean μ and variance σ^2
$\mathcal{N}(\xi;\mu,\ \sigma^2)$	Normal distribution with mean μ and variance σ^2 evaluated in ξ
$\mathcal{N}(\mathbf{a},\mathbf{Q}_t)$	Multivariate normal distribution with mean $\mathbf{a}$ and covariance $\mathbf{Q}_t$
$\mathcal{U}(a,b)$	Uniform distribution between a and b
$()^T$	Transpose operator
$\lfloor\cdot\rfloor$	Floor function, takes the integer which is less or equal than the value inside the operator
$\mathbf{0}$	Vector of zeros
$\mathbf{1}$	Vector of ones
$\mathbf{I}$	Identity matrix
$\mathbf{a}\cdot\mathbf{b}$	Dot product between $\mathbf{a}$ and $\mathbf{b}$
$\mathbf{S}$	Sender position vector in $\mathbb{R}^2$ or $\mathbb{R}^3$, depending on the context
$\mathbf{M}$	Receiver position vector in $\mathbb{R}^2$ or $\mathbb{R}^3$, depending on the context
N	Number of receivers
B	Number of senders

z	Measurement vector
t	Sending time
I	Time interval
T	Reception time
c	Sound velocity

2 Related Work

In this chapter we provide an overview of publications from other authors which are relevant to the field of localization and present important contributions to the topics that will be described in this thesis. We also mention our own publications and how they relate to the other approaches.

Multiple researchers have shown the possibility of locating a speaker using acoustic or ultrasonic measurements with an error in the order of centimeters. Some of them assume the sending time is known (*time of arrival*, e.g. [8, 9]) and others assume it is unknown (*time difference of arrival*, e.g. [10]).

The speaker of a smartphone can be used to emit inaudible acoustic signals [1, 11]. Liu et. al showed how one can also locate its microphone by using static speakers [12].

By locating moving receivers instead of senders one can have unlimited targets estimating their own position without interfering with each other. Several systems have been presented to locate moving receivers with time difference of arrival [13–16] and time of arrival [17–19]. In these systems the speakers are synchronized so that they can emit signals at predefined time instants. In [20] the speakers are not synchronized. However, time-of-arrival (TOA) and angle-of-arrival (AOA) measurements are combined to locate the target. In [4] we presented a cost effective location system which locates receivers with TDOA and does not require synchronization between senders.

Non-line-of-sight Mitigation

Various approaches have been presented for locating a target in mixed non-line-of-sight and line-of-sight conditions. Some approaches focus on the channel and their aim is to select the line-of-sight timestamps which will be later used for localization. Multiple researchers use the correlation between the received signals and the emitted signal pattern. Values such as the mean excess delay and the skewness of the correlation are used [21–24]. In some cases reflections can cause destructive interferences, which might lead to erroneously selecting reflections as line-of-sight measurements. In [25] and [26] the contributions of reflections in the time correlation are iteratively eliminated.

Other approaches assume the timestamps are given and they can be either line-of-sight or non-line-of-sight measurements. With TOA measurements, one can define a closed region were the target is likely to be, even when there are non-line-of-sight measurements. Due to this fact certain authors [27, 28] use constrained optimization when estimating the position of the target. Robust regression approaches which give lower weights to unlikely measurements have been presented in [29]. Feng Yin shows in his PhD thesis [30] how the NLOS measurement

noise can be characterized using a Gaussian mixture or a Kernel density estimate.

When the sending time is unknown, it is more challenging to identify the non-line-of-sight measurements, as the subspace which defines the possible location of the target is not bounded. The only limitation is the maximum distance at which a receiver is capable of receiving a signal. In [31] the NLOS measurements are modeled as an additive mean shifted Gaussian distribution. Wang et. al [32] use robust optimization and a convex approximation method which assumes the distance traveled by a reflection is limited. In [33] robust regression is used to continuously track the sender. Concretely, an interacting multiple model is used, which continuously switches from an only-LOS measurement model to a model which assumes LOS and NLOS measurements. In [5] we used robust regression , fusing the data of an inertial measurement unit, which provided additional information and improved the identification of the line-of-sight signals.

Probabilistic Data Association

Using a recursive Bayesian estimator, one can estimate the position and velocity of the target. By doing this, one can predict which timestamps are more likely to be line-of-sight measurements and discard the non-line-of-sight measurements [29, 34].

Li and Krolik [35] present an approach which simultaneously estimates the position of the target and the location of the walls. The data association between walls and measurements is done with a multi-hypothesis probabilistic algorithm. They are provided with AOA and TOA measurements.

In [6] we use a joint probabilistic data association filter to track the relative motion of the target to the receivers. This is done by assuming the emitter sends signals at regular intervals. The sending time is unknown. The main benefit of this approach is that the data association is done without knowing the positions of the sender and receivers. Then, one does not need to know the position of the receivers in order to associate the data.

In [3] we propose a probabilistic data association algorithm which selects the most likely line-of-sight measurements depending on the predicted position and velocity of the target, the interval at which the signals are emitted and the modulated symbols. In this case, the moving target is a receiver and the static anchor nodes are senders. For this reason, when the target moves the number of constraints is lower than the number of variables. Fusing multiple sources of information one can predict the most likely line-of-sight signal of every sender. This was initially presented in [4] by Traub-Ens et. al. However, we extended it in [3] so that the estimated position and velocity of the target can be used for better data association.

Using Reflections as Additional Information

Dijk et. al [36] use the knowledge about the dimensions of the room as additional information for localization. The method assumes a box-shaped room and uses a signature matching algorithm, which estimates the target position by predicting which reflections will be received and at which time instant. In [37] Dokmanic et. al show that the shape of a room can be reconstructed

using acoustic echoes. There, five microphones and one speaker are used. The positions of the microphones and the sending time of the signals are known in advance. In [38] Kim et. al use the knowledge of the room dimensions and the positions of static speakers in order to predict the motion of a receiver. In [39], a floor plan is used to map ultra-wideband reflections to virtual anchor nodes. In [40] the three-dimensional position of a speaker and microphone is estimated without any a-priori knowledge about the room dimensions nor the positions of the nodes. The data association is done with a random sample consensus (RANSAC) algorithm and the receivers are located sequentially, using first the receivers with a high number of measurements. Reflections are used as additional receivers in order to reduce the localization error. Kuang et. al [41] show how UWB reflections from walls can be used to track a target and estimate the dimensions of the room. Only one anchor node and one mobile device are used. They use a far-field assumption. In [7] we showed that ground reflections can be used to improve the height estimation. We also showed in [6] how the far-field assumption can be used to locate a speaker in three dimensions using acoustic reflections as additional receivers at unknown positions. The positions of the real and virtual receivers are estimated simultaneously to the trajectory of the target. Then, the location system can be self-calibrated.

Self-Calibration

Usually, a moving speaker is located and multiple microphones are installed at static positions. Knowing these positions reduces the number of unknowns and therefore provides a more reliable target position estimation. However, measuring them manually increases the installation cost and requires a considerable amount of time. Automatically calibrating the microphone positions in order to reduce the installation effort has been proved to be feasible by multiple researchers. In [42] a maximum likelihood estimator is used. In [43] a direct solution is found using matrix factorization, requiring ten receivers. In [44] an iterative algorithm is presented. Johannes Wendeberg provides multiple solutions for TDOA self-calibration in his PhD thesis [45]. Among others, a branch-and-bound algorithm is presented which continuously divides the subspace of possible solutions until the positions are found within an error bound ϵ. He also shows how a far-field approximation can be used to initialize the variables of the scenario and avoid that non-linear optimization algorithms get stuck in local minima.

Locating receivers and senders in mixed environments with line-of-sight and non-line-of-sight is challenging due to the high number of variables and nonlinearities. However, it is often the case in real-life environments. In [46] a self-calibration algorithm for dual-microphone arrays is presented. Outliers are eliminated using a random sample consensus (RANSAC) approach.

Far-Field Assumption

Some authors already used the assumption that a single or multiple senders are far from the receivers for self-calibration. This was used in order to simplify the equations and provide an initial estimate, which was often refined with local optimization methods such as gradient

descent.

In [47], Thrun presents a method for locating far sound sources and microphones by using a far-field assumption and affine geometry. He assumed that the sound sources were much farther than the distance between the microphones. It was shown that this assumption can be used as an initial estimate for local optimization algorithms in order to avoid local minima and reduce the localization error.

The approach by Kuang et. al [41] locates a target with an error of around one centimeter without knowing the anchor node positions and using multipath as additional information. The method in [47] is used to estimate the target positions. The data association of the measurements previously to that step is done with a semi-automatic approach.

The ellipsoid method is a method to localize a group of microphones using far sound sources and was presented in [48] by Schindelhauer et. al and extended in [49]. The algorithm worked under the assumption that the sender was much farther than the distance between the receivers.

Approach	Assumption	Data Association	Experiments	Far-field algorithm
Thrun, 2005 [47]	Small distance between static microphones*	N.A.	2D	Affine geometry
Schindelhauer et. al, 2010 [48, 49]	Small distance between static microphones*	N.A.	2D	Ellipsoid method
Kuang et. al, 2013 [41]	Small distance between target positions**	Semi-automatic	2D	Affine geometry
here and in Bordoy et. al, 2020 [6]	Small distance between moving speaker positions *	JPDAF	3D	Modified version of [47] and [48] for 3D

* compared to speaker-microphone distance
** compared to target-anchor node distance

Table 2.1: Comparison between approaches using the far-field assumption. N.A stands for not available. [6]

Table 2.1 shows a comparison of the existing methods which use a far-field approximation for localization. The assumption that the static microphones are close compared to the distance between the speaker and the microphones used in [48] and [47] has the main drawback that all the received timestamps need to be in line-of-sight. Otherwise, the sender cannot be located. One could assume a reflection is like a virtual receiver, but then the far-field assumption would not necessary hold anymore.

The approach which we presented in [6] assumes the sender is moving, and during a certain period of time, its positions are closer than the distance to the receivers. Then, reflections can be used as additional receivers, as the virtual receivers generated by the acoustic echoes will also hold the far-field assumption, because they will be farther than the actual receivers.

A new formulation of the method in [47] was presented which requires less receivers and provides a more robust estimation when locating a moving target. In addition, the ellipsoid method presented in [48] was modified for better performance in this scenario, using constrained optimization.

3 Acoustic Position Estimation

This chapter serves as an introduction to the task of locating a target. It describes how acoustic signals can be used for localization and how can they be related to the position of the target.

We show how a sound signal propagates and is reflected. Of special interest is the description of how reflection from walls can be seen as virtual senders or receivers, which will simplify certain calculations in the following chapters.

In addition, we provide a description of the localization systems used in this thesis. Two localization systems will be used which offer different characteristics and challenges when locating a target in mixed line-of-sight and non-line-of-sight conditions.

3.1 Fundamentals of Acoustics

In order to locate a target using signals in the inaudible range, one needs to comprehend what is a sound wave and how it interacts with the medium. A sound wave is a local deviation of the local atmospheric pressure [50] over time. Then, if one uses a speaker to emit an harmonic with angular frequency ω in the x coordinate, there will be a variation of pressure $p_i(t, x)$ that depends on the time t and the position x. The particles will also move at a certain particle velocity $u_i(t, x)$. The product of the particle velocity and pressure is the intensity of the sound signal. Their variation over the x coordinate will be:

$$p_i(t, x) = \widehat{p}_i e^{-j(\omega t - k_1 x)} \tag{3.1}$$

$$u_i(t, x) = \frac{\widehat{p}_i}{\rho_1 c_1} e^{j(\omega t - k_1 x)} \tag{3.2}$$

where the wave number k_1 depends on the sound velocity of the medium c_1:

$$k_1 = \omega / c_1 \tag{3.3}$$

The variable ρ_1 is the equilibrium density of the material in which the sound wave propagates and j denotes the imaginary unit.

Reflections

When a sound wave changes the material in which it travels, part of the signal is transmitted and part is reflected. The acoustic phenomenon which plays a crucial role in indoor localization is the reflection. As mentioned before, reflected signals can be mistaken for line-of-sight signals,

which can lead to high errors in the position estimates. One can distinguish two cases: when the sound signal hits a material with a direction perpendicular to the plane which contains the material (*normal incidence*) and when it does it with a different angle (*oblique incidence*). We provide a short introduction to these scenarios based on the description in [50].

Normal Incidence

In this case we assume the sound source is located just in front of a wall. The wall dimensions are assumed to be much larger than the wavelength and the dimensions of its irregularities are assumed to be much smaller than the wavelength of the acoustic wave. The wavelength λ at which the sound waves travel in the experiments carried out in this thesis is between one and two centimeters approximately, depending on the frequency f and the sound velocity in the medium c_1:

$$\lambda = \frac{2\pi c_1}{\omega} = \frac{c_1}{f} \tag{3.4}$$

If we assume a wall is located at $x = 0$, there will be two new sound waves, a reflected wave:

$$p_r(t, x) = \widehat{p}_r e^{-j(\omega t + k_1 x)} \tag{3.5}$$

$$u_r(t, x) = -\frac{\widehat{p}_r}{\rho_1 c_1} e^{j(\omega t + k_1 x)} \tag{3.6}$$

and a transmitted wave:

$$p_t(t, x) = \widehat{p}_t e^{-j(\omega t - k_2 x)} \tag{3.7}$$

$$u_t(t, x) = \frac{\widehat{p}_t}{\rho_2 c_2} e^{j(\omega t - k_2 x)} \tag{3.8}$$

Here ρ_2, c_2 and k_2 are the properties of the material in which the sound wave hits.

In order to learn how these waves are related, one must apply the boundary conditions in $x = 0$:

$$p_i(t, 0) + p_r(t, 0) = p_t(t, 0) \tag{3.9}$$
$$u_i(t, 0) + u_r(t, 0) = u_t(t, 0) \tag{3.10}$$

By dividing these two equations:

$$\rho_1 c_1 \frac{\widehat{p}_i + \widehat{p}_r}{\widehat{p}_i - \widehat{p}_r} = \rho_2 c_2 \tag{3.11}$$

Then, one can calculate the amount of reflected pressure divided by the amount of incident pressure, the so-called reflection coefficient, which is noted as R:

$$R = \frac{\widehat{p}_r}{\widehat{p}_i} = \frac{\frac{\rho_2 c_2}{\rho_1 c_1} - 1}{\frac{\rho_2 c_2}{\rho_1 c_1} + 1} \tag{3.12}$$

One can observe how the amount of pressure which is reflected depends on the relation between the two materials.

For a reflected sound wave by a concrete wall, the first material is air and the second material is concrete. Then at 20 degrees Celsius the reflection coefficient is approximately 0.9999. This means, the reflected signal by a wall will barely lose any pressure after being reflected.

Oblique Incidence
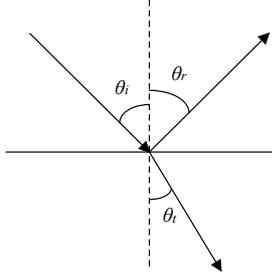
Figure 3.1: A sound wave hitting a material with an angle θ_i, is reflected with an angle θ_r and transmitted with an angle θ_t.

Having a wall at $x = 0$, the incident sound wave can hit the wall in an arbitrary angle. A schematic about the angles involved can be seen in Fig. 3.1. Then, if one defines the angle θ_i as the angle formed by the normal of the wall and the incident wave, the incident sound pressure and particle velocity are:

$$p_i(t, x) = \widehat{p}_i e^{-j(\omega t - k_1 x \cos(\theta_i) - k_1 y \sin(\theta_i))} \tag{3.13}$$

$$u_i(t, x) = \frac{\widehat{p}_i}{\rho_1 c_1} e^{j(\omega t - k_1 x \cos(\theta_i) - k_1 y \sin(\theta_i))} \tag{3.14}$$

which is reflected with an angle θ_r which is defined as the angle formed by the normal of the wall and the reflected wave:

$$p_r(t, x) = \widehat{p}_r e^{-j(\omega t + k_1 x \cos(\theta_r) - k_1 y \sin(\theta_r))} \tag{3.15}$$

$$u_r(t, x) = -\frac{\widehat{p}_r}{\rho_1 c_1} e^{j(\omega t + k_1 x \cos(\theta_r) - k_1 y \sin(\theta_r))} \tag{3.16}$$

If one defines θ_t as the angle at which the wave is transmitted to the new material, the transmitted sound pressure and particle velocity are:

$$p_t(t, x) = \widehat{p}_t e^{-j(\omega t - k_2 x \cos(\theta_t) - k_2 y \sin(\theta_t))} \tag{3.17}$$

$$u_t(t, x) = \frac{\widehat{p}_t}{\rho_2 c_2} e^{j(\omega t - k_2 x \cos(\theta_t) - k_2 y \sin(\theta_t))} \tag{3.18}$$

Applying the boundary conditions:

$$\widehat{p}_i e^{j(\omega t - k_1 y \cos(\theta_i))} + \widehat{p}_r e^{-j(\omega t - k_1 y \cos(\theta_r))} = \widehat{p}_t e^{-j(\omega t - k_2 y \cos(\theta_t))} \tag{3.19}$$

This leads to a well-known group of equations of the Snell's law:

$$\theta_i = \theta_r \tag{3.20}$$

$$c_2 \sin(\theta_i) = c_1 \sin(\theta_t) \tag{3.21}$$

Then, the incident angle is the same as the reflected one. This statement will be used during this thesis in order to model the reflected sound waves.

The reflection coefficient depends on the incident angle:

$$R = \frac{\frac{r_2}{\cos(\theta_t)} - \frac{r_1}{\cos(\theta_i)}}{\frac{r_2}{\cos(\theta_t)} + \frac{r_1}{\cos(\theta_i)}} \tag{3.22}$$

where

$$r_1 = \rho_1 c_1 \tag{3.23}$$

$$r_2 = \rho_2 c_2 \tag{3.24}$$

The transmission angle is:

$$\theta_t = \sqrt{1 - \left(\frac{c_2}{c_1}\right)^2 \sin^2(\theta_i)} \tag{3.25}$$

We focus on the case when $c_1 < c_2$, as the sound velocity inside a wall will be always faster than the sound velocity in air.

Looking at Eq. 3.25, one can realize that when $\sin(\theta_i)$ is above a certain value, the reflection coefficient is not real. The angle which sets the threshold is known as *critical angle* and can be calculated as:

$$\sin(\theta_c) = \frac{c_1}{c_2} \tag{3.26}$$

When the incident angle θ_i is above θ_c, the reflection coefficient is:

$$R = e^{j\phi} \tag{3.27}$$

where:

$$\phi = 2 \arctan\left(\frac{\rho_1}{\rho_2}\sqrt{\frac{\cos^2(\theta_c)}{\cos^2(\theta_i)} - 1}\right) \tag{3.28}$$

Then, the reflected sound wave will have the same sound pressure as the incident one.

Sound Velocity

During this thesis, the sound velocity c is assumed to be a known constant, which is independent on the traveled wave path. This is a simplification, as the velocity depends on the temperature of the medium and the properties of the medium itself. This relation can be expressed as:

$$c = \sqrt{\frac{K}{\rho}} \tag{3.29}$$

where K is the bulk modulus of elasticity, is measured in Pascals and depends on the elastic properties of the material.

Propagation in Air

In air, the sound velocity can be written as:

$$c = \sqrt{\frac{\gamma R T}{M}} \tag{3.30}$$

where γ is the adiabatic index, R is the molecular gas constant, T is the absolute temperature and M is the molecular mass of the gas.

Temperature

The most noticeable environmental factor which changes the sound velocity is the temperature [51]. In dry air, one can rewrite Eq. 3.30 as:

$$c = 331.3\sqrt{1 + \frac{\theta}{273.15}} \quad \text{m/s} \tag{3.31}$$

where θ is the temperature in degrees Celsius.

A room has rarely the same temperature everywhere. Warm air tends to rise to the highest parts of the room. Therefore, the sound velocity will change and the traveled trajectory by the sound waves will no longer be a straight line (see Eq. 3.21).

In the experiments performed in this thesis the sound velocity is assumed to be 343 m/s, which is the velocity at 20 degrees Celsius with dry air. This will be assumed to be part of the measurement noises together with the synchronization errors and the sensor noises.

Humidity

Using Eq. 3.30, one can rewrite γ as:

$$\gamma \approx \frac{1.005 + 1.82H}{1.005 + 1.82H - \frac{R}{M}} \tag{3.32}$$

where H is the humidity in kg water vapor per kg dry air. The molecular weight M also depends on the humidity.

The humidity has a limited influence in the sound velocity. The maximum difference between the highest and the lowest sound velocities is approximately 0.6 %.

Ray Acoustics

In order to simplify the equations, one can assume that the medium in which the sound waves propagate is homogeneous and that the magnitude of the irregularities of the reflectors are much smaller than the wavelength. In this case, the waves are assumed to propagate in a straight line

from speaker to microphone. Having a microphone $\mathbf{M}$ and a sender $\mathbf{S}$ in $\mathbb{R}^3$, if the emitter sends a signal at time t_s, it will be received at time T_r:

$$T_r = t_s + \frac{1}{c}\|\mathbf{M} - \mathbf{S}\| \tag{3.33}$$

Here c is assumed to be constant during the traveled sound wave path.

Image Source Model
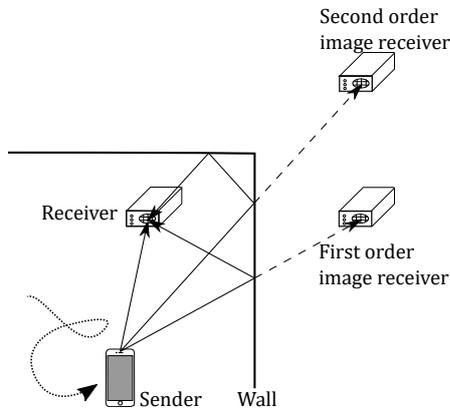
Figure 3.2: Acoustic reflections can be seen as virtual receivers. Reflections which are reflected once are first order image receivers and second reflections produce second order image receivers. [6]

The image source model [52] simplifies the reflections by considering a reflected signal by a wall acts like a virtual sender which is placed symmetrically at the same distance on the other side of the wall. The same can be applied for a receiver (see Fig. 3.2). Higher order reflections are also considered by this model. Then, a virtual sender can have also a virtual sender on another wall, which will be the second reflection. This is due to the fact that the incident angle must be equal to the reflected angle, as shown in Eq. 3.21.

A wall plane can be defined with a normal vector $\widehat{\mathbf{n}}$ and a point of the plane $\mathbf{P}$. Then, a reflection from a sender at position $\mathbf{S}$ will act like a sender $\mathbf{S}_v$ emitting at the same time:

$$\mathbf{S}_v = 2(-\widehat{\mathbf{n}}\|\mathbf{D}^T\widehat{\mathbf{n}}\|) + \mathbf{S} \tag{3.34}$$

$$\mathbf{D} = \mathbf{P} - \mathbf{S} \tag{3.35}$$

This model is only valid for planar surfaces and does not consider effects related to the limited dimensions of the walls. Nonetheless, this model is very effective for first and second reflections, which are the ones that can be mistaken as line-of-sight signals and the ones of interest in this thesis.

3.2 Time Difference of Arrival and Time of Arrival

A common scenario is when a speaker position is unknown and the receiver positions are manually measured and therefore assumed to be known in advance. Then, three variables need to be estimated in three dimensions. The available knowledge results in different kind of mathematical constraints. In this section we describe four scenarios:

- Time Difference of Arrival (TDOA): In this case the sending time is assumed to be unknown.
- Time of Arrival (TOA): The sending time is assumed to be known.
- Periodical Impulses: Here the sending time is unknown but the speaker emits signals at regular intervals which are known in advance.
- Frequency Difference of Arrival (FDOA): The frequency shifts created due to motion are used to estimate the position of the target.

Time Difference of Arrival

Time difference of arrival is a commonly used method for locating a sender $\mathbf{S}_k$ when the sending time is unknown. Having N receivers, one can subtract the timestamp of a reference receiver in order to eliminate the dependence on the sending time. In order to avoid estimating the sending time, the reception time $T_{k,r}$ of a reference receiver with index r is subtracted to the other timestamps:

$$T_{k,i} - T_{k,r} = \frac{1}{c}\|\mathbf{M}_i - \mathbf{S}_k\| - \frac{1}{c}\|\mathbf{M}_r - \mathbf{S}_k\|, \qquad i \in [1, ..., N] \mid i \neq r \tag{3.36}$$

where $\mathbf{M}_i$ is the position of the receiver with index i. The position of the sender $\mathbf{S}_k$ corresponds to its k-th emitted signal.

With two timestamps, the set of possible solutions for $\mathbf{S}_k$ results in an hyperbola in two dimensions and an hyperboloid in three dimensions. Then, having at least three receivers well distributed in a two-dimensional space, one can estimate the position of the target in two dimensions. In three dimensions, one requires at least four.

It is important to note that the noise of the reference receiver has a larger effect on the final estimation than the noise of the other receivers. Some approaches use the receiver which receives the first signal as reference, as it is more likely to be in line-of-sight. If all the receivers need to have the same importance, one can use a weighted least squares approach, which balances the noise with a weighting matrix. Another simple but computationally expensive solution is to use all receivers as reference.

Time of Arrival

Time of arrival (TOA), is a method used when the sending time t_s is known. Then, one knows the distance to the receiver:

$$c(T_{k,i} - t_s) = \|\mathbf{M}_i - \mathbf{S}_k\|, \qquad i \in [1, ..., N] \tag{3.37}$$

The set of possible solutions for $\mathbf{S}_k$ when having one timestamp is a circle in two dimensions and a sphere in three dimensions. Then, having at least three receivers in two dimensions or four in three dimensions, one can estimate the position of the target.

Periodical Impulses

In this case, one can assume the sending time is unknown, but the speaker emits signals at regular intervals I, which are known in advance. Then, when the target is at its k-th position $\mathbf{S}_k$, a timestamp is received:

$$T_{k,i} = \frac{1}{c}\|\mathbf{M}_i - \mathbf{S}_k\| + t_0 + kI, \qquad i \in [1, ..., N], \qquad k \in [1, ..., B] \tag{3.38}$$

where t_0 is the very first time at which the speaker emitted a signal. One can observe that, while this starts being a time difference of arrival problem, it becomes a time of arrival problem once the uncertainty estimating the variable t_0 is reduced.

If the target is emitting signals periodically, one can predict when the next signal will be emitted, and therefore gain more knowledge about the target position.

The number of elapsed intervals between two measurements is unknown. However, if the interval I is much larger than the propagation time, the number of elapsed intervals between two measurements $T_{k,i}$ and $T_{l,i}$ can be estimated straightforward by:

$$k - l = \left\lfloor \frac{T_{k,i} - T_{l,i}}{I} + \frac{1}{2} \right\rfloor \tag{3.39}$$

If the elapsed intervals are always subtracted, only one sending time needs to be estimated.

Frequency Difference of Arrival

Frequency Difference of Arrival (FDOA) uses the Doppler effect in order to track the target. The Doppler effect is defined as the frequency shift caused by the relative velocity between the receiver and the emitter. For an emitted signal with frequency f_0, the detected frequency is:

$$f = \frac{c + v_r}{c + v_s} f_0 \tag{3.40}$$

where v_r is the relative velocity of the receiver and v_s the relative velocity of the sender. In three dimensions, the relative velocities can be written as:

$$v_r = \mathbf{V}_r \cdot \frac{\mathbf{S} - \mathbf{M}_i}{\|\mathbf{S} - \mathbf{M}_i\|} \tag{3.41}$$

$$v_s = \mathbf{V}_s \cdot \frac{\mathbf{S} - \mathbf{M}_i}{\|\mathbf{S} - \mathbf{M}_i\|} \tag{3.42}$$

where $\mathbf{V}_r$ and $\mathbf{V}_s$ are the velocity vectors of the receiver and sender respectively. The operator $\cdot$ denotes the dot product. Usually FDOA is used when two anchor nodes are moving at known

velocities. Then, subtracting the frequency shift of a reference anchor node with the other nodes, one can estimate the position of the target.

While FDOA will not be used in this thesis, using periodical impulses one can also get a notion of the relative movement of the target. If the receivers are static:

$$c(T_{k,i} - T_{k-1,i} - I) = \|\mathbf{M}_i - \mathbf{S}_k\| - \|\mathbf{M}_i - \mathbf{S}_{k-1}\| \tag{3.43}$$

This difference is approximately equal to $I v_s$ if the target moves at constant velocity and the receiver is much farther than the distance moved by the sender. It is a discrete version of the Doppler shift. In Chapter 6 it will be used to estimate the movement of the target.

3.3 Description of the Systems

During this thesis, two localization systems are used to test the proposed algorithms:

- A **receiver localization system**. In this system the target is the receiver and the anchor nodes are the senders. The main benefit of this system is that there are unlimited number of targets, as every receiver can calculate its position independently. Moreover, in this system the speakers are designed to be inexpensive and to operate with a low energy consumption [4].

- A **speaker localization system**. This system consists of static receivers which are designed to receive signals from 18 kHz to 22 kHz, which is the maximum frequency range that a conventional speaker is able to emit with low attenuation [1] and it is also in the limit of the audible range. Therefore, a smartphone emitting inaudible sound signals can be located.

In addition, a high-priced localization system with sub-millimeter accuracy and precision is used as a reference for measuring the localization error.

Speaker Localization

This localization system is named ASSIST [1], which is the acronym for Acoustic Self-calibrating System for Indoor Smartphone Tracking. It consists of a smartphone and stationary receivers, which are placed in the room where the target needs to be located (see Fig. 3.3). The conventional speaker of a smartphone is used to emit acoustic signals above the human audible range. These signals are received by the static sensors and used to locate the smartphone.

The smartphone speaker emits periodical chirp signals from 18 to 22 kHz, which have been proven to be received at a distance of 30 m [1]. In order to analyze the maximum frequency limitation, several commercial off-the-shelf (COTS) smartphones were tested.

The emitted signal is a chirp in which the frequency increases linearly with time. The chirp impulse is defined between $0 \leq t \leq T$ with a start frequency of f_0 and an end frequency of f_1. It can be described according to the following equation:

$$s(t) = \sin\left(2\pi\left(f_0 + \frac{f_1 - f_0}{2T}t\right)t\right) \tag{3.44}$$

Figure 3.3: Sender localization system (ASSIST). The speaker of an out-of-the-box smartphone is used as a sender. [7]

The smartphone is not synchronized with the receivers. Only the receivers share a common time base.

The receivers use microphones constructed with a MEMS (microelectromechanical system) to sense the signals emitted by the smartphone. The digital signals are correlated with a reference signal which is the same that has been transmitted from the smartphone, in order to find the detection times. One could use the following equation in time domain to correlate two real signals of size T:

$$p_t(\tau) = \frac{1}{2T+1} \sum_{t=-T}^{T} r(t)s(t+\tau) \tag{3.45}$$

where $r(t)$ is the received signal.

However, it is computationally less intensive to work in the frequency domain, as shown in [53] by Valin et. al:

$$p_p(\tau) = \text{IFFT}\left\{\frac{R(f)S^*(f)}{|R(f) \circ S^*(f)|}\right\} \tag{3.46}$$

where IFFT$\{\cdot\}$ denotes the inverse of the fast Fourier transform. The operator $\circ$ is the Hadamard product (entry-wise product) and the absolute value is taken also from every pair of values. The complex conjugate operator is denoted by $()^*$. The values of $R(f)$ and $S(f)$ are calculated as follows:

$$R(f) = \text{FFT}\{r(t)\} \tag{3.47}$$
$$S(f) = \text{FFT}\{s(t)\} \tag{3.48}$$

where FFT$\{\cdot\}$ denotes the fast Fourier transform.

The highest correlation peaks of $p_p(\tau)$ are then used to estimate the reception times of the received signals. Then, these timestamps are used to estimate the position of the target.

This system is used in Chapter 5, Chapter 6 and Chapter 8.
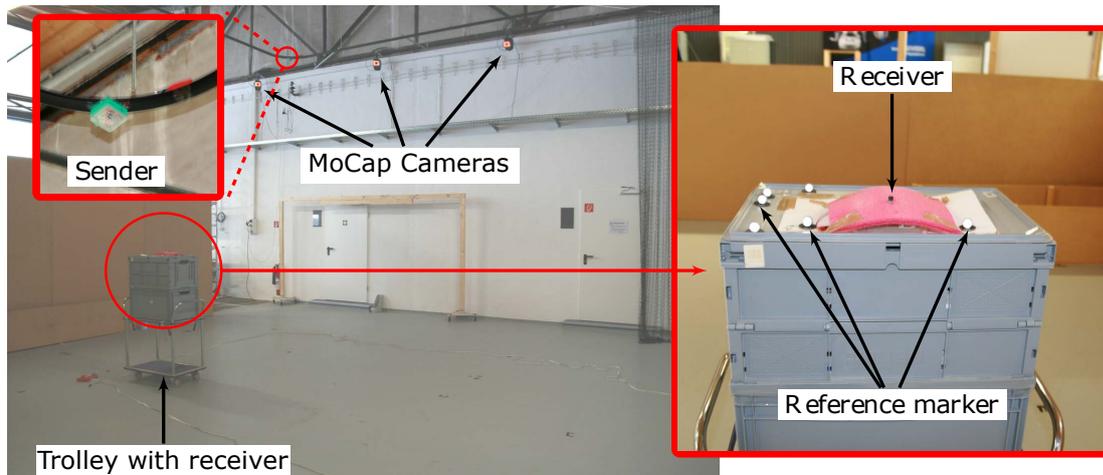
Figure 3.4: Receiver localization system (eCULTS). Multiple reference markers are used in order to locate the receiver with the motion capture system. [3]

Receiver Localization

This location system is part of the industrial project eCULTS, which is an acronym for energy-autarkic Calibration-free ULtrasound Tracking System. The project was funded by the German Federal Ministry of Education and Research (BMBF) and its main aim was to develop an indoor localization system for intralogistics which was energy-efficient.

The system [4] is designed to be easy to install and calibrate and is required to track multiple targets simultaneously with a high update rate. To achieve these purposes, the election was made to track moving microphones (see Fig. 3.4). By doing this, the number of targets which can be tracked simultaneously is unlimited. The anchor nodes were chosen to be inexpensive speakers which did not require synchronization, in order to reduce the energy requirements. Every target is then equipped with a microphone and is required to be capable of estimating its own position and calibrating the system.

Each speaker emits its signals with a different time period. Then, some senders emit signals more frequently than others, which helps distinguishing them. Additionally, the signals are modulated with a $\pi/4$-DQPSK (Differential Quadrature Phase Shift Keying) and two OFDM carrier frequencies and contain the identification number of the sender. The separate carrier frequencies are at 38.8 kHz and 40.74 kHz. Every packet emitted by a sender contains four symbols and therefore eight bits of information. Three of these bits are used for error correction. Therefore, there are 32 possible sender ids. The details of the modulation process are out of the scope of this dissertation. One can refer to the dissertation of Alexander Traub-Ens [54].

The algorithms used on this system will be explained in Chapter 4 and Chapter 7.

Reference System

The reference system used in this thesis is a motion capture (MoCap) system with sub-millimeter accuracy and precision (the standard deviation while standing still is 0.3 mm). This system is a high-priced localization system that uses several cameras to track the positions of reflective markers (see Fig. 3.4). The experiments in this thesis took place in a reduced area due to the limited range of the motion capture system. More details about the experimental environment can be seen in the Appendix of this thesis.

Error Calculation

In order to synchronize the result of the MoCap and the acoustic systems, both systems share a common time base. The positions of the anchor nodes are measured with a total station theodolite.

If the positions of the anchor nodes are known, these positions are rotated and translated according to the MoCap measurements. If the positions of the anchor nodes are unknown and only the sender positions are estimated (Chapter 6), then the estimated positions are translated and rotated using the MoCap positions which are closer in time. If the positions of the anchor nodes are estimated, then they are rotated and translated using the measured positions by a theodolite as a reference. This is because some anchor nodes cannot be measured by MoCap, as they are out of range.

In order to align two sets of positions we use the approach in [55]. Assuming each of the sets has n points, we first estimate the mean:

$$\boldsymbol{\mu}_g = \frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \qquad \boldsymbol{\mu}_h = \frac{1}{n} \sum_{i=1}^{n} \mathbf{h}_i \tag{3.49}$$

where $\mathbf{g}_i$ is the i-th reference point and $\mathbf{h}_i$ is the i-th estimated point.

Afterwards we estimate the covariance between both set of points:

$$\mathbf{W} = \sum_{i=1}^{n} (\mathbf{g}_i - \boldsymbol{\mu}_g)(\mathbf{h}_i - \boldsymbol{\mu}_h)^T \tag{3.50}$$

By subtracting the mean from both sets of points, one only needs to estimate the rotation. If both set of points were equal, $\mathbf{W}$ would be a diagonal matrix. Therefore, one can do the singular value decomposition (SVD) of $\mathbf{W}$:

$$\mathbf{W} = \mathbf{U}\mathbf{D}\mathbf{V}^T \tag{3.51}$$

Then, $\mathbf{D}$ is a diagonal matrix which contains the singular values. The matrices $\mathbf{U}$ and $\mathbf{V}$ contain the singular vectors.

If one wants to find the rotated and translated version of $\mathbf{h}_i$, one can calculate it as:

$$\overline{\mathbf{h}}_i = \mathbf{U}\mathbf{V}^T(\mathbf{h}_i - \boldsymbol{\mu}_h) + \boldsymbol{\mu}_g \tag{3.52}$$

Then the error is estimated by calculating the distance between the rotated and translated estimated position $\overline{\mathbf{h}}_i$ and the reference position $\mathbf{g}_i$.

4 Probabilistic Localization

When locating a target one can make certain assumptions about its motion. Its maximum velocity is often limited and it can spend large periods of time with the same velocity (e.g. a robot moving or a person walking). This means, one can make use of previous estimations to predict where is currently most likely to be the target. In order to do so, one needs to be able to update the target position fast enough so that the region of likely positions becomes as small as possible.

During this thesis, probabilistic algorithms are used for different purposes, always assuming a target position depends on its previous one. These algorithms fulfill the Bayesian filtering scheme, which is a probabilistic approach to recursive state estimation based on the Markov assumption, i.e. the assumption that the current state depends only on the previous state, not on the previous trajectory.

In this chapter, the particle filter, the unscented Kalman filter and the extended Kalman filter are explained for better understanding of the following chapters. In addition, we show how probabilistic methods can be used for locating a moving receiver, using the eCULTS system. While the number of constraints is lower than the number of variables when the receiver is moving, one can use a probabilistic motion model to track the movement of the target.

During this chapter, we formulate the sensor and motion models when the measurements are assumed to be in line-of-sight and with Gaussian distributed noise. Further improvements for mixed line-of-sight and non-line-of-sight measurements will be introduced in the following chapters.

4.1 Recursive Bayesian Estimation

The recursive Bayesian estimation model assumes that there is a hidden Markov process, which contains the variables that need to be estimated. Each of these states leads to a measurement vector. Then, at the time instant t, there is a hidden state $\mathbf{x}_t$, which contains the variables that need to be estimated and a measured vector $\mathbf{z}_t$.

