Technische Universität Wien

Dissertationsschrift

Generic Methods for Ultrasonic Scene Analysis based on Sensor Motion

unter Anleitung von Ass.Prof. Dipl.-Ing. Dr.techn. Herbert Schweinzer
Institute of Electrodynamics, Microwave and Circuit Engineering
Dipl.-Ing. Christian Walter MSc

Studiendekan: Univ.Prof. Mag.rer.nat. Dr.rer.nat. Gottfried Strasser

Gutachter: 1. Em.O.Univ.Prof. Dipl.-Ing. Dr.techn. Gottfried Magerl
2. Prof. Dr.-Ing. Bernd Henning

Rigorosendatum: November 2016

Kurzfassung

Die Verwendung von Ultraschall als physikalisches Messprinzip findet in vielen industriellen, medizinischen aber auch kommerziellen Applikationen eine breite Anwendung. In vielen Anwendungen ist die Messung von Laufzeit (engl. Time of Flight – ToF) das zugrundeliegende Messprinzip. Hierbei wird jene Zeit bestimmt, welche ein Ultraschallsignal benötigt, um eine bestimmte Strecke zu durchlaufen. Die Verwendung von Ultraschall ist vorteilhaft, da mit relativ geringem Aufwand eine hochauflösende Laufzeitmessung erfolgen kann. Weitere Vorteile von Ultraschall sind die geringen Kosten, Verwendbarkeit bei Nebel oder im Dunkeln, als auch die hohe Robustheit der Schallwandler. Dem gegenüber stehen genauigkeitsreduzierende Einflüsse wie Feuchtigkeit, Temperatur, sowie unterschiedliche Ausbreitungsmedien.

Die hier vorgestellte Arbeit befasst sich mit der Anwendung von Ultraschall in Luft zur Realisierung eines kompakten 3D-Sensors. In einer Einzelmessung bestimmt der hier vorgestellte Sensor die Position einer oder mehrerer Reflexionspunkte im Raum. Die Reflexion erfolgt hierbei typischerweise spiegelnd, da die Oberflächenrauigkeit von technischen Oberflächen, im Vergleich zur Wellenlänge von ~7mm (50kHz), gering ist. Aufgrund der daraus stark reduzierten Information ist die relative Bewegung von Messobjekt zum Sensor unerlässlich um ausreichend Messdaten zu erhalten. Ob diese relative Bewegung durch Bewegung des Sensors oder Bewegung des Messobjektes durchgeführt wird, ist hierbei von der jeweiligen Applikation abhängig. Für eine universelle Anwendbarkeit eines kompakten 3D-Sensors ergeben sich folgende Anforderungen an seine Eigenschaften: Messdaten sollen möglichst unabhängig vom Objekttyp sein, um eine vereinfachte und universelle Anwendbarkeit in den Verarbeitungsalgorithmen zu ermöglichen. Im Rahmen einer Einzelmessung soll eine maximal mögliche Anzahl an Informationen ermittelt werden, um umfassende Hinweise auf vorhandene Objekte zu erhalten. Die Anzahl an Ausreißern als auch die Messunsicherheit soll soweit als möglich reduziert werden. Diese Eigenschaften werden von dem vorgestellten Sensor, bestehend aus einem emittierenden und mehreren empfangenden Schallwandlern, weitgehend erfüllt. Die divergente Schallausbreitung von der Quelle und die Auswertung aller Echos, in Kombination mit entsprechender Signalverarbeitung zur Echoseparation und Zuordnung, maximiert die Anzahl an gewonnener Information. Durch eine entsprechende geometrische Konstruktion und angepasste Methoden werden Messfehler an Objektunstetigkeitsstellen, d.h. Objektübergängen, effizient vermieden. Die Kompaktheit des Sensors erlaubt eine nahezu objektunabhängige Messung von Reflexionspunkten.

Die Vorteile des in dieser Arbeit dargestellten universellen Sensors werden in mehreren, praxisnahen Applikationen demonstriert. Eine Szenenanalyse klassifiziert und lokalisiert mit Hilfe des kompakten 3D Sensors verschiedene Objekte im Raum, wobei hier üblicherweise geometrische Grundobjekte wie Ebene, Zylinder, etc. verwendet werden. Voraussetzung ist hierbei, dass die Position und auch Orientierung des Sensors in einem globalen Koordinatensystem bekannt ist. Die zweite vorgestellte Applikation umfasst bewegte Objekte, wie sie z.B. auf einem Förderband auftreten. Mittels einer Gruppe von kompakten 3D-Sensoren werden Parameter dieser Objekte, wie z.B. Lage, Länge und Größe bestimmt. Damit Messdaten einer Gruppe von Sensoren gemeinsam verarbeitet werden können, wird ein gemeinsames Koordinatensystem benötigt. Für die Applikation Szenenanalyse wird ein entsprechendes Verfahren zur automatischen Selbstkalibration eines Ultraschall Ortungssystems vorgestellt, welches zur Ortung des Sensors verwendet werden kann. Hierbei wird im speziellen die zu erreichende Unsicherheit analysiert und minimiert, da diese direkt Einfluss auf die Qualität der Szenenanalyse hat. Für die Kalibration einer Gruppe von Sensoren wird eine Methode vorgestellt, die unter Verwendung eines passiven oder aktiven Kalibrationsmittels ein gemeinsames, lokales Koordinatensystem erzeugen kann.

Die vorliegende Arbeit soll den Sensor detailliert beschreiben und die Möglichkeit der universellen Anwendbarkeit des kompakten 3D-Senors auf Basis von Ultraschall zur Lösung einer Vielzahl an technischen Problemen aufzeigen.

Abstract

Usage of ultrasound as physical measurement principle is common in industrial, medical and commercial applications. A large subclass of these applications employs Time-of-Flight measurements (ToF) as primary source of information. ToF measures the time an ultrasonic signal requires to travel across a given distance. Usage of ultrasound for ToF measurement is beneficial as high resolution ToF measurements are possible with relatively inexpensive equipment. Additional benefits of ultrasound are low cost, independence of illumination and the high robustness of ultrasonic transducers. Possible drawbacks are a reduction of accuracy due to humidity, temperature and changing sound propagation media – all affecting the speed of sound.

This work describes a compact 3D ultrasonic sensor using ultrasound in air. Within a single measurement the sensor obtains the position of multiple reflection points in space. Reflection for ultrasound in air is usually specular, as surface roughness is small compared to the acoustic wavelength of ~7mm (50 kHz). Due to specular reflection the amount of information is limited and relative movement between the sensor and the environment is necessary to obtain sufficient information. For this relative movement it does not matter if the movement is due to the sensor or the object. Which one is employed is most likely defined by the application context. For universal applicability of a compact 3D ultrasonic sensor the following requirements can be stated: Measurements shall be independent of the type of reflecting object to allow simplified and universal applicability of data in successive algorithms. Within a single measurement a maximal possible amount of information shall be obtained to get comprehensive information about objects. Furthermore, the number of outliers and uncertainty shall be reduced. These properties are largely fulfilled by the proposed sensor, consisting of a sound emitting device and multiple microphones. Divergent sound propagation from the source and evaluation of all returned echoes, using suitable signal processing algorithms for echo separation and the echo correspondence problem, maximizes the amount of information. Using a specific geometric construction and methods the sensor can recognize measurement errors typically occurring on object discontinuities. Compactness of the sensor results in object independence already at close proximity to the objects.

The benefits of the proposed universal sensor are shown in multiple practical applications. Scene analysis is the task of classifying and localizing objects in space. Typically simplified geometric primitives like planes, cylinders, etc. are used. A prerequisite for scene analysis is that the position and orientation of the sensor is known in a global system of coordinates. The second application covers moving objects, similar to objects on a moving conveyor belt. Using a group of compact 3D sensors, parameters of a moving object like length, width and height are determined. Combining measurements from multiple sensors in a group requires a local, common system of coordinates. For this task an automatic method for self-calibration of a group of sensors is presented, requiring only a small passive or active calibration device. Scene analysis also requires an initial calibration where a method for automatic calibration of an indoor localization system is proposed. Special attention is drawn to the aspect of measurement uncertainty as any uncertainty in the localization system directly affects the quality of scene analysis.

This works demonstrates the universal applicability of the proposed compact 3D sensor system for solving common technical problems encountered in industrial and robotic applications.

Table of Contents

1	Introduction	1
1.1	Motivation	1
1.2	Outline	1
2	State of the Art	3
2.1	Ultrasonic sensors for map building and object recognition	3
2.2	Comparison to Existing Work	10
3	Physical Background	12
3.1	Acoustic Waves	12
3.2	Sound attenuation	13
3.3	Geometrical Acoustics	16
3.4	Speakers	20
3.4.1	Piston Membrane Model	20
3.4.2	Senscomp 600 Electrostatic Transducer	28
3.4.3	Practical system identification	30
3.4.4	Other Suitable Technologies for Sound Generation	36
3.5	Microphones	38
4	Signal Processing	40
4.1	Matched Filtering	40
4.2	Linear Chirp	42
4.3	Limits on the uncertainty of range estimation	44
4.4	1-Bit binary correlation for efficient signal processing	45
4.5	Choice of sample rate	47
5	Compact 3D Sensor	49
5.1	Basic Sensor Construction	49
5.2	Elementary localization of 3D reflection points	50
5.2.1	Formula for 3D localization	50
5.2.2	Simplified formula for 3D localization	51
5.2.3	Localization in spherical coordinates	51
5.3	Uncertainty of localization	51
5.3.1	Uncertainty of localization for standard formulas	52
5.3.2	Uncertainty of localization for simplified formulas	53
5.3.3	Object and distance dependent systematic error	53
5.4	Practical localization of 3D reflection points	56
5.5	Echo correspondence problem	58
5.5.1	Matching to time difference of arrivals (m_{TDoA})	59
5.5.2	Combined correlation quality	60
5.5.3	The midpoint criterion	61
5.6	Example for 3D echo localization	62

5.7	Benefits of additional receivers	65
6	Applications	67
6.1	Application Categories	67
6.2	Moving Sensor with Global Reference System (scene analysis)	68
6.2.1	Plane reflectors	69
6.2.2	Cylindrical reflectors	70
6.2.3	Example for scene analysis	73
6.2.4	Practical Problems	77
6.3	Moving Objects with Common Sensor Reference System	83
6.3.1	Example for object classification by a static sensor array	83
7	Calibration	92
7.1	Self-Calibration of an array of sensors	92
7.2	Indoor Localization System with automatic self-calibration	95
7.2.1	Self-calibration of indoor local positioning system *LOSNUS*	95
7.2.2	Uncertainty of localization and calibration	97
7.3	Fix-Point Navigation using retro-reflectors	100
8	Conclusion	102
8.1	Applicability of presented methods	102
8.2	Outlook	102
9	Glossary	104
9.1	Notation	104
9.2	Symbols	104
9.3	Abbreviations	105
10	References	106

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken die mir während der Arbeit mit Rat und Tat zu Hilfe gestanden haben. Ohne die hervorragende Unterstützung wäre die Arbeit nicht in diesem Umfang und dieser Qualität umsetzbar gewesen.

Professor Dipl.-Ing. Dr.techn. Herbert Schweinzer danke ich für die freundliche Aufnahme in der Arbeitsgruppe Messtechnik, für die zur Verfügung gestellte Infrastruktur, und im speziellen für die Vielzahl an interessanten und anregenden Gesprächen im Bereich Akustik, Messtechnik und Ultraschall. Seine Ideen und Anregungen waren stets ein wichtiger Antrieb für den erfolgreichen Abschluss dieser Arbeit.

Em.O.Univ. Professor Dipl.-Ing. Dr.techn. Gottfried Magerl danke ich für die Vergabe des Dissertationsthemas, der Möglichkeit an Institut für Electrodynamics, Microwave and Circuit Engineering mitzuarbeiten und für die vielen wertvollen Korrekturhinweise beim Lesen der Arbeit.

Professor Dr.-Ing. Bernd Henning von der Universität Paderborn, Institut für Elektronik und Informationstechnik danke ich ganz besonders für sein Engagement im Bereich Ultraschall und Akustik. Die Tagungen im Kloster Drübeck zur messtechnischen Anwendung von Ultraschall liefern einen essentiellen Beitrag zum universitätsübergreifenden Austausch von Forschern in Bereich Akustik. Weiteres möchte ich mich für die Beurteilung und Kontrolle dieser Arbeit als Zweitbegutachter bedanken.

1 Introduction

A large subclass of industrial and commercial applications employs Time-of-Flight measurements (ToF) in air as their primary source of information. Based on this information the presence and/or position of a reflector can be determined. Ultrasound is beneficial for a large subclass of these applications due to inherent, ultrasound specific properties. First of all the low propagation speed of 343 m/s for sound waves in air, in combination with suitable signal processing, allows high resolution ToF measurements. Furthermore, typical objects within an industrial or commercial environment can be considered as sound-hard boundaries with a surface roughness much smaller than the typical wavelength of a few millimeters. Assuming specular reflection does not only provide additional insight on the reflector but also provides inherent information reduction to relevant features of the environment. Drawbacks of ultrasonic sensors are the dependence of the speed of sound on temperature, humidity and the propagation media itself. Practical issues encountered outside a well-controlled environment, i.e. the laboratory, include multipath-propagation, diffraction, echo overlapping, and non-ideal behavior of transducers. As a consequence a lot of proposed sensor systems are optimized for a single specific task and can therefore not be considered to be of general applicability. The most notable exception would be the simple ultrasonic proximity sensor where the only obtained information is the ToF for the first echo. But even this information is not well defined as the ultrasonic beam can also result in echoes outside of the acoustic axis which is center of the main lobe.

1.1 Motivation

Motivation for this work was mostly driven by the following requirements as they deemed to be essential for a sensor of general applicability: The sensor needs to be active as it must be able to decide when and how often a measurement shall take place. Furthermore, the information obtained by the sensor shall be 3D as it is only approximately possible to emit 2D waves using special transmitters. 3D is the most general form of environment and can always be simplified later. Due to the low propagation speed of acoustic waves it deems essential to obtain the maximum number of information within a single measurement. As simple this requirement initially might sound it requires spherical sound wave propagation from the sensor to “illuminate” the whole environment and processing of all returned echoes. The sensor design must therefore be able to separate different, potential overlapping echoes and also needs to be able to associate different echoes to different objects. Furthermore, the obtained data shall be of high quality which can be summarized as having low uncertainty and high robustness, i.e. only a small number of outliers. Size of the sensor shall be as compact as possible as space is also restricted by most applications.

The proposed compact 3D sensor in this work fulfills all of the requirements defined above. It is of compact size with only 80 x 80mm and consists of a centered electrostatic transmitter and four or more ultrasonic MEMS microphones. Within a single measurement the sensor can cover a field-of-view extending up to 90° in which multiple targets are detected with an uncertainty of only a few millimeters. Using a special geometric construction and suitable signal processing methods the proposed system effectively solves the echo correspondence problem and can also detect any potential outliers on object discontinuities. Global applicability, either as a group of sensors or as part of a larger system is ensured by automatic self-calibration methods avoiding any human or manually introduced errors.