These dynamic systems are assumed to be governed by two models:

- The measurement model $\mathbf{z}_t = h(\mathbf{x}_t)$, which relates the state $\mathbf{x}_t$ and the measurement $\mathbf{z}_t$.
- The motion model, which predicts a state given a previous state $\mathbf{x}_t = f(\mathbf{x}_{t-1})$

These algorithms intent to find the probability distribution of the a posterior probability $p(\mathbf{x}_t|\mathbf{z}_{1:t})$:

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) = \frac{p(\mathbf{z}_{1:t}|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{z}_{1:t-1})}{p(\mathbf{z}_t|\mathbf{z}_{1:t-1})} \tag{4.1}$$

The value of $p(\mathbf{z}_t|\mathbf{z}_{1:t-1})$ is modeled as a constant. The other two terms are:

- The term $p(\mathbf{x}_t|\mathbf{z}_{1:t-1})$ corresponds to the **prediction step**. This step uses the known motion model to predict the next state given the previous one:

$$p(\mathbf{x}_t|\mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})\,d\mathbf{x}_{t-1} \tag{4.2}$$

- The term $p(\mathbf{z}_{1:t}|\mathbf{x}_t)$ is the likelihood of the state $\mathbf{x}_t$ given the measurements $\mathbf{z}_{1:t}$, this corresponds to the **correction step** and uses the measurement model to transform the variables from the state space to the measurement space.

Kalman Filter

The Kalman filter [56] is a simple solution for Bayesian recursive state estimation, when the motion and measurement model which govern the system are linear and have additive Gaussian noise. The motion and measurement model are described as follows:

$$\mathbf{z}_t = \mathbf{H}\mathbf{x}_t + \mathbf{w}_t \tag{4.3}$$
$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{u}_t + \mathbf{v}_t \tag{4.4}$$
$$\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t) \tag{4.5}$$
$$\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t) \tag{4.6}$$

The variable $\mathbf{u}_t$ is a control vector. The matrices $\mathbf{A}$ and $\mathbf{B}$ depend on the chosen motion model. The matrix $\mathbf{H}$ depends on the sensor model.

The Kalman filter finds the optimal state when the variances $\mathbf{Q}_t$ and $\mathbf{R}_t$ are perfectly known, the noise is white and the models perfectly match the real behavior of the system.

During the prediction step, a new state mean $\widehat{\mathbf{x}}_t$ and covariance $\widehat{\mathbf{P}}_t$ are predicted using the motion model:

$$\widehat{\mathbf{x}}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{u}_t$$
$$\widehat{\mathbf{P}}_t = \mathbf{A}\mathbf{P}_{t-1}\mathbf{A}^T + \mathbf{Q}_t$$

During the correction step, the measurement vector is used to correct the predictions and find the optimal state mean $\mathbf{x}_t$ and variance $\mathbf{P}_t$.

The covariance of the predicted measurement can be estimated as:

$$\mathbf{S}_t = \mathbf{H}\widehat{\mathbf{P}}_t\mathbf{H}^T + \mathbf{R}_t \tag{4.7}$$

The Kalman filter minimizes the mean-squared error. In order to do so, it uses a so-called Kalman gain matrix $\mathbf{K}_t$, which finds the optimal weight of the measurements:

$$\mathbf{K}_t = \widehat{\mathbf{P}}_t\mathbf{H}^T\mathbf{S}_t^{-1} \tag{4.8}$$

Then, the state at time t and its covariance are:

$$\mathbf{x}_t = \widehat{\mathbf{x}}_t + \mathbf{K} \left(\mathbf{z}_t - \mathbf{H}\widehat{\mathbf{x}} \right)$$
$$\mathbf{P}_t = \left(\mathbf{I} - \mathbf{K}_t\mathbf{H} \right) \widehat{\mathbf{P}}_t$$

The main advantages of the Kalman filter are its simplicity and its low computational power requirements. These factors make it suitable for scenarios where the number of measurements is elevated or where there are multiple hypotheses which need to be examined in a limited amount of time. These situations will be further explained in the following chapters.

Extended Kalman Filter

The extended Kalman filter is a modification of the Kalman filter for non-linear systems. When the functions $h(\mathbf{x}_t)$ and $f(\mathbf{x}_{t-1})$ are not linear, they are substituted by their Jacobian matrix. Then, the matrices $\mathbf{H}$ and $\mathbf{A}$ of the linear Kalman filter are substituted by the partial derivatives of the non linear functions, which are denoted as $\mathbf{H}_t$ and $\mathbf{A}_t$. This linearizes the functions around the evaluation points. The main drawbacks of this approach are that it can diverge when the initial estimations are far from the actual real value and that the Jacobian matrices need to be recomputed every time the sensor model or motion model are changed.

Unscented Kalman Filter

The unscented Kalman filter [57] uses the so-called *sigma points* for linearization. These points are spread around the mean depending on the variance of the estimations and evaluated using the nonlinear functions. Then, it provides a better representation of the non-linearities than the extended Kalman filter. For our implementation we follow the description in [58].

The previous state mean $\mathbf{x}_{t-1}$ and variance $\mathbf{P}_{t-1}$ are used to generate a set of $2L$ sigma points, where L is the dimension of the state vector:

$$\boldsymbol{\eta}_{t-1}^x = \begin{bmatrix} \mu_{t-1}^T \\ \left(\mathbf{x}_{t-1} + \sqrt{(L+\lambda)\mathbf{P}_{t-1,[1]}} \right)^T \\ \vdots \\ \left(\mathbf{x}_{t-1} + \sqrt{(L+\lambda)\mathbf{P}_{t-1,[L]}} \right)^T \\ \left(\mathbf{x}_{t-1} - \sqrt{(L+\lambda)\mathbf{P}_{t-1,[1]}} \right)^T \\ \vdots \\ \left(\mathbf{x}_{t-1} - \sqrt{(L+\lambda)\mathbf{P}_{t-1,[L]}} \right)^T \end{bmatrix}^T \tag{4.9}$$

where $\mathbf{P}_{\bullet,[\ell]}$ denotes the ℓ-th column of $\mathbf{P}_\bullet$. The scaling parameter

$$\lambda = \alpha^2(L+\rho) - L \tag{4.10}$$

determines how far the sigma points are from the mean, where α and ρ are tuning parameters of the filter.

The sigma points matrix has a dimension of $L \times (2L+1)$. Each row has $2L+1$ sigma points. The sigma points of the previous state and the process noise are passed to a function g, which is the motion model, to predict the new state:

$$\boldsymbol{\eta}_t^x = g(\mathbf{u}_t, \boldsymbol{\eta}_{t-1}^x) \tag{4.11}$$

After that, the Gaussian statistics of the new points are computed by

$$\widehat{\mathbf{x}}_t = \sum_{l=0}^{2L} w_m(l) \boldsymbol{\eta}_{l,t}^x \tag{4.12}$$

$$\widehat{\mathbf{P}}_t = \sum_{l=0}^{2L} w_c(l) \left(\boldsymbol{\eta}_{l,t}^x - \widehat{\mathbf{x}}_t\right) \left(\boldsymbol{\eta}_{l,t}^x - \widehat{\mathbf{x}}_t\right)^T + \mathbf{Q}_t \tag{4.13}$$

where

$$w_m(0) = \frac{\lambda}{L+\lambda}$$

$$w_c(0) = \frac{\lambda}{L+\lambda} + (1 - \alpha^2 + \beta)$$

$$w_m(l) = w_c(l) = \frac{1}{2(L+\lambda)} \qquad (1 \leq l \leq 2L)$$

The parameter β is a tuning parameter of the filter which has to be set depending on the a priori knowledge of the state probability distribution.

In the correction step, the Gaussian statistics of the observations at the sigma points and the predicted state are calculated to correct the measurement. Using the nonlinear sensor model $h(\mathbf{x}_t)$:

$$\overline{\mathbf{z}}_t = \sum_{l=0}^{2L} w_m(l) \, h(\boldsymbol{\eta}_{l,t}^x)$$

$$\mathbf{S}_t = \sum_{l=0}^{2L} w_c(l) \left(h(\boldsymbol{\eta}_{l,t}^x) - \overline{\mathbf{z}}_t\right) \left(h(\boldsymbol{\eta}_{l,t}^x) - \overline{\mathbf{z}}_t\right)^T + \mathbf{R}_t \tag{4.14}$$

In this step also the cross-covariance between the predicted state and the predicted measurement is calculated:

$$\boldsymbol{\Sigma}^{x,z} = \sum_{l=0}^{2L} w_c(l) \left(\boldsymbol{\eta}_{l,t}^x - \widehat{\mathbf{x}}_t\right) \left(h(\boldsymbol{\eta}_{l,t}^x) - \overline{\mathbf{z}}_t\right)^T \tag{4.15}$$

Finally, the mean and the covariance of the state are updated by

$$\mathbf{K}_t = \boldsymbol{\Sigma}^{x,z} \mathbf{S}_t^{-1} \tag{4.16}$$

$$\mathbf{x}_t = \widehat{\mathbf{x}}_t + \mathbf{K}_t(\mathbf{z}_t - \overline{\mathbf{z}}_t) \tag{4.17}$$

$$\mathbf{P}_t = \widehat{\mathbf{P}}_t - \mathbf{K}_t \mathbf{S}_t \mathbf{K}_t^T \tag{4.18}$$

Particle Filter

The particle filter, also known as Monte Carlo localization [59] uses a set of n_p particles to represent the probability distribution $p(\mathbf{x}_t \mid \mathbf{z}_{1:t}, \mathbf{u}_{1:t})$. This is the probability of a certain state given the measurements and control vectors up to the time instant t. Each particle $(\mathbf{x}_t^{[h]}, w_t^{[h]}), (1 \le h \le n_p)$ represents an hypothesis and has an importance weight associated to it $w_t^{[h]}$ and a state vector $\mathbf{x}_t^{[h]}$ which contains an hypothesis for the state variables.

In the prediction step, every particle predicts its next state by using the motion model. Having a state $\mathbf{x}_{t-1}^{[h]}$ and a control variable $\mathbf{u}_t$, the next state is predicted using the assumed probability distribution of the motion model $p(\mathbf{x}_t^{[h]} \mid \mathbf{u}_t, \mathbf{x}_{t-1}^{[h]})$.

In the correction step, the measurement $\mathbf{z}_t$ is used to update the weight, $w_t^{[h]} \propto w_{t-1}^{[h]} p(\mathbf{z}_t \mid \mathbf{x}_t^{[h]})$, of each particle. Therefore the likelihood of the state hypothesis is computed using the sensor model and $\mathbf{z}_t$.

In the resampling step new particles are created and old particles are eliminated. A total of n_p particles are drawn. Particles with low weights are less likely to disappear than particles with higher weights, which propagate into new particles.

In our implementation, the resampling step is executed if the effective number of particles n_{eff}, is smaller than the number of $\frac{n_p}{2}$ particles, as shown in [60], where

$$n_{\text{eff}} = \left(\sum_{h=1}^{n_p} (w_t^{[h]})^2 \right)^{-1} \tag{4.19}$$

4.2 Application Case: Moving Receiver

As stated in the previous chapter, locating a receiver has multiple benefits. The number of targets that can be located simultaneously is unlimited, due to the fact that the receivers do not interfere with each other. In addition, every receiver can estimate its own position without requiring a central unit. This increases the privacy of the users. The main drawback is that when a receiver moves, every signal from a static sender is received at a different position (see Fig. 4.1).

With only one timestamp per position and no further knowledge one cannot estimate the position of the target, since the system of equations is undetermined. However, one can use probabilistic methods in order to predict the target movement and correct the predictions with the received measurements.

Probabilistic methods can be used to estimate the variables depending on the measurements and the previous estimation. Nonetheless, the variables need to be initialized when their uncertainty is still too large. Especially when the system starts, as the position of the target and the time offsets between senders are unknown. Therefore, we propose a *calibration phase* [4], which detects when the target is standing at different positions and uses the timestamps received during that time to estimate the variables of the system. By doing this, the system of equations is overdetermined.

After the variables have been initialized, our implementation estimates the changes of such variables over the time in the *tracking phase* [4, 61]. We study how the particle filter and the unscented Kalman filter (UKF) perform in this phase.
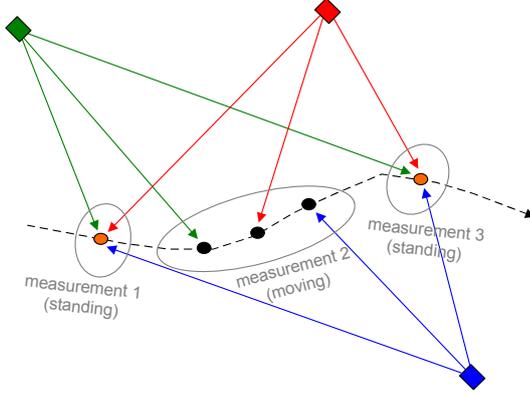
Figure 4.1: Schematic of the under-determined equation system. If the receiver moves continuously in a two-dimensional space, for every new measurement there is two new variables only for its position. Three if it moves in three dimensions. [4]

Problem Formulation

A receiver moves continuously in a three-dimensional space. The localization scenario consists of B stationary senders which are placed at positions $\mathbf{S}_j$ $(1 \leq j \leq B)$ in a three-dimensional Euclidean space. Every sender emits discrete signals at regular points in time at a fixed interval I_j. The interval may differ from sender to sender. The sending time of the k_j-th signal at sender position $\mathbf{S}_j$ is then described by

$$t_{k_j j} = t_{0j} + k_j I_j \ , \qquad (k_j > 0) \tag{4.20}$$

The senders are assumed to be unsynchronized, i.e the intervals I_j and the initial send time t_{0j} varies from sender to sender. Consequently, there is an unknown time offset which relates to the send time between the senders y and j :

$$\delta_{yj} = t_{0,y} - t_{0,j} = (t_{k_y y} - k_y I_y) - (t_{k_j j} - k_j I_j) \tag{4.21}$$

Since the offsets are transitive, only $B - 1$ offsets need to be estimated relative to one sender.

Considering the case where the receiver is continuously moving, signals are received at different positions. This results in the following hyperbolic equation in which two signals, originating from two different senders $\mathbf{S}_y$ and $\mathbf{S}_j$, are received at the positions $\mathbf{M}_{k_y y}$ and $\mathbf{M}_{k_j j}$:

$$\frac{1}{c}(\|\mathbf{M}_{k_y y} - \mathbf{S}_y\| - \|\mathbf{M}_{k_j j} - \mathbf{S}_j\|) = \Delta t_{yj} + \delta_{yj} \tag{4.22}$$

where Δt_{yj} represents the unsynchronized time difference of arrival of the two signals originated by $\mathbf{S}_y$ and $\mathbf{S}_j$, which may be calculated based on the reception times and the intervals as

$$\Delta t_{ij} = (T_{k_y y} - T_{k_j j}) - (k_y I_y - k_j I_j) \tag{4.23}$$

where $T_{k_y y}$ and $T_{k_j j}$ are the reception times.

The length of the intervals I_y, I_j can be easily computed by receiving two or more successive signals $k_1, k_2, k_1 \neq k_2$, emitted by the same sender while it is temporarily stationary:

$$I_y = \frac{1}{k_1 - k_2}(T_{k_1 y} - T_{k_2 y}) \tag{4.24}$$

$$I_j = \frac{1}{k_1 - k_2}(T_{k_1 j} - T_{k_2 j}) \tag{4.25}$$

We also assume the height of the target is known in advance. Assuming the sender positions and the intervals are also known, still exists $2n + (B-1)$ unknown variables after n received signals. Consequently, the equation system is under-determined and cannot be solved without further information or assumptions on the scenario. Therefore, it is required either to know the initial values of the variables and model their changes (*tracking phase*) or to make special assumptions on the scenario (*calibration phase*).

Stop-and-go Motion

When recursive state estimation algorithms lack information and are not capable of tracking the position of a moving receiver, we assume it stops in N_s different positions $\mathbf{H}_i$, then we have time to receive B_r signals per receiver position (*stop-and-go motion*). Doing this, it is only required to estimate one receiver position for every B_r received signals, which reduces notably the uncertainty and makes possible an uniquely determined system of equations (cf. Fig. 4.1).

Being $\mathbf{H}_i$ the receiver position when the u-th signal from the sender one is received and $\mathbf{H}_p$ the receiver position when the v-th signal from the sender j is received, we obtain a system of hyperbolic equations of the form:

$$f_{u,v} = \|\mathbf{H}_i - \mathbf{S}_1\| - \|\mathbf{H}_p - \mathbf{S}_j\| - c(T_{u,k_1} - T_{v,k_j}) + \Delta t_{1j} \tag{4.26}$$

where $2 \leq j \leq B$ and $1 \leq p, i \leq N_s$. The unsynchronized time difference of arrival between two signals, originated by the sender 1 and the sender j, is represented with Δt_{1j}:

$$\Delta t_{1j} = c(k_1 I_1 - k_j I_j + \delta_j) \,. \tag{4.27}$$

If one signal is received per sender, the system of equations has $N_s B$ independent equations, which has to be higher or equal than the number of variables:

$$N_s B \geq \underbrace{2N_s}_{\text{Receiver}} + \underbrace{B}_{\text{Offsets}} \tag{4.28}$$

which means the system of equations can be solved if the number of standing still positions N_s is higher or equal than:

$$N_s \geq \frac{B}{B-2} \tag{4.29}$$

Then, the minimum number of static receiver positions is two.

One can detect that the receiver is not moving by ensuring that the time difference between two measurements of the same sender is a multiple of the interval plus a certain error due to the measurement noise.

$$\rho_{j,t} = |(T_{k_j j} - T_{(k-a)_j j}) - aI_j| \tag{4.30}$$

If the receiver is standing during a intervals $\rho_{j,t}$ will be close to zero for a high number of senders j and received intervals.

Calibration Phase

Assuming the *stop-and-go* motion and having a number of standing positions and senders fulfilling Eq. 4.29, the system of hyperbolic equations can be solved with local optimization algorithms. We use both the gradient descent and the Gauss-Newton method, the two are first-order methods that use the derivative of the system of hyperbolic error equations.

Once the timestamps corresponding with the time when the receiver is standing are selected, one must find the variables which minimize the error function in Eq. 4.26. For simplicity and better understanding we assume that there are G selected signals from every sender. Then, the Eq. 4.26 results in a quadratic objective which can be formulated as follows:

$$\sum_{u=1}^{G} \sum_{v=1}^{(B-1)G} \underset{\mathbf{H}_{1:N_s}, \delta_{2:B}}{\arg\min} (f_{u,v})^2 . \tag{4.31}$$

which in vector notation is proportional to $w = \frac{1}{2}\mathbf{b}^T\mathbf{b}$ with $\mathbf{b} = (f_{1,1}, ..., f_{G,(B-1)G})^T$.

We calculate the direction of the steepest ascent:

$$\nabla w = \nabla\left(\frac{1}{2}\mathbf{b}^T\mathbf{b}\right) = \mathbf{D}^T\mathbf{b} \tag{4.32}$$

where $\mathbf{Q}$ is the Jacobian matrix:

$$\mathbf{D} = \begin{bmatrix} \dfrac{\partial f_{1,1}}{\partial \mathbf{H}_1} & \cdots & \dfrac{\partial f_{G,(B-1)G}}{\partial \mathbf{H}_1} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_{1,1}}{\partial \mathbf{H}_q} & \cdots & \dfrac{\partial f_{G,(B-1)G}}{\partial \mathbf{H}_q} \\ \dfrac{1}{c}\dfrac{\partial f_{1,1}}{\partial \delta_2} & \cdots & \dfrac{1}{c}\dfrac{\partial f_{G,(B-1)G}}{\partial \delta_2} \\ \vdots & \ddots & \vdots \\ \dfrac{1}{c}\dfrac{\partial f_{1,1}}{\partial \delta_B} & \cdots & \dfrac{1}{c}\dfrac{\partial f_{G,(B-1)G}}{\partial \delta_B} \end{bmatrix}^T \tag{4.33}$$

The partial derivative with respect to a vector is defined as the derivative with respect to each of its components:

$$\frac{\partial f_{u,v}}{\partial \mathbf{H}_i} = \left(\frac{\partial f_{u,v}}{\partial H_{i,x}}, \frac{\partial f_{u,v}}{\partial H_{i,y}} \right)^T \tag{4.34}$$

The partial derivative with respect to δ_j can be calculated as follows:

$$\frac{\partial f_{u,v}}{\partial \delta_j} = c \tag{4.35}$$

When the compared timestamps correspond to the same receiver position ($\mathbf{H}_i = \mathbf{H}_p$) the partial derivative with respect to the receiver position is:

$$\frac{\partial f_{u,v}}{\partial \mathbf{H}_i} = \frac{\partial f_{u,v}}{\partial \mathbf{H}_p} = \frac{\mathbf{H}_i - \mathbf{S}_1}{\|\mathbf{H}_i - \mathbf{S}_1\|} - \frac{\mathbf{H}_p - \mathbf{S}_j}{\|\mathbf{H}_p - \mathbf{S}_j\|} \tag{4.36}$$

In all other cases the partial derivatives are:

$$\frac{\partial f_{u,v}}{\partial \mathbf{H}_i} = \frac{\mathbf{H}_i - \mathbf{S}_1}{\|\mathbf{H}_i - \mathbf{S}_1\|} \tag{4.37}$$

$$\frac{\partial f_{u,v}}{\partial \mathbf{H}_p} = -\frac{\mathbf{H}_p - \mathbf{S}_j}{\|\mathbf{H}_p - \mathbf{S}_j\|} \tag{4.38}$$

$$\tag{4.39}$$

All the variables which need to be estimated are components of the state vector $\mathbf{m}$:

$$\mathbf{m} = (\mathbf{H}_1^T, ..., \mathbf{H}_q^T, c\delta_2^T, ..., c\delta_B^T)^T \tag{4.40}$$

Every iteration the state vector is updated using $\mathbf{D}$ and $\mathbf{b}$. The methods used are:

The gradient descent method: In every iteration step l the Gradient Descent method updates the state vector in direction of the steepest descent. The adaptive factor γ sets the step width.

$$\hat{\mathbf{m}} = \gamma \nabla w = \gamma \mathbf{D}^T \mathbf{b} \tag{4.41}$$

$$\mathbf{m}^{l+1} = \mathbf{m}^l - \hat{\mathbf{m}} \tag{4.42}$$

$$\tag{4.43}$$

The Gauss-Newton algorithm: Instead of relying on an adaptive factor γ it calculates the step size using the inverse $(\mathbf{Q}^T\mathbf{Q})^{-1}$ for every iteration:

$$\mathbf{m} = (\mathbf{D}^T\mathbf{D})^{-1}(\mathbf{D}^T\mathbf{b}) \tag{4.44}$$

The Gauss-Newton Algorithm is faster, nevertheless it is very prone to divergence when applied to initial positions which are far from the true values. However, it can be used when the Gradient Descent error function has become steady to reduce notably the number of iterations.

Tracking Phase

During the tracking phase we use recursive state estimation algorithms. We test the performance of a particle filter and an unscented Kalman filter.

In our case the state vector contains the position of the receiver $\mathbf{M}_t$ and the receiver velocity $\mathbf{V}_t$. Moreover, in order to estimate the reception time, the offsets relative to one sender $(\delta_{12}, ..., \delta_{1B})$ and its sending time t_{k_11} are also estimated. Without loss of generality is defined $\delta_j = \delta_{1j}$ where $\delta_1 = 0$. In conclusion, the state vector is formulated as follows:

$$\mathbf{x}_t = \left(\mathbf{M}_t^T, \mathbf{V}_t^T, t_{k_11}, \delta_2, ..., \delta_B\right)^T \tag{4.45}$$

Sensor Model

The measurement is estimated by the following sensor model, which relates the predicted measurement $\widehat{z_{k_j}}$ and the state vector:

$$\widehat{z_{k_j}} = \frac{1}{c}\|\mathbf{M}_t - \mathbf{S}_j\| + (t_{k_11} + \delta_j) \tag{4.46}$$

In the UKF, this is the nonlinear function $h(\mathbf{x}_t)$ in Eq. 4.14.

In the particle filter, this sensor model uses the measurement z_j to compute the probability that the observed measurement matches the current belief. Based on a known sender positions $\mathbf{S}_j$ and the estimated values $\mathbf{M}_t$, t_{k_11}, and δ_j, a hypothesis of the observation is calculated by

$$d_{tj} = \frac{1}{c}\|\mathbf{M}_t - \mathbf{S}_j\| + (t_{k_11} + \delta_j) \tag{4.47}$$

Using this hypothesis d_{tj} the likelihood of a measurement is calculated by

$$p(z_j \mid \mathbf{x}_t) = \mathcal{N}\left(z_j, d_{tj}, \sigma_{\text{sensor}}^2\right) \tag{4.48}$$

where σ_{sensor}^2 is the variance of the sensor noise.

Motion Model

As explained before, the motion model predicts how a state will change over the time. In this case, the position is predicted according to the predicted velocity of the target. After the system has been initialized, the slight changes of the time offsets δ_{ij} need to be estimated as well. This changes appear due to the clock drift of the senders.

In order to calculate the offset between $\mathbf{S}_i$ and $\mathbf{S}_j$ with respect to the sending times t_{0i}, t_{0j} we use the equation

$$\delta_{ij} = (k_i I_i + t_{0i}) - (k_j I_j + t_{0j}) + \zeta \tag{4.49}$$

the variable ζ depends on the assumed noise of such prediction. In the particle filter, we use a uniform distribution

$$\zeta \sim \mathcal{U}(-t_{\text{dist}}, t_{\text{dist}}) \tag{4.50}$$

The unscented Kalman filter only allows Gaussian distributions. Therefore, in this case ζ is a normal distribution with variance σ_ζ^2.

As the offsets are transitive, we settle to estimate $m - 1$ offsets. Without loss of generality we estimate only the offset relative to sender $j = 1$ and define $\delta_i = \delta_{1i}$, where $\delta_1 = 0$.

Also the sending time of one sender must be estimated, as in TDOA only relative distances are measured. To propagate the estimated sending time of the latest received signal of sender i we use the latest estimated sending time and the interval length:

$$t_{k_i i} = t_{(k_i-1)i} + I_i + \nu \tag{4.51}$$

In the particle filter ν is defined as a uniformly distributed noise:

$$\nu \sim \mathcal{U}(-t_{\text{dist}}, t_{\text{dist}}) \tag{4.52}$$

In the UKF it is a normal distribution with variance σ_ν^2. As the send times $t_{k_i i}$ are all relative we estimate only $t_{k_1 1}$.

The reason uniform distributions are used is because after experimentation, in the eCULTS system one can observe that the time shifts are more similar to a uniform distribution than a normal distribution. The error is limited but it is spaced almost equally around the mean.

The control command represents just the time which has passed since the last computation. For the movement, and therefore the next estimated position of the receiver, we use a constant velocity model. This model assumes that the receiver moves with constant velocity, while changes in the velocity are undetermined, which is modeled by Gaussian noise with a covariance matrix Σ_V. In this model, the position and velocity of the receiver are updated according to the following equations:

$$\begin{aligned} \mathbf{M}_{t+1} &= \mathbf{M}_t + h_t \mathbf{V}_t \\ \mathbf{V}_{t+1} &= \mathbf{V}_t + \zeta_t \qquad \zeta_t \sim \mathcal{N}(0, \Sigma_V) \end{aligned} \tag{4.53}$$

where $h_t = T_{k_i i} - T_{k_j j} > 0$ is the elapsed time between two estimations. In this case, the current reception time from sender i and the previous signal, which was emitted by sender j.

4.3 Experimental Results

In order to check the feasibility of the proposed application scenario and analyze the error achieved we do a real experiment with the motion capture system as reference. The ultrasound system is the eCULTS system. For this experiment, four static senders are used. One sender has a height of 2 m and the other three senders have a height of 3.4 m.

In order to find the offset between the senders, we stand still in five different positions. The local optimization algorithms are constantly updated in order to simulate a real-time system and analyze the number of required timestamps to initialize the system. Fig. 4.2 shows the median error of the estimated standing positions for every received timestamp. The error is reduced from 2.286 m with two standing positions to 0.041 m with three standing positions.

Having more than three standing positions reduces only slightly the median error (0.028 m with four and 0.030 m with five). Therefore, two standing positions would not be sufficient for initialization. This is due to the fact that with two standing positions there are the same number of constraints than variables. This increases the error, as the result depends highly on the assumed height of the target. Moreover, the geometry of the standing positions also plays a role. With three positions one can have more noticeable distance changes in all directions, which reduces the error. This is related with the dilution of precision that will be discussed in Chapter 8.
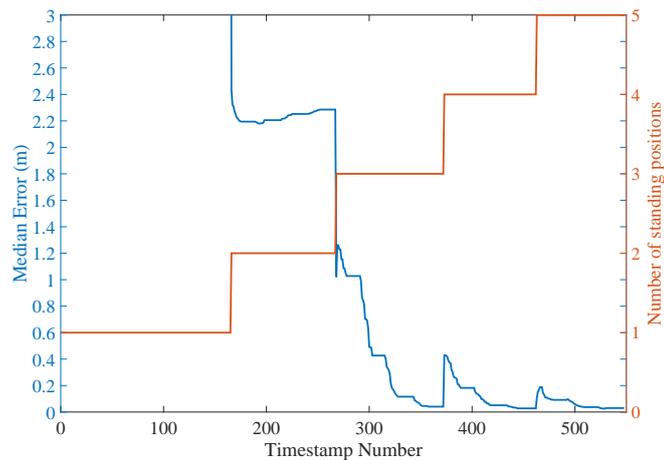
Figure 4.2: The local optimization algorithm estimates the positions where the receiver is standing H_1, H_2, H_3, H_4 and H_5. This graph shows the median error between the estimated and the real positions. Note that every time that there is a new standing position, the error is suddenly increased due to the fact that there is a reduced number of timestamps for that position, which increases the effect of the noise. Subsequently, as the number of timestamps increases the error is reduced again. Median errors above 3 m are not shown in this figure. [3]

Once the offsets have been estimated, one can track the moving target. The receiver is placed on a trolley and moved by a person with an arbitrary trajectory (see Fig.4.3), stopping only at the start and end position. In this experiment the target moves at low velocity (maximum velocity 0.6 m/s). At this velocity, the position only changes slightly between two consecutive measurements. When the target moves faster, the data association becomes challenging. This will be studied in Chapter 7. We use the unscented Kalman filter and the particle filter in order to compare their performance. The particle filter uses 10 000 particles. In order to have a proper comparison of both algorithms we evaluate the error for different values of assumed process noise and measurement noise. The result can be seen in Fig. 4.4 for the UKF and Fig. 4.5 for the particle filter. Note that while underestimating the measurement noise drastically increases the error in the particle filter, this fact has less effect in the UKF. On the other hand, overestimating the process noise variance, has less effect in the particle filter than in the UKF. Choosing the parameters which minimize the median error, the particle filter achieves a median error of 0.040 m with a standard deviation of 0.050 m. The minimum median error of the UKF is
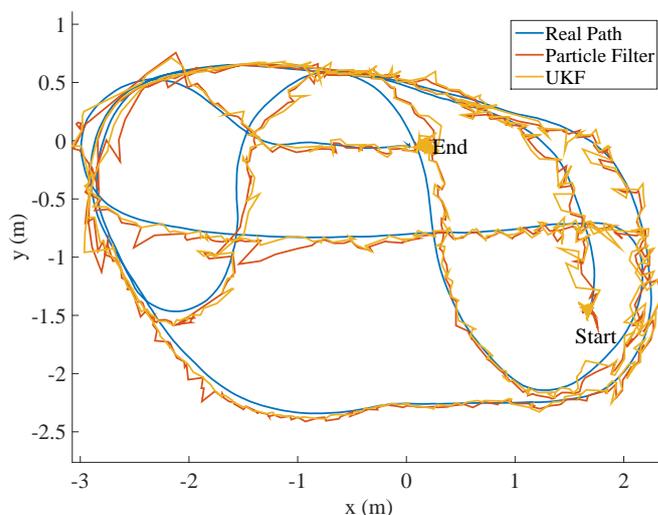
Figure 4.3: Real and estimated trajectory by the unscented Kalman filter and the particle filter with a target moving at a maximum velocity of 0.6 m/s. [3]

0.039 m with a standard deviation of 0.055 m. Therefore, they have a similar performance. The estimated trajectories can be seen in Fig. 4.3.
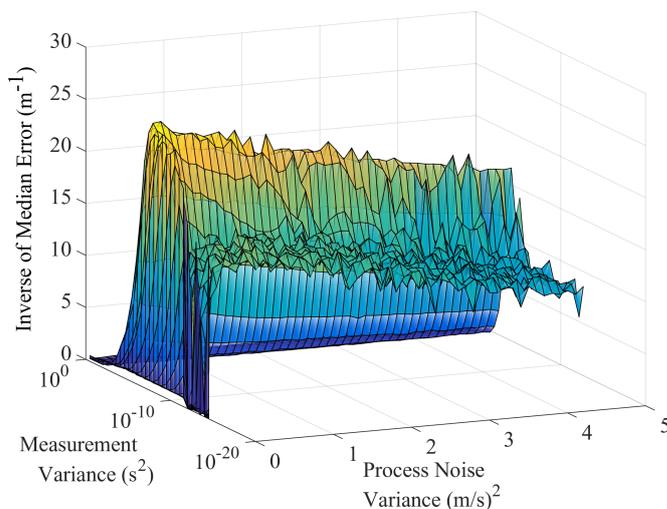
Figure 4.4: Median error achieved by the unscented Kalman filter with different parameters for the trajectory in Fig. 4.3. The measurement noise is plotted in log scale. Note that if the noise is underestimated the error increases drastically whereas if it is overestimated the error increases slower. The minimum error is 0.039 m. [3]

In the particle filter, if the measurement noise is underestimated, $p(z_j \mid \mathbf{x}_t)$ does not properly represent the likelihood of a measurement given a state. Therefore, the weights of the particles are reduced and the resampling step is executed often without succeeding in correctly finding
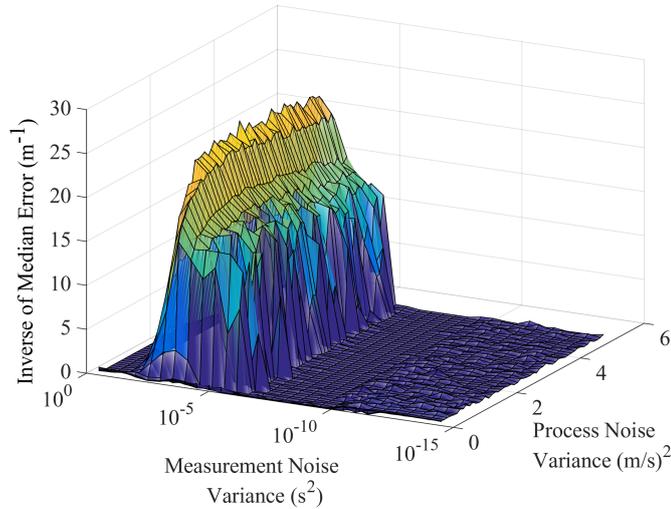
Figure 4.5: Median error achieved by the particle filter with different parameters for the trajectory in Fig. 4.3. The measurement noise is plotted in log scale. While the election of the process noise covariance is not critical, the measurement noise covariance has more effect on the final result than in the UKF (see Fig. 4.4). The minimum error is 0.040 m. [3]

the most likely particles. On the other hand, the motion model is only used to propagate the states, which has less influence in the final estimation.

In the UKF, underestimating the measurement noise or overestimating the process noise have a similar effect, which is to rely more on the correction step and less on predictions. This can be easily observed when calculating the Kalman gain in the UKF. If one looks at Eq. 4.18 one can realize that the Kalman gain will be larger when the assumed measurement noise decreases or when the assumed process noise increases. When the Kalman gain is larger, the UKF trusts more the measurement and less the predictions.

The error distribution (Fig.4.6) for the experiment in Fig. 4.3 shows an error lower than 10 cm for 85% of the timestamps in both algorithms. Then, both algorithms achieve similar performance when the parameters are properly chosen.

As a conclusion, we have observed how, while the system of equations is undetermined, the trajectory of the target can be estimated with a median error of around four centimeters making use of an unscented Kalman filter or a particle filter. This is because the small changes of the variables during consecutive measurements can be modeled using a probabilistic motion model. Moreover, it is interesting to observe how the particle filter and the UKF are capable of achieving similar results even though the particle filter requires more computational power. In Chapter 7 we will propose an algorithm which allows this system to locate the sender at higher velocities by improving the data association of the timestamps.
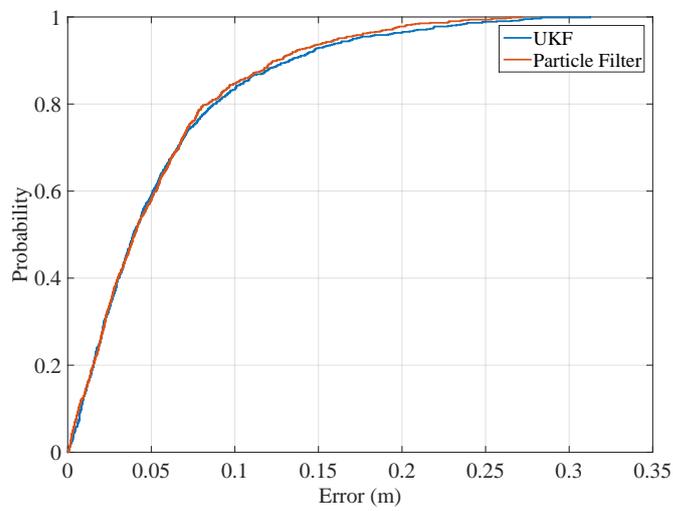
Figure 4.6: Error cumulative distribution of the unscented Kalman filter and the particle filter. Both filters achieve an error lower than 10 cm for 85% of the timestamps. The median error of the particle filter is 0.040 m with a standard deviation of 0.050 m. The minimum median error of the unscented Kalman filter is 0.039 m with a standard deviation of 0.055 m. [3]

5 Detection and Discarding of Reflections

In the previous chapter we have assumed the received signals are in line-of-sight. This is often not true in real scenarios. Then, one cannot assume a normally distributed measurement noise. This chapter shows two strategies for overcoming this by discarding or giving lower weights to non-line-of-sight measurements.

Having a target emitting ultrasound signals, a reflection can be seen as a measurement which does not fit the line-of-sight sensor model. If a sender $\mathbf{S}$ emits a signal at a time t_0, the reception time T_i will depend on the distance that the sound needs to travel to reach the receiver $\mathbf{M}_i$. If the signal is reflected, this distance will increase and therefore the reception time will be larger than expected. This can be modeled as follows:

$$T_i = \frac{1}{c}\|\mathbf{M}_i - \mathbf{S}\| + t_0 + n_i + b_i, \qquad 1 \leq i \leq N \tag{5.1}$$

where n_i is modeled as a Gaussian process with zero mean and standard deviation σ_n and b_i is a value greater than zero that depends on the position of the sender and the reflections. In literature, this has been modeled in different ways, such as a uniform distribution [28] or an exponential distribution [22, 62].

The aim of the algorithms presented in this chapter is to mitigate the effect of the reflections by identifying and discarding the measurements which are not likely to be in line-of-sight. When t_0 is known and n_i is neglected, the subspace of possible positions is bounded. This means, if a receiver at position $\mathbf{M}_i$ measures a distance $c(T_i - t_0)$, the target must be inside of a sphere of radius $c(T_i - t_0)$ and center $\mathbf{M}_i$. This is because a reflection will always arrive later than a line-of-sight signal. When t_0 is unknown, the subspace of possible solutions is not bounded. However, it is a common scenario, as synchronizing a speaker and a microphone increases the cost of the senders and is not always feasible. In this chapter we explore two solutions for identifying non-line-of-sight measurements when the sending time is unknown:

- A **combinatorial approach**, which uses combinations of receivers in order to identify which are the line-of-sight measurements. This approach can be used without making any assumptions about the movement of the target. One can imagine a real-life application where a high number of senders need to be localized and they only emit sound signals when the position is demanded.

- An approach using **robust regression**, which weights the measurements according to their likelihood. While robust regression has been used before for localization by other researchers [33, 63–65], we discuss its limitations and propose an approach for sensor data fusion with an inertial measurement unit (IMU).

All of the approaches in this chapter have in common that they use, in one way or the other,

the residual error of the position estimations to detect non-fitting measurements. Under the assumption that there are more constraints than variables, the residual error is the remaining error after the variables have been estimated. This chapter serves also as an introduction to the peculiarities of the NLOS measurements and the challenges which need to be faced to identify them.

5.1 Combinatorial Approach

A common approach is to calculate the residual error in order to identify erroneous measurements. Having an estimation of the sender position $\mathbf{S}_e$ and an estimated sending time $t_{0,e}$, the residual error r_i of a measurement T_i is:

$$r_i = T_i - \left(\frac{1}{c} \|\mathbf{M}_i - \mathbf{S}_e\| + t_{0,e} \right), \qquad i \in [1, ..., N] \tag{5.2}$$

Then, if the number of constraints is larger than the number of variables, one can use the values of $|r_i|$ as a measure of how precise is the estimation.

The first naive approach that one could think of is trying combinations of measurements in order to identify which measurements are reflections and which measurements are line-of-sight. Then, having N timestamps, it is possible to try all combinations of b receivers $\binom{N}{b}$. For every combination, a position is estimated using time difference of arrival. Then, for every candidate position, one can count the number of timestamps which have a low residual error $|r_i|$.

A timestamp T_i would be considered to be valid if:

$$\left| T_i - \frac{1}{c} \|\mathbf{M}_i - \mathbf{S}_e\| - t_{0,e} + n_i \right| = 0 \tag{5.3}$$

Then, knowing the distribution of n_i one could set a threshold which would identify some large measurements which are not likely to be caused by n_i.

This approach, while simple, presents some inconveniences. The first one is that a wall reflection acts like a virtual sender, which also fulfills Eq. 5.3 for a certain number of receivers. In certain cases, this can be solved by ensuring that b_i has to be greater than zero. If one has an estimation $\mathbf{S}_e$, estimated with a subset of receivers, one needs to ensure that:

$$c(T_i - t_{0,e} + n_i) \geqslant \|\mathbf{M}_i - \mathbf{S}_e\| \qquad \forall i \in [1, \dots, N] \tag{5.4}$$

where $t_{0,e}$ is the estimated sending time. In other words, one needs to ensure that the timestamps which are considered non-line-of-sight measurements arrive later than if they were in line-of-sight, as they have traveled larger paths.

Another interesting observation, is that the maximum time difference between two line-of-sight signals is limited by the distance between the receivers. Then, having two receivers i and j:

$$|T_i - T_j + n_i + n_j| \leqslant \frac{1}{c} \|\mathbf{M}_i - \mathbf{M}_j\| \tag{5.5}$$

This inequality can be used to reduce the number of combinations that need to be computed and therefore reduce the requirement of computational power.

There are other factors that should be considered to ensure the measurements are in line-of-sight. Using Eq. 5.3 for all measurements would mean that the residual errors are independent. This is not the case. Even if the measurement noise for the different receivers is independent, the residual errors are certainly not independent, because the noise has been affected by the position estimation. Then, depending on how the receivers and the sender are distributed in space, one of the measurements can have more influence in the position estimation than others. If all receivers but one are placed in such manner that they do not allow a proper localization in certain coordinates, the result will be highly dependent on the receiver which is properly placed. Therefore, its residual error will be low even though it can drastically change the position estimation.

In the following subsections we introduce three metrics which are widely used in regression analysis [66, 67] which consider this fact: *leverage matrix*, *studentized residuals* and *Cook's distance*.

Hat Matrix

As commented before, the noise of the residual errors is not independent. Therefore, one needs to consider how the measurement noise is affected by the position estimation.

If $\mathbf{H}$ is a linear matrix which relates the measurement vector $\mathbf{z}$ and the estimations $\mathbf{x}$, one can minimize the squared error:

$$\arg \min_{\mathbf{x}} (\mathbf{Hx} - \mathbf{z})^T (\mathbf{Hx} - \mathbf{z}) \tag{5.6}$$

Then, the least squares estimation will be:

$$\mathbf{x} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{z} \tag{5.7}$$

In order to calculate the residual error, one needs to transform again the estimated vector $\mathbf{x}$ to the measurement state:

$$\widehat{\mathbf{z}} = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{z} \tag{5.8}$$

Then, $\widehat{\mathbf{z}}$ is the measurement that would be received if the variables in the vector $\mathbf{x}$ were estimated without any error. In other words, $\widehat{\mathbf{z}}$ would be equal to $\mathbf{z}$ if there would be no noise. The residual error $\mathbf{e}$ can be calculated as the difference between these values and the measured ones:

$$\mathbf{e} = \widehat{\mathbf{z}} - \mathbf{z} \tag{5.9}$$

The variance of the residuals is:

$$\mathbb{E}[(\mathbf{e} - \mathbb{E}[\mathbf{e}])(\mathbf{e} - \mathbb{E}[\mathbf{e}])^T] = \sigma_n (\mathbf{I}_{N \times N} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T)$$
$$= \sigma_n (\mathbf{I}_{N \times N} - \mathbf{L}_v) \tag{5.10}$$

where $\mathbf{I}_{N \times N}$ is a $N \times N$ identity matrix. The matrix $\mathbf{L}_v$ is the so-called *hat matrix* or *leverage matrix*:

$$\mathbf{L}_v = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T \tag{5.11}$$

The diagonal elements of this matrix l_{ii} are the leverage values. Then, looking at the value l_{ii} one knows how large is the contribution of the measurement i into the final estimation. By looking at the equations, one can also see how the measurements which have more influence in the final estimation will also have lower residual error. This means, not only they can change drastically the positions, but they will have low residual errors too. Therefore, it is likely that they will not be detected using Eq. 5.3.

For the moment we have assumed that $\mathbf{H}$ is a linear matrix. However, our sensor model is not linear. One can overcome this issue by linearizing the equations using the Jacobian matrix. Then, one can estimate a position with non-linear optimization and evaluate the Jacobian matrix in that point. By doing this, one can gather knowledge about which measurements had more influence in the estimation.

In Fig. 5.1 one can see an example. For simplicity and clarity in the figure, we consider time of arrival measurements in two dimensions. In the figure, one can observe how while three receivers are aligned and cannot properly estimate the y coordinate, the other two receivers are better distributed. In this case, if the receiver with leverage 0.85 would measure a reflection, it would have a greater impact in the final estimate than if the receivers were better distributed. Moreover, it would have lower residual error. One can observe this also by looking at the geometry of the TOA measurements. If the receiver with leverage 0.85 changes slightly its measurement, the estimated target position will *move* in the y direction. On the other hand, if another measurement changes, the residual error will be higher, as the estimated position cannot change without contradicting other measurements.

Studentized Residuals

The studentized residuals compare the expected standard deviation of the measurements and the residual error. The studentized residual s_i of the i-th observation is:

$$s_i = \frac{e_i}{\widehat{\sigma_n}\sqrt{1 - l_{ii}}} \tag{5.12}$$

where e_i is the i-th component of the residual vector $\mathbf{e}$. The residual vector can be calculated using a position estimated with non-linear optimization. The variable $\widehat{\sigma_n}$ is the estimated noise standard deviation. A common approach to estimate it is:

$$\widehat{\sigma_n} = \sqrt{\frac{1}{N - p} \sum_{i=1}^{N} e_i^2} \tag{5.13}$$

where p is the number of dimensions which are estimated.

These residuals compare the estimated standard deviation of the residuals and their actual values. Then, these values contain more information about whether a measurement is valid or not than just using e_i.
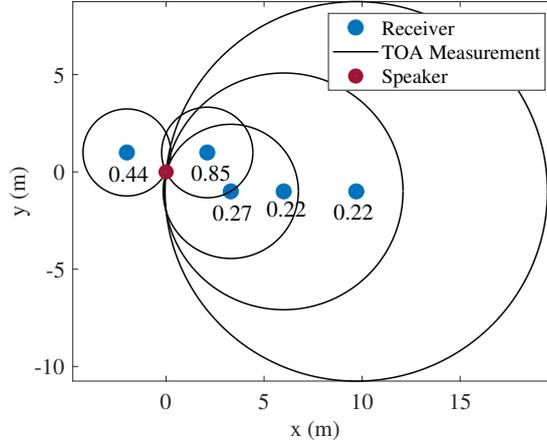
Figure 5.1: Values of the hat matrix corresponding to the TOA measurements showed by the black circles. One can observe how some measurements have more importance in the final estimation than others. In this case, the receiver with the value 0.85 is the one which carries the most amount of information about the position of the target in the y coordinate.

Cook's Distance

While the studentized residuals are already a useful manner of detecting NLOS measurements, there are some cases where a measurement with a small studentized residual r_i can change drastically the position [68]. Therefore, a metric which is more convenient is the Cook's distance, which measures how much an estimation changes by a single measurement.

Concretely, the Cook's distance D_i measures the squared distance between the estimation including the measurement i, which we note as $\hat{\mathbf{x}}$ and the estimation without the measurement i, which we note as $\hat{\mathbf{x}}^{(-i)}$:

$$D_i = \frac{(\hat{\mathbf{x}} - \hat{\mathbf{x}}^{(-i)})^T (\hat{\mathbf{x}} - \hat{\mathbf{x}}^{(-i)})}{(p+1)\widehat{\sigma_n}^2} \tag{5.14}$$

The denominator normalizes this distance using the noise variance and the number of estimated variables p, in order to have always a comparable value.

One can rewrite D_i in a way that it is not necessary to explicitly estimate $\hat{\mathbf{x}}^{(-i)}$:

$$D_i = \frac{e_i^2}{(p+1)} \frac{l_{ii}}{(1-l_{ii})^2} \tag{5.15}$$

Then, the distance D_i is a metric of how much an estimation changed by using the measurement T_i. Ideally, a single measurement should not change drastically the position estimated by the other receivers. If it does, it is likely that it is a non-line-of-sight measurement. It can also happen that the receivers and sender are poorly distributed in space. In this case, a single measurement can also change drastically the final measurement estimation. However, if this happens, one cannot be sure that it is a line-of-sight measurement either, as the position estimation depends mostly on a single measurement.

Proposed Algorithm

We propose an algorithm which uses combinations of three receivers in order to detect the line-of-sight measurements. In [69] a closed-form solution is presented for locating a target in three dimensions using four satellites and TDOA measurements. In our case, we assume the height is approximately known, therefore only three receivers are required (two TDOA measurements). We follow a similar approach to find the closed-form solution.

In order to reduce the number of operations required, we compare the TDOA measurements with the maximum TDOA allowed by their distance (Eq. 5.5) to identify invalid combinations.

Afterwards, if a remaining combination leads to one or two solutions, the leverage matrix is calculated (Eq. 5.11) using the Jacobian matrix. In order to identify which receivers lead to this same position, we extend this matrix using the other receivers. For example, if we want to know whether the timestamp received by the microphone i would lead to approximately the same position $\mathbf{S}_e$ than the timestamps received by the receivers 1, 2 and 3, we extend the Jacobian matrix:

$$\mathbf{H} = \begin{pmatrix} \frac{x-x_2}{R_2} - \frac{x-x_1}{R_1} & \frac{y-y_2}{R_2} - \frac{y-y_1}{R_1} \\ \frac{x-x_3}{R_3} - \frac{x-x_1}{R_1} & \frac{y-y_3}{R_3} - \frac{y-y_1}{R_1} \\ \frac{x-x_i}{R_i} - \frac{x-x_1}{R_1} & \frac{y-y_i}{R_i} - \frac{y-y_1}{R_1} \end{pmatrix} \tag{5.16}$$

being R_i the distance from the sender $\mathbf{S}_e$ to the receiver i:

$$R_i = \|\mathbf{M}_i - \mathbf{S}_e\| \tag{5.17}$$

Having $\mathbf{H}$ one can calculate the leverage matrix $\mathbf{L}_v$ and the Cook's distance D_i for all four measurements. If all Cook's distances are below a certain threshold, the receiver i is considered to have a measurement which is in the same situation as the receivers 1,2 and 3. They are either all *line-of-sight* or *non-line-of-sight*. In order to remove virtual senders from the set of possible solutions, Eq. 5.4 is used, which ensures that b_i is positive.

The algorithm follows the following steps:

1. Consider all possible combinations of 3 receivers. The number of combinations is:

$$\binom{N}{3} = \frac{N!}{3!(N-3)!} = \frac{N(N-1)(N-2)}{6} \tag{5.18}$$

2. If one of the time differences does not fullfil Eq. 5.5, go to the next combination, if not, compute the position using the closed-form algorithm.
3. If there is one or two solutions, use Eq. 5.4 to remove virtual senders from the possible position estimations. If none of the solutions fulfills Eq. 5.4, go to the next combination.
4. Extend the Jacobian matrix using a receiver which was not used in the combination (Eq. 5.16).
5. Calculate the Cook's distance
6. Repeat steps 4 and 5 until all receivers have been used.

7. Count the number of receivers with a low maximum Cook's distance. As the Cook's distance follows an F distribution, one can select a threshold according to the requirements of probability and the number of measurements and variables.

8. From all possible candidates, select the one which has the greatest number of timestamps with low Cook's distance and re-estimate the position using all the line-of-sight signals.

The algorithm used for estimating the final position is the Levenberg–Marquardt algorithm. This algorithm interpolates between the gradient descent and the Gauss-Newton algorithm. The reason to re-estimate the final position is that the estimations with only three receivers will have a higher error than the estimation using all the line-of-sight signals.

5.2 Robust Regression

The least-squares solution in Eq. 5.7 is optimal only if the error is Gaussian. The solution $\mathbf{x}$ which minimizes this equation can be then highly affected by non-Gaussian errors. Then, a better solution is to minimize a weighted version of the measurements, where $\mathbf{W}$ is a diagonal matrix with weights w_i in its diagonal.

$$\arg\min_{\mathbf{x}} (\mathbf{Hx} - \mathbf{z})^T \mathbf{W}(\mathbf{Hx} - \mathbf{z}) \tag{5.19}$$

If we define a vector of errors $\mathbf{r} = \mathbf{Hx} - \mathbf{z}$:

$$\arg\min_{\mathbf{x}} \mathbf{r}^T \mathbf{W} \mathbf{r} \tag{5.20}$$

Then, $\mathbf{x}$ can be estimated as:

$$\mathbf{x} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{z} \tag{5.21}$$

This provides a more robust estimation against measurements which are non Gaussian, especially for heavy-tailed distributions.

There are multiple possibilities for calculating the weights. We make use of the hyperbolic tangent weighting function.

Hyperbolic Tangent Weighting Function

Huber [70] introduced the *M-Estimators*, which minimize a function $\rho(r_i)$ instead of the squares:

$$\arg\min_{\mathbf{x}} \sum_{i=1}^{N} \rho(r_i) \tag{5.22}$$

where r_i is the i-th component of $\mathbf{r}$. Then, setting the partial derivatives with respect to the state $\mathbf{x}$ to zero would give the result. However, the derivative of ρ is not always linear. Therefore, one can define the derivative of the function ρ as

$$\rho'(r_i) = r_i w(r_i) \tag{5.23}$$

There are multiple approaches to calculate the weights depending on r_i. In [5] we use the hyperbolic tangent function proposed by Hambel et. al [71]. This is because it rejects measurements which are extremely large and it has been proven to perform well in TDOA localization before [33]. The weights are defined as:

$$\begin{cases} w(r_i) = 1 & |r_i| \leq a \\ w(r_i) = \frac{d \cdot \tanh\left(\frac{d(b-r_i)}{2}\right)}{r_i} & a < |r_i| \leq b \\ w(r_i) = 0 & |r_i| > b \end{cases} \tag{5.24}$$

where a and b are the *clipping points* and d is defined in order to have a continuous function. The value of a and b depends on the knowledge of the noise. A common approach is to take $a = \frac{1.4826 \text{MAD}}{\theta}$ and $b = 4a$, where MAD=median($|\mathbf{r} - \text{median}(\mathbf{r})|$)) and θ is a tuning parameter. In Fig. 5.2 one can see a graphical representation of the weighting function.
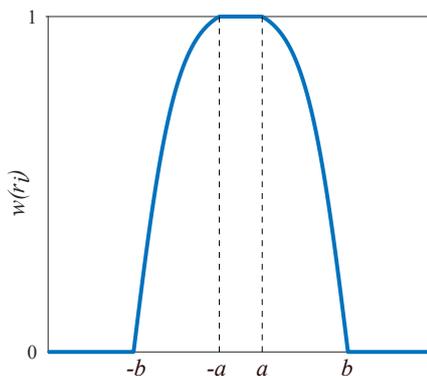
Figure 5.2: Hyperbolic tangent weighting function (see Eq. 5.24).

As we have shown during this thesis, the measurement model is not linear. Therefore, one cannot use such estimators with a linear matrix $\mathbf{H}$. One can linearize the measurement model using a Jacobian matrix. However, the linearization is only valid for points which are close to the actual values of $\mathbf{x}$. Nonetheless, one can rewrite the extended Kalman filter as a robust regression problem. Then, one has a prediction, which is assumed to be close to the actual values of $\mathbf{x}$. Moreover, the probabilistic motion model provides information about the measurements which must be given low weights.

Robust Extended Kalman Filter

The extended Kalman filter (EKF) assumes Gaussian noise. However, it can be formulated in the form of weighted least squares [65]. Using the notation of Section 4.1, the equations of the

Kalman filter can be reformulated as follows:

$$\mathbf{x}_t = (\mathbf{N}^T \mathbf{W}_t \mathbf{N}_t)^{-1} \mathbf{N}_t^T \mathbf{W}_t \mathbf{Y}_t \tag{5.25}$$

$$\mathbf{N}_t = \mathbf{V}_t^{-1} \begin{bmatrix} \mathbf{I} \\ \mathbf{H}_t \end{bmatrix} \tag{5.26}$$

$$\mathbf{Y}_t = \mathbf{V}_t^{-1} \begin{bmatrix} \widehat{\mathbf{x}}_t \\ \widehat{\mathbf{z}}_t \end{bmatrix} \tag{5.27}$$

$$\widehat{\mathbf{z}}_t = \mathbf{z}_t - h(\widehat{\mathbf{x}}_t) + \mathbf{H}_t \widehat{\mathbf{x}}_t \tag{5.28}$$

where $\mathbf{W}_t$ is a diagonal matrix which contains the different weights. The matrix $\mathbf{V}_t \mathbf{V}_t^T$ is given by:

$$\mathbf{V}_t \mathbf{V}_t^T = \begin{bmatrix} \widehat{\mathbf{P}}_t & 0 \\ 0 & \mathbf{R}_t \end{bmatrix} \tag{5.29}$$

Then, $\mathbf{V}_t$ can be calculated using the Cholesky decomposition. The weight j at the iteration k is calculated using Eq. 5.24, where:

$$w(r_i) = w(|y_j - \mathbf{n}_j^T \mathbf{x}^{k-1}|), \qquad j \in [1, ..., N+p] \tag{5.30}$$

the variable y_j is the j-th element of $\mathbf{Y}_t$ and $\mathbf{n}_j$ is the j-th row of $\mathbf{N}_t$. The variable p represents the number of components of the state vector $\mathbf{x}_t$. Every iteration, the weights can be computed using Eq. 5.24.

Data Fusion

While the robust extended Kalman filter (REKF) has been used in the past to mitigate NLOS measurements, this filter runs under the assumption that during a certain time, enough LOS timestamps will be received to locate the target. However, in reality often this is not the case. In certain situations, a target moves to a region which is not well covered by the receivers and the received signals are only NLOS. The robust extended Kalman filter would use the predicted velocity to estimate the positions of the target, which would eventually diverge from the real position. Moreover, in certain scenarios, the speaker might point to a wall and the estimated position will be a virtual sender, and all measurements will have a low residual error. However, it will not be the actual position of the target (see Fig. 5.3). In order to avoid this, other information needs to be used. IMU data can be fused with the TDOA measurements in order to overcome this issue.

In [5] we propose to use a REKF for fusing the data from the acoustic localization system with an inertial measurement unit (IMU). An IMU is inside most of the commercial smartphones in the market. It provides information about the smartphone acceleration, rotation and surrounding magnetic field. IMUs have been used by many researchers as a standalone solution for localization. However, as their estimations do not have any reference fixed in space, their error is accumulative. Moreover, the magnetic field is not a reliable source of information
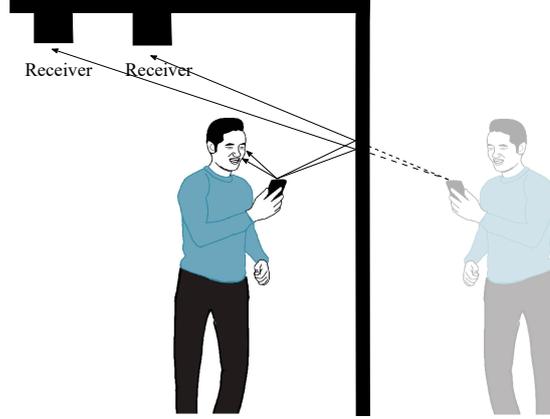
Figure 5.3: In certain situations, a person might block the line-of-sight signals and multiple reflections from one wall might be received. In the case shown in this figure, the measurements from both receivers would have a low residual error for the virtual sender position. If the only signals received by the microphones are the reflected by this wall, the estimated position would have a high error and never be detected using the residual error of the TDOA estimations.

indoors. Therefore, we propose to fuse the IMU data with the acoustic data with robust regression. By doing this, the REKF gives lower weights to unlikely measurements and can be used to correct the accumulative error of the IMU when it deviates from the estimated trajectory using the acoustic measurements.

Motion and Sensor Models

The state of the filter $\mathbf{x}_t$ contains the position $\mathbf{S}_t$ and the velocity $\mathbf{V}_t$ of the target in $\mathbb{R}^2$. We use the following acceleration model:

$$\begin{bmatrix} \mathbf{S}_t \\ \mathbf{V}_t \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{S}_{t-1} \\ \mathbf{V}_{t-1} \end{bmatrix} + \mathbf{G}\boldsymbol{\Phi}_{t-1} \qquad \boldsymbol{\Phi}_{t-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \tag{5.31}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{G} = \begin{bmatrix} \frac{\Delta t^2}{2} & 0 \\ 0 & \frac{\Delta t^2}{2} \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix} \tag{5.32}$$

where Δt is the elapsed time between two estimations.

Then, the motion model is used to predict the next state $\widehat{\mathbf{x}}_t$ and its covariance matrix $\mathbf{P}_t^-$:

$$\widehat{\mathbf{x}}_t = \mathbf{A}\mathbf{x}_{t-1} \tag{5.33}$$

$$\mathbf{P}_t^- = \mathbf{A}\mathbf{P}_{t-1}\mathbf{A}^T + \mathbf{G}\mathbf{Q}\mathbf{G}^T \tag{5.34}$$

When fusing the data, the measurement model contains the TDOA measurements and the IMU measurements. The measurements from the IMU are used to estimate the position of the target $\mathbf{S}_t$. We remove the effect of the gravity and extract the x and y components of the acceleration by combining the acceleration and the angular rate. Afterwards, a zero-velocity update approach [72] is used to estimate the position of the target. The measurement vector then, contains the TDOA measurements $\mathbf{z}_{TDOA}$ and the estimated position using the IMU data $\widehat{\mathbf{S}}_{IMU}$. If one does not use the IMU data, the measurement vector $\mathbf{z}_{FUS}$ would contain only $\mathbf{z}_{TDOA}$:

$$\mathbf{z}_{FUS} = \begin{bmatrix} \mathbf{z}_{TDOA} \\ \widehat{\mathbf{S}}_{IMU} \end{bmatrix} \tag{5.35}$$

The i-th component of the TDOA measurement is modeled as:

$$z_{TDOA,i} = \frac{1}{c}\|\mathbf{S}_t - \mathbf{M}_i\| - \frac{1}{c}\|\mathbf{S}_t - \mathbf{M}_{ref}\| \tag{5.36}$$

This can be rewritten in vector notation as:

$$\mathbf{z}_{TDOA} = g(\mathbf{x}_t) + \rho_t \qquad \rho_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_m) \tag{5.37}$$

Being $\mathbf{T}_t$ the Jacobian matrix of the TDOA measurements, the matrix $\mathbf{H}_t$ which relates the measurements and the state, can be written as:

$$\mathbf{H}_t = \begin{bmatrix} \mathbf{T}_t \\ \mathbf{O}_t \end{bmatrix} \tag{5.38}$$

where $\mathbf{O}_t$ is:

$$\mathbf{O}_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{5.39}$$

Then, the REKF can be used with these motion and sensor models.

5.3 Experimental Results

In order to test the behavior of the proposed algorithms we use the ASSIST localization system to locate a moving sender. Some wooden walls and furniture are placed in the localization environment in order to increase the number of reflections.

Comparison of NLOS Mitigation Algorithms

First we do an experiment without IMU. In order to compare the performance of the different algorithms presented in this chapter we test them with an extensive dataset, consisting of 128 337 samples, more than 213 minutes of TDOA measurements under different NLOS conditions and velocities of the target. More details about the experimental environment can be seen in the Appendix of this thesis.

We compare the robust extended Kalman filter, the EKF, the proposed combinatorial algorithm and the result of using nonlinear least squares with random initialization and all available measurements. The result can be seen in Fig. 5.4. One can observe how the combinatorial algorithm clearly outperforms the other algorithms. This is because no geometrical constraints are used in the other algorithms. This means that the REKF, EKF and the nonlinear optimization approach do not use Eq. 5.5 to discard measurements which have a time difference larger than the maximum allowed by the distance between receivers. Therefore, in multiple occasions the REKF is not capable of locating the target with a reduced error, as the number of non-Gaussian measurements is too elevated. This shows the importance of finding problem-specific constraints which can discard unlikely measurements.
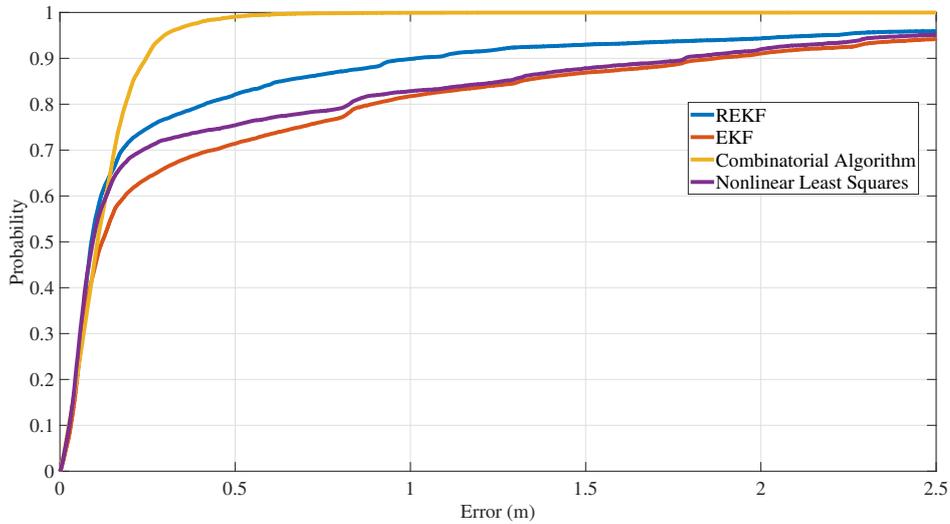
Figure 5.4: Cumulative distributions of the different algorithms for mitigating NLOS measurements. The probability for errors above 2.5 m is not shown. In this case no geometrical constraints are imposed over the measurements. The cumulative distribution using Eq. 5.5 can be seen in Fig. 5.5.

When deleting the measurements which have a larger TDOA than the one allowed by their geometrical distance (Eq. 5.5), the REKF provides the most robust estimation, as it can be seen in Fig. 5.5. One can observe how the proposed combinatorial algorithm has a higher error than the other algorithms when the target is in line-of-sight. This is because some of the timestamps are erroneously discarded and therefore the error is larger than using all line-of-sight measurements. However, when looking at the 98 percentile one can observe how it provides a similar result as the REKF. On the other hand, nonlinear least squares has an error above 2 m for 0.82% of the positions. It is more affected by the NLOS measurements than the others, as it assumes no motion model and it does not use any method for detecting those measurements. The EKF performs better due to its motion model, which does not allow large accelerations.

If one considers one meter as the maximum allowed error, the combinatorial error would not succeed in 0.08% (97 positions) of the cases. The REKF provides a similar result, it would
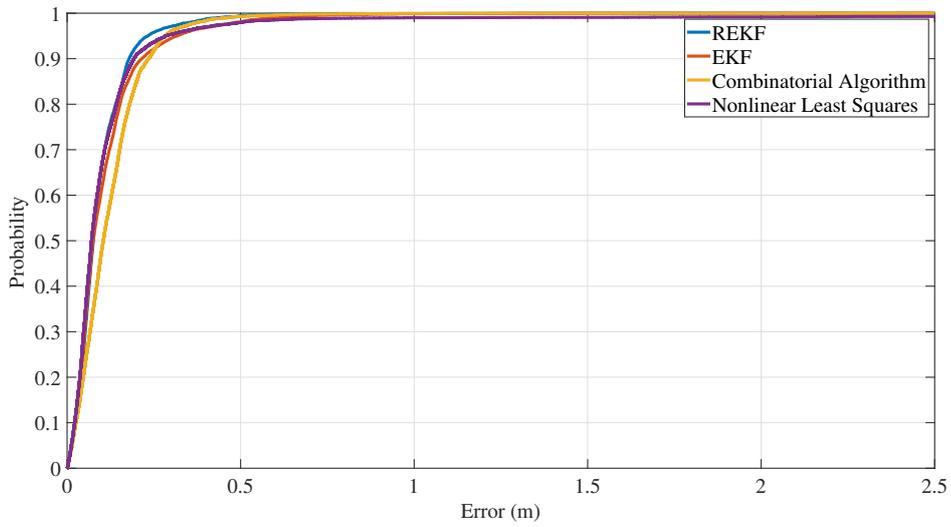
(a) Cumulative distributions of the different algorithms for mitigating NLOS measurements. The probability for errors above 2.5 m is not shown. One can observe how generally the combinatorial algorithm has a higher error, although it performs better under NLOS conditions than the EKF or the nonlinear least squares approach. This can be seen more clearly in the figure below.
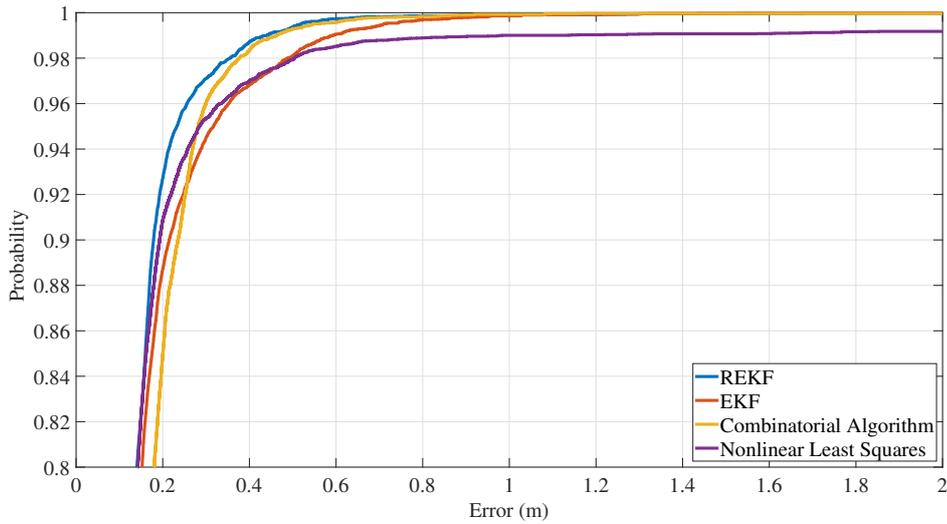
(b) Closer look to the highest probabilities of the cumulative distribution. In these estimated positions one can observe how the REKF and the combinatorial algorithm are capable of reducing the number of errors above 0.4 m. It is interesting to observe how the combinatorial approach leads to a considerable improvement compared to the nonlinear least squares approach, although no assumptions over the movement of the target are made.

Figure 5.5: Cumulative distribution of the localization error when the measurements which have a larger TDOA than the one allowed by their geometrical distance (Eq. 5.5) have been eliminated.

not succeed in 0.09% of the cases (113 positions). The EKF achieves a worse performance than the previous ones, it does not succeed in 0.14% of the cases (177 positions). While the EKF achieves a worse performance than the others, it still proves to be a feasible and simple solution for localization when the number of NLOS measurements is reduced. The worst result is clearly achieved by the nonlinear least squares estimator, with 1.00% (1277 positions). Then, using nonlinear least squares with all valid data is not a feasible solution for localization. Algorithms such as the combinatorial approach are then necessary for locating a target when no assumptions about the movement of the target can be made.

We have seen how, assuming a certain motion model, and using robust regression, a target can be located in mixed line-of-sight/non-line-of-sight measurements. When no assumption about the movement of the target is made, it is also possible to locate a target with the same level of robustness. However, the price to pay is the required computational power.

Data Fusion

In the previous experiments one can observe how even using NLOS mitigation algorithms, there are estimations which have a high error due to the NLOS measurements. This is because if there are only a limited number of received measurements, the TDOA hyperboloids can intersect in one point with low error. One can think of a speaker pointing to a wall in a way that only the reflections from the wall are received by the speakers. In such case, the reflections are seen as a virtual sender, which behaves like a real sender, and cannot be identified. A method to overcome this challenge is to use another source of information, such as the IMU.

In order to prove the feasibility of fusing IMU data and acoustic measurements we proceed with a simulation and a real-world experiment. We provide a proof of concept of how the inertial measurement unit can be used to reduce the localization error in situations where the previously mentioned algorithms are not capable of doing it.

It should be noted that after a certain time, the IMU would become unusable if it is not re-aligned to the acoustic coordinate system. However, we show how the estimated positions remain accurate even when the IMU starts to drift. One could notice the IMU is drifting by looking at the weighting of the measurements and re-align it again trying to increase those weights. This is out of the scope of this dissertation. The aim of these experiments is to show the feasibility of using IMU when the TDOA measurements are not enough to correctly identify the NLOS measurements.

First we test the algorithm with synthetic data, assuming there are 10 receivers available. The sound measurements have a Gaussian noise of 0.3 ms. Furthermore, 40% of the measurements of the receivers are NLOS measurements. The NLOS measurements are simulated by adding a uniformly distributed error from 3 ms to 13 ms. The IMU data is simulated by assuming a constant angle error. We test the REKF and the EKF fusing the IMU data and the timestamps. In addition, we test the REKF using only the timestamps. The estimated and real positions can be seen in Fig. 5.6. The resulting error is shown in Table 5.1. One can see how the REKF with data fusion achieves a lower error than the other alternatives.

In order to test the proposed approach in a real scenario, we use the ASSIST system and

Algorithm	Mean Error (m)	Std. (m)
REKF with data fusion	0.13	0.08
REKF without data fusion	0.16	0.11
EKF with data fusion	0.38	0.20

Table 5.1: Error with simulated data. [5]
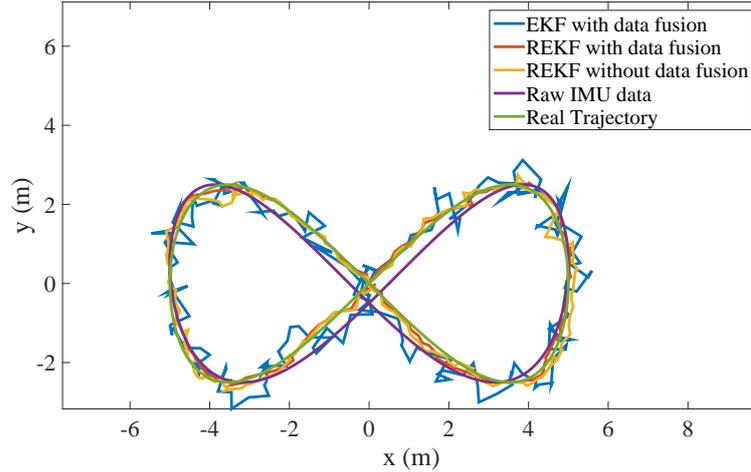
Figure 5.6: Real and estimated positions of a target using synthetic data. In this simulation, 40% of the measurements are NLOS. The REKF which fuses the data from the receivers and the IMU achieves the lowest error: 0.13 m of mean error and 0.08 m of standard deviation. [5]

the built-in smartphone IMU. The target is located in an area with a high number of obstacles. The result can be seen in Fig. 5.7. The mean error achieved using a REKF without sensor data fusion is 0.19 m and the standard deviation is 0.19 m. The best result is achieved by the REKF with sensor data fusion: 0.16 m mean error and 0.14 m standard deviation. The cumulative distribution of the error can be seen in Fig. 5.8. One can see how fusing the IMU data, there are no positions with errors above 0.61 m. Without the IMU, there are noticeable position errors up to 1.31 m.

In conclusion, the presented combinatorial approach and the REKF prove to be capable of reducing the impact of NLOS measurements in the final estimations. However, in certain cases, such as when only reflections from a wall are received, they are not capable of locating the target with a reduced error. For these special situations, one needs an additional source of information. Fusing the IMU data and the timestamps proves to be a feasible solution. In the next chapters we show how emitting signals at regular intervals, the movement of the target can be used for data association. In addition, we show how reflections can be seen as an additional source of information and do not need to be discarded.
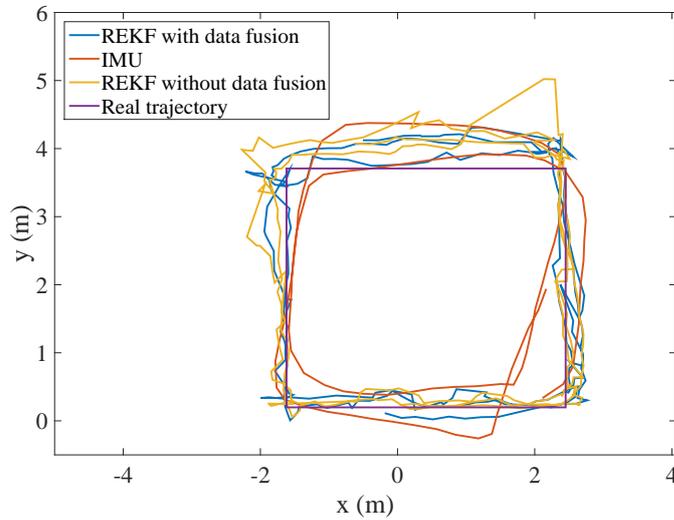
Figure 5.7: Real and estimated positions of a smartphone emitting inaudible sound signals. The signals are detected by static receivers. The REKF which fuses the data from the receivers and the IMU achieves the lowest error: 0.16 m of mean error and 0.14 m of standard deviation. [5]
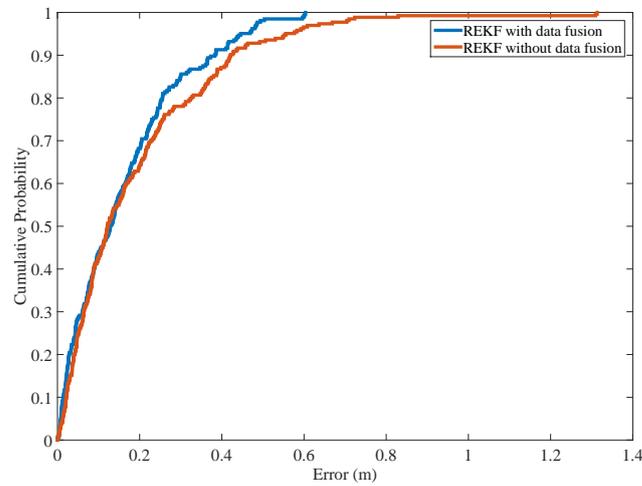
Figure 5.8: Cumulative distribution function for the real experiment in Fig. 5.7. Fusing the IMU data, the error is below 0.37 m for 90% of the measurements. Without the IMU data, the error is below 0.42 m for 90% of the measurements. Moreover, fusing the IMU data, the maximum error is reduced from 1.31 m to 0.61 m. [5]

6 Using Sender Movement for Identification

In the previous chapter we have seen how the time of arrival (TOA) and time difference of arrival (TDOA) equations can be used to estimate the position of a target and reject non-line-of-sight measurements. However, in certain occasions, one requires another source of information, as the reflections can act like a virtual sender which cannot be distinguished from a real sender. In this chapter we use the relative movement of the sender to the static receivers to estimate the target positions and the positions of the receivers. By doing this, the reflections can be modeled as additional virtual receivers which help finding the positions of the target. Then, the target is located regardless of whether the received signals are reflections or line-of-sight signals.