1.2 Outline

This work is organized as follows: the state-of-the-art (chapter 2) gives an introduction to the topic of scene analysis, map building and object classification and covers most of the relevant literature in this field. Chapter 3 gives a brief introduction to the physical principles of acoustics

relevant in the scope of this work covering theory of sound waves, sound attenuation and geometrical acoustics. A large part of chapter 3 is dedicated to modelling and identification of the electrostatic transducer. High resolution ToF measurements are introduced in chapter 4 with a focus on 1-bit binary correlation. Limits on the uncertainty of range estimation are derived and compared to practical measurements. Using the results of chapters 3 and 4 the compact 3D sensor is introduced in chapter 5. It starts with a basic description of the construction, how ToF measurements are used for localizing 3D reflectors, and how uncertainty of the sensor system can be estimated for correlated and uncorrelated noise. The remainder of the section focuses on practical aspects for localization including solving the echo correspondence problem, outlier avoidance, object dependence and object discontinuities. Chapter 6.1 covers two different applications which can be solved by the compact 3D sensor. Chapter 6.2 covers scene analysis and map building where a moving sensor is used to analyze the environment. Suitable methods for identification of basic geometric shapes in full 3D are presented and verified in a practical test setup. A second application is introduced in chapter 6.3 where an array of compact 3D sensors is used to classify moving objects. Classification includes determination of orientation, location and object properties like width, length, height, etc. As overall system accuracy is often reduced by manual calibration chapter 7 proposes two automatic calibration methods suitable for the applications introduced in chapter 6.1. First a suitable method for establishing a common reference frame among a group of sensors is presented in chapter 7.1. The calibration method requires only an additional passive reflector which can be removed after calibration. Chapter 7.2 gives a brief introduction to the indoor localization system *LOSNUS* developed by the ultrasonic working group at the Institute of Electrodynamics, Microwave and Circuit Engineering (EMCE). This indoor localization system can be used to determine the position and orientation of the compact 3D sensor with respect to a world coordinate system. As locating uncertainty is remarkably low (in the range of millimeters) applications like map building and scene analysis are possible. Chapter 8 concludes the work with a summary of the sensor properties and identifies areas for potential future work.

2 State of the Art

Many researches have investigated the use of ultrasonic sensors attracted by the low cost, suitability in industrial environments, applications where optical sensors fail to perform well due to low visibility, mirrors or transparent objects, and their ease of use. Sadly the ease of use of ultrasonic sensors vanishes quickly if such systems are deployed in real applications. Typically, issues encountered include, but are not limited to, unknown speed of sound, low measurement repetition rate due to slow wave propagation in air, multipath propagation in complex environments, large and non-well-defined beamwidth of ultrasonic transducers resulting in echoes outside the acoustic axis, etc. All together the resulting information obtained by such sensors is extraordinary hard to interpret. Nevertheless, despite those difficulties, various sensor concepts have evolved over time being able to solve more and more complex tasks. Common to all systems is that they employ ultrasound to measure the distance between a reflector and the sensor using the time-of-flight (ToF) and the speed of sound. For ToF estimation various methods are available where [Bar98] compares a few of them with respect to bias and uncertainty. Maybe the most important distinction criterion between systems is the number of transducers used and whether the system is able to determine the bearing in 2D or 3D, either by using amplitude information, multiple ToF measurements or statistical methods. A summary of popular old and new systems is given in the following section.

2.1 Ultrasonic sensors for map building and object recognition

For mobile robot applications typical occurring tasks are obstacle avoidance, localization and recognition of obstacles. Using commercially available range finder modules, such as the popular Senscomp 600 module, distance information is easily obtained. One major problem with such sensors is the non-well-defined mapping between observed data and physical objects. This is mostly due to the wide beam width as the measured ToF only provides indirect information about the location of reflectors. Nevertheless, multiple range-finder based systems have been presented such as [Mor85, Leo91]. In [Mor85] 24 Polaroid electrostatic transducers, spaced at 15°, are used to obtain a 360° surround view of the environment. Using the known directional characteristics of the transducer probability profiles for the cases empty and occupied can be defined for a received echo. This is shown in Fig 1(a) where the probability that a space is occupied is maximal at distance R along the acoustic axis. R is the estimated distance of the reflector calculated from the ToF measurement. Using the probability information from all sensors at different positions a probability map is obtained as shown in Fig 1(b) where occupied areas are represented by “x”.

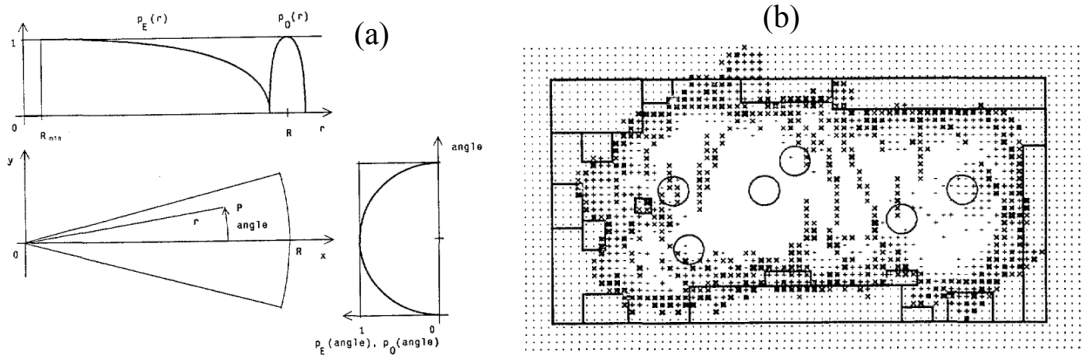


Fig 1: (a) Probability map along the acoustic axis of the sensor (top) for distance measurement R. P_O corresponds to the probability that a space is occupied and is maximal at distance R. P_E is the probability that a space is empty which is the case for distances less than R, as otherwise the echo would have been blocked. Similarly the probability profile is defined over the beam width where positions out of the acoustic axis have smaller probability. (b) Obtained map using data from all sensors at different positions. Occupied areas are represented by “x” and unknown areas by “.”. Robot positions are marked by larger circles.

The system presented in [Leo91] deals with robot navigation using geometric beacons. Geometric beacons are reliably observed, stable, naturally occurring features of the environment useful for navigation. If observed by sonar such geometric features can be used for improving navigation. For example by using an Extended Kalman filter (EKF) to estimate the robot position where the state is updated, if available, from data by the geometric beacons. The sensor uses a densely sampled 360° sonar scan as shown in Fig 2(a). It is seen that for certain headings the range values are approximately constant – hence such sequences are called regions of constant depth (RCD). These RCDs are due to the broad beam-width of the transmitter, where an echo is received, although the acoustic axis of the transmitter is not perpendicular to the reflector surface. The visibility angle over which such a sequence is detected depends on the type of object, where planes and corners result in wider regions than edges. Applying a threshold to the visibility angle the amount of information is reduced as shown in Fig 2(b). Using the observed beacons and matching them with an a-priori known map of beacon locations allows updating the state of the EKF.

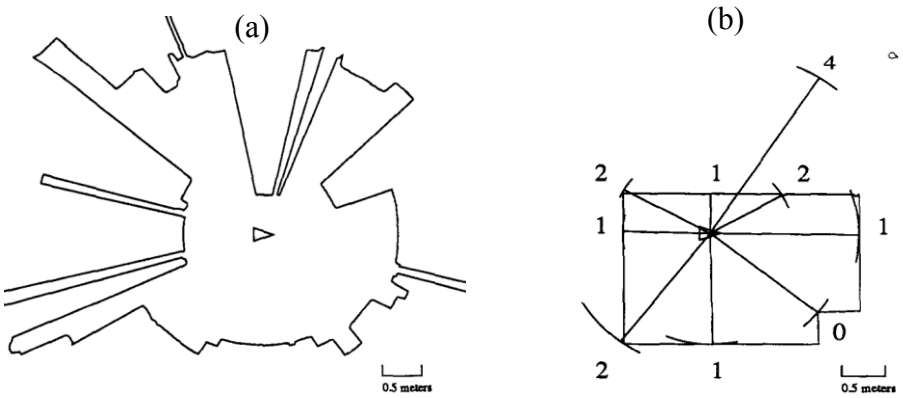


Fig 2: (a) 360° sonar scan. Strong geometric features like planes and corners result in a long sequence of sonar readings with identical range. (b) Obtained features after applying thresholding.

A limiting factor of the systems presented in [Mor85, Leo91] is that the direction of the echo is not known. Therefore, only coarse information is provided about the environment. For obtaining bearing information the system presented in [Boz91] employs a rotational scan of the device. Using a threshold the angles where the echo is first detected (start angle) and eventually vanishes (end angle) are determined. The mean of these two angles is an unbiased estimator for the direction of the reflector. Furthermore, it is possible to distinguish between planes/corners and edges as an edge results in a smaller scan arc. Both cases are shown in Fig 3.

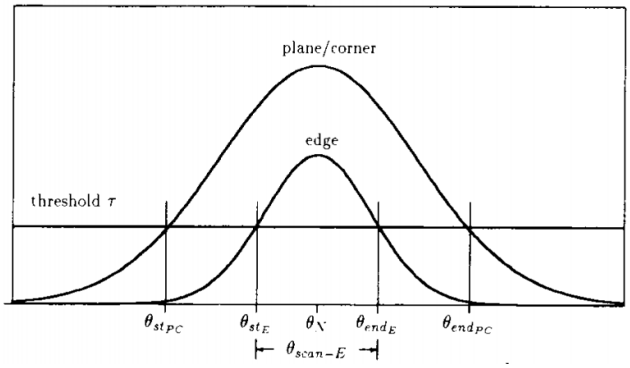


Fig 3: Orientation of the reflector is estimated using the average of θ_{st} and θ_{end}. Using the angular extent alone it is possible to differentiate planes/edges. To differentiate between a plane and a corner two different observation positions are required.

Differentiation of planes and corners is only possible by observation from two different positions. In case of a plane the arc has the same orientation but a different center. In case of a corner the orientation of the arc changes but not its center. As reported in [Boz91] the system suffers from limitations in cluttered environments due to the large beam width of the sensor and the mixture of reflected/diffracted echoes.

A more recent scanning sensor is reported in [Gia12] using four transducers. This system is able to recognize L-shaped surfaces, i.e. intersections of two planes, by sweeping the sensor such that reflections on both planes and their intersection are detected. Using an indicator which uses the ratio of the received energy level and distance, the distance itself, as well as the angular position of the swept sensor, the data points are partitioned into three sets using a fuzzy-clustering algorithm (FCM). An example for the indicator is shown in Fig 4(a). It has local maxima when the sensor is pointing towards the vertical plane (0° to 10°), when pointing towards the corner (35°-40°) and when pointing to the horizontal plane (85°-95°). After partitioning the data into three clusters, the cluster including the data points of the corner is removed. The remaining data shown in Fig 4(b) is used for reconstructing the planes using a RANdom Sample Consensus (RANSAC) algorithm. The quality of the reconstructed planes is acceptable but the main drawback of the proposed system is the limited applicability and robustness, e.g. the lack of ability to cope with any obstacles within the measurement path. The work has been continued in [Gia13] by extending it to three planes and by using Principal Component Analysis (PCA) for determination of the normal vectors. The required time for analyzing the scene is 226s due to the scanning process.

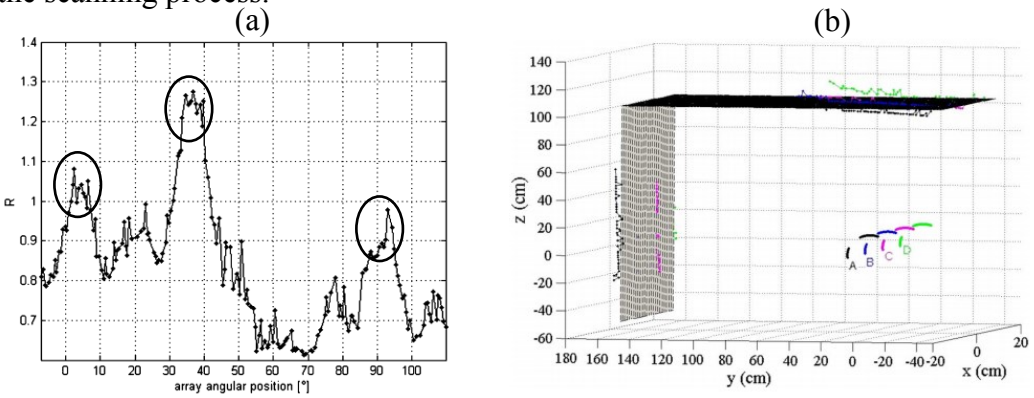


Fig 4: (a) The used indicator calculates the ratio between the received energy and the distance of the echo. This indicator is maximized when the sensor is pointing towards a plane or corner. (b) Data points after clustering used for identification of the planes. A, B, C and D correspond to the four transducers.

Because systems employing scanning are slow and on the other hand systems using sonar rings are expensive and huge, new methods, inspired by the hearing ability of bats, have evolved. Having two ears the bat can determine the bearing easily. This gave rise to the development of bi-aural and tri-aural sensor configurations. Using systems employing multiple transducers and combining their ToF measurements allows objects to be located more precisely, implement spatial filtering, and directly classify objects from a single position. It is important to distinguish this from ring-type sensors where multiple transducers are only used to increase the measurement rate and coverage. A well-known system is presented in [Per93a] which can locate and discriminate objects from a single position. It consists of three sensors lined up and spaced 16cm apart, where the central sensor is used as transmitter and receiver as shown in Fig. 5(a). Within a single measurement multiple objects being present in the vicinity of the sensor result in multiple echoes. Using the data from all three channels triplets of ToF information are obtained. Due to the spatial extent of the sensor these triplets are not only used for localization but can also be used to distinguish between planes/corners and edges. To differentiate between planes and corners movement of the sensor is necessary [Per93b]. This is shown in Fig 5(b)

where the directions of echoes stays parallel in case of a plane reflector. The sensitivity of the discrimination parameter increases with spacing and decreases with increasing distance of the object [Per93a].

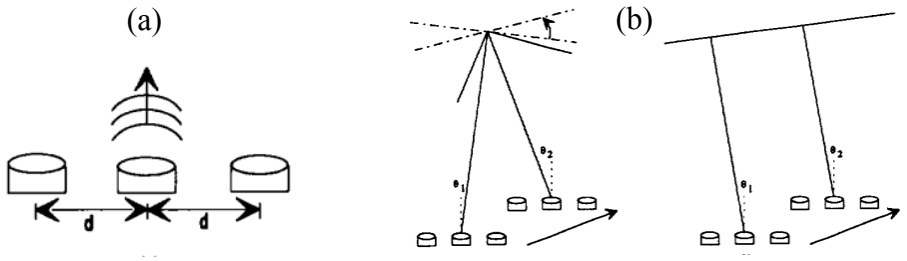


Fig 5 (a) tri-aural sensor consisting of three sensors lined up and spaced 16cm apart. The central element is used as transceiver while the left and right ones are receivers only. (b) Discrimination of planes/corners by sensor movement. In case of a plane the direction of the echo remains parallel and moves with the sensor.

Using multiple transducers for localization and classification of reflectors introduces a new problem termed "echo correspondence problem". As each receiver receives multiple ToF information during a single measurement, groups of ToF information for each receiver channel and object have to be built. This is a challenging task because information can be missing, e.g. due to obstacles or low reflection amplitude. Furthermore, it can happen that reflecting objects are in close vicinity resulting in multiple possible groups, whereas it is not possible to resolve this ambiguity only by physical sensor constraints. In addition overlapping can occur where only a single ToF information is present for multiple reflectors. This problem is illustrated in Fig 6 for the sensor used in [Per93a, Per93b]. Multiple echoes are present in each channel and proper triplets have to be selected.