In this chapter we assume the traveled distance by the sender between two emitted signals is much smaller than the distances to the receivers. By doing this, one can simplify the equations. Moreover, one can associate the measurements by keeping track of the small motions of the target.

When a speaker is emitting periodical signals, one can observe a Doppler shift, as we have explained in previous chapters. This means that the signals are received sooner or later than expected depending on the relative velocity between the speaker and the microphone. Then, every receiver and echo can be considered independently, as they have a different relative velocity and are received at a different time instant. Consequently, every microphone can keep track of the small time differences due to the Doppler effect and distinguish them.

In the first section we present a probability data association filter which runs independently on every receiver and estimates the probability of a certain timestamp belonging to the same sender or echo. Afterwards, the data associations from different receivers need to be combined. This is achieved by assuming the movement of the target during a certain time is much lower than the distance to the receivers (far-field assumption).

6.1 Localization of a Moving Sender

In order to obtain more information about the target than just the TDOA measurements, we configure the speaker to emit chirp impulses at regular intervals I. By knowing this time interval, one can estimate the relative motion of the target to the receivers, as explained in Chapter 3.

Then, when the sender is at the position $\mathbf{S}_k$, it emits its k-th signal. The signal is received by the receiver i at position $\mathbf{M}_i$ at time $T_{k,i}$:

$$T_{k,i} = \frac{1}{c}\|\mathbf{M}_i - \mathbf{S}_k\| + t_0 + kI \qquad (6.1)$$

where t_0 is the time at which the first signal was emitted. After one interval, a signal will be

received at time $T_{k+1,i}$:

$$T_{k+1,i} = \frac{1}{c}\|\mathbf{M}_i - \mathbf{S}_{k+1}\| + t_0 + (k+1)I \qquad (6.2)$$

Then, one can already observe that the time difference will be I plus the relative movement from the target to the receiver. If the interval is known and the relative movement is small, one can keep track of such movements in order to know which timestamps come from the same reflector.

Then, having a target moving at velocity $\mathbf{v}$, the difference between two measurements of the receiver i is:

$$T_{k,i} - T_{l,i} = \frac{1}{c}\|\mathbf{M}_i - (\mathbf{S}_l + \mathbf{v}(k-l)I)\| - \frac{1}{c}\|\mathbf{M}_i - \mathbf{S}_l\| + I(k-l) \qquad (6.3)$$

Therefore, subtracting the elapsed intervals $I(k - l)$, the relative movement v_r between the receiver i and the sender $\mathbf{S}_l$ can be estimated:

$$v_r = \frac{c(T_{k,i} - T_{l,i} - I(k-l))}{I(k-l)} = \frac{\|\mathbf{M}_i - (\mathbf{S}_l + \mathbf{v}(k-l)I)\| - \|\mathbf{M}_i - \mathbf{S}_l\|}{I(k-l)} \qquad (6.4)$$

The velocity v_r and sending time t_0 will depend on the traveled path by the sound wave. Then, reflections from different walls will lead to different values. Moreover, the line-of-sight signals will have a lower t_0 than the sound reflections (see Fig. 6.1).
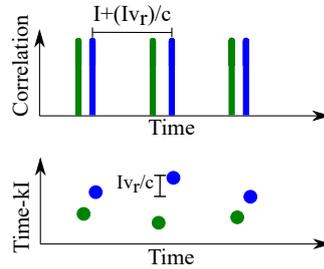
Figure 6.1: Schematic of the relative movement of a target (green) and a reflection (blue). The relative movement observed by the receiver (v_r) is different than the one reflected to a wall (blue).

6.2 Joint Probabilistic Data Association Filter

Wall reflections can be seen as virtual receivers. Then, every receiver will observe a different velocity v_r and sending time t_0 from the sender. By keeping track of the velocity v_r, one can aggregate the timestamps which belong to the same receiver. Every one of these aggregations is a *track*.

In order to keep track of the relative movement to every receiver, in [6] we implemented a Joint Probabilistic Data Association Filter (JPDAF) following the description in [73]. Every

physical receiver is independent and has its own JPDAF, tracking its multiple virtual receivers. Then, at the time k, every track predicts a state $\widehat{\mathbf{x}}_k^t$ that corresponds to a different virtual or real receiver, where t is the track number.

Being Θ^{k-1} a set of hypotheses, and $\Theta_{p(h)}^{k-1}$ the parent hypothesis of Θ_h^k, one can denote the resulting hypothesis at time k as:

$$\Theta_h^k = \{\Theta_{p(h)}^{k-1}, \; \theta_h(k)\} \tag{6.5}$$

where $\theta_h(k)$ is a measurement-to-track association for the set of received measurements Z^k. Using Bayes' theorem we can calculate the probability of the hypothesis Θ_h^k given the set of measurements Z^k:

$$
\begin{aligned}
P(\Theta_h^k | Z^k) &= \frac{P(\Theta_{p(h)}^{k-1}, \theta_h(k), Z_k, Z^{k-1})}{P(Z^k, Z^{k-1})} \\
&= P(Z_k | \Theta_{p(h)}^{k-1}, \; \theta_h(k), \; Z^{k-1}) P(\theta_h(k) | \Theta_{p(h)}^{k-1}, \; Z^{k-1}) \frac{P(\Theta_{p(h)}^{k-1} | Z^{k-1})}{P(Z^k | Z^{k-1})}
\end{aligned}
\tag{6.6}
$$

The JPDAF considers only the data association probabilities at a certain time window k. Then, the probability that a measurement j belongs to a certain track t can be formulated as the sum of the hypothesis probabilities where the measurement j is assigned to the track t:

$$\beta_j^t = \sum_{\theta \in \Theta_{jt}^k} P(\theta | Z^k) \tag{6.7}$$

where Θ_{jt}^k are all the hypotheses in which the measurement j is associated to the track t.

The JPDAF fulfills the Markov assumption, therefore, association probabilities at the instant k are calculated only relative to the previous hypothesis $\theta(k-1)$. The probability of the hypothesis Θ_h^k given the set of measurements Z^k is:

$$P(\Theta_h^k | Z^k) = \frac{1}{\lambda} P(Z^k | \theta_h(k), \; \theta(k-1), \; Z^{k-1}) P(\theta_h(k) | \theta(k-1), \; Z^{k-1}) \tag{6.8}$$

where λ is a normalizing constant.

The term $P(\theta_h(k) | \theta(k-1), \; Z^{k-1})$ can be modeled as a constant. Then:

$$P(Z_k | \theta_h(k), \; \theta(k-1), \; Z^{k-1}) = P(Z_k | \theta_h(k)) = \prod_{j=1}^{N_k} P(z_j | \theta_h(k)) \tag{6.9}$$

and:

$$P(z_j | \theta_h(k)) = \begin{cases} P_F, & \text{if } z_j \text{ outside validation gate} \\ P_D P(z_j | \widehat{\mathbf{x}}_k^t), & \text{if } z_j \text{ inside validation gate} \end{cases} \tag{6.10}$$

where P_D is the detection probability, P_F the false alarm probability, and N_k is the number of timestamps available during the time interval k. A validation gate is defined such that only

measurements whose likelihood is above a certain threshold are used for updating the tracks. With Eq. 6.7, Eq. 6.8, and Eq. 6.9:

$$\beta_j^t = \frac{1}{\lambda} \sum_{\theta \in \Theta_{jt}^k} \prod_{j=1}^{N_k} P(z_j|\theta) \tag{6.11}$$

These association probabilities are used to weight the different measurements for each track. Each track is estimated using a Kalman filter, which is modified to consider the probabilities of each measurement belonging to it.

Kalman Joint Probabilistic Data Asociation Filter

The duration of every time window k corresponds to one time interval I. During this time, every microphone (or virtual microphone) is expected to receive either one or no timestamps. The state of the filter $\mathbf{x}_k^t$ of the track t contains the reception time τ_t and the relative velocity $v_{r,k}^t$ of the target during this specific time interval. Being Δt the elapsed time between two estimations, the acceleration model is:

$$\begin{bmatrix} \tau_k^t \\ v_{r,k}^t \end{bmatrix} = \mathbf{A} \begin{bmatrix} \tau_{k-1}^t \\ v_{r,k-1}^t \end{bmatrix} + \mathbf{G}\Phi_{k-1}^t, \qquad \Phi_{k-1}^t \sim \mathcal{N}(0, Q) \tag{6.12}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \tag{6.13}$$

$$\mathbf{G} = \begin{bmatrix} \frac{(\Delta t)^2}{2} \\ \Delta t \end{bmatrix} \tag{6.14}$$

Then, the motion model is used to predict the next state $\widehat{\mathbf{x}}_k^t$ and its covariance matrix $\widehat{\mathbf{P}}_k^t$:

$$\widehat{\mathbf{x}}_k^t = \mathbf{A}\mathbf{x}_{k-1}^t \tag{6.15}$$

$$\widehat{\mathbf{P}}_k^t = \mathbf{A}\mathbf{P}_{k-1}^t\mathbf{A}^T + \mathbf{G}Q\mathbf{G}^T \tag{6.16}$$

The sensor model predicts the measurement z_k^t as follows:

$$z_k^t = \tau_k^t + \rho_k^t, \qquad \rho_k^t \sim \mathcal{N}(0, R_m) \tag{6.17}$$

Therefore, in matrix notation:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{6.18}$$

The innovation vector is the difference between the predicted state and the measurement:

$$\nu_j^t = z_j - \mathbf{H}\widehat{\mathbf{x}}_k^t \tag{6.19}$$

and the covariance is:

$$\Lambda_k^t = \mathbf{H}\widehat{\mathbf{P}}_k^t\mathbf{H}^T + R_m \tag{6.20}$$

The Eq. 6.10 of the JPDAF can now be estimated by evaluating a normal distribution with zero mean and covariance Λ_k^t, due to the fact that its value is proportional to the probability $P(z_j|\widehat{\mathbf{x}}_k^t)$:

$$P(z_j|\widehat{\mathbf{x}}_k^t) \propto \mathcal{N}(\nu_j^t; 0, \ \Lambda_k^t) \tag{6.21}$$

Only values with a high likelihood $\mathcal{N}(\nu_j^t; 0, \ \Lambda_k^t)$ are used for estimation. With such n measurements, a new weighted innovation ν^t is calculated depending on the association probabilities of the track:

$$\nu^t = \sum_{j=1}^n \beta_j^t \nu_j^t \tag{6.22}$$

Then, instead of using a single innovation vector like the traditional Kalman filter, multiple innovation vectors are weighted according to their likelihood. The sensor model is used to correct the predicted state as follows:

$$\mathbf{K}_k^t = \widehat{\mathbf{P}}_k^t\mathbf{H}^T(\Lambda_k^t)^{-1} \tag{6.23}$$

$$\mathbf{x}_k^t = \widehat{\mathbf{x}}_k^t + \mathbf{K}_k^t\left(\nu^t\right) \tag{6.24}$$

The covariance matrix $\mathbf{P}_k^t$ is calculated as:

$$\mathbf{P}_k^t = \overline{\beta}^t\widehat{\mathbf{P}}_k^t + \left(1 - \overline{\beta}^t\right)\left(\mathbf{I} - \mathbf{K}_k^t\mathbf{H}\right)\widehat{\mathbf{P}}_k^t + \mathbf{K}_k^t\mathbf{P}_{\nu^t}(\mathbf{K}_k^t)^T \tag{6.25}$$

where:

$$\overline{\beta}^t = 1 - \sum_{j=1}^n \beta_j^t \tag{6.26}$$

$$\mathbf{P}_{\nu^t} = \sum_{j=1}^n \beta_j^t\nu_j^t\nu_j^{t^T} - \nu^t\nu^{t^T} \tag{6.27}$$

The probability of not having a measurement which belongs to t is then calculated by $\overline{\beta}^t$. If this value is one, the covariance of the state will be the predicted one, and the measurement model will not be considered. On the other hand, when this value is close to zero, the covariance is similar to a standard Kalman filter, with the addition of the estimated covariance of the weighted innovations.

When the covariance of a track becomes too large, the track is eliminated, as it is not anymore capable of finding the timestamps which belong to it. This will happen when a receiver does not receive any timestamp from a sender during a certain period of time. Then, the next received timestamp cannot be predicted reliably, as the target can be far from the position at which it was before. New tracks are continuously created with the unassociated timestamps.

6.3 Far-Field Assumption

Now, we assume the knowledge of the data association between timestamps and receivers, which can be either virtual or real. The positions of these receivers are assumed to be unknown. One could try to estimate the position of the target and the receivers with non-linear optimization, in a similar way as it has been done in previous chapters. If the number of constraints (measurements) is higher than the number of variables, the algorithm could succeed to find all the variables. However, estimating these variables is a challenging task even with highly overdetermined systems of equations, as the estimation is prone to local minima. This has been proved by multiple researchers [45, 47]. Moreover, as the target moves, the receivers which are in range change. This leads to tracks being created and eliminated. Then, one does not know which tracks belong actually to the same receiver. One only has a limited number of timestamps per receiver.

In order to overcome these issues, we use a far-field assumption, which simplifies the equations and provides an initial estimation for non-linear optimization algorithms [47]. Then, assuming the movement of the target during a short period of time is much smaller than the distance to the receivers, one can locate the target with a reduced number of timestamps per receiver. This fits our purpose. First, the number of timestamps per receiver is limited, which means the target did not move large distances. Moreover, the receivers are usually placed at the ceiling, at large distances from the sender. In addition, the reflections come after the line-of-sight signals, which means the distance to the virtual receivers will be even higher.

If we assume a target is moving from a certain point A to B during one time interval I with speed v, the movement of the target is assumed to be much lower than the distance from the sender to the receiver:

$$c(T_{l,i} - T_{k,i} - (l - k)I) \approx (l - k)vI \cos(\varphi) \tag{6.28}$$

where φ is the angle formed by the points A, B and the microphone position, being A the angle vertex. This assumption simplifies the equations and is a common situation in many real-life scenarios where the receivers are mounted at a very high altitude compared to the distance moved by the target during a few hundred milliseconds.

Ellipsoid Method

In one dimension, the movement needs to be coherent. Then, if a target moves towards the receiver i, and away from the receiver j, the velocity vector needs to have the opposite sign (see Fig. 6.2). In three dimensions, this becomes more complicated to visualize. However, using the far-field assumption, one can use the observed change of distances by the receivers during four sender positions to describe an ellipsoid [48]. If the target is moving in a two-dimensional scenario, an ellipse is described.

From now on, in order to simplify the explanation, we assume the elapsed intervals have been already subtracted from the reception times $T_{t,i}$.
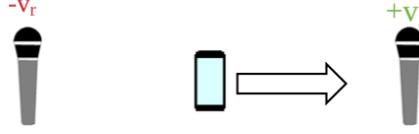
Figure 6.2: In one dimension, two receivers in opposite sides should observe the same relative velocity but with opposite sign. In two dimensions, an ellipse can be defined using multiple receivers. In three dimensions, an ellipsoid. [6]

Two Dimensions

Consider the target moves to three positions which form a triangle $\triangle \mathbf{S}_1 \mathbf{S}_2 \mathbf{S}_3$ in $\mathbb{R}^2$, where $d_1 = \|\mathbf{S}_1 - \mathbf{S}_2\|$ and $d_2 = \|\mathbf{S}_2 - \mathbf{S}_3\|$ denote the distances, and α denotes the angle $\angle_{\mathbf{S}_1 \mathbf{S}_2 \mathbf{S}_3}$ between the line segments $\mathbf{S}_1 \mathbf{S}_2$ and $\mathbf{S}_2 \mathbf{S}_3$

In the positions $\mathbf{S}_1$ $\mathbf{S}_2$ and $\mathbf{S}_3$, a signal is sent and is received by a certain receiver $\mathbf{M}_i$ at times $T_{i,1}, T_{i,2}$, and $T_{i,3}$, that yield the time differences:

$$\begin{aligned} \tau_1 &= T_{i,2} - T_{i,1} \\ \tau_2 &= T_{i,2} - T_{i,3} \end{aligned} \tag{6.29}$$

We define $\gamma_i = \gamma_1 + \alpha/2 = \gamma_2 - \alpha/2$ as the angle between $\mathbf{M}_i$ and the bisection of $\mathbf{S}_1 \mathbf{S}_2$ and $\mathbf{S}_2 \mathbf{S}_3$, as can be seen in Fig. 6.3. Then:

$$\begin{aligned} x_i &= c(T_{i,2} - T_{i,1}) = d_1 \cos(\gamma_1) = d_1 \cos(\gamma_i - \alpha/2) \\ y_i &= c(T_{i,2} - T_{i,3}) = d_2 \cos(\gamma_2) = d_2 \cos(\gamma_i + \alpha/2) \end{aligned} \tag{6.30}$$

Then, γ_i points to the receiver and α defines the moved angle by the target.
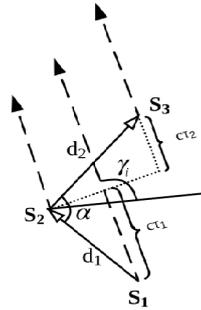
Figure 6.3: The target moves to the positions $\mathbf{S}_1$, $\mathbf{S}_2$ and $\mathbf{S}_3$. The received timestamps by every receiver form an ellipse that depends on α, d_1 and d_2. Figure adapted from [74].

Every receiver, virtual or real, will generate a point with coordinates x_i and y_i. The points of all the receivers form an ellipse in the form $ax^2 + by^2 + cxy = 1$. The relation between the three coefficients a, b, c, and the parameters d_1, d_2, and α, which uniquely determine the

triangle of senders $\triangle \mathbf{S}_1 \mathbf{S}_2 \mathbf{S}_3$ is:

$$d_1 = 2\sqrt{\frac{b}{4ab - c^2}} \tag{6.31}$$

$$d_2 = 2\sqrt{\frac{a}{4ab - c^2}} \tag{6.32}$$

$$\cos(\alpha) = \frac{-c}{2\sqrt{ab}} \tag{6.33}$$

Then, without knowing where the receivers are, one can find the parameters a, b and c in order to know how the target moved and at which angle were the receivers. Only three receivers are required.

It is quite common that a target is moving with a linear movement during three consecutive measurements. In this case, the ellipse degenerates in a line:

$$\begin{aligned} x_i &= d_1 \cos(\gamma_i) \\ y_i &= d_2 \cos(\gamma_i) \end{aligned} \tag{6.34}$$

Then, the slope m of the line will be

$$m = \frac{d_2}{d_1} \tag{6.35}$$

Then, linear movements will lead to lines. Also movements which are quasi-linear will lead to long ellipses. In these cases it is more challenging to estimate the parameters of the ellipse.

Three Dimensions

A target moving to four positions $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_3$ and $\mathbf{S}_4$ describes an ellipsoid. The angles are defined as in Fig. 6.4. The distances from the point $\mathbf{S}_1$ to the points $\mathbf{S}_2$, $\mathbf{S}_3$ and $\mathbf{S}_4$ are defined as d_B, d_C and d_D, respectively. The angles $\gamma_B, \gamma_C, \gamma_D$ are the angles between the point $\mathbf{S}_1$, the receiver $\mathbf{M}$ and the points $\mathbf{S}_2$, $\mathbf{S}_3$ and $\mathbf{S}_4$ respectively. The point $\mathbf{S}_1$ is the angle vertex. Then, having:

$$\begin{aligned} x_i &= d_B \cos(\gamma_B) \\ y_i &= d_C \cos(\gamma_C) \\ z_i &= d_D \cos(\gamma_D) \end{aligned} \tag{6.36}$$

the ellipsoid is in the form of:

$$(ax)^2 + (bx + cy)^2 + (dx + ey + fz)^2 = 1 \tag{6.37}$$

where:

$$a = \frac{1}{d_B} \qquad\qquad b = -\frac{\cos(\phi_C)}{d_B \sin(\phi_C)}$$

$$c = \frac{1}{d_C \sin(\phi_C)} \qquad\qquad e = \frac{\cos(\lambda_D)}{d_C \sin(\phi_C) \sin(\lambda_D)}$$

$$f = \frac{1}{d_D \sin(\phi_D) \sin(\lambda_D)} \qquad\qquad d = -\frac{cos(\phi_D)}{d_B sin(\phi_D) sin(\lambda_D)} -$$

$$\frac{\cos(\phi_C)\cos(\lambda_D)}{d_B \sin(\phi_C)\sin(\lambda_D)}$$

In order to fit the ellipsoid to the data we use constrained optimization. This was not necessary in [48], as the microphones and speakers were assumed to be well distributed in the three-dimensional space. In our approach [6], the sender can move in a quasi-linear trajectory and the receivers do not need to be perfectly distributed over the three-dimensional space.
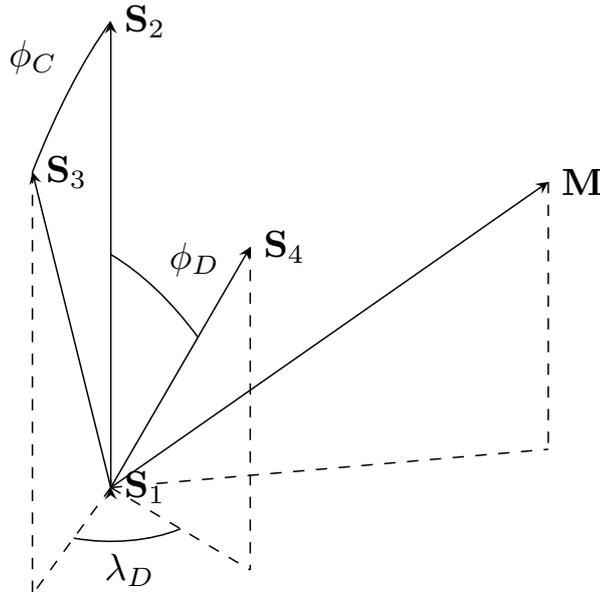
Figure 6.4: The target moves to the positions $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_3$ and $\mathbf{S}_4$ sending a signal in each of that positions. The signals are received by the microphone $\mathbf{M}$. Using the timestamps of multiple receivers, an ellipsoid is defined. [6]

Least-squares optimization can erroneously result in another shape such as an hyperboloid or a paraboloid. In order to avoid this, we use the constraint presented by Li and Griffiths in [75]. Having an ellipsoid such as

$$ax^2 + by^2 + cz^2 + 2dxy + 2exz + 2fyz + 2gx + 2hy + 2iz = 1 \qquad (6.38)$$

then:

$$j = a + b + c \tag{6.39}$$

$$k = ab + bc + ac - f^2 - e^2 - d^2 \tag{6.40}$$

$$4k - j^2 > 0 \tag{6.41}$$

The algorithm used is interior point optimization. This algorithm optimizes the squared error function with an additional logarithmic barrier function which approaches to infinity when the estimation does not meet the constraints.

Affine Geometry

One can express the far-field assumption as a function of the projection of the velocity vector over the vector defined by the sender and the receiver positions:

$$c(T_{k,i} - T_{k-1,i}) \approx \mathbf{v}I \cdot \frac{\mathbf{M}_i - \mathbf{S}_{k-1}}{\|\mathbf{M}_i - \mathbf{S}_{k-1}\|} \tag{6.42}$$

where $\mathbf{v}$ is the velocity vector and $\cdot$ denotes the dot product. Having a reference position $\mathbf{S}_0$ which is assumed to be at the origin of the coordinate system, the time difference between the timestamp received at that position and the timestamp received at the position $\mathbf{S}_k$ is:

$$T_{k,i} - T_{0,i} = \frac{1}{c}\left(\mathbf{S}_k \cdot \frac{\mathbf{M}_i}{\|\mathbf{M}_i\|}\right) \tag{6.43}$$

Having N receivers and B sender positions, we define a measurement matrix $\mathbf{L}$:

$$\mathbf{L} = \begin{bmatrix} T_{1,0} - T_{0,0} & \dots & T_{1,N-1} - T_{0,N-1} \\ T_{2,0} - T_{0,0} & \dots & T_{2,N-1} - T_{0,N-1} \\ \vdots & \ddots & \vdots \\ T_{B-1,0} - T_{0,0} & \dots & T_{B-1,N-1} - T_{0,N-1} \end{bmatrix} \tag{6.44}$$

The matrix $\mathbf{L}$ can be expressed as the multiplication of two matrices $\mathbf{X}\boldsymbol{\Omega}$:

$$\mathbf{L} = \begin{bmatrix} \mathbf{S}_1 \\ \vdots \\ \mathbf{S}_{B-1} \end{bmatrix} \begin{bmatrix} \frac{\mathbf{M}_0}{\|\mathbf{M}_0\|} & \cdots & \frac{\mathbf{M}_{N-1}}{\|\mathbf{M}_{N-1}\|} \end{bmatrix} \tag{6.45}$$

Then, by doing a singular value decomposition of $\mathbf{L}$ the result is $\mathbf{U}\mathbf{V}\mathbf{W}^T$. The three greater eigenvalues correspond to the three dimensions of the senders and receivers. The other eigenvalues would be zero in an ideal case, but this is often not true due to the noise and violations of the far-field assumption. The matrix $\mathbf{X}$ can be then estimated by $\mathbf{U}\mathbf{V}$ and the matrix $\boldsymbol{\Omega}$ by $\mathbf{W}^T$. In both cases, only the dimensions corresponding to the three larger eigenvalues are taken. Nonetheless, the matrix $\mathbf{X}$ is not euclidean, as no constraints have been imposed to it. In

the original approach [47], an auxiliary matrix $\mathbf{C}$ is used, which minimizes the error between the estimation and the euclidean version of it. Then, the euclidean version of $\mathbf{X}$ is found by $\mathbf{UVC}^{-1}$ and the euclidean version of $\mathbf{\Omega}$ as $\mathbf{CW}$. In [47] they present it in two dimensions. The equivalent in three dimensions would be a 3×3 matrix, which would mean estimating nine variables, requiring at least nine timestamps per position. If the matrix is assumed to be symmetric, then six timestamps are required. One has to realize that this means at least six receivers are needed for estimating such values. Moreover, it is a nonlinear optimization problem which can get stuck in local minima.

In [6] we proceed with a different strategy to recover the euclidean constraints. We know that $\frac{\mathbf{M}_i}{\|\mathbf{M}_i\|}$ is a normalized vector. Therefore, first we take the dimensions corresponding to the three largest eigenvalues of $\mathbf{W}$ and normalize the columns of the estimated $3 \times N$ matrix, getting $\mathbf{\Omega_n}$. Afterwards, we know that:

$$\mathbf{L} = \mathbf{X_n}\mathbf{\Omega_n} \tag{6.46}$$

Then, in order to find $\mathbf{X_n}$ we use the pseudo-inverse:

$$\mathbf{X_n} = \mathbf{L}\mathbf{\Omega_n}^T(\mathbf{\Omega_n}\mathbf{\Omega_n}^T)^{-1} \tag{6.47}$$

The number of variables must be lower or equal than the number of constraints. In a three dimensional space:

$$BN \geq 3(B - 1) + 3N \tag{6.48}$$

which simplifies to:

$$N \geq \frac{3B - 3}{B - 3} \tag{6.49}$$

Then, the minimum number of receivers is four, two less than using the matrix $\mathbf{C}$.

While using the auxiliary matrix $\mathbf{C}$ aims to find the optimum rotation and translation using nonlinear optimization, the presented approach finds a sub-optimal euclidean version of $\mathbf{\Omega}$, by normalizing its columns. This requires less receivers and does not rely on nonlinear optimization. Moreover, this approach allows the receivers to be affected by different sources of error, whereas with the matrix $\mathbf{C}$ all receivers are normalized in the same manner. This is important in acoustics, as sound can experience different measurement noises on different receivers depending on the traveled path (e.g. due to different temperatures at which the sound travels, or whether there is a reflection involved). Moreover, as we show in the Appendix of this thesis, the error generated by assuming far-field will be different for every receiver, as it depends on how the sender moves relative to them and how far they are.

To summarize, the ellipsoid method uses the equality $\mathbf{F}^T\mathbf{L}^T\mathbf{L}\mathbf{F} = 1$, where $\mathbf{F}$ contains the parameters of the ellipsoid, which contain information about the movement of the target. This is done for four sender positions using constrained least squares. On the other hand, the affine geometry approach aims to find the solution to $\mathbf{L} = \mathbf{X}\mathbf{\Omega}$, where $\mathbf{X}$ has the sender positions and $\mathbf{\Omega}$ the unitary vectors corresponding to the receivers.

6.4 Simulations and Experiments

First, we do a simulation to compare the performance of the different algorithms here exposed. The sender is assumed to move to the following points (x,y,z): $(0,0,0)$, $(\varepsilon,0,0)$, $(0,\varepsilon,0)$ and $(0,0,\varepsilon)$, where ε is equal to $0.3\,$m (see Fig. 6.5). Nine receivers are randomly placed at the same distance from the origin, four meters.
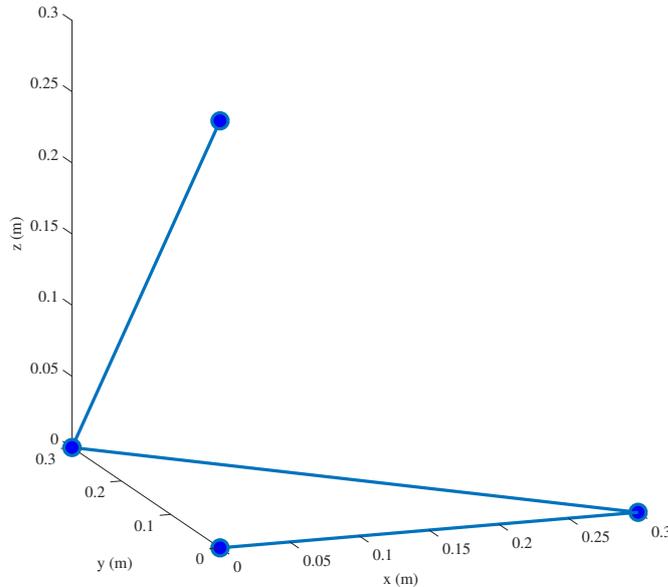
Figure 6.5: Simulated positions of the target for evenly distributed positions in the three-dimensional space. The achieved errors can be seen in Fig. 6.6.

In order to test how resilient are the different algorithms to noise, Gaussian noise is added to the measurements. The median error achieved by every algorithm in presence of different magnitudes of error can be seen in Fig. 6.6. Every point is the median of 1000 estimations. One can observe how the new method using affine geometry is more resilient to noise, although it performs worse than the others when the standard deviation of the noise is below $0.07\,$ms. When the noise is greater than this value, the algorithms get stuck in local minima estimating the matrix $\mathbf{C}$ or do not succeed estimating the ellipsoid parameters. On the other hand, when the noise is below this value, one can find the optimal matrices which provide a lower error than normalizing $\mathbf{\Omega}$. Another interesting observation is that the traditional affine geometry method and the ellipsoid method obtain a similar result. This is due to the fact that the sender is perfectly distributed in the three-dimensional space, which facilitates the task of fitting an ellipsoid to the data. In addition, one can see how these methods can be used as an initial estimate for a local optimization algorithm, such as gradient descent.

In order to see the effect of the distribution of sender positions in the final estimation, we perform another simulation with a more realistic sender movement (see Fig. 6.7). The result can be seen in Fig. 6.8. As expected, the ellipsoid method is the method which is more affected
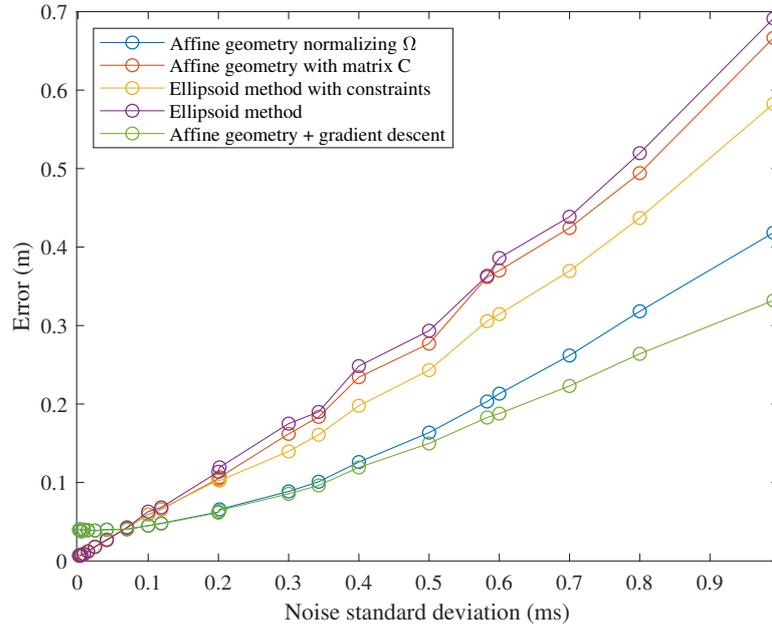
Figure 6.6: Median error of the presented algorithms when a target is moving in four positions aligned to the x, y and z axis. One can observe how the ellipsoid method and the method presented by Thrun [47] present similar performance, as the ellipsoid generated is well spread around the three coordinates. Moreover, one can observe how the methods presented in this chapter present an improvement when the standard deviation of the measurement noise is above 0.1 ms. [6]

by the quasi-linear movement. It is interesting to observe also how in this case normalizing Ω provides the best result even with a noise standard deviation below 0.05 ms.
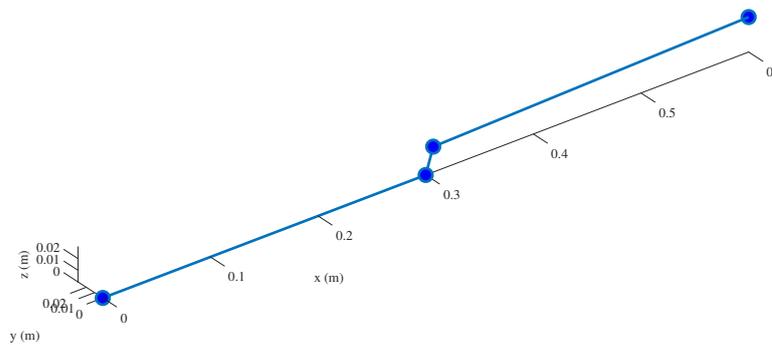
Figure 6.7: Simulated positions of the target for quasi-linear movements. The positions of the target are not well distributed in the three-dimensional space, which means the movement in certain dimensions can be masked by the noise. The achieved errors can be seen in Fig. 6.8.

In order to test the localization error in a real environment, we use the ASSIST system. In total 12 receivers are used, above a surface of approximately 15 m × 11 m. Nine receivers
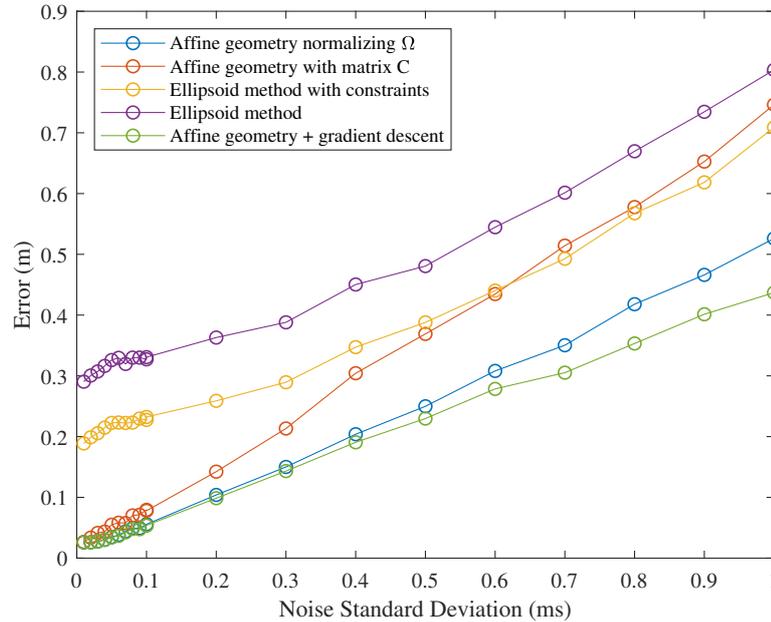
Figure 6.8: Error for the simulated positions in Fig. 6.7. One can observe how the ellipsoid method presents the worse performance, as the long-thin ellipsoids generated by the movement of the target are not properly estimated. This is because a small change in the estimated points of the ellipsoid can lead to a high change in the estimated ellipsoid parameters.

are at 4.9 m height and three of them at 3.4 m height. In addition, we use the motion capture system as a reference. The microphone positions are precisely measured using a total station theodolite. The estimated positions are rotated and translated to the same coordinate system as the reference system. The motion capture system and therefore the estimated positions cover only a section of the hall surface. The localization area has tables and wooden walls which emulate a real environment. More details about the experimental environment can be seen in the Appendix of this thesis.

First we test the capability of the proposed approach to find the target in three dimensions using only four receivers, as with the standard method using the auxiliary matrix $\mathbf{C}$ was not possible. The result can be seen in Fig. 6.9. The error has a mean of 0.20 m and a standard deviation of 0.10 m in two dimensions and a mean of 0.31 m and standard deviation of 0.14 m in three dimensions. In this experiment, groups of 30 samples are aligned to the same coordinate system by using samples that belong to more than one block. Due to this fact, certain parts of the trajectory have higher error than others.

In order to test the capability of the proposed approach to track a target in a large trajectory we use a block of 80 positions and seven receivers. The result can be seen in Fig. 6.10. The comparison of the estimation errors can be seen in Table 6.1. It can be seen that normalizing Ω results in a lower error than using the auxiliary matrix $\mathbf{C}$.

For a better comparison of the methods presented in this chapter, we use the algorithms with
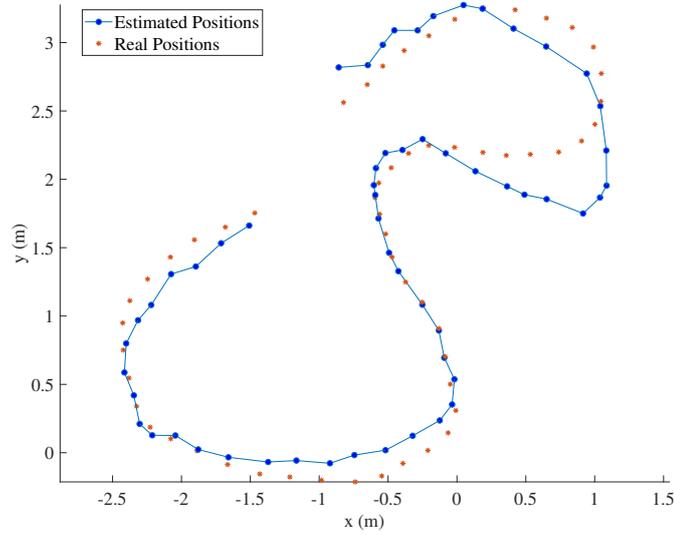
Figure 6.9: Estimated positions using only four receivers and groups of 30 samples. The localization error has a mean of 0.31 m and a standard deviation of 0.14 m in three dimensions. [6]

	Median Error in 3D (m)	Standard Deviation in (m)	Median Error in 2D (m)	Standard Deviation in 2D(m)
Normalizing Ω	0.12	0.08	0.09	0.07
No Normalization	1.31	0.38	1.31	0.39
Using Matrix **C**	0.36	0.19	0.35	0.19

Table 6.1: Comparison of the localization errors when estimating 80 sender positions and using the measurements of seven receivers. [6]

2894 groups of four timestamps. The cumulative distribution function when the target moved at least one centimeter can be seen in Fig. 6.11. The median movement $d_1 + d_2 + d_3$ is 0.71 m. The proposed changes in this chapter prove to be effective. Using constrained optimization for the ellipsoid method improves the result and normalizing the matrix Ω has lower error than using the matrix **C**.

Using affine geometry the error is lower than using the ellipsoid method, as fitting an ellipsoid to the data when the target is moving in a line or the receivers are not well distributed is a challenging task. The ellipsoid method is, however, an elegant method to display the movement of the target and can achieve a low error when the receivers are well distributed and the sender positions are not in a line. In Fig. 6.12 the estimated ellipsoid for a sender moving in three dimensions is shown. The mean error of the eleven points with the ellipsoid is 0.12 m and the standard deviation is 0.10 m. The estimated sender positions had a mean error of 0.17 m and the distance moved was estimated with only 0.01 m error.
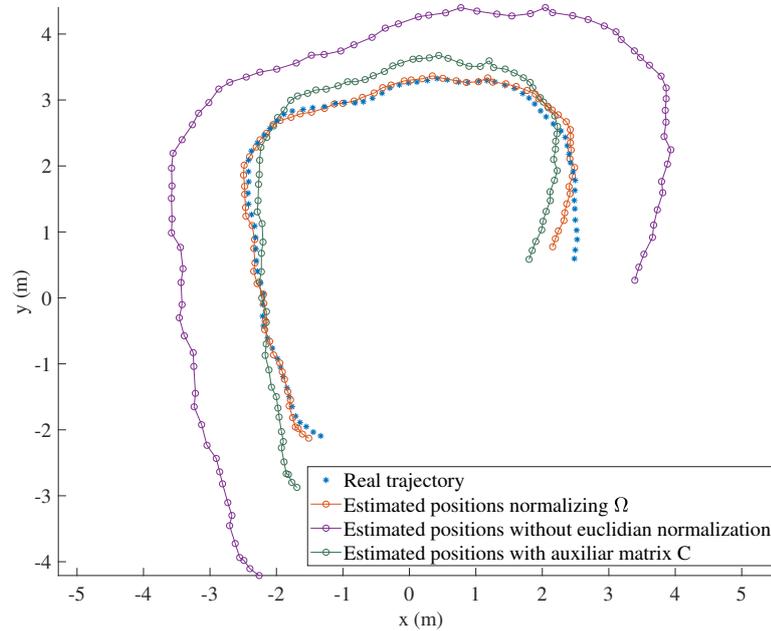
Figure 6.10: Estimated positions using a block of 80 sender positions and seven receivers. Normalizing Ω the lowest error is achieved. [6]
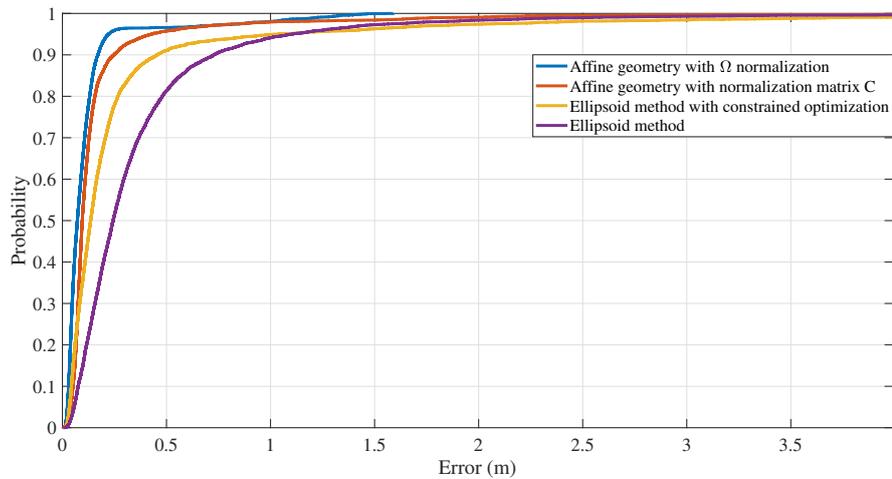
Figure 6.11: Cumulative distribution function of error of the different algorithms estimating the sender position in three dimensions. [6]

Another experiment is done in order test the capability of the presented affine geometry method to estimate the receiver positions. For this experiment we assume the sending times are known. If this is not the case, one would only know the angles at which the receivers are. One could then make use of other sender positions in other regions of the localization area or could make use of the knowledge of the height of the receivers. If the heights of the
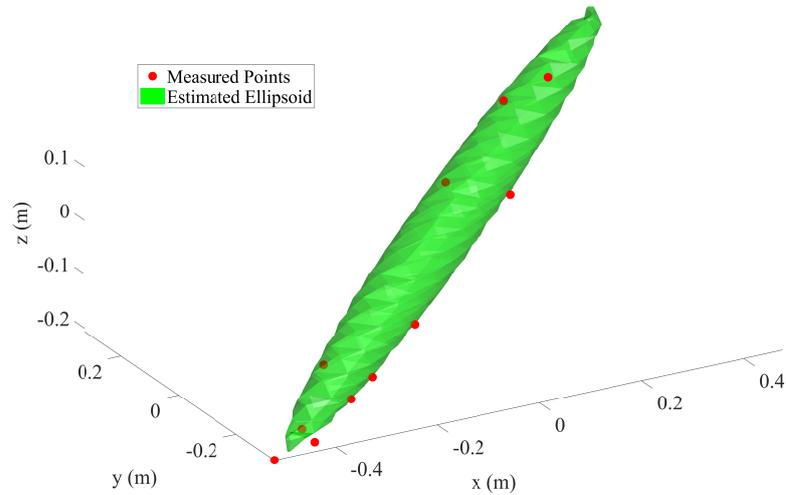
Figure 6.12: One can see how a speaker moving to four positions generates an ellipsoid in three dimensions. The mean error of the eleven points with the ellipsoid is 0.12 m and the standard deviation is 0.10 m. [6]

receivers are known, the estimated angles would intersect in one point with the plane located at the height of the receiver. The sending times used in this experiment are estimated using the reference positions provided by the motion capture system. The result can be seen in Fig. 6.13 and Fig. 6.14. One can see how the target positions are estimated with a median error of 0.04 m and standard deviation 0.02 m. Using the knowledge of the sending time, one can observe how two of the receivers used are actually virtual receivers. They are clearly second reflections from the ceiling and the ground. We suspect the reason why the first reflections are not present is because the receivers are placed on metal bars which block the signals which come from the ceiling. However, they are strong reflections, as the speaker was pointing upwards. The error in the position of the virtual receivers is 1.69 m and 2.68 m. The height error is 0.98 m and 0.75 m. The positions of the LOS receivers are estimated with a median error of 0.62 m in two dimensions and 0.88 m in three dimensions. The standard deviation is 0.24 m and 0.17 m respectively. Regarding the angle error, the angle to the LOS receivers is estimated with a median error of 3.55° and the angle to the virtual receivers with a median error of 5.78°.

In conclusion, the presented modifications of existing methods prove to achieve better results for three dimensional acoustic localization than the state-of-the-art, especially when the sender positions are not well distributed in the three-dimensional space. In the case of affine geometry, the presented approach requires two receivers less than the state of the art, locating a target with a mean error of 0.31 m and a standard deviation of 0.14 m.

The ellipsoid method is the method which performs worse when the target positions are not well distributed. Using constrained optimization the error is reduced, although it is still higher than the other approaches. This is because, as explained in [75], the constraint used in this chapter does not include a subset of ellipsoids which are long-thin or compressed. An algorithm to consider also this subset is presented in [75] which could be used to achieve better results.
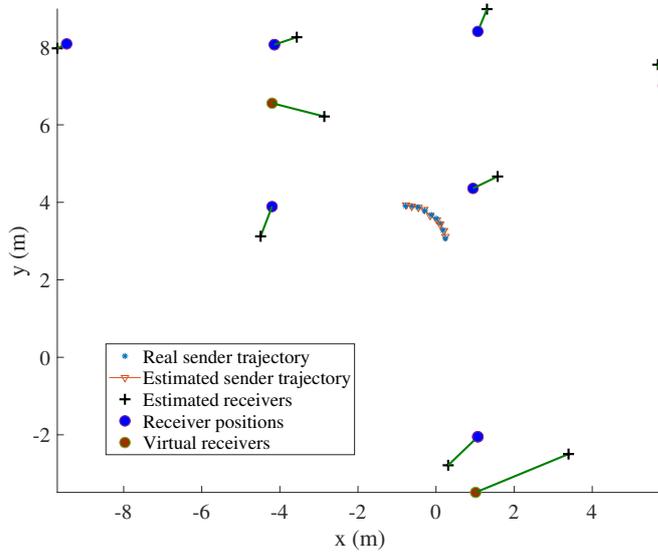
Figure 6.13: Real and estimated sender and receiver positions. The positions of the LOS receivers are estimated with 0.62 m median error in two dimensions. The result in three dimensions is shown in Fig. 6.14. [6]
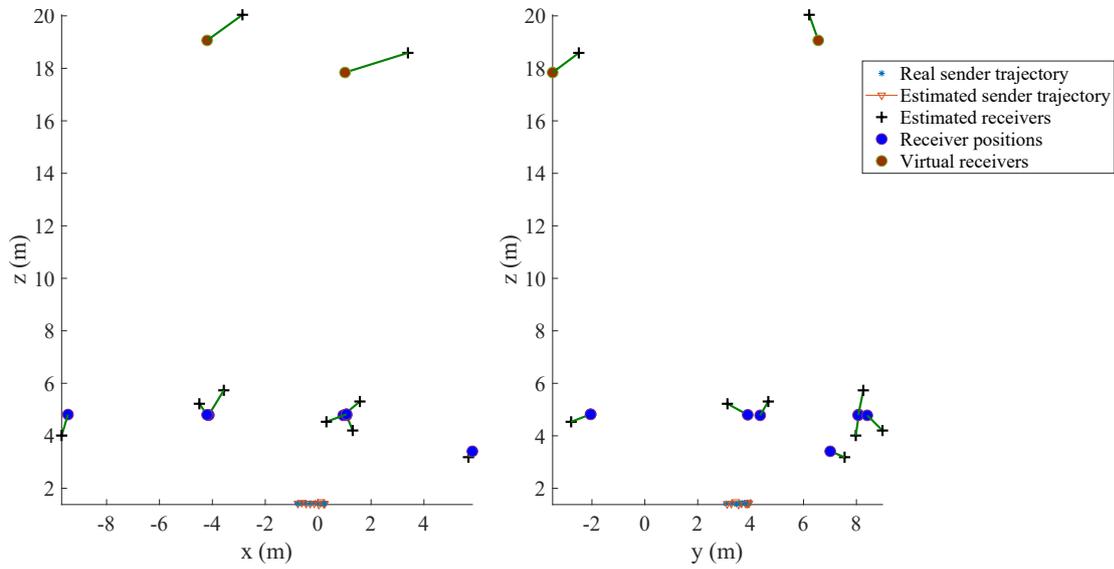
Figure 6.14: Real and estimated sender and receiver positions. The positions of the LOS receivers are estimated with 0.88 m median error in three dimensions. One can see how there are two virtual receivers due to the echoes from the ground and the ceiling. The result in two dimensions is shown in Fig. 6.13. Figure adapted from [6].

7 Using Receiver Movement for Identification

In Chapter 4 we have shown how a moving receiver can be located using static speakers. This has the benefit that the number of targets that can be tracked is unlimited, as they can estimate their own position independently of the other users. The main drawback is that, when the target is moving, only one timestamp is received per position, which is less than the number of constraints required to estimate a position using TDOA or TOA. Having only one timestamp per position, it is challenging to know whether the timestamp is a reflection or a line-of-sight signal. In order to do so, we exploit the movement of the target to predict the most likely line-of-sight signals, similar as it is done in the previous chapter. However, in this case every target position corresponds to an emitted signal by a different sender (see Fig. 7.1). Therefore, one can only measure the relative velocity of the target to a certain sender every time that it emits a signal. In order to predict the relative movement of the target and do a proper data association, one needs to fuse data from different senders.

In this scenario there is an added challenge which is how to identify which signal comes from which sender. Multiple senders are emitting signals and one needs to know which signal comes from which emitter and whether they are line-of-sight signals. The approach presented in this chapter fuses the identification that is provided by every sender and the predicted position and velocity of the target. Every speaker encodes an id and emits signals with a different time period. Then, fusing this information and the estimated velocity of the target one can decide which timestamps belong to which emitter and whether they are line-of-sight signals or reflections (see Fig 7.2).

In this chapter we extend the system presented in Chapter 4 so that the system can be robust in mixed line-of-sight/non-line-of-sight environments and is capable of locating a target moving at higher velocities.

7.1 Problem Formulation

A receiver moves continuously in a three-dimensional space and B static senders are installed on the ceiling. The height of the receiver and the position of the senders are assumed to be known. Additionally, we assume the time offset between the senders has been already initialized with the stop-and-go motion explained in Chapter 4.

Moreover, as explained in Chapter 3 and 4, the senders emit signals with different time intervals. In addition, every sender has a unique ID, which is modulated using a $\pi/4$-DQPSK modulation. This means one symbol is equivalent to two bits of data. Every signal emitted by a sender contains four symbols and therefore a total of eight bits. Three of them are used for
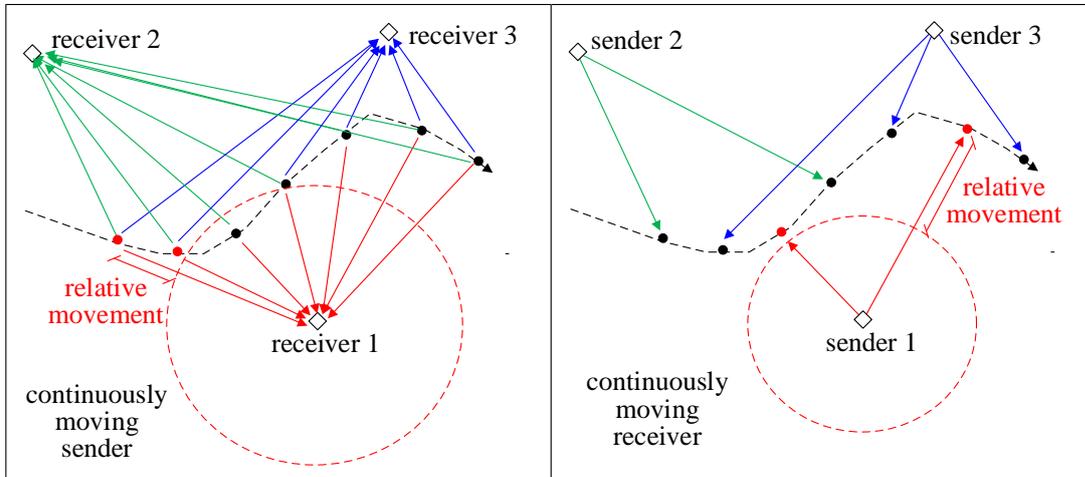
Figure 7.1: When a speaker is moving, the relative movement to an anchor in range can be measured every time the target sends a signal (left figure). On the other hand, when a receiver is moving, the relative motion to an anchor node can only be measured when the anchor emits a signal (right figure). In this chapter we focus on the second case. We show how the position and velocity estimated using the measurements from one sender can be used to predict the next timestamp from another sender.
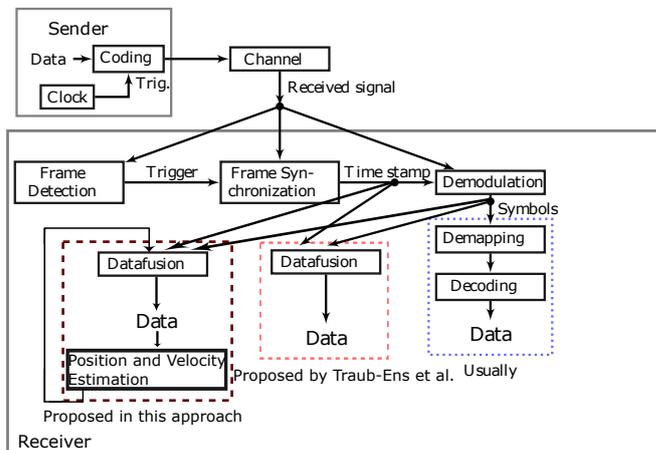
Figure 7.2: In this chapter the estimated position and velocity of the target are used to estimate the most reliable timestamps and IDs. In [76] and in Chapter 4 the interval diversity and the received symbols are used to identify the senders, independently from the position and velocity estimations. [3]

error correction, which means there are 32 possible identification numbers.

Fusing the different time intervals I_j ($1 \leq j \leq B$) with the symbol information has been proven to be more effective than just using the symbol information [77]. However, as we have seen in the previous chapters, when a target moves, the observed time period by the receiver is different than when it is standing.

In [77], a FDOA value is calculated with the information of the current timestamp n for the ID number j, the latest received timestamp $k_j - 1$ and the time interval I_j as follows:

$$\tau_{\text{FDOA},\,n}(j) = T_{k_j j} - T_{(k-1)_j j} - I_j \tag{7.1}$$

This value is assumed to be zero if the received timestamps $T_{k_j j}$ and $T_{(k-1)_j j}$ come from the same sender j. Therefore, it is using the information about the known interval I_j to guess from which sender the timestamps have been emitted. However, note that this value will be zero only if the receiver is not moving or if the current position of the target $\mathbf{M}_t$ is exactly the same as it was in $T_{(k-1)_j j}$. If the target moved, the value of $\tau_{\text{FDOA},\,n}(j)$ will be:

$$\tau_{\text{FDOA},\,n}(j) = \frac{1}{c}\|\mathbf{M}_t - \mathbf{S}_j\| - \frac{1}{c}\|\mathbf{M}_{(k-1)_j j} - \mathbf{S}_j\| \tag{7.2}$$

where $\mathbf{M}_{(k-1)_j j}$ is the latest position with a measurement of sender j. Therefore, the movement of the target can lead to a wrong identification of the sender and can also deteriorate the LOS signal selection, especially when the target is moving at high velocities. For instance, if the receiver moves towards the sender, Eq. 7.1 would be zero for a later reception time than the actual one. Then, a reflection would be easily mistaken for a line-of-sight signal, as it would arrive later.

As mentioned in Chapter 4, the position and velocity of the moving receiver can be tracked with an unscented Kalman filter. Then, using this information one can have a better prediction of the next timestamp emitted by the sender j than just using Eq. 7.1.

In [3] we use the latest estimated velocity $\mathbf{V}_{t-1}$ and estimated position $\mathbf{M}_{t-1}$ by the unscented Kalman filter to predict the position of the target $\widehat{\mathbf{M}}_t$ in the current time instant:

$$\widehat{\mathbf{M}}_t = \mathbf{M}_{t-1} + h_t \mathbf{V}_{t-1} \tag{7.3}$$

Note that $\mathbf{M}_{t-1}$ can be estimated with the timestamp of another sender (see Fig. 7.1). This means that, as the receiver is assumed to receive constantly signals from different senders, the position and velocity estimates are updated frequently enough so that the current position of the receiver $\widehat{\mathbf{M}}_t$ can be predicted reliably.

Then, instead of using Eq. 7.1 we propose to use:

$$\tau_{\text{FDOA},\,n}(j) = T_{k_j j} - T_{(k-1)_j j} - I_j - \left(\frac{1}{c}\|\widehat{\mathbf{M}}_t - \mathbf{S}_j\| - \frac{1}{c}\|\mathbf{M}_{(k-1)_j j} - \mathbf{S}_j\| \right) \tag{7.4}$$

At the risk of abusing the notation, in this case $\mathbf{M}_{(k-1)_j j}$ refers to an estimation and not the actual position of the target. This estimation is done by the unscented Kalman filter.

The FDOA value is a real number $\tau_{\text{FDOA},n} \in \mathbb{R}$ and the received symbols are in the symbol space $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^4 : \|x\|_\infty < \pi\}$. Hence, before fusing the data, the FDOA values are mapped by a function $f_{\text{map}} : \mathbb{R} \to \mathbb{S}$ into the symbol space. Then the mapped timestamps and the symbols can be fused. This procedure was already defined in [77]. We reproduce it here for a better understanding of the algorithm.

The timestamps are mapped into the symbol space by the function $f_{\text{map}} : \mathbb{R} \to \mathbb{S}$. The time intervals I_m are defined as:

$$I_m = \tau_{min} + m\tau_d \tag{7.5}$$

where m is the id of the sender. The values of τ_{min} and τ_d are constant and can be set depending on how often the target needs to be located.

The FDOA value $\tau_{\text{FDOA},n}$ estimated using Eq. 7.4 is normalized using the difference between two intervals:

$$r_{\text{TS}} = \frac{\tau_{\text{FDOA},n}}{\tau_d} \tag{7.6}$$

Then, the value of r_{TS} will be zero if the sender j has the id m. However, if it is close to one it will mean that the sender is likely to have the id $m+1$ and if it is close to -1 is likely to be $m-1$. This can be used to map $\tau_{\text{FDOA},n}$ to the symbol space:

$$\mathbf{s}_{\text{TS}} = f_{\text{map}}(m, r_{\text{TS}}) = \mathbf{s}_{\text{m}} + |r_{\text{TS}}| \cdot \begin{cases} \mathbf{s}_{\text{m}-1} - \mathbf{s}_{\text{m}} & r_{\text{TS}} < 0 \\ \mathbf{s}_{\text{m}+1} - \mathbf{s}_{\text{m}} & \text{else} \end{cases} \tag{7.7}$$

The vector $\mathbf{s}_{\text{TS}}$ points to the direction of the nearest ID. The ID is always taken modulo 32 (ID=-1 is ID=31 and ID=32 is ID=0) and the subtraction is phase corrected by

$$\phi = \begin{cases} \phi - 2\pi & \phi > \pi \\ \phi + 2\pi & \phi < -\pi \\ \phi & \text{else} \end{cases} \tag{7.8}$$

7.2 Data Fusion

We propose an approach [3] where the position and velocity estimated by the unscented Kalman filter are used to predict the most likely line-of-sight signal emitted by every sender. Then, we have B Kalman filters, each one predicting the next line-of-sight timestamp of every sender. The received timestamps which are closer to the predictions are used to estimate the position and velocity of the target, which are used again to predict the next timestamps (see Fig. 7.3). Every sender has a Kalman filter with a state vector $\mathbf{x_n} = [T_e \ \mathbf{s}_{\text{DF}}]^T$ which contains the time $T_e \in \mathbb{R}$ and the fused ID symbols $\mathbf{s}_{\text{DF}} \in \mathbb{S}$. The time T_e is the estimated reception time and the symbols contain the ID of the sender. Note that we are not estimating the relative velocity of the target in the Kalman filters, as it was done in the previous chapter. The reason is that we are using the predicted position and velocity from the UKF to do so, as it is updated more frequently (see Fig. 7.1).
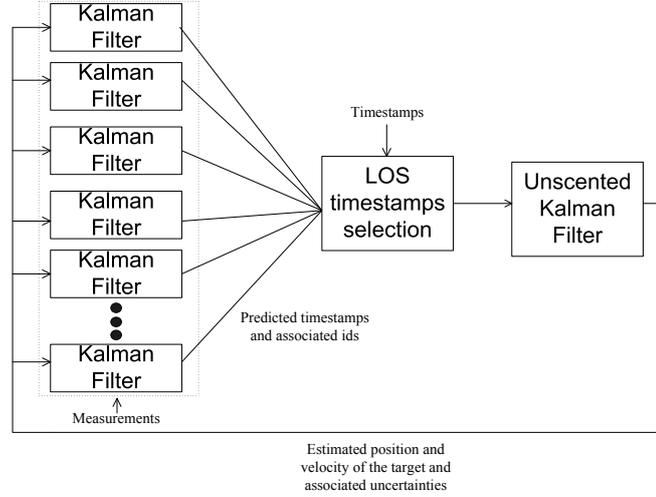
Figure 7.3: Each Kalman filter of the bank of Kalman filters tracks a signal of an specific sender. Using the predicted timestamps, the LOS signals are selected and forwarded to the Unscented Kalman filter, which estimates the position and velocity of the target. These estimations are used to predict the next LOS signals. [3]

One can predict the next timestamp that will be received $\widehat{T_{e,n}}$ using Eq. 7.4 as:

$$\widehat{T_{e,n}} = T_{e,n-1} + I_j + \left(\frac{1}{c} \|\widehat{\mathbf{M}}_t - \mathbf{S}_j\| - \frac{1}{c} \|\mathbf{M}_{(k-1)_j j} - \mathbf{S}_j\| \right) \tag{7.9}$$

where $T_{e,n-1}$ is the previously estimated reception time. Note that the position of the target $\mathbf{M}_{(k-1)_j j}$ has been estimated already by the UKF and the current position $\widehat{\mathbf{M}}_t$ can be estimated using the current estimate of the UKF:

$$\widehat{\mathbf{M}}_t = \mathbf{M}_{t-1} + h_t \mathbf{V}_{t-1} \tag{7.10}$$

where h_t is the elapsed time between this estimation and the previous one. Then, one can rewrite the predicted time $\widehat{T_{e,n}}$ as a function of two UKF estimates:

$$\widehat{T_{e,n}} = T_{e,n-1} + I_j + g_t(\boldsymbol{\chi}_{t-1}) + f_t(\boldsymbol{\chi}_j) \tag{7.11}$$

being $\boldsymbol{\chi}_{t-1}$ the latest state estimated by the unscented Kalman filter and $\boldsymbol{\chi}_j$ the latest estimated UKF state with the sender j. The function g_t can be written as:

$$g_t(\boldsymbol{\chi}_{t-1}) = \left(\frac{1}{c} \|\mathbf{M}_{t-1} + h_t \mathbf{V}_{t-1} - \mathbf{S}_j\| \right) \tag{7.12}$$

The function f_t can be written as:

$$f_t(\boldsymbol{\chi}_j) = -\frac{1}{c} \|\mathbf{M}_{(k-1)_j j} - \mathbf{S}_j\| \tag{7.13}$$

Then, the process model can be formulated in matrix notation:

$$\widehat{\mathbf{x}}_n = \mathbf{A}\mathbf{x}_{n-1} + \mathbf{B}\mathbf{u}_n + \begin{bmatrix} g_t(\boldsymbol{\chi}_{t-1}) + f_t(\boldsymbol{\chi}_j) \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{7.14}$$

The first row, which is the estimated reception time, is propagated with the estimated UKF states and the control vector $\mathbf{u}_n$. The next four rows of the state, which contain the symbol information, are predicted without any changes. More in detail, the matrix $\mathbf{A}$ represents the transition model and is set to an identity matrix $\mathbf{A} = \mathbf{I}_5$. The matrix $\mathbf{B}$ represents the control input model, which is also set to an identity matrix $\mathbf{B} = \mathbf{I}_5$, with the control vector $\mathbf{u} = [I_j\ 0\ 0\ 0\ 0]^{\mathrm{T}}$. Then, the next timestamp is assumed to be received after one interval I_j plus the corrections due to the movement of the target, which are done using the UKF estimations.

As we are using the estimations from the UKF we need to consider the uncertainty of these variables. The functions g_t and f_t are not linear, one needs to linearize them in order to propagate the uncertainty of the variables $\boldsymbol{\chi}_{t-1}$ and $\boldsymbol{\chi}_j$ to the uncertainty of $\widehat{\mathbf{x}}_n$. We linearize them by using the Jacobian matrix, as it is done in the extended Kalman filter. Then, every column will have the derivative of the function Eq. 7.12 with respect to a variable of the UKF state. As g_t is only used for estimating $\widehat{T_{e,n}}$ and not the symbols, the latest four rows of the matrix will contain zeros:

$$\mathbf{J}_{t-1} = \begin{bmatrix} \frac{\partial g_t}{\partial \mathbf{M}_{t-1}} & \frac{\partial g_t}{\partial \mathbf{V}_{t-1}} & \frac{\partial g_t}{\partial t_{k_1 1}} & \frac{\partial g_t}{\partial \delta_2} & \cdots & \frac{\partial g_t}{\partial \delta_m} \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix} \tag{7.15}$$

The derivative with respect to the offsets δ_j and sending time $t_{k_1 1}$ is zero. The derivative with respect to the receiver position $\mathbf{M}_{t-1}$ and velocity $\mathbf{V}_{t-1}$ are:

$$\frac{\partial g_t}{\partial \mathbf{M}_{t-1}} = \frac{\mathbf{M}_{t-1} + h_t \mathbf{V}_{t-1} - \mathbf{S}_j}{\|\mathbf{M}_{t-1} + h_t \mathbf{V}_{t-1} - \mathbf{S}_j\|} \tag{7.16}$$

$$\frac{\partial g_t}{\partial \mathbf{V}_{t-1}} = h_t \frac{\mathbf{M}_{t-1} + h_t \mathbf{V}_{t-1} - \mathbf{S}_j}{\|\mathbf{M}_{t-1} + h_t \mathbf{V}_{t-1} - \mathbf{S}_j\|} \tag{7.17}$$

Now we need to find a matrix which relates $\boldsymbol{\chi}_j$ and $\widehat{\mathbf{x}}_n$, as Eq. 7.13 is also not linear. The Jacobian matrix of f_t has the same structure of Eq. 7.15:

$$\mathbf{D}_j = \begin{bmatrix} \frac{\partial f_t}{\partial \mathbf{M}_j} & \frac{\partial f_t}{\partial \mathbf{V}_j} & \frac{\partial f_t}{\partial t_{k_1 1}} & \frac{\partial f_t}{\partial \delta_2} & \cdots & \frac{\partial f_t}{\partial \delta_m} \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix} \tag{7.18}$$

The derivative with respect to the offsets δ_j and sending time t_{k_11} is also zero. The derivative with respect to the velocity $\mathbf{V}_j$ is also zero. The derivative with respect to the position $\mathbf{M}_j$ is

$$\frac{\partial f_t}{\partial \mathbf{M}_j} = -\frac{\mathbf{M}_j - \mathbf{S}_j}{\|\mathbf{M}_j - \mathbf{S}_j\|} \tag{7.19}$$

Then, one can write the covariance of the predicted state $\widehat{\mathbf{x}}_n$ using the matrices $\mathbf{J}_{t-1}$ and $\mathbf{D}_j$:

$$\mathbb{E}[(\widehat{\mathbf{x}}_n - \mathbb{E}[\widehat{\mathbf{x}_n}])(\widehat{\mathbf{x}_n} - \mathbb{E}[\widehat{\mathbf{x}_n}])^T] = \mathbf{A}\mathbf{P}_{n-1}\mathbf{A}^T + \mathbf{J}_{t-1}\mathbf{\Sigma}_{t-1}\mathbf{J}_{t-1}^T + \mathbf{D}_j\mathbf{\Sigma}_j\mathbf{D}_j^T \tag{7.20}$$

where $\mathbf{\Sigma}_{t-1}$ is the covariance matrix of $\mathbf{x}_{t-1}$, $\mathbf{\Sigma}_j$ is the covariance matrix of $\boldsymbol{\chi}_j$ and $\mathbf{P}_{n-1}$ is the covariance matrix of the state $\mathbf{x}_{n-1}$. One comes to this result if all cross correlations between the variables $\mathbf{x}_{n-1}$, $\boldsymbol{\chi}_j$ and $\mathbf{x}_{t-1}$ are zero. The Kalman filters fulfill the Markov assumption. This means a state depends only on its previous state. Therefore, $\boldsymbol{\chi}_j$ and $\mathbf{x}_{t-1}$ are clearly uncorrelated, as they are not consecutive states. The variables $\mathbf{x}_{t-1}$ and $\mathbf{x}_{n-1}$ are also uncorrelated, as they correspond to different time instants. The predicted symbols and time in $\mathbf{x}_{n-1}$ are actually used to select the most likely LOS measurement from the sender j in order to estimate $\boldsymbol{\chi}_j$. Therefore, they are correlated. However, if one would want to consider it one would have to consider other data associations at that time step, which would mean building a data association tree that would soon become untreatable due to the number of possibilities. Therefore, we consider only the data associations in one time step and assume the previous data associations were correct.

The measurement vector consists of the measured timestamp τ, the measured symbols $\mathbf{s}_{\mathrm{MS}} \in \mathbb{S}$ and the mapped symbols $\mathbf{s}_{\mathrm{TS}} \in \mathbb{S}$

$$\mathbf{z} = \begin{bmatrix} \tau & \mathbf{s}_{\mathrm{MS}} & \mathbf{s}_{\mathrm{TS}} \end{bmatrix}^{\mathrm{T}} . \tag{7.21}$$

Here the mapped symbols depend also on the movement of the target, which is estimated by the unscented Kalman filter. We can rewrite Eq. 7.7 as:

$$\mathbf{s}_{\mathrm{TS}} = f_{\mathrm{map}}(m, r_{\mathrm{TS}}) = \mathbf{s}_{\mathrm{m}} + r_{\mathrm{TS}} \cdot \boldsymbol{\Psi} \tag{7.22}$$

where $\boldsymbol{\Psi}$ depends on the symbol distances.

$$\mathbf{s}_{\mathrm{TS}} := f_{\mathrm{map}}(m, r_{\mathrm{TS}}) = \mathbf{s}_{\mathrm{m}} + \boldsymbol{\Psi}\frac{T_{k_j j} - T_{(k-1)_j j} - I_j - (g_t(\boldsymbol{\chi}_{t-1}) + f_t(\boldsymbol{\chi}_j))}{\tau_d} \tag{7.23}$$

The covariance of the measurement vector is denoted by $\mathbf{R}_m$:

$$\mathbf{R}_m = \mathbb{E}[(\mathbf{z} - \mathbb{E}[\mathbf{z}])(\mathbf{z} - \mathbb{E}[\mathbf{z}]^T)] = \begin{bmatrix} \sigma_n^2 & 0 & 0 \\ 0 & \mathbf{R}_{MS} & 0 \\ 0 & 0 & \mathbf{R}_{TS} \end{bmatrix} \tag{7.24}$$

The variance of the measured timestamp σ_n^2 and the covariance of the measured symbols $\mathbf{R}_{MS}$ depend on the measurement noises. Using Eq. 7.22 one can observe how the covariance of the mapped symbols depends on the measured noise and the UKF uncertainty:

$$\mathbf{R}_{TS} = \frac{1}{\tau_d^2}\left(2\sigma_n^2\boldsymbol{\Psi}\boldsymbol{\Psi}^T + \boldsymbol{\Psi}\mathbf{J}_{t-1}\mathbf{\Sigma}_{t-1}\mathbf{J}_{t-1}^T\boldsymbol{\Psi}^T + \boldsymbol{\Psi}\mathbf{D}_j\mathbf{\Sigma}_j\mathbf{D}_j^T\boldsymbol{\Psi}^T\right) \tag{7.25}$$

The fusion of the symbols is described by the observation model

$$
\mathbf{H} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & \mathbf{I}_4 & \\ \vdots & & & \\ 0 & & \mathbf{I}_4 & \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{7.26}
$$

Therefore, the upper identity matrix $\mathbf{I}_4$ maps the fused symbols $\mathbf{s}_{\mathrm{DF}}$ to the measured $\mathbf{s}_{\mathrm{MS}}$ and the lower identity matrix maps the fused symbols $\mathbf{s}_{\mathrm{DF}}$ to the mapped symbols $\mathbf{s}_{\mathrm{MS}}$.

Then, the Kalman filter for a certain sender j is implemented as follows:

$$
\widehat{\mathbf{x}}_n = \mathbf{A}\mathbf{x}_{n-1} + \mathbf{B}\mathbf{u}_n + g_t(\boldsymbol{\chi}_{t-1}) + f_t(\boldsymbol{\chi}_j) \tag{7.27}
$$

$$
\widehat{\mathbf{P}}_n = \mathbf{A}\mathbf{P}_{n-1}\mathbf{A}^T + \mathbf{J}_{t-1}\boldsymbol{\Sigma}_{t-1}\mathbf{J}_{t-1}^T + \mathbf{D}_j\boldsymbol{\Sigma}_j\mathbf{D}_j^T \tag{7.28}
$$

$$
\mathbf{S}_n = \mathbf{H}\widehat{\mathbf{P}}_n\mathbf{H}^T + \mathbf{R}_m \tag{7.29}
$$

$$
\mathbf{K}_n = \widehat{\mathbf{P}}_n\mathbf{H}^T\mathbf{S}_n^{-1} \tag{7.30}
$$

$$
\mathbf{x}_n = \widehat{\mathbf{x}}_n + \mathbf{K}\left(\mathbf{z}_n - \mathbf{H}\widehat{\mathbf{x}}\right) \tag{7.31}
$$

$$
\mathbf{P}_n = \left(\mathbf{I} - \mathbf{K}_n\mathbf{H}\right)\widehat{\mathbf{P}}_n \tag{7.32}
$$

Then, the UKF predictions are used two times. The first time to predict the next timestamp and the second time to create a mapped symbol which contains information about the sender ID, as it depends on the interval I_j.

Each tracker for the ID number j consists of one Kalman-Filter, which tracks the symbols and fuses the data. In order to estimate which timestamp belongs to which sender and whether they are in LOS, we select all the measurements which are closer than a certain threshold δ to the predicted timestamp $\widehat{T_{e,n}}$. Among those timestamps, the best one is selected by calculating the angle error between the symbols:

$$
\phi_{\mathrm{Error}} = \frac{\|\widehat{\mathbf{s}_{DF}} - \mathbf{s}_{MS}\|_1 + \|\widehat{\mathbf{s}_{DF}} - \mathbf{s}_{TS}\|_1}{2} \tag{7.33}
$$

where $\|\cdot\|_1$ is the ℓ_1 norm. The timestamp which has the lowest error ϕ_{Error} is used as a measurement for the unscented Kalman filter. The ID is estimated with the estimated symbols $\widehat{\mathbf{s}_{DF}}$. If the lowest angle error ϕ_{Error} associated to a timestamp is above a certain threshold ϕ_{MAX} it is considered invalid. If it is below the threshold, it is forwarded to the unscented Kalman filter in order to estimate the state at time t.

Note that in this case we are not considering other Kalman filters when selecting the most likely timestamps. We are also not using the covariances of the states for data association. The reason behind this is that in this system there is only one Kalman filter per real sender.

Reflections are not tracked in order to reduce the computational requirements of the receiver. This is because the amount of data that need to be processed is higher than in the ASSIST system. Moreover, in this localization system all the computational effort relies on the moving receiver. In the ASSIST system used in the previous chapter, every receiver and a central unit share the computational load.

As no reflections are tracked in this system, the Kalman filter of every sender is responsible of finding the most likely line-of-sight signal according to the symbol error and the timestamp.