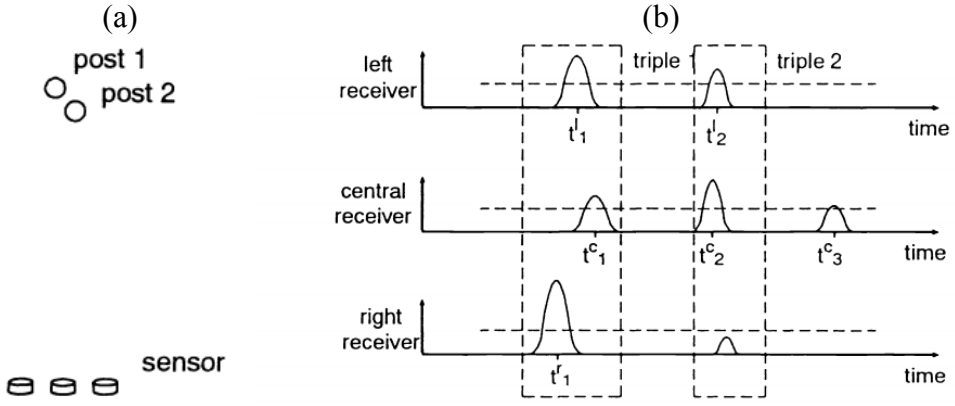


Fig 6: (a) Binaural sensor performing a measurement in a scenario with multiple reflectors. (b) Obtained data for each receiver channel. The challenge consists in choosing a proper triplet. For example the triplet t_1, t^c_2, t^r_1 would not be a sound one as it combines information from different reflectors. Furthermore, if reflectors are close echoes can overlap or can be missing.

The method suggested in [Per92] maximizes the likelihood of a given observation by seeking well-chosen objects at different positions that could have produced this data. The used hypotheses are H_1 = all echoes result from the same object, H_2 = one echo results from a different object and H_3 = one echo coincides with another object producing only a single ToF. [Per92] points out that an important advantage of such a binaural sensor system is that much less measurements have to be made to construct a reliable model of the environment.

Another well-known system is presented in [Kle94] which can classify planes, corners and edges without sensor movement. The sensor configuration is shown in Fig 7 consisting of two transmitters and two receivers. Using two transmitters is equivalent to performing two measurements at different sensor positions, yielding the ability to resolve the plane/corner ambiguity

from a single location. Contrary to other systems [Kle94] already argued that closely spacing the receivers is beneficial as the echo correspondence problem can be solved more easily. With the small spacing of only 35mm only reflectors closer than 9mm can cause a correspondence problem.

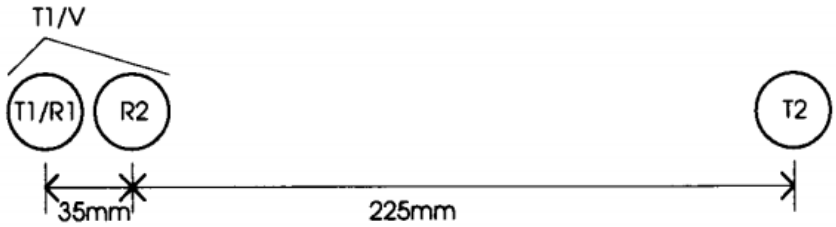


Fig 7: Vector sensor using two transmitters and two receivers. This configuration allows classification of planes, corners and edges from a single position in 2D.

If the sensor is moving during measurement this imposes additional challenges as the position of transmission and reception is no longer the same. Systems employing scanning completely lack this ability as the sweep is assumed to be performed from a single position. A system employing a binaural configuration, which includes the direction of motion, is presented in [Kle02]. It employs two transmitters and two receivers yielding four ToFs within a measurement. The difference in arrival angles is used for target classification into planes, corners and edges. The effect on movement was investigated for speeds up to 1m/s and the reported errors are less than 0.4°. Another system inherently considering motion is presented in [Bro09]. It consists of a ring of 24 receivers and a centered transmitter mounted above a parabola shown in Fig 8. Using this parabola an omni-directional horizontal wave is emitted from the sensor covering 360°. Using a high speed FPGA matched filtering is performed for every receiver pair. The time differences are used to obtain bearing information of the reflectors. The reported measurement rate of 30Hz in combination with a field of view of 360° is remarkable.

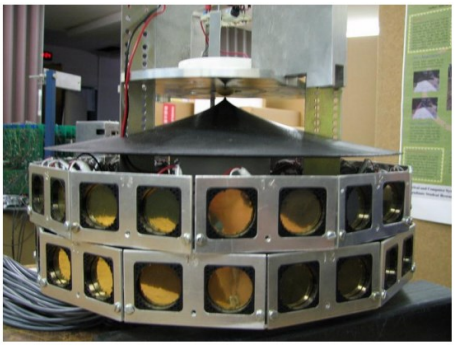


Fig 8: Sonar ring with 24 pairs of receivers used to obtain bearing. A transmitter is mounted above a parabola emitting an omnidirectional horizontal wave. The sensor covers 360° within a single measurement.

Previous systems focused on 2D problems only. In [Kle95] the sensor system of [Kle94] was extended to 3D. The basic sensor arrangement shown in Fig 9 is an orthogonal arrangement of the sensor shown in Fig 7 where one transmitter/receiver pair is shared. One challenge associated with the extension to 3D is that classification from a single position is even more complex. Whereas each 2D vector sensor only classified planes, corners and edges now combinations of such elementary objects are possible. For example a plane/plane combination is classified as a 3D-plane. But also more complex combinations are possible like an Edge/Corner classified as “Complex Point”. From the sixteen possible combinations nine are fully classified. The reported uncertainty is low with about 200µm for ranging and less than 0.2° for bearing.

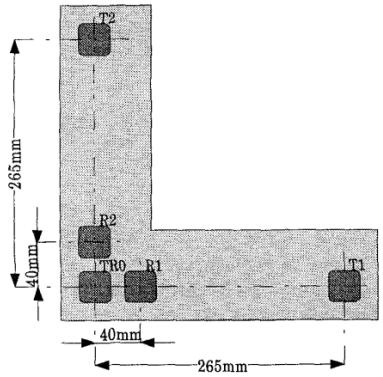


Fig 9: 3D vector sensor capable of localizing and classifying basic geometric shapes in 3D.

A different approach of a 3-D sensor is presented in [Och09]. It consists of four transducers all of them working as transmitters and receiver as shown in Fig 10. Hence a single measurement yields a total of 16 ToFs. The field of view of the sensor is partitioned into a discrete space of hexagons where the numbers of elements correspond to the spatial resolution of the sensor. In a (simulated) learning step edges, planes and corners are assumed at these locations and the corresponding ToFs are calculated. From these ToFs multiple optimal transformation matrices for principal component analysis (PCA) are calculated. Determining the type of object, distance, and bearing is performed by transforming the measured data using these matrices, recovering the feature vector, and searching the corresponding original feature vector which minimizes the recovery error. Although the basic measurement seems to work the reported performance of the system, especially for corners, is poor with only a 50% success rate for corners at a distance of 2m. Furthermore, the description of the corner-type reflector is not very generic as it assumes a special orientation.

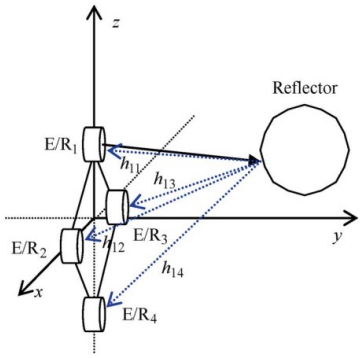


Fig 10: 3D sensor using four transducers where each is used as receiver and transmitter. Within a single measurement 16 ToFs are obtained for a single reflector. Using a simulated learning phase and partitioning the space into discrete hexagons allows localization and classification of reflectors by matching them to known data.

The availability of large processing power also resulted in the development of 3D sensors using beam forming for target localization. Such a system is presented in [Ste13] where an array of 32 randomly placed broadband receivers is used in combination with a transmitter emitting a hyperbolic chirp. The received signals are processed by a matched filter and a delay-and-sum beam former as shown in Fig 11(a). After beamforming the envelope of the signals is calculated resulting in a matrix of energy density signals. Each such signal corresponds to a small narrow band beam pointing into a given direction. By evaluating the amplitude of the energy signal it can be decided if a reflector exists at a given position or not. Reported range and bearing errors are low with an uncertainty of 2° for azimuth and elevation and 0.5mm for ranging. Separation

of multiple targets is also possible as the beam former and the used pulse compression result in a good spatial resolution. An example is shown in Fig 11(b) and (c).

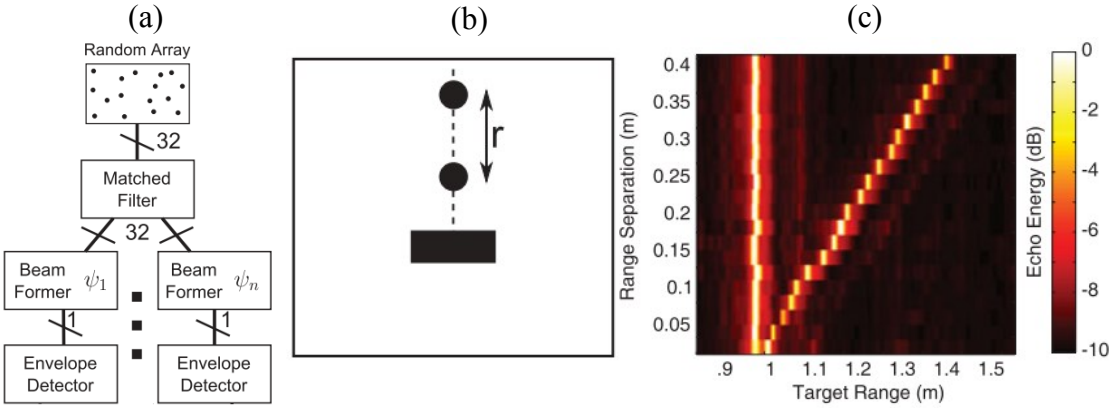


Fig 11: Echoes are sensed by a randomly placed array of 32 microphones. After matched filtering a beamforming process is applied which can steer the array in arbitrary directions. For each such direction the energy envelope is calculated and by application of a threshold objects can be detected within the energy plot. (b) Shows an example for two reflectors placed in identical direction but with varying separation. This scene was analyzed in (c) where the echo separation capabilities of the sensor system can be seen. Echo separation capability is good due to the pulse compression properties of the hyperbolic chirp.

Another approach for object recognition is using neural networks (NN) or genetic algorithms (GA). Such a system is presented in [Bab14] which uses three arrays with 16 ultrasonic sensors and two transmitters. The acoustic sound field produced by a transmitter is disturbed by an obstacle where different types of obstacles result in different kind of disturbances. Identification is performed on features from trial objects. As practical measurements are performed under different conditions the geometric features are expanded, reduced or rotated to find the corresponding reference. The system was able to identify different objects like cylinders, cones, etc. under laboratory conditions. Nevertheless, such a system cannot be directly used for collision avoidance and localization tasks.

Tab. 1 summarizes the most important properties of the different sensors concepts introduced in this chapter.

	[Mor85]	[Leo91]	[Boz91]	[Gia12]	[Per93a]	[Kle94]	[Kle02]	[Bro09]	[Kle95]	[Och09]	[Ste13]	[Bab14]
Suitable for indoor navigation	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No
2D or 3D operation	2D	2D	2D	2D	2D	2D	2D	2D	3D	3D	3D	No
Can determine bearing	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Can classify planes	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No*5
Can classify edges	No	No	Yes	No	Yes	Yes	Yes	No	Yes	Yes	No	No*5
Can classify corners	No	No	Yes	No	No	Yes	Yes	No	Yes	Yes *3	No	No*5
Classification requires movement	-	-	Yes	No	No	No	No	-	No	No	No	No
# Transmitters						1	2	1	2		1	
# Receivers					2	1	2	48	2		32	
# Transceivers	24	1	1	4	1	1			1	4	-	48

Measurement rate H = High (> 10Hz) M = Medium (> 1Hz) L =Low (< 1Hz)	L	L	L	L	M	M	M	H	M	M	M	L
Scanning	No	Yes	Yes	Yes	No	No	No	No	No	No	No	No
Range estimation MF = Matched Filter TH = Threshold CS = Coded Signal	TH	TH	TH	TH	MF	MF	MF	MF	MF	MF	MF *4	-
Field of View	360°	360°	360°	180°	<90°	<90°	<90°	360°	<90°	<30°	<65°	Fix
Aware of correspondence problem	No	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No
Stationary operation only*2	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Spacing of transducers	15°	-	-	8cm	15cm	35mm 225mm *1	40mm	15°	40mm 265mm	14cm	-	-
Resolution D=Discrete C=Continuous	D	C	C	C	C		C	C	C	D	D	D
Applicability N =Indoor navigation S = Scene analysis R = Object recognition	N,S	N,S	N,S	R	N,S,R	N,S,R	N,S,R	N	N,S,R	S,R	N,S	R

Tab. 1: Comparison of different ultrasonic sensor systems with respect to construction, localization properties and applicability to different tasks.

*1: The receivers are closely spaced to reduce the echo correspondence problem

*2: Measurements have to be performed while sensor is not moving. This is in contrast to a system which can move which results in the problem that the sensor position at the time of emission is not the same as at the time of reception.

*3: The orientation of the corner is fixed

*4: Matched filter and delay-and-sum beam former

*5: Directly classifies more complex objects like cones, cylinders, etc.

2.2 Comparison to Existing Work

The sensor we would like to propose follows a different concept than the sensors presented in the state of the art. Systems employing a surround array [Mor85, Bro09] are limited to certain applications due to cost and space constraints. On the other hand systems using a scanning approach [Leo91, Boz91, Gia12] suffer from a low measurement rate. A beamforming approach as in [Ste13] requires a huge number of microphones and processing power. Therefore, none of these approaches was followed in the design of the sensor due to above reasons.

The proposed sensors in [Per93a, Kle94, Kle95, Och09] offer high resolution and can perform measurements in a short amount of time. What is common to these systems is that they perform localization and object recognition, if possible, in a single step. This, in the author's opinion, is also the main conceptual drawback. As reported already in [Per93b] not all objects can be discriminated from a single position requiring sensor (or object) movement. Furthermore, the fact has to be stressed that although an environment typically consists of multiple primitive geometric objects, their intersection can form arbitrary complex objects. Therefore, most likely the systems presented in [Per93a, Kle94, Kle95] will not perform well in that case, as this is not included in the system model. The system in [Kle95] is partially able to resolve this situation for planes, corners and edges. Another issue is that the significance of the object discrimination

criterion depends on the spacing of the transducers and distance. A good overview is given in [Kre10] where the dominant result is that larger spacing of transducers or close proximity to reflectors eases the task. The distance of reflectors is given by the environment and cannot be chosen. Increasing the spacing seems a valuable approach but only works well if single reflectors are considered. The larger the spacing the more severe is the channel correspondence problem, that is which echoes belong to the same reflector. Also the problem of object discontinuities becomes larger, as it is more likely that reflection points are on different objects, preventing their identification and localization.

The sensor presented in this work features a compact design using four or more microphones and a single transmitter for determination of bearing and distance. As the sensor design is compact the localization becomes object independent at a reasonable distance (50cm and more) as it is the reverse result of [Kre10]. The chosen spacing of d=8cm between microphones limits the maximum time difference between ToAs. Echoes from reflectors at a different distance are therefore easily distinguished. Echoes from reflectors with distances in close proximity can also be distinguished by a simple mathematical expression as compared to [Per92] where a more complex statistical method was used. Furthermore, for convex and planar objects, the reflection points on an arbitrary surface are spaced at most by 4cm. The problem of object discontinuities therefore is limited to an area of 4x4cm. By employing redundancy the sensor is able to detect this problem and can avoid any artifacts in the results. Object recognition is no longer possible from a single position and the sensor requires relative movement of the environment to distinguish among objects. The proposed sensor is capable of identifying planes, corners, edges and cylinders in full 3D with high accuracy.

3 Physical Background

This chapter gives a brief introduction on the most important physical phenomena of acoustic wave propagation relevant within the scope of this work. Furthermore, as the proposed sensor requires a sound generating (speaker) and sensing devices (microphones), their technology is described herein as well.