7.3 Experimental Results

In order to test the presented approach, we do an experiment with the eCULTS localization system. The MoCap system is used as the reference system. As the approach in Chapter 4 proves to have a reduced error at low velocity, in this new experiment we move at different velocities, from 0.1 m/s to 1.8 m/s (see Fig. 7.4). Five static senders are used for locating the receiver. The positions of the senders can be seen in the Appendix of this thesis.
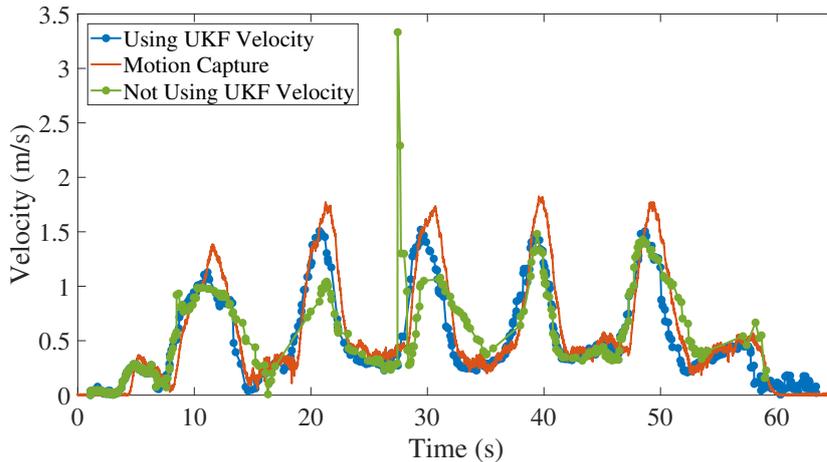
Figure 7.4: Real and estimated velocities for a target moving within the track defined in Fig. 7.5. One can observe how the presented approach is capable of tracking the target velocity whereas the previous approach starts diverging when the target moves above 1 m/s. Note that in order to locate the target using the approach in [77] one needs to allow higher angle errors ϕ_{Error}, which leads to an increase of wrong data associations.

The target is tracked with a median error of 0.061 m and a standard deviation of 0.064 m. The estimated positions can be seen in Fig. 7.5. The estimated and the real velocity can be seen in Fig. 7.4. There can be synchronization errors between the estimated velocity and the real velocity, due to different delays in the reception of the data. In these figures one can observe how using the estimated velocity by the UKF for data association allows tracking the target at high velocities. On the other hand, the approach presented in [77] is incapable of predicting the correct LOS measurements at velocities above 1 m/s. Therefore, wrong data associations

are done and the error is increased. These figures show the result using the UKF velocity and ϕ_{MAX}=2.3 rad versus the result not using the UKF velocity and ϕ_{MAX}=3 rad. The threshold ϕ_{MAX} needs to be increased, as the high velocities lead to high angles errors if Eq. 7.1 is used. In Table 7.1 one can observe how increasing ϕ_{MAX} the number of estimated positions is higher but it allows also more erroneous data associations. The case where Eq. 7.1 is used and ϕ_{MAX}=2.3 rad has the highest error. This is because there were no estimations during 9.56 s, as the angle error was above the threshold. This led to the algorithm divergence and not being capable of converging to the correct solution. In these experiments no additional NLOS mitigation algorithms were used when estimating the positions. This is because the aim of these experiments was to directly show the effect of wrong data associations on the position estimations.

	Median Error (m)	Standard Deviation (m)	Number of Positions per Second
ϕ_{MAX}=2.3 rad, using UKF velocity	0.061	0.064	7.85
ϕ_{MAX}=2.3 rad, not using UKF velocity	0.115	2.554	4.61
ϕ_{MAX}=3 rad, using UKF velocity	0.063	0.082	10.10
ϕ_{MAX}=3 rad, not using UKF velocity	0.088	0.253	5.35

Table 7.1: Comparison of localization errors and number of estimated positions.
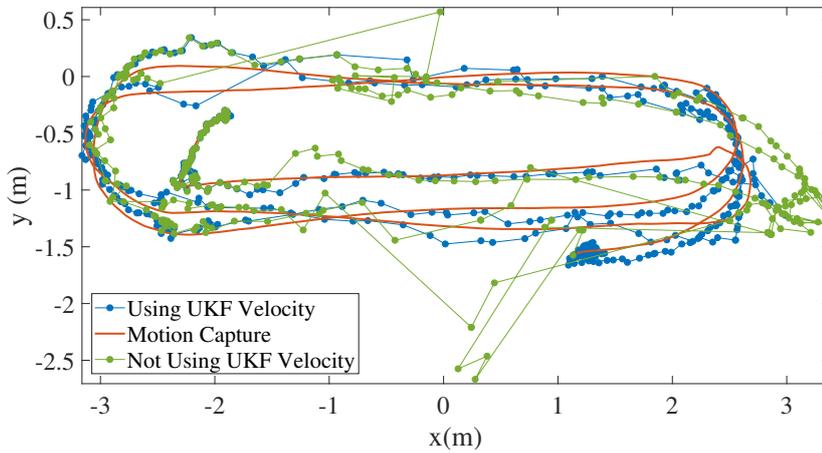
Figure 7.5: Estimated positions using the bank of Kalman filters in closed-loop with the UKF compared with the previous existing approach. The target was moving at velocities from 0.1 m/s to 1.8 m/s (see Fig. 7.4). When the UKF velocity is not used, the median error is 0.088 m and the standard deviation 0.253 m. On the other hand, the proposed approach proves to be capable of tracking the target with a median error of 0.061 m and a standard deviation of 0.064 m.

Then, in Chapter 4 the receiver was tracked with 0.039 m median error and 0.055 m standard deviation. The target was moving at a maximum velocity of 0.6 m/s. In this chapter we show how the target can be located with 0.061 m error and 0.064 m standard deviation when the target is moving at a maximum velocity of 1.8 m/s. Then, even though the target is moving much faster, a similar error is achieved. The error is slightly higher due to the fact that the distance between measurements is higher and therefore the uncertainty is increased. Looking at Fig. 7.5 one can realize that some consecutive positions are more than 0.20 m apart. As the system of equations is underdetermined, the error depends on how well the motion model can predict the trajectory of the target. If the target moves large distances between two measurements, the error is increased.

In the Appendix of this thesis we show how the timestamps of an specific sender are associated for the trajectory in Fig. 7.5.

8 Using Ground Reflections to Reduce the Localization Error

In Chapter 6, we have seen how acoustic reflections can be used for locating a sender. However, it remains unclear whether they improve or deteriorate the localization estimation. Therefore, it remains an open question whether they should be eliminated or used as additional information. Reflections can be less accurate than line-of-sight signals, due to the simplified image model. It is also due to the fact that they travel larger paths than the line-of-sight signals that they are more vulnerable to temperature gradients and scattering. However, in certain cases, they can considerably reduce the localization error. This is the case for ground reflections, which can be used to estimate the height of the target. Most of the location systems do not estimate the height of the target or they do it with less precision than the other coordinates. This is because of the dilution of precision (DOP). Usually the receivers are not well distributed in the three-dimensional localization space. The receivers are often placed at similar heights, which deteriorates the height estimation. One could place the anchor nodes at different heights. However, this would increase the installation effort. Moreover, if some receivers were installed at low heights, the line-of-sight signals would be easily blocked by furniture or other obstacles. Using the ground reflection, the receivers can be mounted on the ceiling without increasing the error in the height estimation, as it will be proved in the following sections.

8.1 Cramer-Rao Lower Bound and Dilution of Precision

In order to study the effect of using ground reflections in scenarios where the receivers are placed at similar heights, we analyze the error in the height estimation. We are interested in finding out whether ground reflections can actually reduce the localization error. In order to do so, consider we have an estimation of the position of the target $\widehat{\mathbf{x}}$ in $\mathbb{R}^3$. We assume our estimator is unbiased. Therefore, the mean of our estimation $\widehat{\mathbf{x}}$ is the actual position of the target $\mathbf{x}$:

$$\mathbb{E}\left[\widehat{\mathbf{x}}\right] = \mathbf{x} \tag{8.1}$$

As our estimator is assumed to be unbiased, the position errors that we observe come from the variance of our estimation. Then, if the variance of the z component of $\mathbf{x}$ is large, the position estimations will have an imprecise height estimate. We are interested in finding the covariance matrix of $\widehat{\mathbf{x}}$:

$$\mathbb{E}\left[(\widehat{\mathbf{x}} - \mathbb{E}[\widehat{\mathbf{x}}])(\widehat{\mathbf{x}} - \mathbb{E}[\widehat{\mathbf{x}}])^T\right] = \mathbb{E}\left[(\widehat{\mathbf{x}} - \mathbf{x})(\widehat{\mathbf{x}} - \mathbf{x})^T\right] \tag{8.2}$$

This covariance depends on the estimator that is used. However, one can make use of the Cramer-Rao bound, which provides a lower bound for the covariance of an unbiased estimator:

$$\mathbb{E}\left[(\mathbf{x} - \widehat{\mathbf{x}})(\mathbf{x} - \widehat{\mathbf{x}})^T\right] \geq \mathbf{J}^{-1} \tag{8.3}$$

where $\mathbf{J}$ is the Fisher information matrix.

$$\mathbf{J} = \mathbb{E}\left[\left(\frac{\partial}{\partial \mathbf{x}} \log p(\mathbf{z}|\mathbf{x})\right)\left(\frac{\partial}{\partial \mathbf{x}} \log p(\mathbf{z}|\mathbf{x})\right)^T\right] \tag{8.4}$$

Assuming the noise sources are Gaussian and independent, the probability of a certain measurement vector $\mathbf{z}$ given $\mathbf{x}$ is:

$$p(\mathbf{z}|\mathbf{x}) = \frac{1}{\sqrt{\|2\pi \mathbf{R}_t\|}} \exp\left(-\frac{1}{2}(\mathbf{z} - h(\mathbf{x}))^T \mathbf{R}_t^{-1}(\mathbf{z} - h(\mathbf{x}))\right) \tag{8.5}$$

where $h(\mathbf{x})$ is the function which relates the position and the measurements. This depends on the measurement model used. It can be, for instance, time of arrival or time difference of arrival, as explained in Chapter 3. The matrix $\mathbf{R}_t$ is the noise covariance matrix. Then, if we define $\mathbf{H}$ as the Jacobian matrix of h:

$$\mathbb{E}\left[(\mathbf{x} - \widehat{\mathbf{x}})(\mathbf{x} - \widehat{\mathbf{x}})^T\right] \geq (\mathbf{H}^T \mathbf{R}_t^{-1} \mathbf{H})^{-1} \tag{8.6}$$

When all the measurements are assumed to be affected by the same measurement covariance σ_n^2:

$$\mathbb{E}\left[(\mathbf{x} - \widehat{\mathbf{x}})(\mathbf{x} - \widehat{\mathbf{x}})^T\right] \geq \sigma_n^2 \boldsymbol{\Sigma_p}^{-1} \tag{8.7}$$

where $\boldsymbol{\Sigma_p} = \mathbf{H}^T \mathbf{H}$.

The Cramer-Rao lower bound can also be interpreted as the minimum achievable root mean square error (RMSE) by an unbiased estimator [30]. The RMSE is used often in literature as a quality measure for the position estimations. If the estimation $\widehat{\mathbf{x}}$ has components $\widehat{x}, \widehat{y}$ and $\widehat{z}$, the RMSE is calculated as follows:

$$\text{RMSE}(\widehat{\mathbf{x}}) = \sqrt{\mathbb{E}\left[(\widehat{x} - x_T)^2 + (\widehat{y} - y_T)^2 + (\widehat{z} - z_T)^2\right]} = \sqrt{\text{Tr}\{\mathbb{E}\left[(\mathbf{x} - \widehat{\mathbf{x}})(\mathbf{x} - \widehat{\mathbf{x}})^T\right]\}} \tag{8.8}$$

where x_T, y_T and z_T are the components of the true position. One can then rewrite the Cramer-Rao lower bound as a lower bound for the RMSE:

$$\text{RMSE}(\widehat{\mathbf{x}}) \geq \sqrt{\text{Tr}\{\sigma_n^2 \boldsymbol{\Sigma_p}^{-1}\}} \tag{8.9}$$

This can be related with the dilution of precision (DOP) which is commonly used in localization to measure the influence of the geometrical distribution of anchor nodes in the final localization error. Then, depending on the geometry, a certain Gaussian noise with standard deviation σ_n can have a different effect on the final position estimation. The dilution of precision in x, y and z is defined as $\frac{1}{\sigma_n}\sqrt{\text{Tr}\{\sigma_n^2 \boldsymbol{\Sigma_p}^{-1}\}}$. Then:

$$\text{RMSE}(\widehat{\mathbf{x}}) \geq \sigma_n \text{DOP} \tag{8.10}$$

Then, the dilution of precision is a measure of the achievable RMSE of the estimations given that all measurements are affected by an identical noise distribution.

One can focus on certain components of the position. Then, the two-dimensional dilution of precision measures the achievable RMSE in the x and y coordinates. It can be measured as $\frac{\sqrt{\sigma_x^2 + \sigma_y^2}}{\sigma_n}$, where:

$$(\mathbf{H}^T \mathbf{R}_t^{-1} \mathbf{H})^{-1} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \tag{8.11}$$

If the target is located in three dimensions, the vertical dilution of precision is a measure of the achievable RMSE in the z coordinate. It is defined as $\frac{\sigma_z}{\sigma_n}$, where:

$$(\mathbf{H}^T \mathbf{R}_t^{-1} \mathbf{H})^{-1} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2 \end{pmatrix} \tag{8.12}$$

The three-dimensional dilution of precision is measured as $\frac{\sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}}{\sigma_n}$.

8.2 Time of Arrival

Having a certain number of receivers $N \geq 4$, the height of the target can be estimated using time of arrival (TOA). However, in certain scenarios, the changes in height might produce changes in time which are below the noise level. The effects of the positions of the anchor nodes in the quality of the localization have been previously studied for different measurement models such as TOA or TDOA [78, 79]. We focus on the effect of the ground reflections on the vertical dilution of precision.

In general case for distance measurements, the matrix $\mathbf{H}$ is:

$$\mathbf{H} = \begin{pmatrix} \frac{S_x - M_{x,1}}{R_1} & \frac{S_y - M_{y,1}}{R_1} & \frac{S_z - M_{z,1}}{R_1} \\ \vdots & \vdots & \vdots \\ \frac{S_x - M_{x,N}}{R_n} & \frac{S_y - M_{y,N}}{R_n} & \frac{S_z - M_{z,N}}{R_N} \end{pmatrix} \tag{8.13}$$

being R_i the distance from the target $\mathbf{S}$ to the receiver i:

$$R_i = \|\mathbf{M}_i - \mathbf{S}\| \tag{8.14}$$

The receiver position $\mathbf{M}_i$ has the components $M_{x,i}$, $M_{y,i}$ and $M_{z,i}$. The target position $\mathbf{S}$ has the components S_x, S_y and S_z.

The matrix $\mathbf{H}$ can be rewritten using unit vectors $\mathbf{u}$ that go from the sender to the receiver position:

$$\mathbf{u}_i = \left(u_{i,x}, u_{i,y}, u_{i,z}\right)^T = \left(\frac{S_x - M_{x,i}}{R_i}, \frac{S_y - M_{y,i}}{R_i}, \frac{S_z - M_{z,i}}{R_i}\right)^T \tag{8.15}$$

Then:

$$\mathbf{H} = \begin{pmatrix} u_{1,x} & u_{1,y} & u_{1,z} \\ \vdots & \vdots & \vdots \\ u_{N,x} & u_{N,y} & u_{N,z} \end{pmatrix} \tag{8.16}$$

We define a matrix $\mathbf{A}$ such that:

$$\mathbf{A} = \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \tag{8.17}$$

When the target and receivers are in $\mathbb{R}^2$, the matrix $\mathbf{A}$ is a 2×2 matrix and the dilution of precision in two dimensions DOP_{2D} is:

$$\text{DOP}_{2D}^2 = \frac{1}{\sigma_n^2} \frac{A_{1,1} + A_{2,2}}{A_{1,1} A_{2,2} - A_{1,2}^2} \tag{8.18}$$

Spirito [80] did an extensive study about the dilution of precision in two dimensions. He proved that the denominator can be simplified as follows:

$$A_{1,1} A_{2,2} - A_{1,2}^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} u_{i,x} u_{j,y} \vec{k} \cdot (\mathbf{u}_i \times \mathbf{u}_j) \sigma_i^{-2} \sigma_j^{-2} \tag{8.19}$$

$$A_{1,1} A_{2,2} - A_{1,2}^2 = \sum_{i=1}^{N} \sum_{j>i}^{N} (\sin^2(\theta_{i,j})) \sigma_i^{-2} \sigma_j^{-2} \tag{8.20}$$

where $\theta_{i,j}$ is the angle between the receiver i, and the receiver j when the sender is the vertex of the angle.

Then,

$$\text{DOP}_{2D}^2 = \frac{1}{\sigma_n^2} \frac{\sum_{i=1}^{N} u_{i,x}^2 \sigma_i^{-2} + u_{i,y}^2 \sigma_i^{-2}}{\sum_{i=1}^{N} \sum_{j>i}^{N} (\|\mathbf{u}_{i,xy}\|^2 \|\mathbf{u}_{j,xy}\|^2 \sin^2(\theta_{i,j})) \sigma_i^{-2} \sigma_j^{-2}} \tag{8.21}$$

If all receivers are affected by the same noise distribution:

$$\text{DOP}_{2D}^2 = \frac{N}{\sum_{i=1}^{N} \sum_{j \in (1,\dots,N), j>i} (\sin^2(\theta_{i,j}))} \tag{8.22}$$

One can already notice how this term does not directly depend on how far the receivers are, only in which angle are they observed by the sender. While two-dimensional localization is out of the scope of this chapter, this term will be important to understand the role that the two-dimensional DOP plays in the height estimation.

We are interested in the vertical dilution of precision, which is measured as $\frac{\sqrt{\sigma_z^2}}{\sigma_n}$. In other words, it is the component in the third column, third row of the matrix $\frac{1}{\sigma_n}(\mathbf{H}^T \mathbf{R}_t^{-1} \mathbf{H})^{-1}$. This can be written as:

$$\sigma_n^2 \text{VDOP}^2 = \frac{\det(\mathbf{A}_{1:2,1:2})}{\det(\mathbf{A})} \tag{8.23}$$

where $\mathbf{A}_{1:2,1:2}$ indicates the 2×2 sub matrix of $\mathbf{A}$ which corresponds to the two-dimensional coordinates of $\mathbf{A}$. This is an interesting result, as one can see how the numerator of this fraction

is related to the dilution of precision in two dimensions, only that here it is the projection in two dimensions of the three-dimensional space. The variable σ_n represents the noise standard deviation which is used to normalize the VDOP. In our case, we assume the noise depends on the receiver, as reflections might be less accurate than line-of-sight signals. From now on, we define $M_{i,k}$ as the coordinate i of the vector $\mathbf{M}_k$, and we define S_i as the coordinate i from the vector $\mathbf{S}$. By doing this, the value in the row i and the column j of the matrix $\mathbf{A}$ is defined as:

$$A_{i,j} = \sum_{k=1}^{N} \left(\frac{S_i - M_{i,k}}{\sigma_k^2 R_k} \right) \left(\frac{S_j - M_{j,k}}{\sigma_k^2 R_k} \right) \tag{8.24}$$

Then:

$$\text{VDOP} = \frac{1}{\sigma_n} \sqrt{\frac{A_{1,1}A_{2,2} - A_{1,2}^2}{A_{3,3}(A_{1,1}A_{2,2} - A_{1,2}^2) - A_{1,1}A_{2,3}^2 - A_{2,2}A_{1,3}^2 + 2A_{1,3}A_{2,3}A_{1,2}}} \tag{8.25}$$

Using Eq. 8.24, we know that:

$$A_{1,1}A_{2,2} - A_{1,2}^2 = \sum_{i=1}^{N} u_{i,x}^2 \left(\sum_{i=1}^{N} u_{i,y}^2 \right) - \left(\sum_{i=1}^{N} u_{i,x}u_{i,y} \right)^2 \sigma_i^{-2} \tag{8.26}$$

which can be rewritten as:

$$A_{1,1}A_{2,2} - A_{1,2}^2 = \sum_{i=1}^{N} \sum_{j>i}^{N} (u_{i,x}^2 u_{j,y}^2 + u_{j,x}^2 u_{i,y}^2 - 2u_{i,x}u_{i,y}u_{j,x}u_{j,y})\sigma_i^{-2}\sigma_j^{-2} \tag{8.27}$$

which can be compacted as:

$$A_{1,1}A_{2,2} - A_{1,2}^2 = \sum_{i=1}^{N} \sum_{j>i}^{N} (u_{i,x}u_{j,y} - u_{i,y}u_{j,x})^2 \sigma_i^{-2}\sigma_j^{-2} \tag{8.28}$$

$$A_{1,1}A_{2,2} - A_{1,2}^2 = \sum_{i=1}^{N} \sum_{j>i}^{N} (\|\mathbf{u}_{i,xy}\|^2 \|\mathbf{u}_{j,xy}\|^2 \sin^2(\theta_{i,j}))\sigma_i^{-2}\sigma_j^{-2} \tag{8.29}$$

where $\theta_{i,j}$ is the angle between the projection of the receivers i and j in two dimensions. The projection of the sender position in two dimensions is the angle vertex. The variable $\mathbf{u}_{i,xy}$ is the projection of the unit vector $\mathbf{u}_i$ in two dimensions. In [80] they assumed a two-dimensional scenario. Then the norm of the unit vectors is one, which simplifies the equation and results in Eq. 8.22. In this case, one can clearly see that the determinant of the first two components is actually the sum of the areas of the parallelograms formed by the projection of the unit vectors in two dimensions:

$$A_{1,1}A_{2,2} - A_{1,2}^2 = \sum_{i=1}^{N} \sum_{j>i}^{N} \text{Area}_p^2(\mathbf{u}_{i,xy}, \mathbf{u}_{j,xy})\sigma_i^{-2}\sigma_j^{-2} \tag{8.30}$$

where $\text{Area}_p(\mathbf{a}, \mathbf{b})$ denotes the area of a parallelogram with edges $\mathbf{a}$ and $\mathbf{b}$. Then:

$$\text{VDOP}^2 = \frac{1}{\sigma_n^2} \frac{\sum_{i=1}^{N} \sum_{j>i}^{N} \text{Area}_p^2(\mathbf{u}_{i,xy}, \mathbf{u}_{j,xy})}{\det(\mathbf{H}^T \mathbf{R}_t^{-1} \mathbf{H})} \tag{8.31}$$

where:

$$\det(\mathbf{H}^T \mathbf{R}_t^{-1} \mathbf{H}) =$$

$$\sum_{i=1}^{N} u_{i,z}^2 \sigma_i^{-2} \sum_{i=1}^{N} \sum_{j>i}^{N} \text{Area}_p^2(\mathbf{u}_{i,xy}, \mathbf{u}_{j,xy}) \sigma_i^{-2} \sigma_j^{-2} - \sum_{i=1}^{N} u_{i,x}^2 \sigma_i^{-2} \left(\sum_{i=1}^{N} u_{i,y} u_{i,z} \sigma_i^{-2} \right)^2 -$$

$$\sum_{i=1}^{N} u_{i,y}^2 \sigma_i^{-2} \left(\sum_{i=1}^{N} u_{i,x} u_{i,z} \sigma_i^{-2} \right)^2 + 2 \left(\sum_{i=1}^{N} u_{i,x} u_{i,y} \sigma_i^{-2} \sum_{i=1}^{N} u_{i,x} u_{i,z} \sigma_i^{-2} \sum_{i=1}^{N} u_{i,y} u_{i,z} \sigma_i^{-2} \right) \tag{8.32}$$

In order to understand Eq. 8.32 and why ground reflections can lead to a low error in the height estimation, in the following sections we analyze the VDOP for different cases.

Special Case: Three Receivers

The TOA measurements of three receivers lead to two solutions. However, in certain cases, one of the solutions can be eliminated using information about the environment. For example, if one knows where the ground is located and what is the range of heights at which the target can be, one can often eliminate one of the solutions due to its unfeasible height. Moreover, the VDOP in this special case can be represented geometrically, which allows a better comprehension of the terms that are involved on its calculation.

As commented before, the vertical dilution of precision can be calculated as:

$$\text{VDOP}^2 = \frac{A_{1,1} A_{2,2} - A_{1,2}^2}{\det(\mathbf{H}^T \mathbf{H})} \tag{8.33}$$

when the standard deviation of the noise is the same for all receivers. The matrix $\mathbf{H}^T \mathbf{H}$ is a so-called Gramian matrix, which means its determinant is the square of the volume of the parallelotope with edges the columns of $\mathbf{H}$. The columns of $\mathbf{H}$ do not have any geometrical meaning. However, every row corresponds to a different receiver and can be viewed as a unit vector with center the position of the target. This is crucial when the matrix is symmetric, as the determinant remains the same when a symmetric matrix is transposed. The matrix $\mathbf{H}$ is symmetric when the number of receivers is three. In this case, the determinant of the Gramian matrix is equal to the volume of the parallelepiped with edges the three unit vectors. This was already used in [81] with pseudo-range measurements to study the dilution of precision of the GPS. In our case, the vertical dilution of precision is related to the area formed by the projections of the unit vectors in two dimensions and the three-dimensional volume generated by them.

Assuming the noise is Gaussian with standard deviation σ_n, and having three receivers with unit vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$, the VDOP can be rewritten as:

$$\text{VDOP}^2 = \frac{\sin^2(\theta_{u,v})\|\mathbf{u}_{xy}\|^2\|\mathbf{v}_{xy}\|^2 + \sin^2(\theta_{u,w})\|\mathbf{u}_{xy}\|^2\|\mathbf{w}_{xy}\|^2 + \sin^2(\theta_{w,v})\|\mathbf{w}_{xy}\|^2\|\mathbf{v}_{xy}\|^2}{(\|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})\|)^2}$$

(8.34)

This can be seen geometrically as:

$$\text{VDOP}^2 = \frac{\text{Area}_p^2(\mathbf{u}_{xy}, \mathbf{v}_{xy}) + \text{Area}_p^2(\mathbf{u}_{xy}, \mathbf{w}_{xy}) + \text{Area}_p^2(\mathbf{w}_{xy}, \mathbf{v}_{xy})}{(\text{Vol}_p(\mathbf{u}, \mathbf{v}, \mathbf{w}))^2}$$

(8.35)

The volume of the parallepiped formed by the vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ is denoted as $\text{Vol}_p(\mathbf{a}, \mathbf{b}, \mathbf{c})$. As we are looking at a minimal case (three receivers) one cannot have one ground reflection for every line-of-sight measurement. However, Eq. 8.35 provides a useful geometrical representation of the VDOP. In Fig. 8.1 one can observe how a ground reflection can be useful when the line-of-sight receivers have a reduced z component in their unit vectors. In this case, the area of the two-dimensional projections is reduced and the volume of the parallepiped is increased. Then, the value of the numerator is increased and the value of the denominator is reduced. Therefore, ground reflections can improve the result by increasing the absolute value of the unit vectors in the z coordinate.

Special case: Equally Distributed Receivers in x and y

If the sender observes that the receivers are equally distributed in the x and y axis, and they are installed at the same height, we know that:

$$\sum_{i=1}^{N} u_{i,x} u_{i,z} = \sum_{i=1}^{N} u_{i,y} u_{i,z}$$

(8.36)

This does not mean that the speaker has to be in the middle of the receivers. An example of this scenario is when the unit vectors in two dimensions are mirrored over the line $y = x$. This allows a high number of configurations, simplifies the equations and results in a VDOP which can be easily comprehended.

Then:

$$\det(\mathbf{H}^T \mathbf{R}_t^{-1} \mathbf{H}) = \sum_{i=1}^{N} u_{i,z}^2 \sigma_i^{-2} \sum_{i=1}^{N} \sum_{j>i}^{N} \text{Area}_p^2(\mathbf{u}_{i,xy}, \mathbf{u}_{j,xy}) \sigma_i^{-2} \sigma_j^{-2} -$$
$$\left(\sum_{i=1}^{N} u_{i,x} u_{i,z} \sigma_i^{-2} \right)^2 \left(\sum_{i=1}^{N} u_{i,x}^2 \sigma_i^{-2} + \sum_{i=1}^{N} u_{i,y}^2 \sigma_i^{-2} - 2\left(\sum_{i=1}^{N} u_{i,x} u_{i,y} \sigma_i^{-2} \right) \right)$$

(8.37)

For further simplification, we assume the standard deviation of the noise is the same for all receivers:

$$\text{VDOP}^2 =$$

$$\frac{\sum_{i=1}^{N} \sum_{j>i}^{N} \text{Area}_p^2(\mathbf{u}_{i,xy}, \mathbf{u}_{j,xy})}{\sum_{i=1}^{N} u_{i,z}^2 \left(\sum_{i=1}^{N} \sum_{j>i}^{N} \text{Area}_p^2(\mathbf{u}_{i,xy}, \mathbf{u}_{j,xy}) \right) - \left(\sum_{i=1}^{N} u_{i,x} u_{i,z} \right)^2 \left(\sum_{i=1}^{N} (u_{i,x} - u_{i,y})^2 \right)}$$

(8.38)
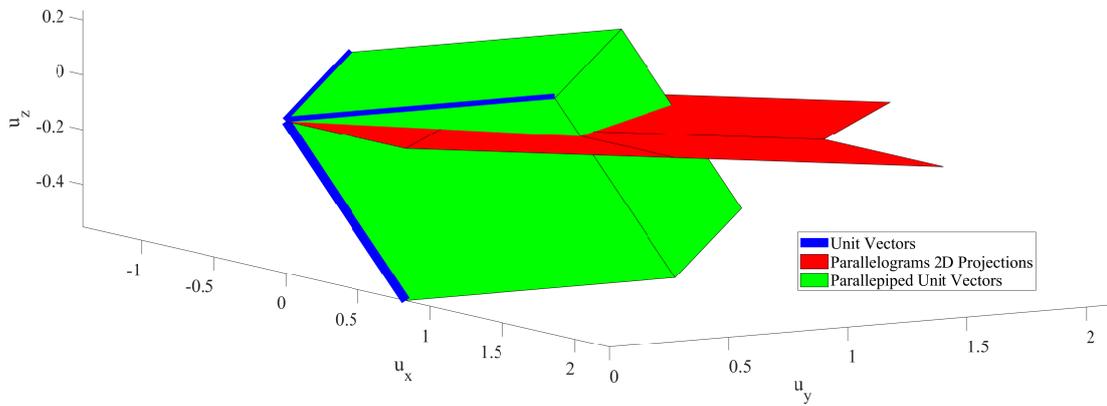
(a) Using a ground reflection, the squared volume of the parallepiped is 0.30, while the sum of the squared area of the two-dimensional projections (in red) is 1.67. Then, using Eq 8.35, the VDOP is 2.35. The dilution of precision in x and y is 1.48.
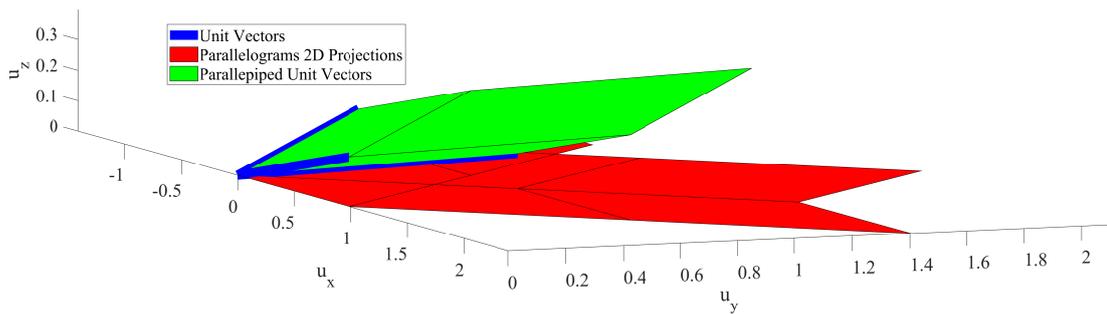
(b) Using only line-of-sight measurements, the squared volume of the parallepiped is 0.03, whereas the squared area of the two-dimensional projections (in red) is 1.93. Then, using Eq 8.35, the VDOP is 8.57. The dilution of precision in x and y is 2.02.

Figure 8.1: Geometrical representation of the vertical dilution of precision, when the target is located with three receivers and time of arrival measurements. The receivers are at 2.5 m height, and the target at 1.5 m height. The first figure shows the result when one of the three measurements is a ground reflection. The second figure shows the result when the three receivers are line-of-sight. One can see how the volume of the parallepiped is considerably increased, while the two-dimensional projection is similar in both cases. This reduces the virtual dilution of precision, as seen in Eq 8.35. Then, in this case, having a ground reflection would considerably improve the height estimation.

It is very clear from this equation, why the ground reflection can highly improve the height estimation. First, the unit vector in the z component $u_{i,z}$ will be increased, as the reflection always comes from a larger distance in z than the LOS signal. Second, the unit vector in z of the ground reflection will have the opposite sign than the LOS signal, which will reduce the component $\sum_{i=1}^{N} u_{i,x} u_{i,z}$.

Special case: Sender in the Center of the Receivers

In this case, we assume a pair number of receivers are placed at the same height and equally distributed in the x and y coordinates. Then, the sender is placed in the center of the receivers. If this happens, the terms $\sum_{i=1}^{N} u_{i,x} u_{i,z}$ and $\sum_{i=1}^{N} u_{i,y} u_{i,z}$ are close to zero. As a result of this, the vertical dilution of precision depends only on the z component of the unit vectors. This leads to the same result as it would be if the sender position in x and y was perfectly known in advance.

If a signal is in line-of-sight signal, the z component of its unit vector is:

$$u_{i,z}^* = \frac{(S_z - z_r)}{R_i^*} \tag{8.39}$$

where R_i^* is the distance from the sender to the receiver. For the ground reflection we define another unit vector $\mathbf{u_i'}$ whose z component is:

$$u_{i,z}' = \frac{(S_z + z_r)}{R_i'} \tag{8.40}$$

where R_i' is the distance from the virtual sender to the receiver. For simplification, from now on we assume the received signals have n_l ground reflections and the same number of line-of-sight signals.

$$\sigma_n \text{VDOP}_{\text{LOS}} = \frac{1}{\sqrt{\sum_{i=1}^{n_l} \sigma_{i,n}^{-2} (u_{i,z}^*)^2}} = \frac{1}{\sqrt{\sum_{i=1}^{n_l} \sigma_{i,n}^{-2} \left(\frac{z - z_i}{R_i^*}\right)^2}} \tag{8.41}$$

Assuming all LOS receivers have the same noise standard deviation σ_n and the coordinate z_i is the same for all the receivers (z_r):

$$\text{VDOP}_{\text{LOS}} = \frac{1}{|S_z - z_r| \sqrt{\sum_{i=1}^{n_l} \left(\frac{1}{R_i^*}\right)^2}} \tag{8.42}$$

Now we introduce the ground reflection. We assume all the reflections have an additive Gaussian noise with variance σ_r^2 and all the line-of-sight measurements have an additive Gaussian noise with variance σ_n^2. Note that the approximation of a virtual receiver will be true only if there is a ground reflection ($S_z, z_r > 0$). Then, if we add the ground reflections of every receiver:

$$\text{VDOP}_{\text{W/GROUND}} = \frac{1}{\sqrt{(S_z - z_r)^2 \sum_{i=1}^{n_l} \left(\frac{1}{R_i^*}\right)^2 + (\frac{\sigma_n}{\sigma_r})^2 (S_z + z_r)^2 \sum_{i=1}^{n_l} \left(\frac{1}{R_i'}\right)^2}} \tag{8.43}$$

Then:

$$\frac{\text{VDOP}_{\text{LOS}}}{\text{VDOP}_{\text{W/GROUND}}} = \sqrt{1 + \frac{(\frac{\sigma_n}{\sigma_r})^2 (S_z + z_r)^2 \sum_{i=1}^{n_l} \left(\frac{1}{R_i'}\right)^2}{(S_z - z_r)^2 \sum_{i=1}^{n_l} \left(\frac{1}{R_i^*}\right)^2}} = \sqrt{1 + \frac{(\frac{\sigma_n}{\sigma_r})^2 \sum_{i=1}^{n_l} (u_{z,i}')^2}{\sum_{i=1}^{n_l} (u_{z,i}^*)^2}}$$

(8.44)

It is interesting to observe how the major improvement in this case comes from the change in the unit vector in the z coordinate, as we showed in [7]. Here the ground reflection does not improve the height estimation by having a good distribution of receivers, it improves it because of having additional receivers which are farther in the z coordinate than the actual receivers. This is because in this case there is no ambiguity in the other components.

Special Case: Receivers Much Higher Than the Sender

If $z_r \gg S_z$, then, $u_{i,z}^* \approx -u_{i,z}'$. Using the ground reflection:

$$(\text{VDOP}_{\text{W/GROUND}})^2 \approx \frac{1}{\sigma_n^2} \frac{1}{\sum_{i=1}^{n_l} (u_{i,z}^*)^2 \sigma_{i,n}^{-2} + \sum_{i=1}^{n_l} u_{i,z}'^2 \sigma_{i,r}^{-2}}$$

(8.45)

We observe how even when the speaker has a low height, the ground reflection removes the dependence from the coordinates x and y. The resulting VDOP is the same than when these variables are known in advance or the sender is in the middle of the receivers, as it has been shown in Eq. 8.41 and Eq. 8.44.

Special Case: Receivers and Sender at the Same Height

If $S_z = z_r$, then:

$$u_{i,z}^* = 0 \qquad \forall i \in [1, \ldots, n_l]$$

(8.46)

$$u_{i,z}' = \frac{(2z_r)}{R_i'} \qquad \forall i \in [1, \ldots, n_l]$$

(8.47)

In this case, without the ground reflection it is not possible to estimate the height of the target:

$$\text{VDOP}_{\text{LOS}} = \infty$$

(8.48)

with ground reflections the height can be estimated:

$$(\text{VDOP}_{\text{W/GROUND}})^2 = \frac{1}{\sigma_n^2} \frac{\sum_{i=1}^{N} \sum_{j>i}^{N} \text{Area}_p^2(\mathbf{u}_{i,xy}, \mathbf{u}_{j,xy}) \sigma_i^{-2} \sigma_j^{-2}}{\det(\mathbf{H}^T \mathbf{R}_t^{-1} \mathbf{H})}$$

(8.49)

and

$$\det(\mathbf{H}^T \mathbf{R}_t^{-1} \mathbf{H}) = \sum_{i=1}^{n_l} u'^2_{i,z} \sigma_{i,r}^{-2} \sum_{i=1}^{N} \sum_{j>i}^{N} \text{Area}_p^2(\mathbf{u}_{i,xy}, \mathbf{u}_{j,xy}) \sigma_i^{-2} \sigma_j^{-2} -$$

$$\left(\sum_{i=1}^{n_l} (u^*_{i,x})^2 \sigma_{i,n}^{-2} + \sum_{i=1}^{n_l} u'^2_{i,x} \sigma_{i,r}^{-2} \right) \left(\sum_{i=1}^{n_l} u'_{i,y} u'_{i,z} \sigma_{i,r}^{-2} \right)^2 -$$

$$\left(\sum_{i=1}^{n_l} (u^*_{i,y})^2 \sigma_{i,n}^{-2} + \sum_{i=1}^{n_l} u'^2_{i,y} \sigma_{i,r}^{-2} \right) \left(\sum_{i=1}^{n_l} u'_{i,x} u'_{i,z} \sigma_{i,r}^{-2} \right)^2 +$$

$$2 \left(\sum_{i=1}^{n_l} u^*_{i,x} u^*_{i,y} \sigma_{i,n}^{-2} + \sum_{i=1}^{n_l} u'_{i,x} u'_{i,y} \sigma_{i,r}^{-2} \right) \sum_{i=1}^{n_l} u'_{i,x} u'_{i,z} \sigma_{i,r}^{-2} \sum_{i=1}^{n_l} u'_{i,y} u'_{i,z} \sigma_{i,r}^{-2} \tag{8.50}$$

where $u'_{i,x}$ and $u'_{i,y}$ are the x and y components of the unit vector to the virtual receiver i. We see the big impact that the ground reflection has when the sender has a similar height as the receivers. Without the ground reflection, it is not possible to estimate the height reliably. With the ground reflections, the result depends on the distribution of the receivers in two dimensions and on how far the speaker is from the ground. If the speaker and the receivers are close to the ground, $(u'_{i,z})^2$ is close to zero and therefore the VDOP is increased. This is because then the virtual receivers, the real receivers and the sender have all approximately the same height.