3.1 Acoustic Waves

Gases, liquids and solids are all suitable media for propagation of mechanical vibrations called sound waves. While transverse waves can propagate within solids, this is not possible for liquids and gases as no shear stress component exists. Therefore, only longitudinal waves are present where the direction of displacement of the molecules is identical to the direction of wave propagation. A sound field can be fully described by the scalar sound pressure p [1 Pascal = 1 Pa = 1 N/m^2 = 94 dBSPL re. 20µPa] and the sound particle velocity **v** [m/s]. Assuming linear acoustics, i.e. no acoustical overload of the media, the governing equations for sound pressure and particle velocity are [Mey79]:

$$\begin{aligned} \rho_0 \frac{\partial \mathbf{v}}{\partial t} &= -\mathrm{grad}\ p \\ \rho_0 \cdot \mathrm{div}(\mathbf{v}) &= -\frac{1}{c^2}\frac{\partial p}{\partial t} \end{aligned} \tag{1}$$

The upper part of (1) is closely related to Newton's equation of motion. The acceleration of the mass enclosed in a small volume is the result of a force due to pressure differences on the volume boundaries. The second part of the equation follows from the adiabatic assumption for sound waves and the hydrodynamic law of conversation of mass.

For ideal gases the sound velocity can be calculated from the static pressure p_0, the density ρ_0 and the adiabatic index γ (often assumed 1.4).

$$c = \sqrt{\gamma \frac{p_0}{\rho_0}} = \sqrt{\gamma \frac{RT}{M}} \tag{2}$$

The second part of the formula uses the molar gas constant R = 8.31446 J/(mol·K), the temperature T in Kelvin, and the molar mass M = 28.97 g/mol (for air). The dependence of the speed on the sound, together with an often used linear approximation of c = 331.4[m/s] + 0.6[m/s°C] is graphically shown in Fig 12.

The set of partial differential equations (1) is a function of the sound pressure p and the sound particle velocity **v**. By calculating the partial derivative with respect to the time t the well-known result given in (3) is obtained.

$$\begin{aligned} \rho_0 \cdot \mathrm{div}\left(\frac{-\mathrm{grad}\ p}{\rho_0}\right) &= -\frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \\ \frac{1}{c^2}\frac{\partial p^2}{\partial t^2} - \mathrm{div}(\mathrm{grad}\ p) &= 0 \end{aligned} \tag{3}$$

As acoustic wave propagation is irrotational [Mey79] the vector field of the sound particle velocity can be expressed by a velocity potential (4). As the velocity potential allows direct calculation of sound pressure and sound particle velocity it is a convenient method for fully describing the sound field.

$$\Delta\Phi = \frac{1}{c^2}\frac{\partial^2 \Phi}{\partial t^2} \text{ with } p = \rho_0 \frac{\partial \Phi}{\partial t}, \mathbf{v} = -\mathrm{grad}\Phi \tag{4}$$

A particular important solution, where the sound field depends only on the distance to the source, is a spherical wave. Assuming an unbounded, homogenous media and a harmonic source of sound a particular solution of the wave equation is given by:

$$\Phi = \frac{\Phi_0}{r} e^{i\omega t} e^{-ikr}, k = \omega/c \tag{5}$$

Using (5) and (4) the pressure field can be calculated where it can be seen that the sound pressure reduces by 1/r. This corresponds to a pressure reduction of 6dB for each doubling of the distance. In a non loss-less propagation media the sound pressure is not only reduced by the diverging spherical wave but also by attenuation of the acoustic media.

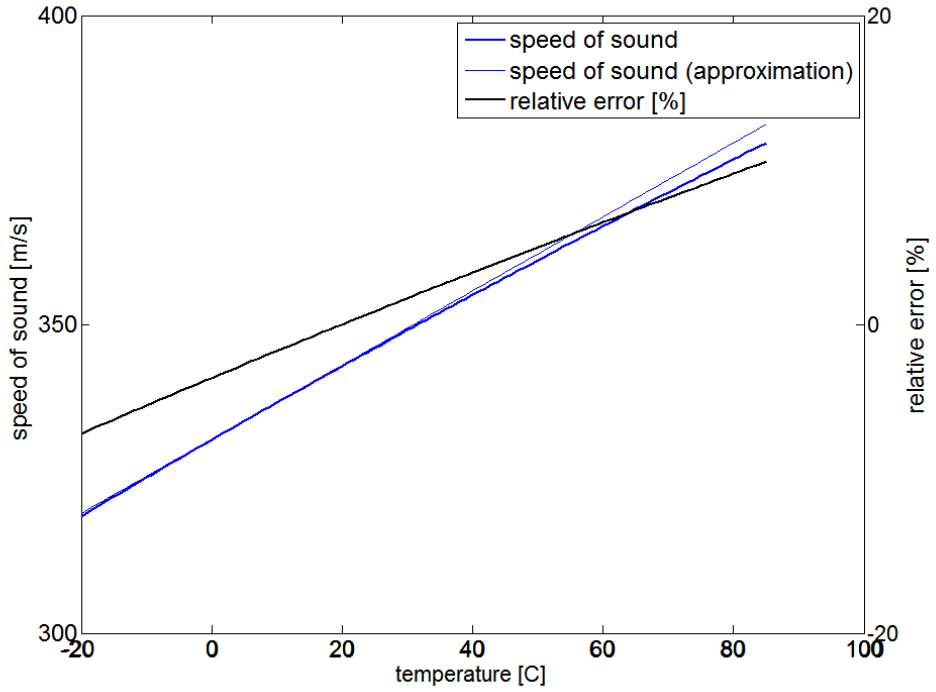


Fig 12: Speed of sound over industrial temperature range from -20° to +85° degree Celsius. Without temperature compensation the error can be as large as +10% to -7%. The linear approximation provides a sufficient approximation with an error less than 0.8%.

3.2 Sound attenuation

Sound propagation in a lossless media is not subject to attenuation. Considering attenuation is important for ultrasound in air at higher frequencies. Attenuation is due to viscosity (internal friction), thermal conduction effects, and molecular relaxation. If the attenuation coefficient α (Nepers per meter) and the total distance r are given, the total attenuation A can be calculated from (6).

$$A = -20 \log(e^{-\alpha r}) \tag{6}$$

Attenuation due to dynamic viscosity η [Pa·s] is sufficiently approximated for air by (7) [Ler09]. Using Sutherland's formula the viscosity η can be calculated for different temperature values. At 20°C η equals 18.4µPa·s. The density of air can be calculated from the molar mass, the universal gas constant and the partial pressure. At 20°C and dry air ρ_0 equals 1.2kg/m^3. Using (7) typical values for different frequencies have been calculated in Tab. 2.

$$\alpha = \frac{8\pi^2}{3} \frac{\eta}{\rho_0 c} \frac{1}{\lambda^2} \tag{7}$$

An acoustic wave not only produces local pressure and density changes but also temperature changes. Due to the limited thermal conductivity of the media, exchange processes between different locations irreversibly weaken the sound field [Ler09]. Attenuation due to thermal conductivity can be approximately calculated by (8). For dry air at 20°C the thermal conductivity ν_t is approx. 0.0257 W/mK and Cv, the heat capacity at constant volume, is approx. 0.718 kJ/(kg·K).

$$\alpha = \frac{\gamma - 1}{\gamma} \frac{\omega^2 \nu_t}{2 C_v \rho_0 c^3} \tag{8}$$

In case of humid air relaxation processes are a significant contributor to the total attenuation. Attenuation can be calculated from the relaxation time, i.e. the time required for the system to return to thermal equilibrium, the partial water vapor pressure and the frequency. Relaxation not only results in a frequency dependent attenuation but also in a frequency dependent speed of sound resulting in dispersion of the acoustic wave. The relaxation time τ_r can be calculated according to (9) where RH is the relative humidity in %, p_{Sat} is the partial pressure of water vapor and p_0 is the total static pressure.

$$\frac{1}{\tau_r} = 1.92 \cdot 10^5 h^{1.3}, \qquad h = RH \cdot \frac{p_{Sat}}{p_0} \tag{9}$$

With the relaxation time the attenuation coefficient due to relaxation can be calculated according to (10) [Ler09] where ϑ is the temperature in degree Celsius and f is the frequency.

$$\alpha_r = \frac{\pi 10^{-5} (0.81 + 0.022 \cdot \vartheta) \tau_r f^2}{1 + 4\pi^2 f^2} \tag{10}$$

Using the individual attenuation coefficients the total attenuation coefficient is calculated from the sum according to (11).

$$\alpha = \alpha_r + \alpha_t + \alpha_v \tag{11}$$

Frequency	Attenuation due to viscosity	Attenuation due to thermal conductivity	Attenuation due to relaxation
1kHz	0.086 [mdB/m]	0.036 [mdB/m]	Not significant for dry air
10kHz	8.63 [mdB/m]	3.601 [mdB/m]	Not significant for dry air
100kHz	863.89 [mdB/m]	360.13 [mdB/m]	Not significant for dry air

Tab. 2: Attenuation due to viscosity and thermal conductivity for dry air at 20°C.

Calculation of sound attenuation is standardized within ISO 9613-1. Using the formulas of the ISO 9613-1 attenuation in dB/m was calculated for different frequencies in dry and humid air, at 20% and 50% relative humidity, and for a static pressure of 101.325kPa at 20°C. The results are shown in Fig 13. At high frequencies attenuation easily exceeds more than 2dB/m, significantly reducing the signal to noise ratio at larger distances.

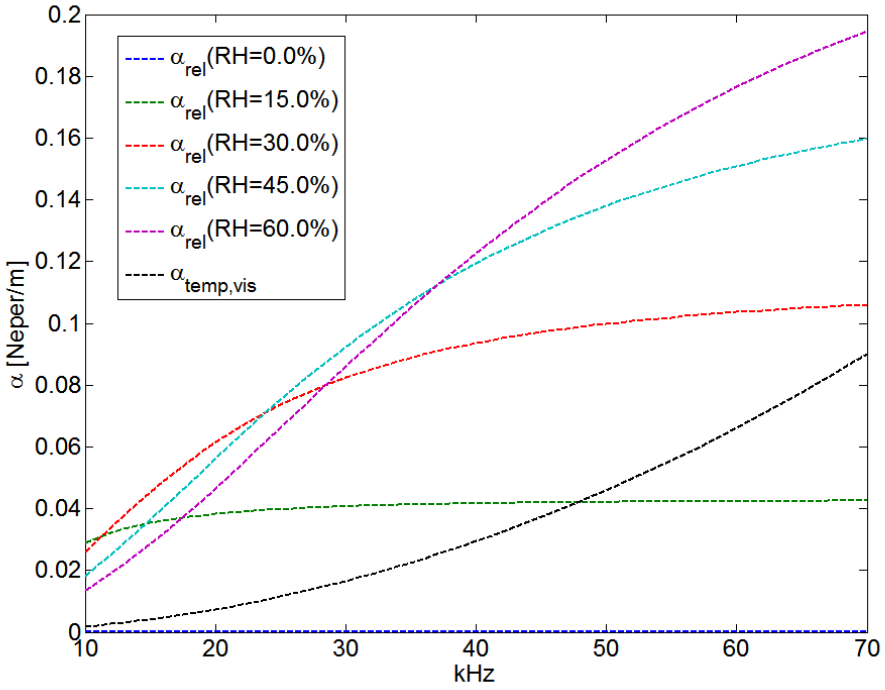


Fig 13: Individual contributors to the attenuation coefficient α. The coefficient $\alpha_{temp,vis}$ is the total attenuation coefficient due to temperature and viscosity. This coefficient is not dependent on the relative humidity but strongly dependent on the frequency. The attenuation coefficient due to relaxation depends on frequency and relative humidity.

Practical use of this data was already taken by [Kle94] where the frequency dependent attenuation and the minimum phase property of physical systems was used to improve spatial prediction of sound waves. Using the attenuation data and a discrete Hilbert transform the minimum phase system was derived, which in turn was used to predict the response at different distances and temperatures. The obtained results are remarkable and shown in Fig 14. It can be seen that the pulses measured at 1m and 5m are quite different. Not only is the amplitude reduced by a factor of approx. 20dB but also the shape of the signal has changed. The red traces are the predicted waveforms using the model described in [Kle94]. The method works remarkably well.

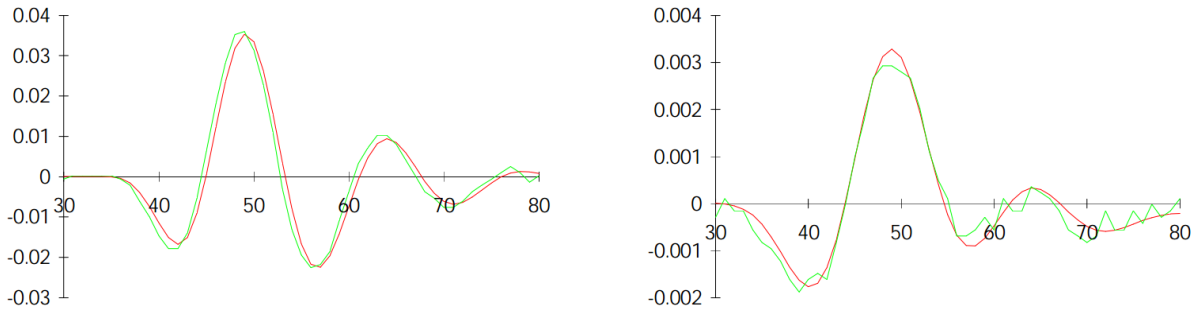


Fig 14: Measured pulse at a distance of 2m (left) and 5m (right). The x-axis is in microseconds and the y-axis in volts. Apart from the attenuation by a factor of 20dB, due to attenuation and a diverging wave, the shape of the waveform is not preserved. Furthermore it can be seen that the extrapolated data (red) fits reasonably well with the measurements (green). Image taken from [Kle94]

3.3 Geometrical Acoustics

If the propagation of sound waves is restricted by finite space effects of transmission, reflection and diffraction have to be considered as well. The effect of transmission and reflection is best explained by a plane wave propagating in positive x-direction incident on a perpendicular wall. The reflected part of the wave is called p_r and together with the complex amplitude of the incident wave p_i defines the reflection factor Γ [Mey79].

$$\Gamma = \frac{p_r}{p_i} \tag{12}$$

Depending on the acoustic impedance Z_m of the media and the acoustic impedance Z_s of the boundary, values of Γ range between -1 and +1. If both impedances are known Γ can be calculated according to (13). For ultrasound in air, with an acoustic impedance of approx. 414 Ns/m^3 most technical surfaces present sound hard boundaries. For example a wooden panel has an acoustic impedance of approx. 1MRayl = 1000000 N·s/m^3 resulting in a value of Γ close to 1. Plastic, steel and other solids have even higher acoustic impedances.

$$\Gamma = \frac{Z_s - Z_m}{Z_s + Z_m} \tag{13}$$

Total reflection on a surface results in a typical pressure increase of +6dB as the incident and reflected wave superpose. This effect is well known in the scope of reference microphones where a free-field correction factor compensates the pressure increase at the diaphragm. Alternatively pressure microphones can be oriented perpendicular to the direction of propagation. The effect is also seen in one of the diffraction simulations shown in Fig 17(b) where the pressure in front of the rod doubles. The part of the wave p_t not reflected by the boundary is called transmitted wave.

If the incident wave is not perpendicular to the wall three different effects are observed: Similar to optics the angle of the reflected wave is identical to the angle of the incident wave, i.e. $\beta_i = \beta_r$. Furthermore, the direction of propagation within the second media follows Snell's law, i.e. $\sin(\beta_i)/c_1 = \sin(\beta_t)/c_2$. Apart from that the coincidence effects plays a major role in sound insulation. If the incident wave is no longer perpendicular to the wall different parts of the surface are no longer excited in phase enforcing a bending wave. If the projected wavelength in air equals the wavelength of the bending wave coincidence occurs and the wave will be transmitted through the media with nearly no attenuation [Mey79].

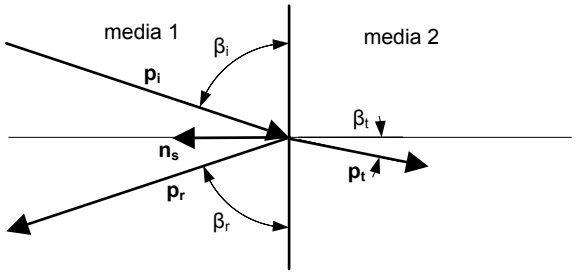


Fig 15: The angle β_r of the reflected wave equals the angle β_i of the incident wave. The angle β_t of the transmitted wave can be calculated according to Snell's law.

The law of reflection can also be formulated in 3D. Let $\mathbf{p_i}$ be the direction vector of the incident wave and $\mathbf{p_r}$ be the direction of the reflected wave. Furthermore let $\mathbf{n_s}$ be the surface normal. Then the law of reflection in vector form can be written according to:

$$\mathbf{p_r} = \mathbf{p_i} - 2(\mathbf{p_i} \cdot \mathbf{n_s})\mathbf{n_s} \tag{14}$$

Due to the large wavelength of acoustic waves in air typical objects extents can be of similar spatial dimension. Therefore, not only reflection but also diffraction of acoustic waves is of concern or better a combination of both physical principles. Analytical results for the amplitude of the diffracted wave are quite complicated. Therefore, only two important practical examples are presented herein. The first example shows scattering of an acoustic wave across a 2D cylinder. An important result is that the amplitude of the diffracted wave depends on the wavenumber k and the radius a of the cylinder, i.e. it is frequency dependent. The basic setup for the simulation is shown in Fig 16. An incident plane wave with wavenumber $k = 2\pi/\lambda$ hits a cylindrical obstacle with diameter a. The solution for the acoustic potential in spherical coordinates is given in (16) [Aky01]. Despite the simplicity of the problem statement, the resulting solution is already quite complex. Still the individual terms can be explained quite simple. The incident plane wave in polar coordinates (left part of (15)) can be expanded in a Fourier-Bessel series and written as an infinite sum. Together with the complex amplitude it forms the first part of (16), the incident plane wave. Setting $\theta_0 = 0$ it only depends on its complex amplitude A, the distance r, and the angle θ. The second part of equation (16) accounts for the diffracted wave. It depends on the distance r, the polar angle θ, and the radius a of the cylinder. J_n and H_n are Bessel and Hankel functions of the first kind where J_n' and H_n' are their respective first-order derivates. The function ε(n) equals 1 for n = 1 and 2 otherwise.

$$e^{j\cdot k\cdot r\cdot\cos(\theta-\theta_0)} = \sum_{n=0}^{\infty} \varepsilon(n) j^n J_n(k\cdot r)\cos\big(n(\theta-\theta_0)\big) \tag{15}$$

$$\Phi(r,\theta) = Ae^{j\omega t}\sum_{n=0}^{\infty} \varepsilon(n) j^n \left(J_n(kr) - \frac{J_n'(ka)}{H_n'(ka)} H_n(kr)\right)\cos(n\theta) \tag{16}$$

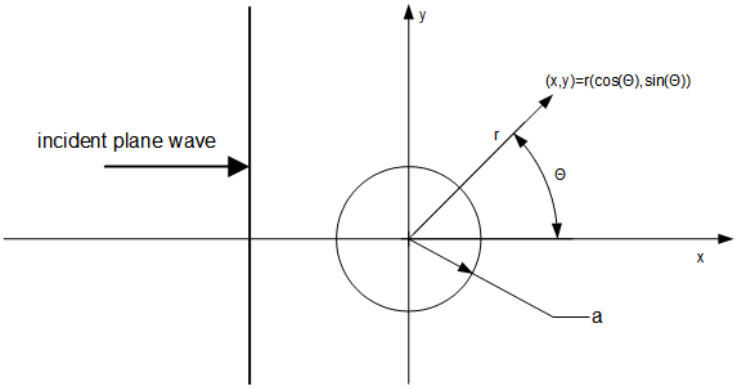


Fig 16: The acoustic potential is calculated for an incident plane wave at θ_0=0° which is scattered on a cylinder with radius a. The solution is given in polar coordinates.

Three basic cases can be identified: First assume the factor k·a is very large, i.e. the cylinder has a much larger radius than the acoustic wavelength. In this case the wave is mostly reflected at the front side surface of the cylinder. The wave field in the shadow region is very small. This is shown in Fig 17 (a) which was obtained from a MATLAB simulation. If k·a is close to 1 then the reflected wave and the wave in the shadow region are both strong. This is shown in Fig 17 (b). For obstacles small compared to the wavelength the acoustic wave is not disturbed at all as shown in Fig 17 (c). Two important observations are made: First of all the amplitude of the reflected wave depends on the factor k·a, i.e. it is frequency dependent. Secondly, as the wave does not transmit through the object, the time required for an acoustic wave to reach a point within the shadow region is larger than the direct path without obstacle. The time required can be calculated by using a string which is pulled taut and dividing its total length by the speed of

sound. This effect and the impact on scene analysis is verified in a practical measurement in 6.2.4.1.

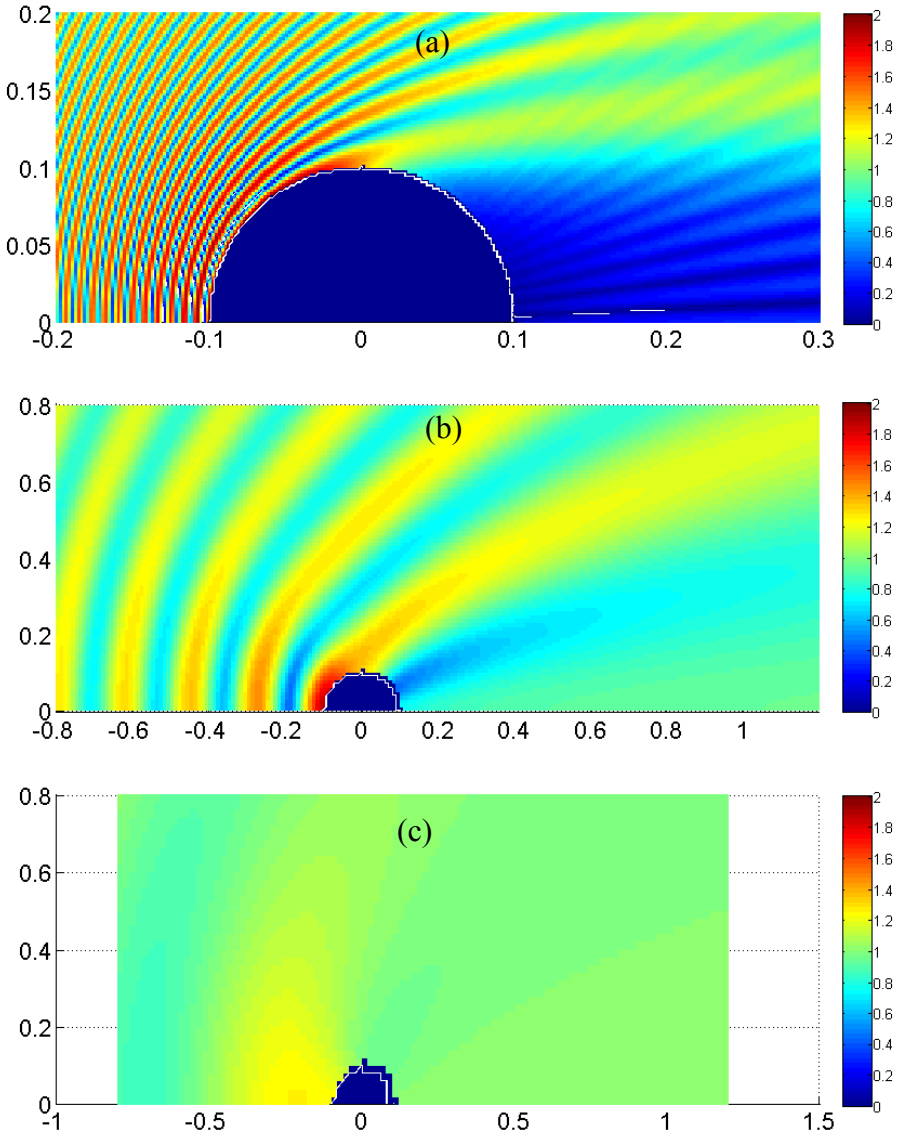


Fig 17: (a) Scattering on a cylinder with diameter = 10cm and wavelength $\lambda \approx 1.7$cm ($k \cdot a \approx 36.6$). The shadow field is very weak and the reflected wave is strong. (b) Scattering for wavelength $\lambda \approx 34$cm ($k \cdot a \approx 1.8$). The shadow field and the reflected wave are strong. (c) Scattering for wavelength $\lambda \approx 1.7$m ($k \cdot a \approx 0.34$). The object is nearly invisible.

To better illustrate the frequency dependence of the sound pressure in the shadow region of the cylinder a second simulation was carried out for a cylinder with a diameter of 5cm. The acoustic potential was evaluated for a constant sound pressure level at different positions along the x-axis. The result is shown in Fig 18. It can be seen that the higher the frequency and the closer the observation point is towards the backside of the cylinder the weaker is the observed signal.

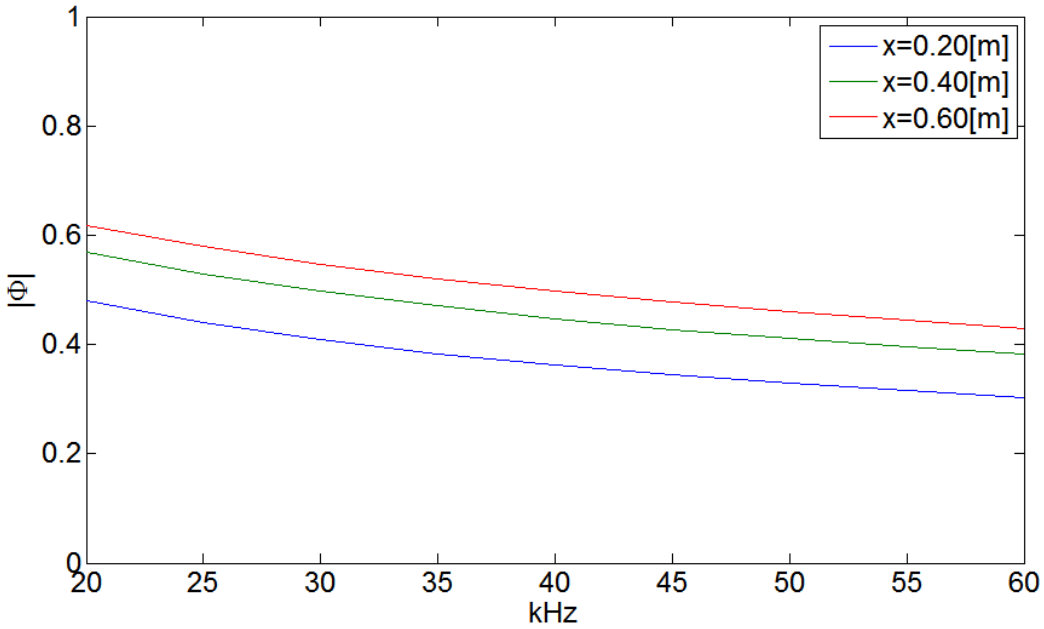


Fig 18: Amplitude of the acoustic potential for different frequencies and an observer at y=0 and different x positions. At close distances, i.e. in the shadow field of the obstacle, the amplitude is decreased.

Another very common situation encountered in practice is diffraction on a wedge. The model is shown in Fig 19. An incident plane wave with angle α strikes a solid wedge. The solid wedge is specified by the parameter ν where for ν = 2 the wedge degenerates to an infinitely small barrier. For ν = 1 the wedge corresponds to a solid wall. Three different zones can be identified. Within Zone 1 the reflected- and incident wave exists. Within Zone 2 only the incident wave is present. Zone 3 solely corresponds to the diffracted wave. This problem can be solved analytically and the solution for the acoustic potential is given in (17)

$$\Phi(r,\theta) = Ae^{j\omega t}\frac{2}{\nu}\left[J_0(kr) + 2\sum_{n=1}^{\infty} e^{-i\frac{n\pi}{2}} J_{\frac{n}{\nu}}(kr)\cos\frac{n\alpha}{\nu}\cos\frac{n\theta}{\nu}\right] \tag{17}$$

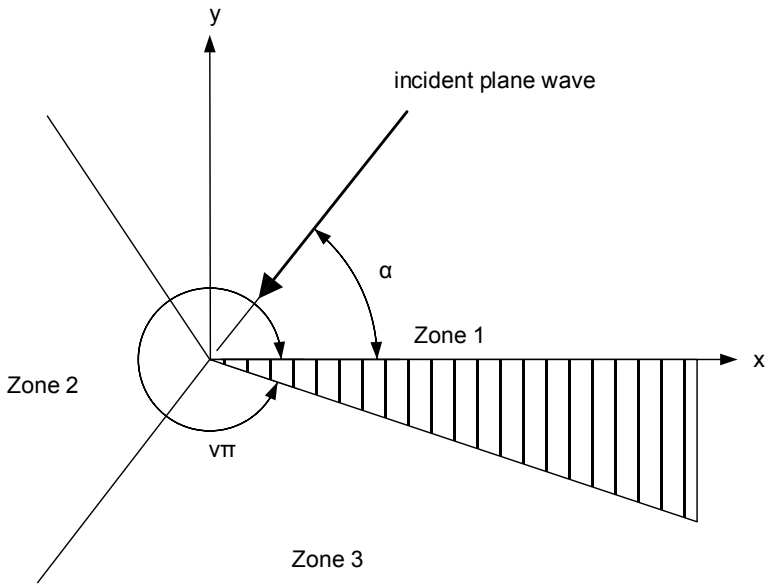


Fig 19: Simulation model for diffraction on a solid wedge. For an incident plane wave with angle α three different zones can be identified. Within Zone 1 there exist the reflected and incident waves. Zone 2 is due to the incident wave and diffracted waves are present in Zone 3.

An example is given in Fig 20 for an incident plane wave with $\alpha = 45°$ and a frequency of 20kHz. The decrease of the amplitude within Zone 3 is obvious. Despite the strong decrease of amplitude in the shadow region, as will be shown in a later chapter, echoes from reflectors within the shadow region can still be detected. This is due to correlative signal processing which greatly enhances the signal to noise ratio.

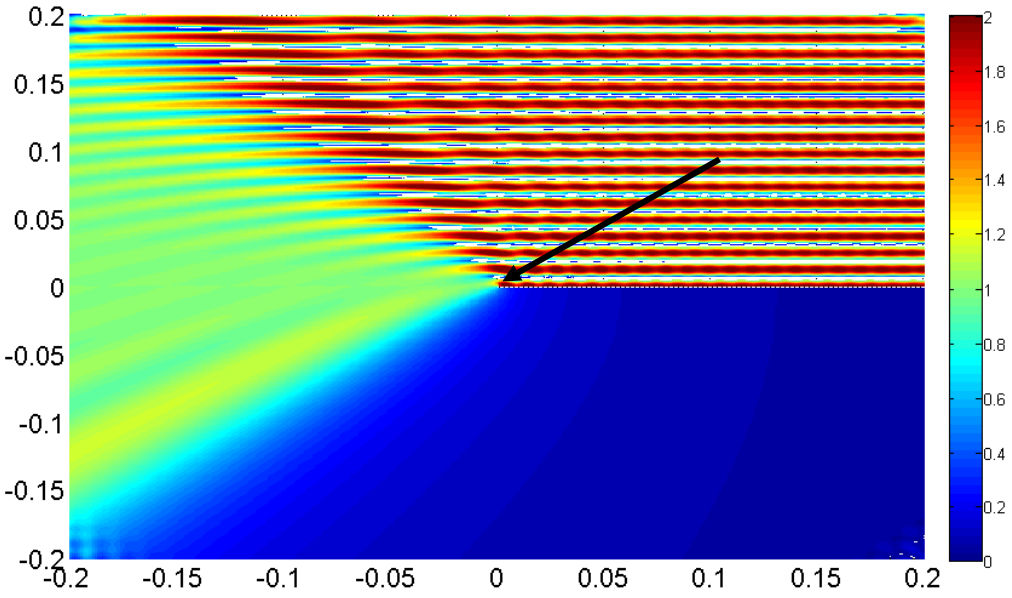


Fig 20: Simulation using a degenerated wedge ($\nu = 2$) and a frequency of 20kHz. The sound pressure decreases quickly in the shadow region.