8.3 Time Difference of Arrival

When using time difference of arrival, the matrix $\mathbf{H}$ is:

$$\mathbf{H} = \begin{pmatrix} \frac{S_x - M_{x,1}}{R_1} - \frac{S_x - M_{x,ref}}{R_{ref}} & \frac{S_y - M_{y,1}}{R_1} - \frac{S_y - M_{y,ref}}{R_{ref}} & \frac{S_z - M_{z,1}}{R_1} - \frac{S_z - M_{z,ref}}{R_{ref}} \\ \vdots & \vdots & \vdots \\ \frac{S_x - M_{x,N-1}}{R_n} - \frac{S_x - M_{x,ref}}{R_{ref}} & \frac{S_y - M_{y,N-1}}{R_{N-1}} - \frac{S_y - M_{y,ref}}{R_{ref}} & \frac{S_z - M_{z,N-1}}{R_{N-1}} - \frac{S_z - M_{z,ref}}{R_{ref}} \end{pmatrix} \tag{8.51}$$

This can be rewritten using unit vectors that go from the sender to the receiver position:

$$\mathbf{H} = \begin{pmatrix} u_{1,x} - u_{ref,x} & u_{1,y} - u_{ref,y} & u_{1,z} - u_{ref,z} \\ \vdots & \vdots & \vdots \\ u_{N-1,x} - u_{ref,x} & u_{N-1,y} - u_{ref,y} & u_{N-1,z} - u_{ref,z} \end{pmatrix} \tag{8.52}$$

For simplicity, we define the auxiliary vector $\mathbf{w}_i = \mathbf{u}_i - \mathbf{u}_{ref}$. Then:

$$\mathbf{H} = \begin{pmatrix} w_{1,x} & w_{1,y} & w_{1,z} \\ \vdots & \vdots & \vdots \\ w_{N-1,x} & w_{N-1,y} & w_{N-1,z} \end{pmatrix} \tag{8.53}$$

When using hyperbolic multilateration, two timestamps are used for every measurement, which means the noise between multiple measurements is correlated. As shown in [82], one can define

a matrix $\mathbf{D}$ as:

$$\mathbf{D} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & 0 & 0 & \cdots & 1 \end{pmatrix} \tag{8.54}$$

The first column corresponds to the reference receiver. The other $N - 1$ columns correspond to the other receivers. Then the $(N - 1) \times (N - 1)$ noise matrix of the TDOA measurements will be:

$$\mathbf{R}_{\text{TD}} = \mathbf{D}\mathbf{R}_t\mathbf{D}^T \tag{8.55}$$

where $\mathbf{R}_t$ is a $N \times N$ matrix which contains the noise covariance of all N receivers. Then, the matrix $\mathbf{A}$ will be calculated as follows:

$$\mathbf{A} = (\mathbf{H}^T\mathbf{R}_{\text{TD}}^{-1}\mathbf{H}) \tag{8.56}$$

which complicates its evaluation. However, it is interesting to observe how now the determinant $\det(\mathbf{A})$ depends on the vector $\mathbf{w}$ instead of $\mathbf{u}$. This means, if the z component of the unit vectors is similar for all receivers, the third column of $\mathbf{H}$ will be close to zero. If the third column is close to zero, changes in the height of the target will lead to small changes in the measurements and therefore a high dilution of precision.

In the following sections we analyze the VDOP for two extreme cases: when the sender is at the same height than the receivers and when the sender is at a much lower height than the receivers.

Special Case: Receivers and Sender at the Same Height

If all receivers are at the same height z_r and $S_z = z_r$, then:

$$w_{i,z} = 0 \qquad \forall i \in [1, \ldots, N - 1] \tag{8.57}$$

assuming only line-of-sight signals are received.

As $\mathbf{H}$ has a column of zeros, the matrix $\mathbf{A}$ is singular. This is because the third column and the third row of $\mathbf{A}$ contain only zeros. Therefore, the determinant is zero and the matrix is not invertible. This means without the ground reflection it is not possible to estimate the height of the target:

$$\text{VDOP}_{\text{LOS}} = \infty \tag{8.58}$$

whereas with ground reflections the height can be estimated.

Special Case: Receivers Much Higher Than the Sender

From the TOA case we know that if $z_r \gg S_z$, then $u_{i,z}^* \approx -u_{i,z}'$. In order to simplify the equations, it is important to note that using pseudo-range measurements the vertical dilution

of precision is exactly the same as using time difference of arrival [83]. Pseudo-range measurements are defined in a way such that one estimates the variables x, y, z and an additional variable d_0 which contains the sending time. Then, a measurement m_i from the receiver i would be:

$$m_i = \|\mathbf{M}_i - \mathbf{S}\| + d_0 \tag{8.59}$$

Then, we can reformulate $\mathbf{A}$ as:

$$\mathbf{A} = (\mathbf{H}_{pr}^T \mathbf{R}_t^{-1} \mathbf{H}_{pr}) \tag{8.60}$$

where:

$$\mathbf{H}_{pr} = \begin{pmatrix} \frac{S_x - M_{x,1}}{R_1} & \frac{S_y - M_{y,1}}{R_1} & 1 & \frac{S_z - M_{z,1}}{R_1} \\ \vdots & \vdots & \vdots & \\ \frac{S_x - M_{x,N}}{R_N} & \frac{S_y - M_{y,N}}{R_N} & 1 & \frac{S_z - M_{z,N}}{R_N} \end{pmatrix} \tag{8.61}$$

One can observe how now we have a similar matrix as in the TOA case, but with an additional column which contains the derivative with respect to the distance offset d_0 corresponding to the estimated sending time. Then we can define:

$$\mathbf{A} = \begin{pmatrix} A_{xx} & A_{xy} & A_{xd_0} & A_{xz} \\ A_{xy} & A_{yy} & A_{yd_0} & A_{yz} \\ A_{xd_0} & A_{yd_0} & A_{d_0 d_0} & A_{zd_0} \\ A_{xz} & A_{yz} & A_{zd_0} & A_{zz} \end{pmatrix} \tag{8.62}$$

We know that:

$$A_{xz} = \sum_{k=1}^{N} \left(\frac{S_x - M_{x,k}}{\sigma_k^2 R_k} \right) \left(\frac{u_{k,z}}{\sigma_k^2} \right) \tag{8.63}$$

Then, as the LOS signals will have a component $u_{k,z}$ with the opposite sign and the same value as their corresponding ground reflection, this value will be close to zero ($A_{xz} \approx 0$) if the noise standard deviation is the same for all measurements. The same can be applied for the y component ($A_{yz} \approx 0$). Regarding the distance offset d_0:

$$A_{zd_0} = \sum_{k=1}^{N} \left(\frac{u_{k,z}}{\sigma_k^2} \right) \tag{8.64}$$

which will be also close to zero if all measurements have an identical noise distribution and we receive the corresponding ground reflection of every line-of-sight signal.

In a similar way as we have seen in Eq. 8.23:

$$\sigma_n^2 \text{VDOP}^2 = \frac{\det(\mathbf{A}_{1:3,1:3})}{\det(\mathbf{A})} \tag{8.65}$$

As the fourth column and fourth row of $\mathbf{A}$ will contain only zeros except for the component A_{zz}, one can simplify the equation above as:

$$(\sigma_n \text{VDOP}_{\text{W/GROUND}})^2 \approx \frac{1}{A_{zz}} \tag{8.66}$$

which is equivalent to:

$$(\text{VDOP}_{\text{W/GROUND}})^2 \approx \frac{1}{\sum_{i=1}^{N}(u_{i,z})^2} \tag{8.67}$$

Then, in this scenario, the VDOP using ground reflections is the same as using TOA when the noise distribution is the same for all measurements. It is the same as it would be if the variables S_x, S_y and d_0 were perfectly known in advance.

We can conclude that in the two extreme cases for the height of the target the ground reflection presents a considerable improvement in the VDOP, as seen in Table 8.1.

	$S_z \ll z_r$	$S_z \approx z_r$
VDOP using ground reflections	Same as knowing S_x, S_y and d_0 in advance	$\leq \infty$
VDOP not using ground reflections	$\leq \infty$	∞

Table 8.1: Comparison of the VDOP in the extreme cases for the height of the sender S_z. We consider the receivers are placed at the same height z_r and all measurements have an identical noise distribution.

8.4 Exploiting Ground Reflection

We have seen how ground reflections can be used to increase the precision of the height estimation. However, one needs to identify which signals are ground reflections in order to use them as additional information. In this section we provide a simple solution [7] for identifying the ground reflections from a smartphone with limited computational power and without relying on position estimations.

The ground can be modeled as an infinite plane. This plane can be defined with a normal vector $\hat{\mathbf{n}}$ and a point of the plane P.

Having a sender at position $\mathbf{S}$, the ground will produce the same measurement as a virtual sender $\mathbf{S}_v$ (see Fig. 8.2). Its position depends on the plane and the sender position:

$$\mathbf{S}_v = 2(-\hat{\mathbf{n}}\|\mathbf{D}^T\hat{\mathbf{n}}\|) + \mathbf{S} \tag{8.68}$$

$$\mathbf{D} = \mathbf{P} - \mathbf{S} \tag{8.69}$$

Then, a microphone at position $\mathbf{M}_i$ will receive the line-of-sight signal $T_{L,i}$ and the reflection from the ground $T_{r,i}$:

$$T_{L,i} = \frac{1}{c}\|\mathbf{S} - \mathbf{M}_i\| + t_s$$
$$T_{r,i} = \frac{1}{c}\|\mathbf{S}_v - \mathbf{M}_i\| + t_s \tag{8.70}$$

where t_s is the sending time. In order to simplify the equations, we define:

$$\overline{T}_{L,i} = T_{L,i} - t_s$$
$$\overline{T}_{r,i} = T_{r,i} - t_s \tag{8.71}$$
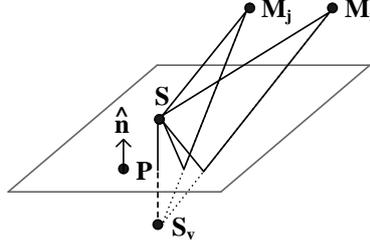
Figure 8.2: Schematic of the ground reflection. The speaker at position $\mathbf{S}$ emits a sound signal. The line-of-sight signal and the reflection from the ground are received by the receivers $\mathbf{M}_i$ and $\mathbf{M}_j$. [7]

Having more than one receiver, one can calculate the difference of two squares. In the case of the line of sight signals:

$$\overline{T}_{L,i}^2 - \overline{T}_{L,j}^2 = \frac{\|\mathbf{M}_i\|^2 - \|\mathbf{M}_j\|^2 - 2\mathbf{S}^T(\mathbf{M}_i - \mathbf{M}_j)}{c^2} \tag{8.72}$$

The relation between the ground reflections is:

$$\overline{T}_{r,i}^2 - \overline{T}_{r,j}^2 = \frac{\|\mathbf{M}_i\|^2 - \|\mathbf{M}_j\|^2 - 2(2(-\widehat{\mathbf{n}}\|\mathbf{D}\widehat{\mathbf{n}}\|) + \mathbf{S})^T(\mathbf{M}_i - \mathbf{M}_j)}{c^2} \tag{8.73}$$

If the normal vector of the plane is perpendicular to the vector $\mathbf{M}_i - \mathbf{M}_j$, then the product $(\widehat{\mathbf{n}}\|\mathbf{D}\widehat{\mathbf{n}}\|)^T(\mathbf{M}_i - \mathbf{M}_j)$ is zero:

$$\overline{T}_{r,i}^2 - \overline{T}_{r,j}^2 = \frac{\|\mathbf{M}_i\|^2 - \|\mathbf{M}_j\|^2 - 2\mathbf{S}^T(\mathbf{M}_i - \mathbf{M}_j)}{c^2} \tag{8.74}$$

Then, the timestamps of the ground reflection and the timestamps of the line of sight signals are related as follows:

$$\overline{T}_{r,i}^2 - \overline{T}_{r,j}^2 = \overline{T}_{L,i}^2 - \overline{T}_{L,j}^2 \tag{8.75}$$

This means, if the receivers are in a plane parallel to the ground, the sending time can be estimated using only the information of the received timestamps:

$$(T_{r,i} - t_s)^2 - (T_{r,j} - t_s)^2 = (T_{L,i} - t_s)^2 - (T_{L,j} - t_s)^2 \tag{8.76}$$

Then, isolating the variable t_s:

$$t_s = \frac{T_{r,i}^2 - T_{r,j}^2 - T_{L,i}^2 + T_{L,j}^2}{-2T_{L,i} + 2T_{L,j} - 2T_{r,j} + 2T_{r,i}} \tag{8.77}$$

LOS and Ground Reflection Detection by RANSAC

Random sample consensus [84] is a recursive method which has been widely used for detecting samples which fulfill a certain model, eliminating the *outliers* and finding the *inliers*. In our

approach [7] we use it in order to identify the LOS signals and the reflections from the ground, eliminating reflections from walls and objects.

First, one has to realize that the sending time is independent of the smartphone position and the receivers. The Eq. 8.77 can be used to determine which samples are outliers for a certain sender position. However, if the sender emits signals at regular intervals I, multiple sender positions can be used to estimate the sending time. The reception times are:

$$
\begin{aligned}
T_{L,i,k} &= \frac{1}{c}\|\mathbf{S} - \mathbf{M}_i\| + t_0 + kI \\
T_{r,i,k} &= \frac{1}{c}\|\mathbf{S}_v - \mathbf{M}_i\| + t_0 + kI
\end{aligned}
\tag{8.78}
$$

If the interval I is much larger than the propagation time, the number of elapsed intervals k can be estimated, as explained in previous chapters. Then, we subtract the elapsed intervals to every received timestamp:

$$
\begin{aligned}
T_{L,i} &= T_{L,i,k} - kI \\
T_{r,i} &= T_{r,i,k} - kI
\end{aligned}
\tag{8.79}
$$

One can use Eq. 8.77 to estimate t_0, which has to be the same for all sender positions. The procedure of the RANSAC method is:

1. Select four timestamps $T_{L,i}, T_{r,i}, T_{L,j}, T_{r,j}$
2. Subtract elapsed intervals kI from all of them
3. Estimate sending time t_0 using Eq. 8.77
4. If t_0 is close in time to a previously estimated sending time, add the timestamps to its pile, otherwise create a new pile
5. Repeat the previous steps for all the possible combinations of timestamps
6. Choose the t_0 pile with most *inliers* and output the median of the estimated t_0 as the sending time for localization.

Position Estimation

Once the timestamps from N receivers are selected, the position of the sender $\mathbf{S}$ can be estimated with non linear least squares algorithms by minimizing:

$$
\sum_{i=1}^{N} \arg \min_{\mathbf{S}} \left((f_i)^2 + (f_i')^2 \right)
\tag{8.80}
$$

where

$$
f_i = \|\mathbf{M}_i - \mathbf{S}\| + c(t_0 - T_{L,i})
\tag{8.81}
$$

and

$$
f_i' = \|\mathbf{M}_i' - \mathbf{S}\| + c(t_0 - T_{r,i})
\tag{8.82}
$$

where $\mathbf{M}_i'$ is the position of the receiver i with inverted height (the virtual receiver). The algorithms used to minimize the error function are gradient descent and Gauss-Newton.

8.5 Experimental Results

Real experiments were done to prove the feasibility of the proposed approach and observe how much improvement can the ground reflection provide in reality. The ASSIST system was used. A total number of nine receivers were mounted on the ceiling at a median height of 4.84 m. Due to the irregularities of the construction, the height was not exactly the same for all of the receivers. The maximum height was 4.879 m and the minimum height was 4.824 m.

Static Sender

First of all, we placed a smartphone at six positions with different heights (1.39 m and 2.12 m). The aim of this first experiment was to show experimentally whether the ground reflections can increase the precision of the height estimations, as predicted with the Cramer-Rao lower bound. Therefore, in this first experiment the reference positions were used to find the most likely ground reflections and line-of-sight signals. The algorithm to detect the ground reflections was not used for this experiment. We estimated the positions of the target using ground reflections and not using them. The resulting root mean square error for the height estimation can be seen in Fig. 8.3.
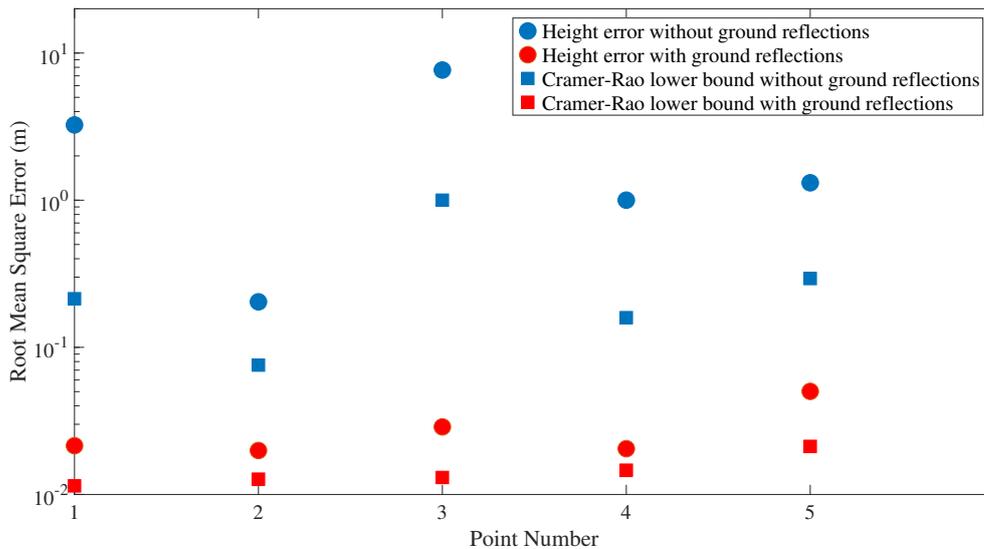
Figure 8.3: Cramer-Rao lower bound for the root mean square error compared with the actual error estimating the height of the sender using TDOA. One can observe how using the ground reflections, the error estimating the height of the target is drastically reduced. A median of 621 points are used per sender position.

The resulting errors are compared with the Cramer-Rao lower bound. A median of 621 estimated positions were used for every position of the sender. Using ground reflections the lower bound is drastically reduced and consequently the measured localization error is reduced too. The standard deviation of the noise is assumed to be 0.044 ms. Nonetheless, the actual

noise was not normally distributed. Especially since it presented a heavy tail due to errors in the detection of the timestamps and in the synchronization. However, a Gaussian distribution is a best-case scenario which allows us to estimate a lower bound for the error. More information about the estimation of the standard deviation can be seen in the Appendix of this thesis.

In order to test the proposed approach for detecting ground reflections another experiment was done with the smartphone standing still in four positions with two different heights (1.39 m and 2.12 m), resulting in eight different positions. A total number of 395 positions were estimated, approximately 50 estimated positions per real position. A picture of the experimental set-up can be seen in Fig. 8.4. The timestamps selected by the RANSAC method were used for estimating the positions using TDOA, TOA with ground reflections and TOA with only line-of-sight signals (see Fig. 8.5). The sending time used for TOA is the one estimated by the RANSAC method. The height estimation is clearly improved by using the ground reflections (see Fig. 8.6). With TOA, the maximum RMSE using ground reflections is 0.17 m whereas the maximum RMSE without them is 1.52 m.

Figure 8.4: A smartphone was placed at eight different positions. Four of them with a height of 1.39 m and four of them with a height of 2.12 m. [7]

One can also calculate the median error and the standard deviation. Using the ground reflections, the height estimation has a median error of 0.042 m and a standard deviation of 0.017 m. Not using the ground reflections, with TOA the median error is 0.249 and the standard deviation 0.131. The greatest error is achieved using TDOA: 0.572 m median error and a standard deviation of 0.270 m.

Moving Sender

Another group of experiments were done in order to test the capability of the presented approach of estimating the target position in a more realistic scenario. To do this, a person was holding a smartphone while walking and standing still at different positions. The person was moving the smartphone up and down to different heights. The motion capture system was used
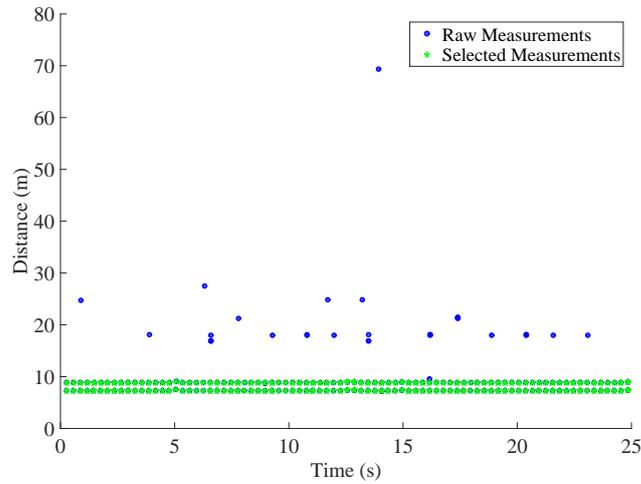
Figure 8.5: The blue dots show the raw timestamps received by a certain receiver with a sender standing still. The RANSAC method selects the timestamps which fulfill Eq. 8.77. The selected timestamps are the line-of-sight signal and the ground reflection. [7]
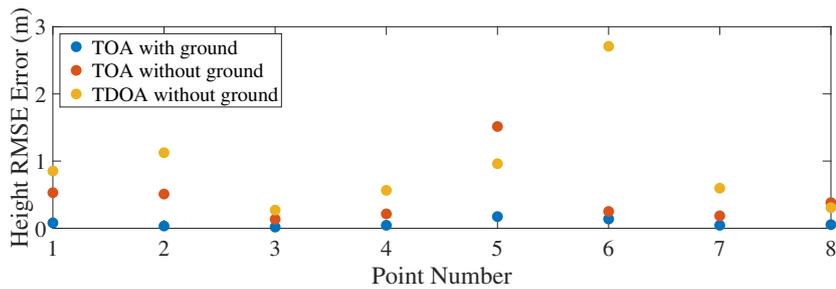
Figure 8.6: Root mean square error in the height estimation using TDOA, TOA and TOA with ground reflection as virtual receivers. The height estimation is improved in all the points by using the ground reflections. The maximum RMSE using ground reflections is 0.17 m. The sending time is estimated using the RANSAC method.

as a ground-truth.

First we analyze again the error in the height estimation. In Fig. 8.7 one can see the cumulative error distribution when the sending time is estimated with RANSAC. Using the ground reflections, the error estimating the height is considerably reduced. When using ground reflections and the estimated sending time, the median error is 0.022 m, which is the same as using time difference of arrival. However, when no reflections are used, the estimated sending time provides a better height estimation (0.141 m median error) than using time difference of arrival (0.289 m median error). This can be easily explained by looking at the virtual dilution of precision of the estimated positions (Fig. 8.8). One can see how, when the ground reflections are used, the dilution of precision in the z coordinate is very similar for TOA and TDOA. Then, using an estimation of the sending time has a reduced impact on the final estimation error.

The cumulative distribution when the real sending time is known can be seen in Fig. 8.9. In this case, the error is slightly reduced when using TOA. When no ground reflections are used, knowing the sending time improves drastically the height estimation.

By looking at the VDOP figure (Fig. 8.8) one can clearly observe how using ground reflections reduces drastically the VDOP. The maximum VDOP using them is 1.17, using time difference of arrival. The maximum VDOP without using the ground reflections and using TOA measurements, is 4.86, when the target is at 2.98 m height. Using TDOA measurements the maximum is at 23.72.
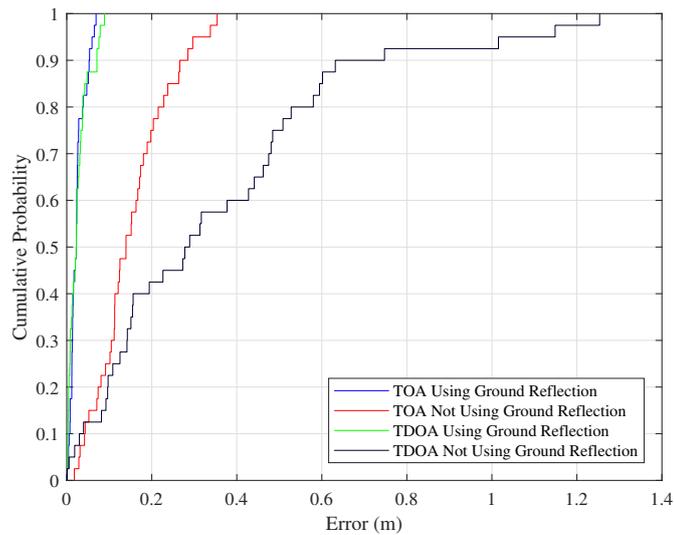
Figure 8.7: Cumulative distribution of the height estimation error when the sending time is estimated by the presented approach. Using ground reflections and the estimated sending time, the median error is 0.022 m, which is the same as using time difference of arrival. When no reflections are used, the estimated sending time provides a better height estimation (0.141 m median error) than using time difference of arrival (0.289 m median error).

In Table 8.2 one can see the median error of the different algorithms. The best result is achieved when using the real sending time and ground reflections.

TOA with reflections and using known sending time	0.014 m
TOA with reflections and using estimated sending time	0.022 m
TDOA with reflections	0.022 m
TOA without reflections and using known sending time	0.098 m
TOA without reflections and using estimated sending time	0.141 m
TDOA without reflections	0.289 m

Table 8.2: Comparison of the height median error.

When the sending time was estimated by the presented approach, the resulting positions can
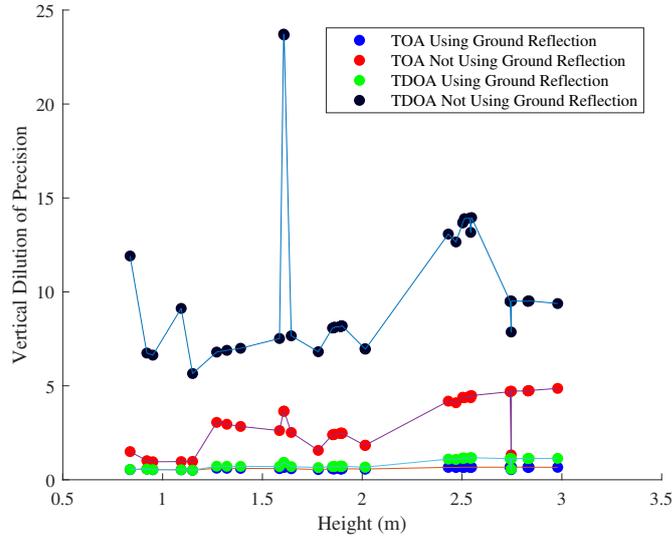
Figure 8.8: We can observe how the vertical dilution of precision is considerably reduced using ground reflections as additional information. This happens for different target heights. The maximum VDOP using ground reflections is 1.17, which allows a precise height estimation.
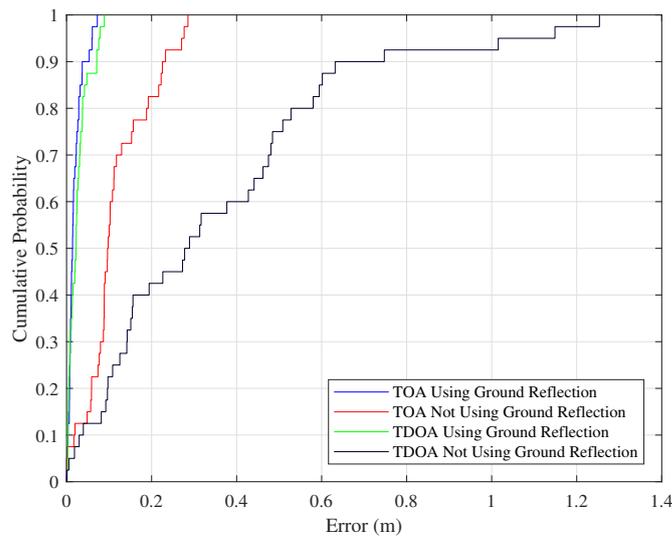
Figure 8.9: Cumulative distribution of the height estimation error when the sending time is estimated using the MoCap data.

be seen in Fig. 8.10. One can observe how the target is not continuously tracked, as the ground reflections are not always present. The error of these estimations in two dimensions can be seen in the cumulative distribution in Fig. 8.10. In this case, the sending time is estimated using the presented approach. One can observe how the ground reflections do not improve the final estimation in two dimensions. This, while it can be seen as unexpected or surprising, it

can be explained. When the sending time is estimated and the ground reflections are not used, the error in the sending time is mitigated by changing the estimated z coordinate. Then, if the estimated sending time is lower than the real one, the estimation will decrease (erroneously) the target height to fit the equations. When reflections are used, the height is constrained by the reflections, which do not allow this. It is also worth mentioning that the estimated sending time provides a lower localization error than just using TDOA. This is because in this case the DOP difference is higher between TOA and TDOA (see Fig. 8.12).

(a) Cumulative distribution of the two-dimensional error.

(b) Estimated two-dimensional positions.
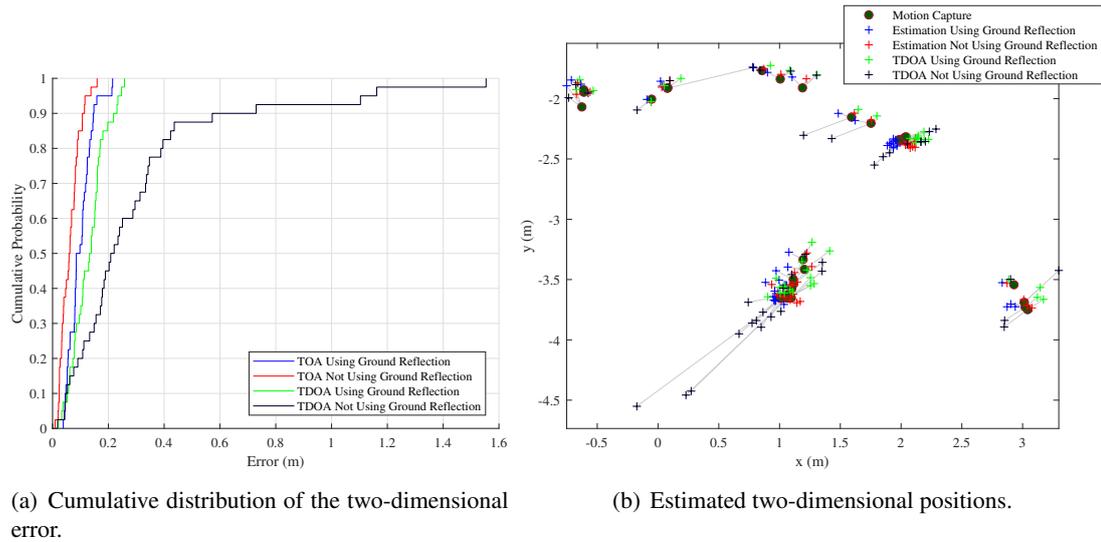
Figure 8.10: Experimental result in two dimensions using the estimated sending time. In this case, when using TOA, the ground reflection slightly increases the error. This is because when no reflection is used, the estimated height can correct errors in the sending time, by giving a wrong height estimation (see Fig 8.7). When reflections are used, the position of the target has more constraints.

Using the sending time calculated with the Mocap data, the ground reflections provide an improvement on the two-dimensional error (see Fig. 8.11). The median error is 0.047 m with ground reflections and 0.091 m without ground reflections.

The dilution of precision in two dimensions can be seen in Fig. 8.12. One can see how the dilution of precision in two dimensions is also reduced when using ground reflections. It is interesting to observe how using TDOA without ground reflections the dilution of precision is much higher than with the other alternatives. An example can be seen in Fig. 8.13. In this position, the DOP in two dimensions has its maximum (28.17). The figure shows how three hyperboloids do not intersect clearly in one point and a small error in the measurements can change dramatically the estimated position.

In two dimensions, for most of the positions, knowing the sending time leads to a higher improvement than the ground reflections, as using TOA without reflections provides a lower DOP than using TDOA and ground reflections. It is also worth mentioning that using the ground reflections the dilution of precision in the coordinates x and y is higher than in the
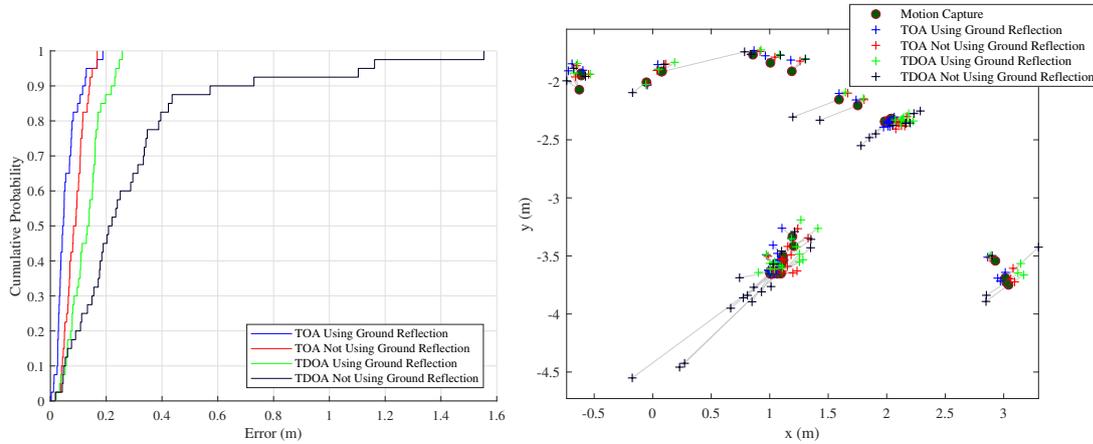
(a) Cumulative distribution of the two-dimensional error.

(b) Estimated two-dimensional positions

Figure 8.11: Using the sending time estimated using the ground-truth data. In this case, using ground reflections also improves the two-dimensional estimations.
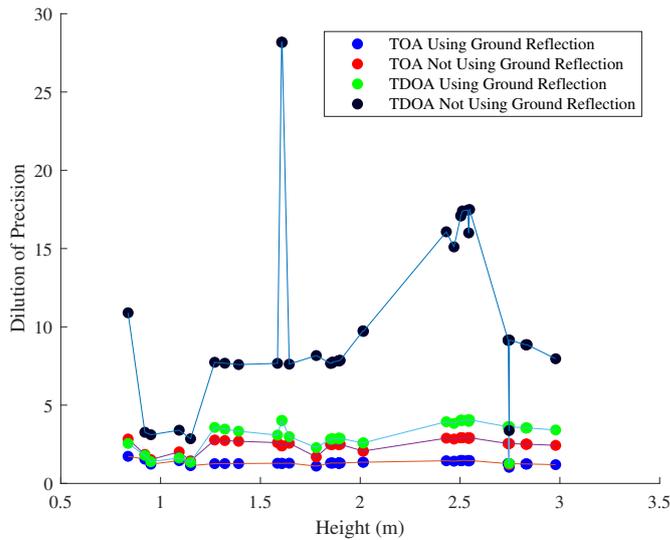
Figure 8.12: Two-dimensional dilution of precision. One can see how the ground reflections also improve the two-dimensional estimation. This is because the ambiguity in the z coordinate can increase the error in the other dimensions (see Fig. 8.13).

coordinate z. The lowest DOP is achieved by using TOA and ground reflections. The median VDOP is then 0.60 whereas the median DOP in x and y is 1.30.

In conclusion, we have seen how using the ground reflections one can considerably reduce the error estimating the height of the target. This is consistent with the lower bounds found in
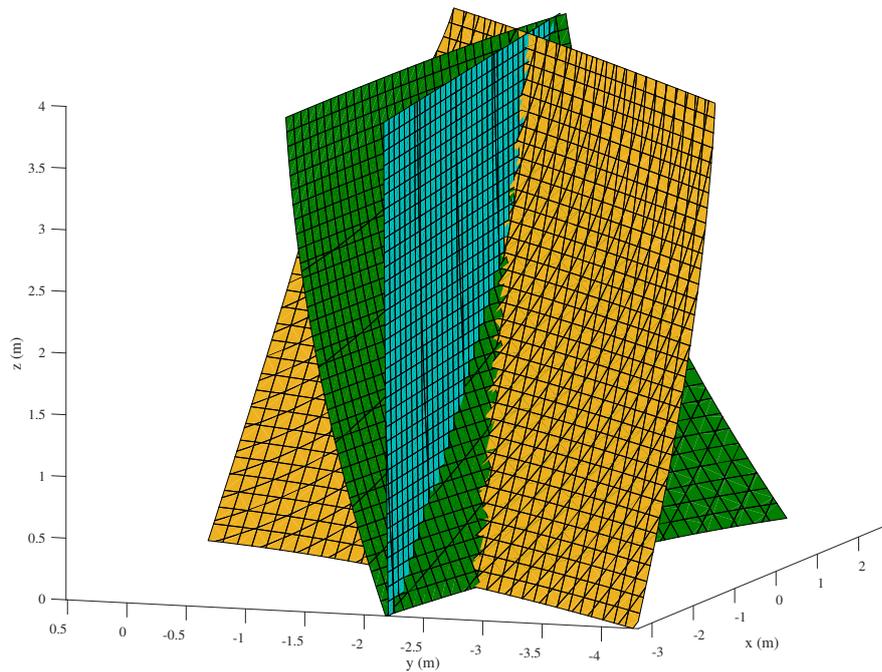
Figure 8.13: It is revealing to look at the case when the TDOA dilution of precision is maximum. We can observe how the intersection between the three hyperboloids can change drastically by changing slightly the measurements. It is interesting to observe how assuming a wrong height would lead to a wrong xy position, as the intersection between the two dimensional hyperbolas would move in the x direction.

the previous sections. The proposed method for finding the ground reflections and line-of-sight signals proves to be capable of doing the data association. The estimated sending time by the method, while it has a certain error, provides a lower three-dimensional error than just using TDOA.

9 Conclusions and Outlook

With this work we have contributed to the field of indoor localization. We have presented multiple approaches for localization in mixed line-of-sight and non-line-of-sight conditions which work under different assumptions and constraints. In addition, we have shown how acoustic reflections can be used as an additional source of information for locating the target.

9.1 Conclusions

In Chapter 5 we have described how NLOS measurements can be detected by using the residual errors. We have proposed a novel combinatorial algorithm which does not assume any particular motion of the target. The price to pay is the computational time required to use multiple combinations of receivers. Another algorithm which obtains similar results is the robust extended Kalman filter. Then, one can use a probabilistic motion model in order to identify measurements which are not likely to be in line-of-sight. In both cases, we show how these algorithms present a limitation, they cannot identify when all the received signals come from the same wall. We also show how this issue can be overcome by fusing the data from an inertial measurement unit.

In Chapter 6 we have shown how when a sender is emitting signals at regular intervals, one can measure the relative movement of the speaker to the receivers. By tracking the small motions with a joint probability data association filter, signals that come directly from the speaker or are reflected by the same reflector can be grouped. Then, one does not need to discard non-line-of-sight signals, as reflectors can be treated as virtual receivers. Nonetheless, the groups of timestamps can have a reduced number of samples. Therefore, we propose to use a far-field approximation approach in order to estimate the position of the target and the receivers. This simplifies the equations and provides an initial solution for non-linear optimization algorithms, which would otherwise be likely to result in a local minima. Two algorithms are presented and tested with experimental and simulated data. The first one uses affine geometry, which is robust against violations of the far-field assumption. However, it requires a normalization in order to recover the euclidean structure. We propose a normalization which proves to be more robust to noise than the state-of-the-art. The second algorithm uses the ellipsoid method. We propose a constrained version of it. This method does not require normalization, although it performs worse when the positions of the target are not well distributed over the three-dimensional space. However, we show how its performance can be improved by using constrained optimization.

In Chapter 7, a moving receiver is located with static senders. By locating receivers instead of senders one can potentially locate unlimited targets. However, as the sound signals are received at different positions, this leads to an undetermined system of equations. In order

to overcome this issue, one can assume a probabilistic motion model. As there is no over-determination, it is challenging to eliminate NLOS measurements or use them as additional information. Therefore, we propose a method which uses the estimated position and velocity to predict the time instants at which the signals will be received. This is fused with the modulated symbols in order to provide a robust localization at velocities up to 1.8 m/s.

Together with the line-of-sight signals there is often a reflection from the ground. These reflections can reduce the localization error, as they provide valuable information about the height of the target. In Chapter 8 we have shown that the reflection from the ground can considerably reduce the error in the height estimation. We have shown analytically how the lower bound of the RMSE is reduced. Moreover, we have shown experimentally how one can estimate the height of the target with only 0.014 m median error.

Table 9.1 shows a comparison of the presented algorithms in this thesis. One can observe how they have different requirements and capabilities. It is also worth noticing that multiple algorithms can be combined. For instance, one could locate a moving receiver and use robust regression when estimating the position of the target. One could also use the combinatorial algorithm and exploit the ground reflection once it is detected.

Algorithm	Motion Model	A Priori Knowledge	Potential Number of Targets	Reflections as Additional Information
Combinatorial Algorithm	None	Position of the Receivers	Limited	No (Discarded)
Exploting Ground Reflection	None	Position of the Receivers, Interval Length	Limited	Yes (Ground Reflection)
Robust Regression (REKF)	Probabilistic	Position of the Receivers	Limited	No (Lower Weights)
Receiver Movement for Identification	Probabilistic	Position, IDs and Interval of the Senders, Modulation	Unlimited	No (Discarded)
Sender Movement for Identification	Probabilistic, Far-Field	Interval Length	Limited	Yes (Virtual Receivers)

Table 9.1: Comparison of the algorithms for localization in mixed line-of-sight and non-line-of-sight conditions.

9.2 Outlook

While we have shown how to locate a target in mixed line-of-sight and non-line-of-sight conditions under different scenarios, there is room for improvement.

When locating a moving sender, one could integrate the presented approach in Chapter 6 into an algorithm which would continuously estimate the position of the target and the receivers, using groups of sender positions.

The ellipsoid method can be improved by using constraints which contain all long-thin and compressed ellipsoids. Iterative solutions such as the presented in [75] can be used. In addition, one could find a formulation for more than four sender positions. With these two improvements, the ellipsoid method could potentially provide a lower localization error than using affine geometry. This is because it does not lose the euclidean constraints when estimating the variables. Therefore, it could provide the optimal solution and at the same time be robust against noise

and quasi-linear movements.

When locating a receiver, one could detect when the receiver is standing still and use such positions to increase the robustness of the system, as then the number of constraints would be larger than the number of variables. One could also use an IMU to estimate the trajectory between two received timestamps when the target is moving.

The results in Chapter 5 suggest that the data of an inertial measurement can be used to facilitate the data association. Its cumulative error would need to be corrected using acoustic measurements. A tightly-coupled fusion should be implemented which is capable of correcting the errors from the inertial measurement unit and identify the non-line-of-sight measurements.

In Chapter 8, we suggest a more detailed analysis about the improvement provided by other acoustic reflections. Signals which are reflected by the ceiling or the walls can increase the localization precision in certain scenarios. For example, a rectangular room can provide virtual receivers which are well distributed over the x and y dimensions.

Overall during this thesis we have provided novel approaches for localization in presence of non-line-of-sight measurements which will hopefully inspire new research in the field. In addition, the results presented in this thesis suggest that using acoustic reflections as additional information for localization is promising and merits more research attention.

Bibliography

[1] Alexander Ens, Fabian Höflinger, Johannes Wendeberg, Joachim Hoppe, Rui Zhang, Amir Bannoura, Leonhard M. Reindl, and Christian Schindelhauer. Acoustic self-calibrating system for indoor smart phone tracking. *International Journal of Navigation and Observation*, 2015.

[2] Lukasz Zwirello, Tom Schipper, Marlene Harter, and Thomas Zwick. UWB localization system for indoor applications: Concept, realization and analysis. *Journal of Electrical and Computer Engineering*, 2012.

[3] Joan Bordoy, Alexander Traub-Ens, Ali Sadr, Johannes Wendeberg, Fabian Höflinger, Christian Schindelhauer, and Leonhard Reindl. Bank of Kalman filters in closed-loop for robust localization using unsynchronized beacons. *IEEE Sensors Journal*, 16(19): 7142–7149, 2016.

[4] Alexander Ens, Joan Bordoy, Johannes Wendeberg, Leonhard Reindl, and Christian Schindelhauer. Unsynchronized ultrasound system for TDOA localization. In *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2014.

[5] Joan Bordoy, Christian Schindelhauer, Rui Zhang, Fabian Höflinger, and Leonhard M. Reindl. Robust Extended Kalman filter for NLOS mitigation and sensor data fusion. In *2017 IEEE International Symposium on Inertial Sensors and Systems (INERTIAL)*, pages 117–120, 2017.

[6] Joan Bordoy, Christian Schindelhauer, Fabian Höflinger, and Leonhard M. Reindl. Exploiting acoustic echoes for smartphone localization and microphone self-calibration. *IEEE Transactions on Instrumentation and Measurement*, 69(4):1484–1492, 2020.

[7] Joan Bordoy, Johannes Wendeberg, Christian Schindelhauer, Fabian Höfflinger, and Leonhard M. Reindl. Exploiting ground reflection for robust 3D smartphone localization. In *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2016.

[8] Antonio R. Jiménez and Fernando Seco Granja. Precise localisation of archaeological findings with a new ultrasonic 3D positioning sensor. *Sensors and Actuators A: Physical*, 123:224–233, 2005.

[9] Nissanka Bodhi Priyantha. *The Cricket indoor location system*. PhD thesis, Massachusetts Institute of Technology, 2005.

[10] Bin Xu, Ran Yu, Guodong Sun, and Zheng Yang. Whistle: Synchronization-free TDOA for localization. In *2011 31st International Conference on Distributed Computing Systems*, pages 760–769. IEEE, 2011.

[11] Patrick Lazik, Niranjini Rajagopal, Oliver Shih, Bruno Sinopoli, and Anthony Rowe. Alps: A bluetooth and ultrasound platform for mapping and localization. In *Proceedings of the 13th ACM conference on embedded networked sensor systems*, pages 73–84. ACM, 2015.

[12] Kaikai Liu, Xinxin Liu, and Xiaolin Li. Guoguo: Enabling fine-grained smartphone localization via acoustic anchors. *IEEE Transactions on Mobile Computing*, 15(5):1144–1156, 2016.

[13] José M Villadangos, J. Urena, M. Mazo, A. Hernandez, C. De Marziani, M.C. Pérez, F. Alvarez, J.J. García, A. Jimenez, and I. Gude. Ultrasonic local positioning system with large covered area. In *IEEE International Symposium on Intelligent Signal Processing (WISP)*, pages 935–940, 2007.

[14] J. Ureña, A. Hernández, A. Jiménez, J. M. Villadangos, M. Mazo, J. C. García, J. J. García, F. J. Álvarez, C. De Marziani, M. C. Pérez, J. A. Jiménez, A. R. Jiménez, and F. Seco. Advanced sensorial system for an acoustic LPS. *Microprocessors and Microsystems*, 31 (6):393–401, 2007.

[15] H. Schweinzer and M. Syafrudin. LOSNUS: an ultrasonic system enabling high accuracy and secure TDoA locating of numerous devices. In *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2010.

[16] Jesus Ureña, Álvaro Hernández, J. Jesús García, José M. Villadangos, M. Carmen Pérez, David Gualda, Fernando J. Álvarez, and Teodoro Aguilera. Acoustic local positioning with encoded emission beacons. *Proceedings of the IEEE*, 106(6):1042–1062, 2018.

[17] D. Ruiz, E. Garcia, J. Urena, D. de Diego, D. Gualda, and J.C. Garcia. Extensive ultrasonic local positioning system for navigating with mobile robots. In *10th Workshop on Positioning Navigation and Communication*, 2013.

[18] C. Medina, J.C. Segura, and A. De la Torre. A synchronous TDMA ultrasonic TOF measurement system for low-power wireless sensor networks. *IEEE Transactions on Instrumentation and Measurement*, 62(3):599–611, 2013.

[19] Seong Jin Kim and Byung Kook Kim. Dynamic ultrasonic hybrid localization system for indoor mobile robots. *IEEE Transactions on Industrial Electronics*, 60(10):4562–4573, 2013.

[20] M. M. Saad, Chris J. Bleakley, T. Ballal, and Simon Dobson. High-accuracy reference-free ultrasonic location estimation. *IEEE Transactions on Instrumentation and Measurement*, 61(6):1561–1570, 2012.

[21] Stefano Marano, Wesley M. Gifford, Henk Wymeersch, and Moe Z. Win. NLOS identification and mitigation for localization based on UWB experimental data. *IEEE Journal on Selected Areas in Communications*, 28(7):1026–1035, 2010.

[22] Ismail Guvenc, Chia-Chin Chong, and Fujio Watanabe. NLOS identification and mitigation for UWB localization systems. In *2007 IEEE Wireless Communications and Networking Conference*, pages 1571–1576, 2007.

[23] Lei Zhang, Danjie Huang, Xinheng Wang, Christian Schindelhauer, and Zhi Wang. Acoustic NLOS identification using acoustic channel characteristics for smartphone indoor localization. *Sensors*, 17(4):727, 2017.

[24] Enrique García, Pablo Poudereux, Álvaro Hernández, Jesús Ureña, and David Gualda. A robust UWB indoor positioning system for highly complex environments. In *2015 IEEE International Conference on Industrial Technology (ICIT)*, pages 3386–3391. IEEE, 2015.

[25] Teodoro Aguilera, Fernando J. Álvarez, David Gualda, José Manuel Villadangos, Álvaro Hernández, and Jesús Ureña. Multipath compensation algorithm for TDMA-based ultrasonic local positioning systems. *IEEE Transactions on Instrumentation and Measurement*, 67(5):984–991, 2018.

[26] Maximo Cobos, Juan J. Perez-Solano, Óscar Belmonte, German Ramos, and Ana M. Torres. Simultaneous ranging and self-positioning in unsynchronized wireless acoustic sensor networks. *IEEE Transactions on Signal Processing*, 64(22):5993–6004.

[27] Xin Wang, Zongxin Wang, and Bob O'Dea. A TOA-based location algorithm reducing the errors due to non-line-of-sight (NLOS) propagation. *IEEE Transactions on Vehicular Technology*, 52(1):112–116, 2003.

[28] Swaroop Venkatesh and R. Michael Buehrer. NLOS mitigation using linear programming in ultrawideband location-aware networks. *IEEE transactions on Vehicular Technology*, 56(5):3182–3198, 2007.

[29] Ulrich Richard Hammes. *Robust positioning algorithms for wireless networks*. PhD thesis, Technische Universität Darmstadt, 2010.

[30] Feng Yin. *Robust wireless localization in harsh mixed line-of-sight/non-line-of-sight environments*. PhD thesis, Technische Universität Darmstadt, 2014.

[31] Hao Jiang, Jie Xu, and Zhen Li. NLOS mitigation method for TDOA measurement. In *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 196–199. IEEE, 2010.

[32] Gang Wang, Anthony Man-Cho So, and Youming Li. Robust convex approximation methods for TDOA-based localization under NLOS conditions. *IEEE Transactions on Signal processing*, 64(13):3281–3296, 2016.

[33] Rui Zhang, Fabian Höflinger, and Leonhard Reindl. TDOA-based localization using interacting multiple model estimator and ultrasonic transmitter/receiver. *IEEE Transactions on Instrumentation and Measurement*, 62(8):2205–2214, 2013.

[34] Nan Hu, Chengdong Wu, Pengda Liu, Hao Wu, Boya Wu, and Long Cheng. Vote selection mechanisms and probabilistic data association-based mobile node localization algorithm in mixed LOS/NLOS environments. *Telecommunication Systems*, 62(4):641–655, 2016.

[35] Li Li and Jeffery L. Krolik. Simultaneous target and multipath positioning. *IEEE Journal of Selected Topics in Signal Processing*, 8(1):153–165, 2013.

[36] Esko Dijk, Kees Van Berkel, Ronald Aarts, and Evert Van Loenen. Estimation of 3D device position by analyzing ultrasonic reflection signals. In *2003 Annual Workshop on Circuits, Systems and Signal Processing (ProRISC 2003)*, page 88, 2003.

[37] Ivan Dokmanić, Reza Parhizkar, Andreas Walther, Yue M Lu, and Martin Vetterli. Acoustic echoes reveal room shape. *Proceedings of the National Academy of Sciences*, 110(30): 12186–12191, 2013.

[38] Kang-Wook Kim, Myung-Gon Park, Takao Hishikawa, Junghee Han, and Chang-Gun Lee. Exploiting ultrasonic reflections for improving accuracy of indoor location tracking. In *Ubiquitous Intelligence & Computing, International Conference on*, pages 87–95. IEEE, 2012.

[39] Paul Meissner, Christoph Steiner, and Klaus Witrisal. UWB positioning with virtual anchors and floor plan information. In *2010 7th Workshop on Positioning, Navigation and Communication*, pages 150–156. IEEE, 2010.

[40] Simayijiang Zhayida, Simon Segerblom Rex, Yubin Kuang, Fredrik Andersson, and Kalle Åström. An automatic system for acoustic microphone geometry calibration based on minimal solvers. *arXiv preprint arXiv:1610.02392*, 2016.

[41] Yubin Kuang, Kalle Astrom, and Fredrik Tufvesson. Single antenna anchor-free UWB positioning based on multipath propagation. In *2013 IEEE International Conference on Communications (ICC)*, pages 5814–5818, 2013.

[42] Rahul Biswas and Sebastian Thrun. A passive approach to sensor network localization. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ*, volume 2, pages 1544–1549. IEEE, 2004.

[43] Marc Pollefeys and David Nister. Direct computation of sound and microphone locations from time-difference-of-arrival data. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2445–2448, 2008.

[44] Johannes Wendeberg, Fabian Höflinger, Christian Schindelhauer, and Leonhard Reindl. Calibration-free TDOA self-localisation. *Journal of Location Based Services*, 7(2):121–144, 2013.

[45] Johannes Wendeberg. *Calibration-Free Localization using Time Differences of Arrival*. PhD thesis, Albert-Ludwigs-Universität Freiburg, 2013.

[46] Simayijiang Zhayida, Simon Burgess, Yubin Kuang, and Kalle Åström. TOA-based self-calibration of dual-microphone array. *IEEE Journal of Selected Topics in Signal Processing*, 9(5):791–801, 2015.

[47] Sebastian Thrun. Affine structure from sound. In *18th International Conference on Neural Information Processing Systems (NIPS)*, pages 1353–1360, 2005.

[48] Christian Schindelhauer and Johannes Wendeberg. Localization solely based on ambient signals. *Technical Report No. 261, University of Freiburg*, 2010.

[49] Johannes Wendeberg, Thomas Janson, and Christian Schindelhauer. Self-localization based on ambient signals. *Theoretical Computer Science*, 453:98–109, 2012.

[50] Lawrence E. Kinsler, Austin R. Frey, Alan B. Coppens, and James V. Sanders. *Fundamentals of acoustics*. Wiley-VCH, 1999.

[51] EA Dean. Atmospheric effects on the speed of sound. *The Journal of the Acoustical Society of America*, 72(2):642–642, 1982.

[52] Jont B. Allen and David A. Berkley. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 65(4):943–950, 1979.

[53] J-M Valin, François Michaud, Jean Rouat, and Dominic Létourneau. Robust sound source localization using a microphone array on a mobile robot. In *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)*, volume 2, pages 1228–1233, 2003.

[54] Alexander Traub. *Ultraschall-Übertragungssystem mit mehreren Trägerfrequenzen für unsynchronisierte Lokalisierungssysteme*. PhD thesis, University of Freiburg, 2018.

[55] K. Somani Arun, Thomas S. Huang, and Steven D. Blostein. Least-squares fitting of two 3-D point sets. *IEEE Transactions on pattern analysis and machine intelligence*, (5): 698–700, 1987.

[56] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[57] Simon J. Julier and Jeffrey K. Uhlmann. New extension of the Kalman filter to nonlinear systems. In *AeroSense'97*, pages 182–193. International Society for Optics and Photonics, 1997.

[58] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics*. MIT Press, 2005.

[59] S. Thrun, D. Fox, W. Burgard, and F. Dellaert. Robust Monte Carlo Localization for Mobile Robots. *Artificial Intelligence*, 128(1-2):99–141, 2000.

[60] Jun S. Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and computing*, 6(2):113–119, 1996.

[61] Joan Bordoy, Patrick Hornecker, F Hoflinger, Johannes Wendeberg, Rui Zhang, Christian Schindelhauer, and Leonhard Reindl. Robust tracking of a mobile receiver using unsynchronized time differences of arrival. In *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2013.

[62] Benoit Denis, J-B Pierrot, and Chadi Abou-Rjeily. Joint distributed synchronization and positioning in UWB ad hoc networks using TOA. *IEEE transactions on microwave theory and techniques*, 54(4):1896–1911, 2006.

[63] Ulrich Hammes and Abdelhak M Zoubir. Robust MT tracking based on M-estimation and interacting multiple model algorithm. *IEEE Transactions on Signal Processing*, 59 (7):3398–3409, 2011.

[64] Tommi Perala and Robert Piché. Robust extended kalman filtering in hybrid positioning applications. In *2007 4th Workshop on Positioning, Navigation and Communication*, pages 55–63. IEEE, 2007.

[65] F. El-Hawary and Yuyang Jing. Robust regression-based EKF for tracking underwater targets. *IEEE journal of oceanic engineering*, 20(1):31–41, 1995.

[66] George A. F. Seber and Alan J. Lee. *Linear regression analysis*, volume 329. John Wiley & Sons, 2012.

[67] John Neter, Michael H. Kutner, Christopher J. Nachtsheim, and William Wasserman. *Applied linear statistical models*, volume 4. Irwin Chicago, 1996.

[68] Andy Field. *Discovering Statistics Using SPSS*. SAGE Publications, 2009.

[69] Ralph Bucher and Durga Misra. A synthesizable ,VHDL model of the exact solution for three-dimensional hyperbolic positioning system. *VLSI Design*, 15(2):507–520, 2002.

[70] Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

[71] Frank R. Hampel, Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel. *Robust statistics*. Wiley Online Library, 1986.

[72] Lauro Ojeda and Johann Borenstein. Non-GPS navigation for security personnel and first responders. *Journal of Navigation*, 60(03):391–407, 2007.

[73] Josip Ćesić, Ivan Marković, Srećko Jurić-Kavelj, and Ivan Petrović. Detection and tracking of dynamic objects using 3D laser range sensor on a mobile platform. In *2014 11th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, volume 2, pages 110–119. IEEE, 2014.

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

[75] Qingde Li and John G. Griffiths. Least squares ellipsoid specific fitting. In *Geometric modeling and processing, 2004. Proceedings*, pages 335–340. IEEE, 2004.

[76] Alexander Traub-Ens, Joan Bordoy, Johannes Wendeberg, Leonhard M. Reindl, and Christian Schindelhauer. Data fusion of time stamps and transmitted data for unsynchronized beacons. *Sensors Journal, IEEE*, 15(10):5946–5953, 2015.

[77] Alexander Traub-Ens, Joan Bordoy, Johannes Wendeberg, Leonhard Reindl, and Christian Schindelhauer. Data fusion of time stamps and transmitted data for unsynchronized beacons. *Sensors Journal, IEEE*, 2015.

[78] Binghao Li, Andrew G Dempster, Jian Wang, et al. 3D DOPs for positioning applications using range measurements. *Wireless Sensor Network*, 3(10):334, 2011.

[79] Harry B. Lee. A novel procedure for assessing the accuracy of hyperbolic multilateration systems. *Aerospace and Electronic Systems, IEEE Transactions on*, (1):2–15, 1975.

[80] Maurizio A. Spirito. On the accuracy of cellular mobile station location estimation. *IEEE Transactions on vehicular technology*, 50(3):674–685, 2001.

[81] Michael Thomas Vodhanel. *Problems in GPS accuracy*. PhD thesis, Claremont Graduate University, 2011.

[82] Max Deffenbaugh, James G. Bellingham, and Henrik Schmidt. The relationship between spherical and hyperbolic positioning. In *OCEANS 96 MTS/IEEE Conference Proceedings. The Coastal Ocean-Prospects for the 21st Century*, volume 2, pages 590–595. IEEE, 1996.

[83] Dong-Ho Shin and Tae-Kyung Sung. Comparisons of error characteristics between TOA and TDOA positioning. *IEEE Transactions on Aerospace and Electronic Systems*, 38(1): 307–311, 2002.

[84] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[85] Sean Victor Hum. Lecture notes in radio and microwave wireless systems. University of Toronto, 2016. URL http://www.waves.utoronto.ca/prof/svhum/ece422/notes/07a-nearfar.pdf.

A Appendix

Receiver Localization

In order to locate a moving receiver in Chapter 7, five senders were used. The positions of the receivers and the actual trajectory of the target can be seen in Fig. A.1. Four of the senders were placed at a height of 3.4 m and one of them at a height of 2.02 m. Every sender emitted a signal with a different period of emission. The median of the periods of emission I_j was 0.301 s.
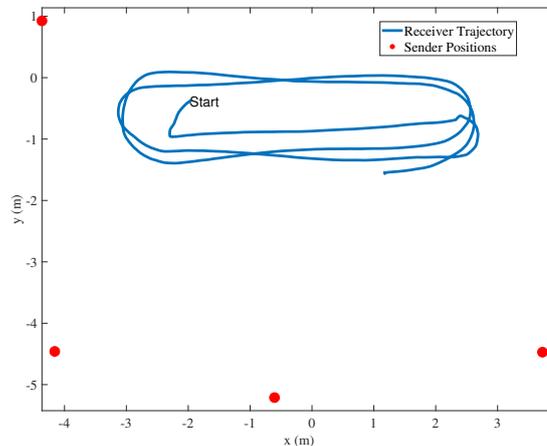
Figure A.1: Static senders and trajectory of the moving receiver. The height of the senders was 3.4 m except for the receiver in the bottom middle, which had a height of 2.02 m.

In Chapter 7 we showed how fusing the estimated velocity and position of the receiver and the detected symbol one can predict which are the line-of-sight signals of every sender. In Fig. A.2 one can observe how the proposed algorithm detects the timestamps which correspond to the estimated position and velocity of the target. One can also observe how using only the angle error of the symbols is not enough for a proper localization, as the line-of-sight signals of the sender do not always have the lowest angle error. It is also interesting to observe how the approach in [77] cannot keep track of the large time differences between the line-of-sight signals produced by the high velocity of the target. This is because the approach in [77] does not use the estimated position and velocity of the target to predict the line-of-sight signals. Therefore, it predicts that the timestamps will be received after one interval I_i, which is only true when the target is standing. The angle error is calculated as the euclidean norm of the vector which contains the angle difference between the measured four symbols $\mathbf{s}_{MS}$ and the

actual four symbols $\mathbf{s}_{GT}$ which contain the identification of the sender:

$$\phi_{\text{RXError}} = \|\mathbf{s}_{GT} - \mathbf{s}_{MS}\| \tag{A.1}$$
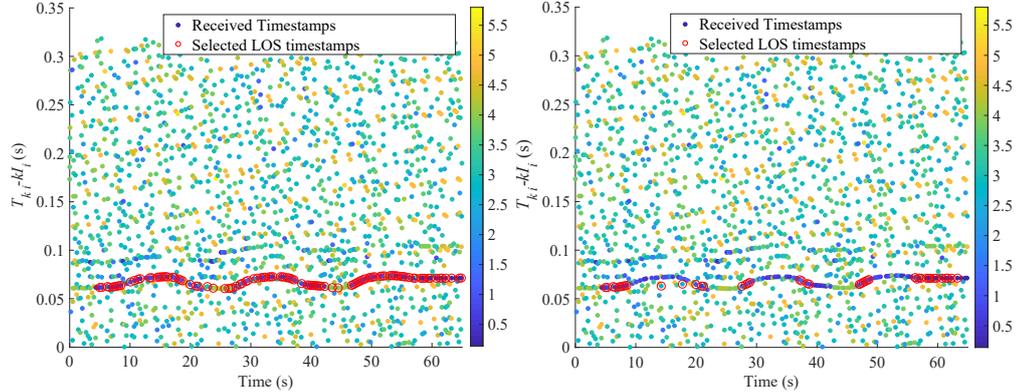
(a) Detected line-of-sight signals of the sender i using the estimated position and velocity of the target as additional information.

(b) Detected line-of-sight signals of the sender i assuming two consecutive LOS signals are received after an interval I_i.

Figure A.2: Selected line-of-sight signals of a receiver moving in the trajectory of Fig. A.1. The points with the same $T_{ki} - kI_i$ value correspond to a receiver standing still. This is why the algorithm used on the right figure tends to detect signals with a similar $T_{ki} - kI_i$ value, whereas the algorithm used in the left figure is capable of tracking time differences caused by the movement of the target.

Sender Localization

The experiments carried out in this thesis for locating an acoustic sender were done with the experimental setup in Fig. A.3. Twelve receivers were used, three of them with a height of approximately 3.40 m and nine of them with a height of approximately 4.84 m. In Chapter 8 only the receivers with larger height were used. This is because of the assumption that the receivers are placed at a similar height. The zone where the motion capture system is capable of locating is limited. Therefore, in order to be able to have a reference system, the experiments were done in that area.
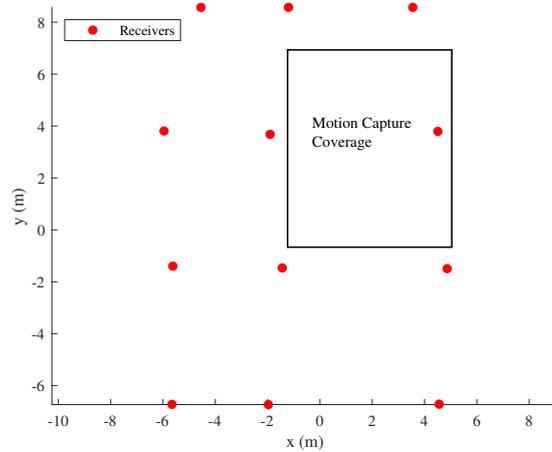
Figure A.3: Position of the receivers used for locating a sender in this thesis. The motion capture system had a limited coverage area. Therefore, the experiments were done in that area. The area marked with a rectangle is an approximation of the maximum range of the system. The three receivers at the top of the picture had a height of approximately 3.40 m whereas the others had a height of approximately 4.84 m.

Joint Probability Data Association Filter

As explained in Chapter 6, a joint probability data association filter (JPDAF) can be used to keep track of the relative movement of the sender to the receivers. This is because the movement of the target between two consecutive measurements is limited. In order to show the performance of such a data association algorithm, we record the timestamps of two different senders separately and we put the measurements together without any information about the sender which produced them. Moreover, we add a time offset to the measurements generated by one of the senders to simulate that the emitted signals by both senders are received at similar times, which makes the data association more challenging. The aim is to use JPDAF to identify which sender generated each signal. The result can be seen in Fig. A.4. One can observe how the estimated relative velocity of the target is used for predicting the next timestamp of every sender.

It is also interesting to observe how the covariance of the state increases when no measurements are received. By limiting the maximum covariance of the state one can limit the maximum amount of time that a track can be without data before being eliminated. In this example, one can see how near the second 25 there are no measurements for one of the senders during approximately 0.9 s. In this case, one could have eliminated the track and created a new one when the next measurement was received. A person with normal walking speed (1 m/s) would have resulted in a maximum time difference of 2.6 ms. Depending on the maximum velocity of the target and the number of measurements to be associated, the maximum covariance can be set differently. This explains why the tracks might have a limited number of measurements.

As every track can have a limited number of measurements, the elapsed time and therefore
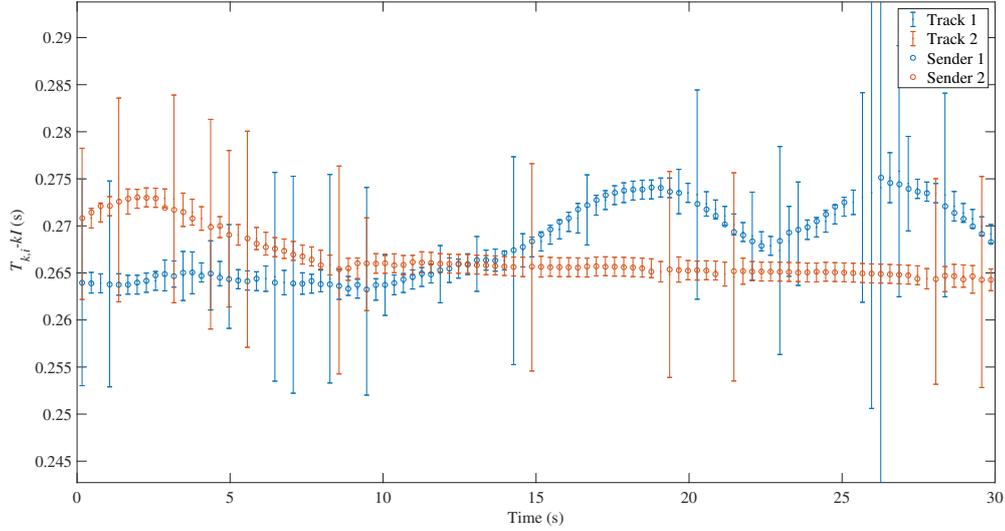
Figure A.4: Associated timestamps of two senders using a JPDAF. Two tracks estimate the relative velocity of the senders and every measurement is associated to the most likely track. The line of every track represents two standard deviations of the state. One can observe how the covariance of the state increases when no measurements are received. It also increases when a measurement is far from the predicted relative velocity. This happens when the target accelerates or the error in the measurements is large.

the movement of the target during that time is limited compared to the distances to the receivers. This allows the possibility of using a far-field assumption for initializing the non-linear optimization algorithms which would otherwise be likely to get stuck in a local minima.

Far-Field Assumption

In Chapter 6 we show how the far-field assumption can be used for locating a sender and estimating the position of the receivers. Then, the question arises of when does this assumption hold. In order to answer this, one can derive analytically the error of this assumption [85].

For simplification and without loss of generality we assume the sender $\mathbf{S}_t$ and the receivers are located in a two-dimensional scenario. A schematic can be seen in Fig. A.5.

Having a sender $\mathbf{S}_t$ and a receiver $\mathbf{M}_i$ in $\mathbb{R}^2$, the distance $d_{i,t}$ between them is:

$$d_{i,t} = \sqrt{(S_{x,t} - M_{i,x})^2 + (S_{y,t} - M_{i,y})^2} \tag{A.2}$$

For simplicity we define the coordinate system in such a way that the previous position of the sender $\mathbf{S}_{t-1}$ is in $x = 0$ and $y = 0$ and it moved in the y direction at velocity v. Then, after a certain time Δt, we know that $M_{i,y} = d_{i,t-1}cos(\varphi)$, $S_{y,t} = (\Delta t)v$ and $S_{x,t} = 0$, where φ is the angle between the velocity vector and the vector formed by $\mathbf{M}_i$ and $\mathbf{S}_{t-1}$. Then:

$$d_{i,t} = \sqrt{((\Delta t)v)^2 - 2(\Delta t)vd_{i,t-1} \cos(\varphi) + d_{i,t-1}^2} \tag{A.3}$$
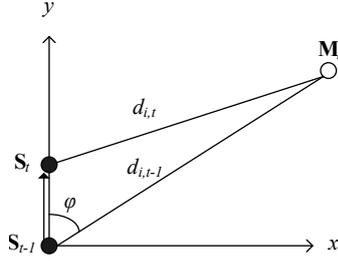
Figure A.5: Schematic of the far-field assumption. A sender moves from $\mathbf{S}_{t-1}$ to $\mathbf{S}_t$. The distance moved is assumed to be much lower than the distance to the receiver $\mathbf{M}_i$.

Using the binomial theorem, one can approximate this as:

$$d_{i,t} = d_{i,t-1} - (\Delta t)v\cos(\varphi) + \frac{((\Delta t)v)^2}{2d_{i,t-1}}\sin^2(\varphi) + ... \tag{A.4}$$

Then:

$$d_{i,t} - d_{i,t-1} \approx -(\Delta t)v\cos(\varphi) + \frac{((\Delta t)v)^2}{2d_{i,t-1}}\sin^2(\varphi) \tag{A.5}$$

One can observe how this is the sum of the far-field assumption and the far-field approximation error. Higher order terms can also increase the approximation error. However, we neglect them and focus on:

$$\epsilon_f = \frac{((\Delta t)v)^2}{2d_{i,t-1}}\sin^2(\varphi) \tag{A.6}$$

which has its maximum when the velocity vector is perpendicular to the receiver direction. This is an interesting result, as it shows that the far-field assumption error will not have the same effect on all the receivers, it will vary depending on how they are located relative to the velocity vector of the sender. Therefore, the proposed normalization method in Chapter 6 increases the robustness against violations of the far-field assumption, as it does not normalize all receivers with the same $\mathbf{C}$ matrix.

The maximum of the error term ϵ_f is:

$$\frac{((\Delta t)v)^2}{2d_{i,t-1}} \tag{A.7}$$

In Fig. A.6 we show an example of the error caused by assuming far-field when a sender moved 0.1 m and was located at a distance of 1 m from the receiver. This would be equivalent to how much would have moved a person walking at normal speed (1 m/s) during one time interval of 0.1 ms. In order to observe only the error produced by assuming a far-field, no measurement noise was added to this simulation. One can observe that even when the receiver is at only ten times the distance moved by the sender, the maximum error is only 5% of the maximum distance difference. One can also observe how the term ϵ_f is the most noticeable term of the far-field assumption error.
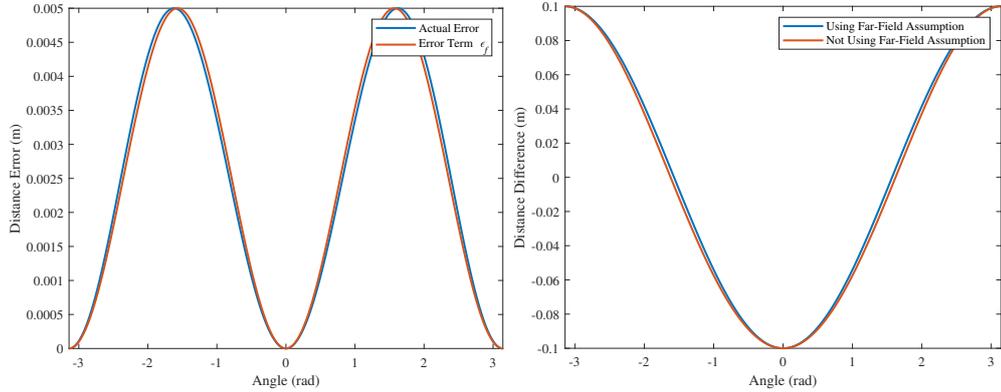
(a) Distance error between the actual values and the ones using the far-field assumption for different values of φ.

(b) In blue, measured distance difference using the far-field assumption and the actual angle φ between the senders. In red, actual distance difference.

Figure A.6: Far-field assumption errors for two sender positions located at 1 m distance from the receiver. The distance between the sender positions is 0.1 m. One can observe how the error of assuming a far-field has its minimum when the target moves in the direction of the receiver and its maximum when it moves perpendicularly to it, as seen in Eq. A.6.

Ground Reflections

In Chapter 8 we present the Cramer-Rao lower bound for locating a sender using ground reflections. In order to do so one needs to know the standard deviation of the measurement noise, which is assumed to be Gaussian distributed. For this reason, we estimated the error of 6878 line-of-sight measurements and 2738 non-line-of-sight measurements. The speaker was located at six positions with a height of 1.39 m and 2.12 m. The error of every measurement was estimated by subtracting the elapsed intervals from the received timestamps. The deviation from the median value was considered to be the error. This means, it does not include systematic errors which are clearly non-Gaussian such as the ones produced by assuming a wrong sound velocity.

In Fig. A.7 one can observe the histogram of the measurement errors for line-of-sight signals and ground reflections. The error is still not Gaussian distributed. Among other reasons, the error presents a heavy tail due to synchronization errors and systematic errors in the detection of the correct reception time. However, as we are interested in a lower bound for the localization error, the Gaussian assumption can be used. This is then a best-case scenario and provides us with the minimum achievable error using ground reflections and not using them. In the presented histograms one cannot observe a remarkable difference between the error in the line-of-sight signals and in the non-line-of-sight signals. This means, the reflected signals and the line-of-sight signals deviate from the expected time intervals in a similar manner.
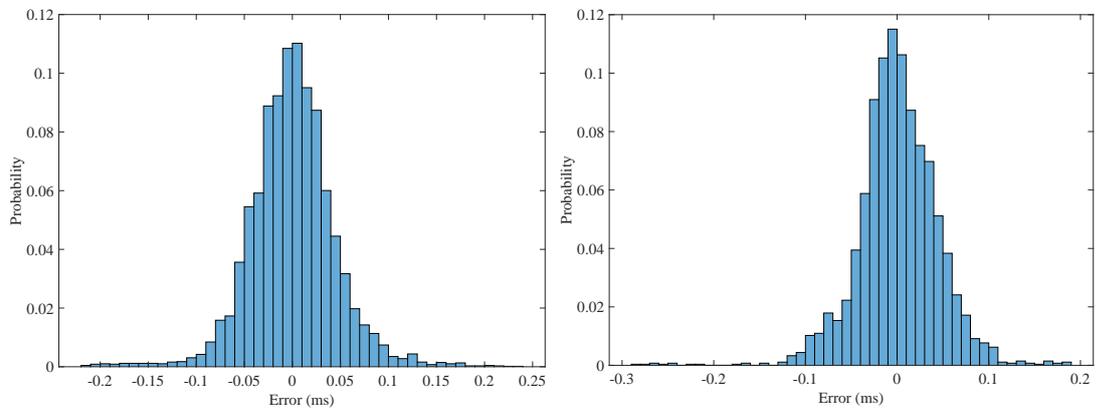
(a) Histogram of the error of the line-of-sight signals using 6878 samples.

(b) Histogram of the error of the ground reflections using 2738 samples.

Figure A.7: Histogram of the error in the received timestamps when locating a sender standing still at six different positions.