3.4 Speakers

3.4.1 Piston Membrane Model

Modelling the acoustic transducer is of general interest as it allows the prediction of the sound field using either measurements or synthetic signals. Depending on the complexity of the modelling different levels of details are possible. The most complex modelling relates the electrical driving signal of the transducer to the acoustic potential for an arbitrary field point in space. From the acoustic potential sound velocity and sound pressure can be calculated. This section covers the first part of the modelling where a spatial filter for a rigid and flexible piston is derived. Combining this with the transducer identification presented in 3.4.3.2 the transducer is fully described.

The rigid piston model is a suitable approximation for a wide range of acoustic transducers. A solid circular harmonically vibrating disc with radius a is mounted in an infinite baffle. The radiated sound pressure $p(r,\theta,\omega)$ for a point $\mathbf{r}$ in the half-space in front of the piston shall be calculated. Due to the symmetric construction of a piston the sound-field by itself is rotational symmetric. It can therefore be treated as a 2D problem with polar coordinates $\mathbf{r} = (r, \theta)$. The problem is illustrated in Fig 21.

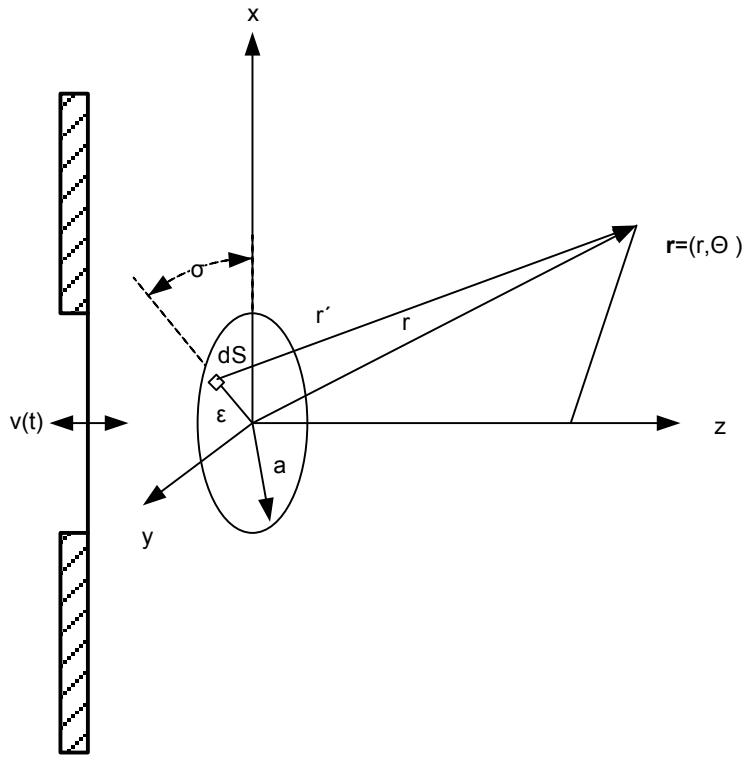


Fig 21: Rigid piston with diameter a mounted in an infinite plate. The circular disc with radius a is harmonically vibrating with surface velocity v(t). The sound pressure p(**r**,ω,t) is calculated.

The sound pressure can be calculated according to (18) [Zol93] by application of Huygens principle and calculation of the surface integral using small point sources as vibrating elements. Within (18) Z_0 is the acoustic impedance of the media ($Z_0 = 414\text{Ns/m}^3$ for air), k is the wave number and V(ω) is the amplitude of the harmonic vibration. To avoid confusion it shall be noted that p(r,θ,ω,t) is a function in the time domain.

$$\begin{aligned} p(r,\theta,\omega,t) &= \frac{jZ_0 \cdot k \cdot V(\omega)}{2\pi} \int_{\varepsilon=0}^{a} \int_{\sigma=0}^{2\pi} \frac{e^{j(\omega t - kr')}}{r'} d\sigma\, \varepsilon d\varepsilon \\ r' &= \sqrt{r^2 + \varepsilon^2 - 2r\varepsilon \cdot \sin\theta \cdot \cos\sigma} \end{aligned} \tag{18}$$

Neglecting the amplitude difference and using an approximation for the phase, which is valid for $r > \varepsilon$, $r > \varepsilon^2/\lambda$, the final result is given in (19) [Zol93] where J_1 is the Bessel function of the first kind.

$$\begin{aligned} p(r,\theta,\omega,t) &= H(r,\theta,\omega)V(\omega)e^{j\omega t} \\ H(r,\theta,\omega) &= \frac{2J_1(ka \cdot \sin\theta)}{ka \cdot \sin\theta} \frac{jZ_0 k}{2r} a^2 \end{aligned} \tag{19}$$

Using (19) the familiar polar patterns for a rigid piston can be calculated by plotting the amplitude of the complex exponential. They are shown in Fig 22 for typical ultrasonic frequencies used in air. It shall be noted that the phase between adjacent side lobes in the polar pattern changes by 180°. Furthermore, it can be seen that the amplitude is greatly reduced outside the mainlobe (20dB or more). Therefore, if such weak echoes are to be used, adequate signal processing techniques need to be employed, due to the degraded signal to noise ratio.

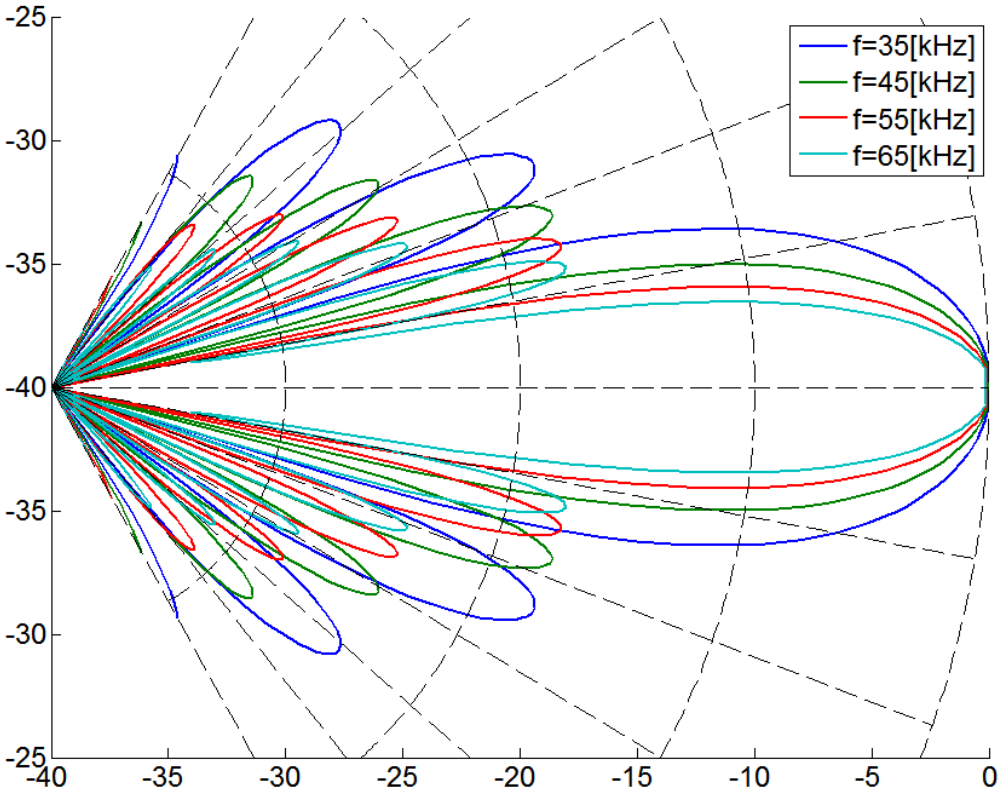


Fig 22: Polar pattern for a rigid piston (a = 2cm, c_0 = 343m/s) at typical ultrasonic frequencies in air for θ = 0° to 60°. The dashed horizontal lines corresponds to angles of 10°. The x- and y-axis is given in dB relative to the maximum sound pressure level at 0°.

Taking the unitary Fourier transform of (19) the following result is obtained:

$$\mathcal{F}\{p(r,\theta,\omega,t)\} = \mathcal{F}\{H(r,\theta,\omega)V(\omega)e^{j\omega t}\} = \frac{\mathcal{F}\{H(r,\theta,\omega)V(\omega)\} * \mathcal{F}\{e^{j\omega t}\}}{\sqrt{2}\pi} \tag{20}$$

As the complex exponential is the only function with a time varying argument the other functions are constants for the Fourier transform. Using ν as frequency domain variable to avoid confusion, the result in (21) is obtained.

$$P(r,\theta,\omega,\nu) = \frac{[H(r,\theta,\omega)V(\omega)] * \left(\sqrt{2}\pi\delta(\nu - w)\right)}{\sqrt{2}\pi} \tag{21}$$

The convolution with the Dirac delta function finally gives:

$$P(r,\theta,\nu) = H(r,\theta,\nu)V(\nu) \tag{22}$$

The result of (22) can be interpreted as following. H(r,θ,ν) is a spatial filter in the frequency domain for calculation the sound pressure given the surface velocity of the piston. As a multiplication in the frequency domain corresponds to a convolution within the time domain a spatial filter h(r,θ,t) can also be calculated by the inverse Fourier transform of H(r,θ,ν). Hence (23) can be used for calculation of the sound pressure given an arbitrary surface velocity in the time domain.

$$h(r,\theta,t) = \mathcal{F}^{-1}\{H(r,\theta,\nu)\} = \frac{1}{\sqrt{2}\pi}\int_{-\infty}^{+\infty} H(r,\theta,\nu)e^{j\nu t}d\nu \tag{23}$$

One issue in modelling the transducer is that the surface velocity of the piston is neither known in the frequency nor in the time domain. This issue can be circumvented by performing a measurement at sufficient distance from the speaker at $\theta = 0°$. By applying a limiting argument for $\theta \rightarrow 0°$, i.e. the response along the acoustic axis, $H(\omega)$ reduces to the simple term shown in (24).

$$P(r,0°,\nu) = H(r,0°,\nu)V(\nu) = \frac{jZ_0 k}{2r}a^2 V(\nu) \tag{24}$$

The ratio between the sound pressure at different angles θ can now be calculated according to (25).

$$\frac{P(r,\theta,\nu)}{P(r,0°,\nu)} = \frac{2J_1(ka\cdot\sin\theta)}{ka\cdot\sin\theta} \tag{25}$$

From (25) it follows that

$$P(r,\theta,\nu) = \frac{2J_1(ka\cdot\sin\theta)}{ka\cdot\sin\theta}P(r,0°,\nu) \tag{26}$$

Or respectively in the time domain

$$p(r,\theta,t) = \mathcal{F}^{-1}\left\{\frac{2J_1(ka\cdot\sin\theta)}{ka\cdot\sin\theta}\right\} * p(r,0°,t) \tag{27}$$

Fig 23 shows the spatial impulse response for different angles θ. Two important corner cases are shown – For $\theta = 90°$ the length of the impulse response is 116µs which corresponds to the diameter of the piston of 4cm. At $\theta = 0°$ the impulse response is a discrete Dirac impulse. Using the spatial impulse response and a measured or generated signal at $\theta = 0°$ the signal at a different polar angle is predicted. An example is given in Fig 24 where an artificially generated linear frequency modulated chirp with B = 30kHz, f_c = 50kHz and T = 500µs is transformed to a different polar angle. It shall be noted that this does not yet include any modelling of the transducer transfer characteristic which relates the electrical driving signal to the surface velocity of the piston. The applied method covering the complete signal chain is continued in 3.4.2.

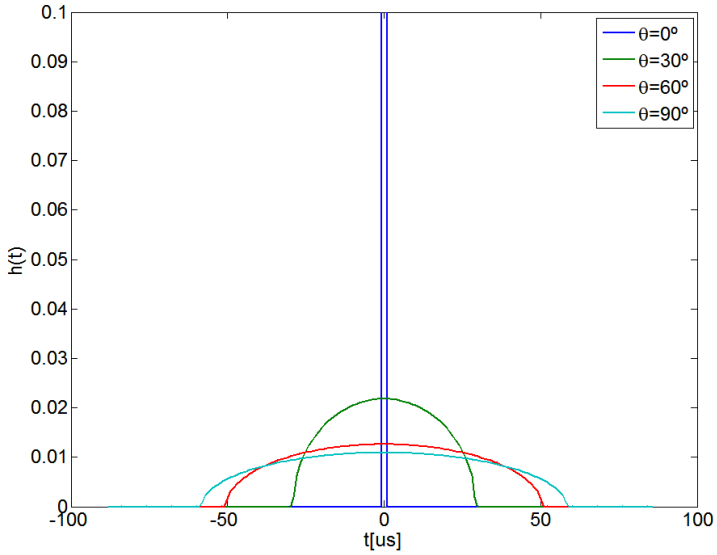


Fig 23: Spatial impulse response in the time domain for different angles θ. At θ = 0° the impulse response corresponds to a discrete Dirac. At θ = 90° the impulse response is maximally broad with ≈ 116µs corresponding to the piston diameter of 4cm used within this simulation.

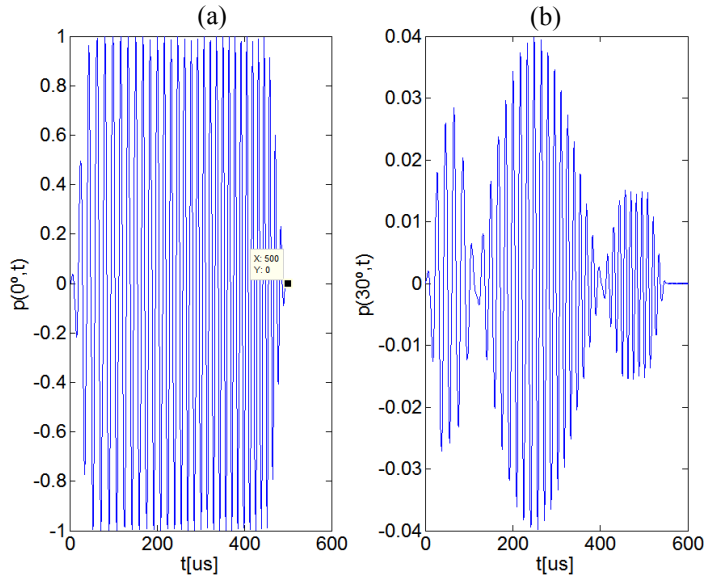


Fig 24: (b) Signal predicted at θ = 30° (right) using the spatial filter of (27) and the mainlobe signal (a) at θ = 0°

The simplified membrane model can be improved by dropping the requirement of a uniform surface velocity profile. This is far more realistic, as in case of the electrostatic speaker used within this work, the membrane is fixed on the outside, hence v(a) = 0. [Aar11] provides a semi-analytic method for the computation of the spatial impulse response of a baffled piston membrane with a radially symmetric velocity profile v(σ), σ < a. Fig 25 shows the system model where $\mathbf{r_S}$ is a vector on the surface of the piston and **r** is a field-point in the half-space right of the piston, for which the velocity potential shall be calculated. The distance r' is calculated as $r' = |\mathbf{r} - \mathbf{r_S}|$, ω is the angular frequency, k is the wave number, c is the speed of sound and ρ_0 is the density of the media. The sound pressure at field point **r** is calculated according to (28) using the Rayleigh integral. With the exception of v(|**rs**|) this form is equivalent to (18).

$$p(\mathbf{r},\omega,t) = \frac{j\rho_0 ck}{2\pi}\int_S v(|\mathbf{r_S}|)\frac{e^{-jkr'}}{r'}dS\cdot e^{j\omega t} \tag{28}$$

Using the inverse of (4), i.e. $p = \rho\cdot\partial\Phi/\partial t$, and noting that the complex exponential is the only time varying argument, the velocity potential is calculated as

$$\Phi(\mathbf{r},\omega,t) = \left(\frac{1}{2\pi}\int_S v(|\mathbf{r_S}|)\frac{e^{-jkr'}}{r'}dS\right)\cdot e^{j\omega t} \tag{29}$$

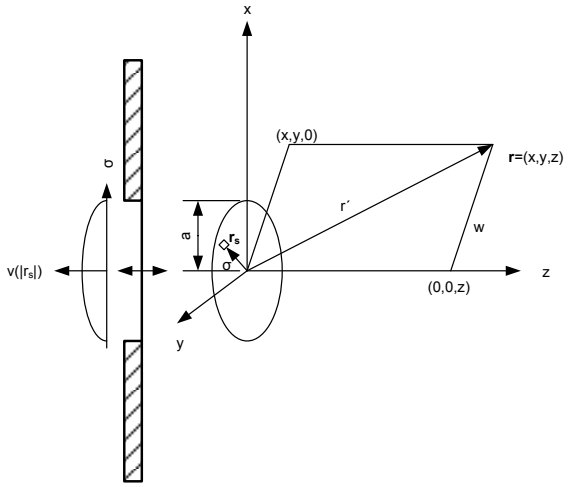


Fig 25: Calculating the velocity potential at the field-point **r** for a baffled piston with surface velocity v(|**rs**|).

Similar to (22) we can interpret the part within the brackets of (29) as a spatial filter in the frequency domain. Again we can apply the inverse Fourier transform and obtain a spatial filter in the time domain which can be convolved with an arbitrary surface velocity. Different to the previous result the surface velocity in the time domain is weighted using the velocity distribution profile on the membrane.

$$\Phi(\boldsymbol{r},t) = (h*v)(t) = \mathcal{F}^{-1}\left\{\frac{1}{2\pi}\int_S v(|\boldsymbol{r_S}|)\frac{e^{-jkr'}}{r'}dS\right\} * v(t) \tag{30}$$

The inverse Fourier transform is given by (31) [Aar11] where Vs equals the average surface velocity, H is the Heaviside step function and w, R and A are auxiliary variables according to (32).

$$h(\mathbf{r},t) = \left(\frac{c}{\pi V_s}H(ct-z)\int_0^A v\left(\sqrt{w^2+R^2(t,z)-2wR(t,z)\cos\alpha}\right)d\alpha\right) \tag{31}$$

$$\begin{aligned}
&H(x) = 0, x<0 \quad H(0) = \frac{1}{2} \quad H(x) = 1, x>0 \\
&w = \sqrt{x^2+y^2} \\
&R(t,z) = \sqrt{c^2t^2-z^2}, ct>z \\
&A = \begin{cases} 0 & |w-R(t,z)|>a \\ \cos^{-1}\left(\dfrac{w^2+R^2(t,z)-a^2}{2wR(t,z)}\right) & |w-R(t,z)|<a<w+R(t,z) \\ \pi & w+R(t,z)<a \end{cases} \\
&V_s = \frac{1}{\pi a^2}\int_S v(|\mathbf{r_S}|)dS
\end{aligned} \tag{32}$$

The main result of [Aar11] is that an analytical result is available if the velocity profile v(σ) is expanded using Zernike polynomials. One benefit of the Zernike polynomials, instead of other basis functions, is that for polynomial velocity profiles the expansion series is finite. For other velocity profiles, e.g. Gaussian, the higher-order terms of the series vanish quickly and can be neglected.

Similar to the rigid piston the time domain response on the acoustic axis (θ = 0°) is very close to a Dirac delta function (for all practical signals used by ultrasound in air and the available bandwidth of the transducers) with the exception of a distance dependent attenuation to account for the spherical spreading. Hence equations (30) and (31) can be interpreted as (33) where K is a frequency dependent constant.

$$\Phi([x_0, y_0, 0], t) = h([x_0, y_0, 0]) * v(t) = K(w_0)\delta(t) * v(t) = K(w_0)v(t) \tag{33}$$

The acoustic potential at an arbitrary position can be calculated as

$$\Phi(\mathbf{r}, t) = h(\mathbf{r}) * v(t) = h(\mathbf{x}) * \frac{\Phi([x_0, y_0, 0], t)}{K(w_0)} \tag{34}$$

Using (4) the sound pressure can be calculated from (34) as

$$p(\mathbf{r}, t) = \rho_0 \frac{\partial \Phi(\mathbf{r}, t)}{\partial t} = \rho_0 \frac{\partial \left(h(\mathbf{r}) * \frac{\Phi([x_0, y_0, 0], t)}{K(w_0)} \right)}{\partial t} \tag{35}$$

Using linearity and the fact that h(**r**) is not time dependent we get the final result

$$p(\mathbf{r}, t) = \frac{1}{K(w_0)} \left(h(\mathbf{r}) * \rho_0 \frac{\partial \Phi([x, y, 0], t)}{\partial t} \right) = \frac{1}{K(w_0)} \left(h(\mathbf{r}) * p([x, y, 0], t) \right) \tag{36}$$

Therefore the spatial filter h(**r**) again relates the sound pressure on the main axis to the sound pressure at an arbitrary position.

It is interesting to compare the two different models used in the context of the electrostatic transducer used within this work. This transducer, by construction, has the membrane clamped at the boundary, i.e. v(a) = 0. An example for the relative velocity profile is given in Fig 26 for a parabolic velocity profile. Intuitively we can expect that the transducer with the clamped piston has a broader polar diagram, as it is similar to a rigid piston with smaller diameter. This can be seen in Fig 27 where the respective polar patterns for 45 kHz are shown for the rigid piston and the piston with the parabolic velocity profile.

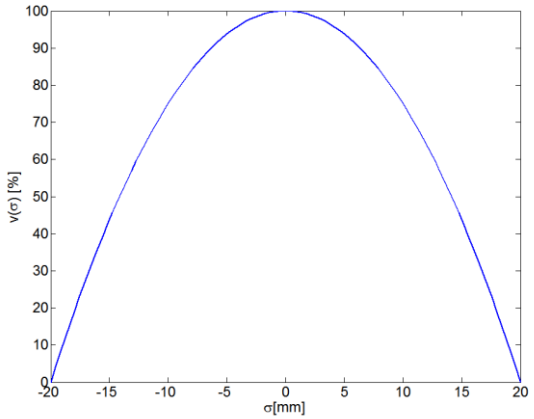


Fig 26: Parabolic velocity profile for a piston membrane with diameter = 4cm using a 2nd order polynomial.

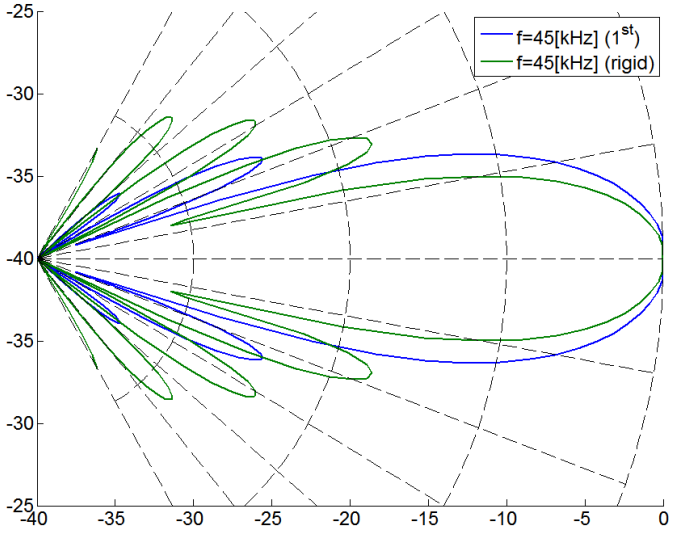
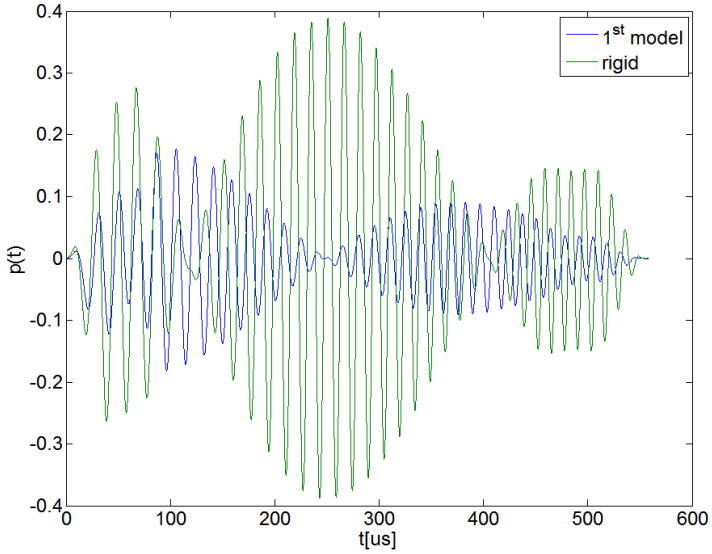


Fig 27: Comparison of radiation pattern for the solid piston (green) and a piston with a parabolic velocity profile.

`Fig 28 compares the different models using the linear chirp already used in Fig 24. The difference in the time domain is even more obvious than in the frequency domain. As the phase between adjacent side-lobes changes by approximately 180° in both models it can be seen that sometimes the time domain signals are in phase and sometimes have opposite polarity. Using the binary correlation, which will be introduced later, and which is only sensitive to the phase of the signal, good correlation results require adequate prediction of the sound field produced by the transducer.



`Fig 28: Comparison of time domain signals for θ = 30° and the linear frequency modulated chirp.

The method for prediction of spatial impulse responses within this work uses (36).

3.4.2 Senscomp 600 Electrostatic Transducer

Electrostatic transducers are well suited for the generation of airborne ultrasound due to their good acoustic coupling on air, the high bandwidth and their low-cost. One of the most common electrostatic transducers is the Senscomp 600 shown in Fig 29(a). It consists of a flexible membrane of approx. 40mm in diameter and a backplate forming the two electrodes of a capacitor. The inner construction is shown in Fig 29(b). Using a V-grooved backplate allows better control of the resonance frequency of the transducers [Mat95]. If used as a receiver, incident sound waves exert a pressure on the surface of the membrane, which results in a small change of capacitance. By applying a DC polarization voltage by a large bias resistor the charge on the capacitor is kept constant, hence the change in capacitance results in a small alternating voltage, which in turn can be measured with a high-impedance amplifier. This mode of operation is not used within this work. Therefore, no further details are given. Instead the transducer is used to generate sound by applying a DC polarization in the range of 100V-200V and an alternating voltage on top of it.

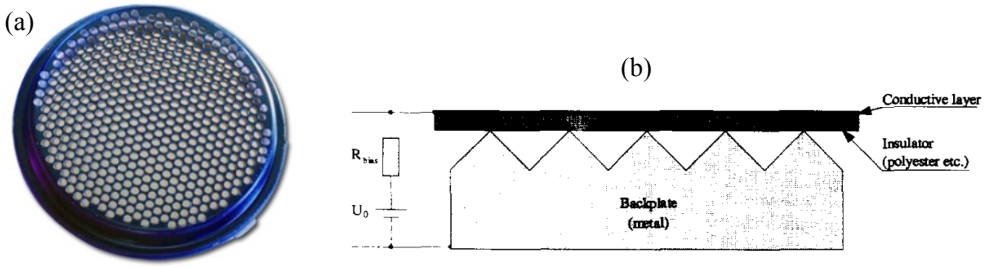


Fig 29: Photo of Senscomp 600 industrial grade transducer with mounted protection grid (a). (b) shows the internal construction of the sensor as well as the DC bias and polarization voltage for the receiving mode.

It is desirable to have a model for the relationship between the electrical quantities at the input of the transducer and the particle speed of sound on the surface of the transducer, as this allows the calculation of the sound field using the results of chapter 3.4.2. Such a model is given in [Mat95] where an electromechanical equivalent circuit as shown in Fig 30 was used. According to [Mat95] C_0 is the static capacitance of the transducer, L_{md} is the mechanical inductance due to the mass of the membrane, C_m is the mechanical capacitance due to the air volume and surface area, R_{mr} is the radiation resistance and R_{md} includes all dissipative elements. The particle velocity can then be calculated directly by $v = F/R_{mr}$ where F is the force in Newton.

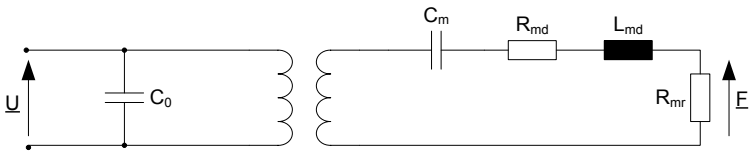


Fig 30: Electromechanical equivalent circuit for an electrostatic transducer with V-grooved backplate.

Instead of modelling the transducer a system identification approach can be applied as well. The basic principle is shown in Fig 31. A known electrical driving signal $U(\omega)$ is applied to an electrostatic transducer, and by measuring the output signal $Y(\omega)$ and successive computation an estimate for the unknown transfer function $T(r,\theta,\omega) = E(\omega) \cdot H(r,\theta,\omega)$ can be obtained. $E(\omega)$ relates the input signal, which exerts a force on the membrane, to the surface velocity of the speaker. $H(r,\theta,\omega)$ is the spatial filter which relates the sound particle velocity at the surface of the speaker to the sound pressure at position r and polar angle θ. S is the sensitivity of the microphone which is assumed to have a flat frequency response.

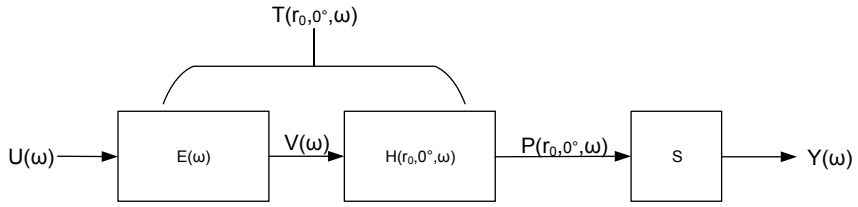


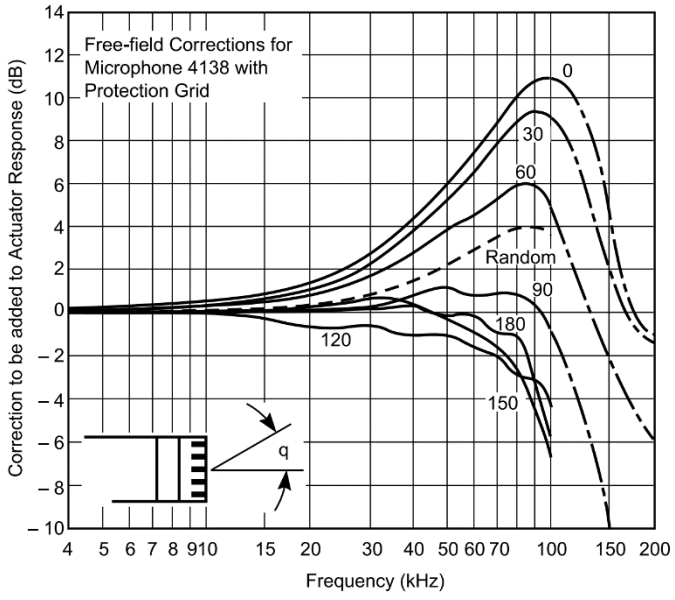
Fig 31: An electrical input signal U(ω) exerts a force on the electrostatic membrane resulting in a particle surface velocity V(ω). The spatial filter H relates the sound particle velocity at the surface to a sound pressure which is measured by a reference microphone with sensitivity S. If the transfer function T=P/U is estimated the sound pressure field for arbitrary angles θ can be predicted using the results of chapter 3.4.1.

If a measurement is performed at the acoustic axis of the transducer, i.e. θ = 0°, the transfer function H(r,θ,ω) according to (24) has no zeros in the frequency domain and is therefore invertible. Furthermore the SNR is maximized and the output signal can easily be related to other polar angles and distances using (26). Using a 1/8" reference microphone (B&K, Type 4138) as shown in Fig 32, having a nearly flat frequency response in the frequency range of interest with sensitivity S, the transfer function of the transducer can be identified using deconvolution according to (37).

$$T(r_0, 0°, \omega) = \frac{Y(\omega)}{S \cdot U(\omega)} \tag{37}$$

Despite the simplicity of this formula system identification imposes a number of problems. First of all the acoustic environment can introduce reflections requiring short pulses or pulse compression techniques for system identification. Secondly the inversion requires proper signals as the division in the frequency domain is highly sensitive to noise. In addition, care has to be taken to not introduce errors by the reference microphone itself. To measure the free-field response the reference microphone shall be oriented at 90° as shown in Fig 32 to avoid any pressure increase at the tip of the microphone.

(a)



(b)

Fig 32: Free-field correction (a) for the B&K 4138 reference microphone with protective cap (b). They microphone shall be mounted in 90° orientation to the sound field to obtain the equivalent sound pressure assuming no microphone is present. The error is ≤ 1dB for frequencies less than 100 kHz. Images taken from [BK08]

Using the MEMS microphone SPM0204UD5 from Knowles within the system identification process is also possible and inherently includes the transfer characteristic of the microphone as

well. This is a sound approach, as the miniature MEMS microphone has omni-directional characteristic, and therefore the transfer function is not dependent on the angle of the incident wave. If for example the electrostatic transducer is used in a pulse-echo combination the spatial filter needs to be applied twice – once for the transmission and once for the reception.

The next section covers more details for the system identification process and also explains the method applied within this work.

3.4.3 Practical system identification

As a first system identification approach an LTI system was assumed. Using averaging of multiple unit step impulses and a discrete derivative an estimator for the transfer function $t[n]$ was obtained. Using the General Standards analog input/output interface card 16AISS16AO2 a step response was created within MATLAB. The step response was chosen such that the DC bias of the electrostatic transducer equals the DC bias used during operation (200V). The amplitude of the unit step was also chosen to be close to the AC amplitudes used during real measurements. The measurement was performed twice – once with the B&K reference microphone and once with the Knowles MEMS microphones. The slight benefit of this approach in using only very short bursts and, therefore, avoiding signal overlapping, are dominated by various other issues. First of all the energy contained in the step response is very low resulting in a weak SNR for the identification process. Secondly, driving a step response into a capacitive transducer with ~500pF (Senscomp 600) requires large currents. Last but not least the discrete derivative is very sensitive to noise and a Savitzky-Golay filter [Sav64] in combination with averaging (N=100) was used. Averaging is especially problematic as it requires strong time invariance and a good mechanical setup. As a general conclusion this method for system identification is not recommended and only included for reference. Other methods often used for system identification are maximum length sequences (MLS) and time domain spectrometry (TDS) [Far00]. Both require perfect linearity and time invariance of the whole setup.

3.4.3.1 System identification considering non-linearities

One of the currently best available techniques is presented by [Far00] and uses an exponential swept chirp. It not only allows the determination of the impulse response, but within the same measurement, also allows separation of the impulse responses for linear, quadratic, cubic, etc. terms. Another benefit is that the measurement does not require any averaging. Thus system identification can be performed in short time, avoiding any potential issues with time instabilities (e.g. by temperature, mechanical displacement, etc.). The system is modeled to have an initial non-linearity and a subsequent linear system [Nov10]. This is a reasonable approach for our electrostatic speaker as non-linearities are due to the driving of the transducer, whereas the actual radiation and sound propagation can be assumed to be a linear process. For an input signal x(t), and using N distinct impulse responses for each harmonic, the output signal y(t) can be calculated according to (38) [Nov10]. The basic concept is also shown in Fig 33.

$$y(t) = (x * g_1)(t) + (x^2 * g_2)(t) + \cdots + (x^N * g_N)(t) \tag{38}$$

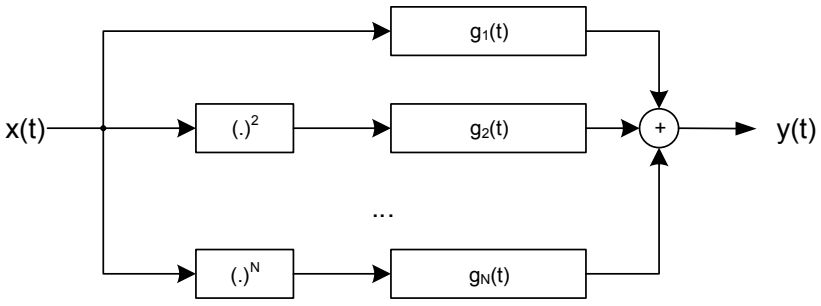


Fig 33: Non-linear system modeled with initial non-linearity and subsequent time invariant linear systems.

By far the most important result of [Far00] is, that the output signal y(t) (40), after convolution with $x^*(t)$, contains temporarily separated impulse responses for different orders of non-linearities if a logarithmic sweep is used as an input signal x(t) (39),. These impulse responses in turn can then be used to calculate the transfer functions g_1 to g_N of Fig 33.

$$x(t) = \sin\left(\omega_1 L\left(e^{\frac{t}{L}} - 1\right)\right), L = \frac{T}{\log\left(\frac{\omega_2}{\omega_1}\right)}, 0 \le t \le T \tag{39}$$

To correct for the high frequency loss in the input signal (39) a time dependent amplitude correction factor is added to the time reversed input signal (See the frequency spectra within Fig 35)

$$x^*(t) = \frac{\omega_1}{2\pi L} e^{-\frac{t}{L}} \cdot x(-t) \tag{40}$$

The temporal spacing of the separated impulse responses can be calculated according to (41) [Far00]. Note that the spacing Δt in (41) depends on the order n of the harmonic, where higher order harmonics are more closely spaced. Furthermore, increasing the length of the identification sequences pushes the different responses further apart.

$$\Delta t(n) = T \frac{\log(n)}{\log(\frac{\omega_2}{w_1})} \tag{41}$$

Fig 34 shows an example for a sending and receiving signal suitable for the identification of the used electrostatic transducer within the frequency range from 30kHz to 100kHz. The total time duration is 1ms. It can be seen in Fig 34(b) that the frequency does not change linearly over time.

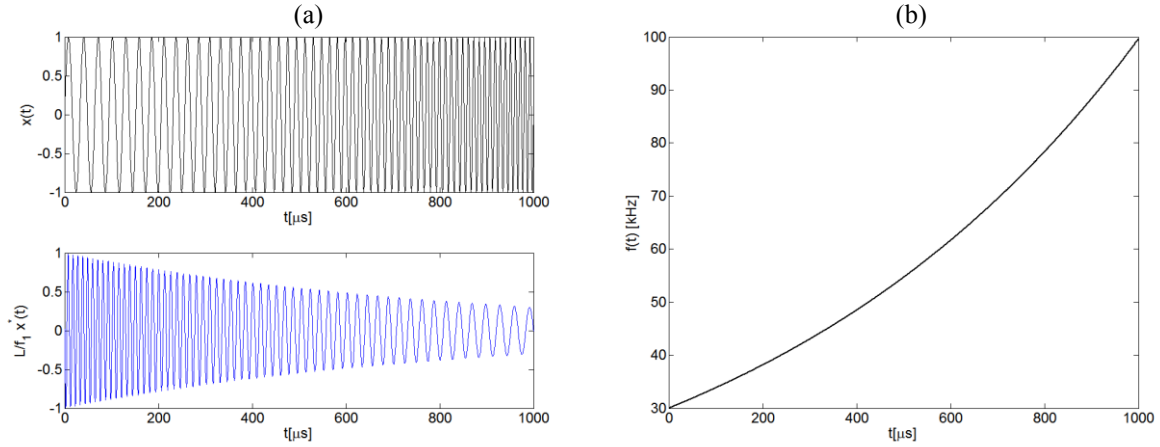


Fig 34: (a) Sending signal x(t) with total length of 1ms. The receiving signal $x^*(t)$ includes a time dependent factor to account for the loss in the high frequency response. (b) Shows the instantaneous frequency of the sending signal x(t).

The sending and receiving frequency spectra are shown in Fig 35. Calculation of the transfer functions $g_1(t)$ to $g_N(t)$ requires that the main peak is identified within the output signal after convolution with $x^*(t)$. Using (41) the time shifted impulses $h_2(t)$ to $h_N(t)$ can be obtained by using the respective part of the output signal as shown in Fig 36. Once these individual responses have been obtained, (42) and (43) [Nov10] can be used to calculate $g_1(t)$ to $g_N(t)$. Note that (42) is a linear problem and a such can be solved easily.

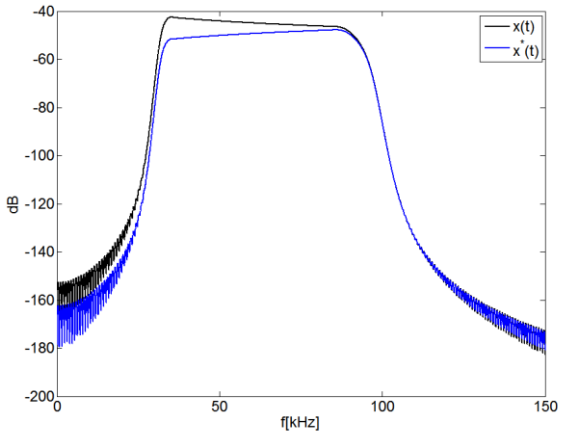


Fig 35: Sending and receiving spectra for the chirp with $f_1 = 30$kHz, $f_2 = 100$kHz and T = 1ms. The product of the individual spectra (convolution in the time domain) has a flat frequency response in the frequency region of interest.

$$h_m(t) = \sum_{n=1}^{N} A_{n,m} g_n(t) \tag{42}$$

$$A_{n,m} = \begin{cases} \dfrac{(-1)^{2n} + (1-m)/2}{2^{n-1}} \binom{n}{(n-m)/2} & n \geq m \text{ and } (n+m) \text{ is even} \\ 0 \end{cases} \tag{43}$$

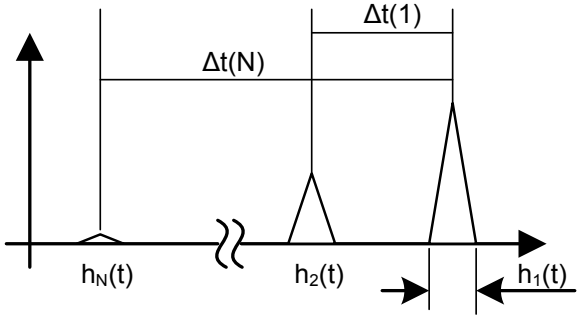


Fig 36: Output of the system during the system identification process if the system is excited with x(t) and the output is convolved with x*(t). The output contains multiple time shifted responses $h_1(t)$ to $h_N(t)$ which can then be used to calculate the impulse responses $g_1(t)$ to $g_N(t)$.

Fig 37 gives an example for a system with an initial non-linearity and a constant time delay. The system described by (44) was identified by a logarithmic swept chirp where the output after convolution is given in Fig 37(a). Using the two time-delayed respones $h_1(t)$ and $h_2(t)$ the impulse responses $g_1(t)$ and $g_2(t)$ for the different harmonics were calculated according to (42).

$$y(t) = \left(U_0 + x(t)\right)^2 \tag{44}$$

Correctness of the identification was verified by calculating the expected THD for a single 50 kHz tone with amplitude A = 0.4 and a DC bias $U_0 = 1$. Manual calculation of the THD by (45) yields results in good agreement with the output of the identified system as shown in Fig 37(b). While this simple example cannot be applied to the transducer it shows the suitability of the identification approach. In fact, the method was applied as is to the electrostatic transducer as shown in 3.4.3.2.

$$x(t) = A \cdot \sin(\omega t) \rightarrow y(t) \text{ showing AC components only}$$
$$y(t) = 2U_0 A \cdot \sin(\omega t) + A^2 \cdot \sin(\omega t)^2 = 2U_0 A \cdot \sin(\omega t) - \frac{A^2}{2}\cos(2\omega t) \quad (45)$$
$$\text{THD} = \frac{A^2/2}{2U_0 A} = \frac{A}{4U_0}$$

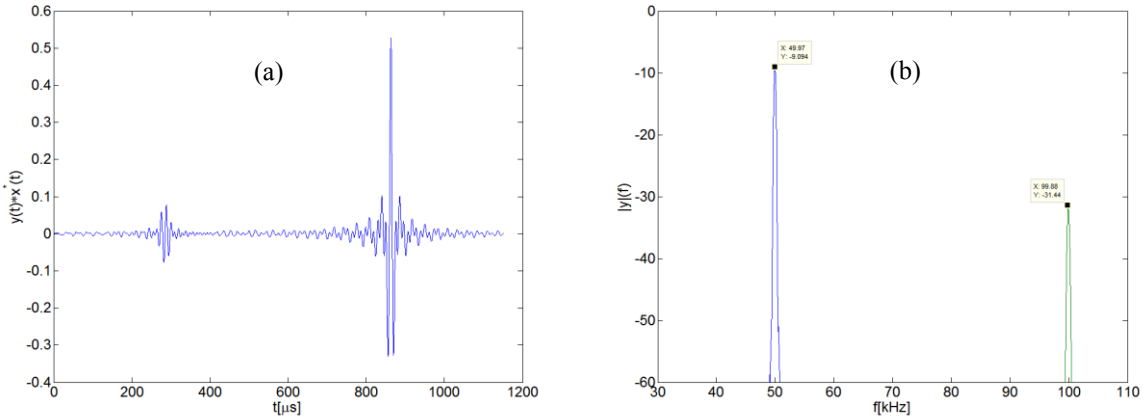


Fig 37: (a) Output after convolution with x*(t) for the non-linear system according to (44). (b) Output of the system for linear extinction with a 50kHz sine wave with A = 0.4 and U_0= 1. The difference in dB between the fundamental and the harmonic is 22.35dB. This is in close agreement with the calculated THD of 20dB.

3.4.3.2 Identification of the testing system

The compact 3D sensor described in this work is built up from the Senscomp 600 electrostatic transducer and multiple MEMS microphones. Given the datasheet of the MEMS microphone and the transducer we wanted to measure the transfer characteristic of the whole system. Based on the very rough data in the datasheet we cannot expect very good correspondence but basic properties should match. Fig 38(a) and Fig 38(b) show the sensitivity of the receivers and transmitter.

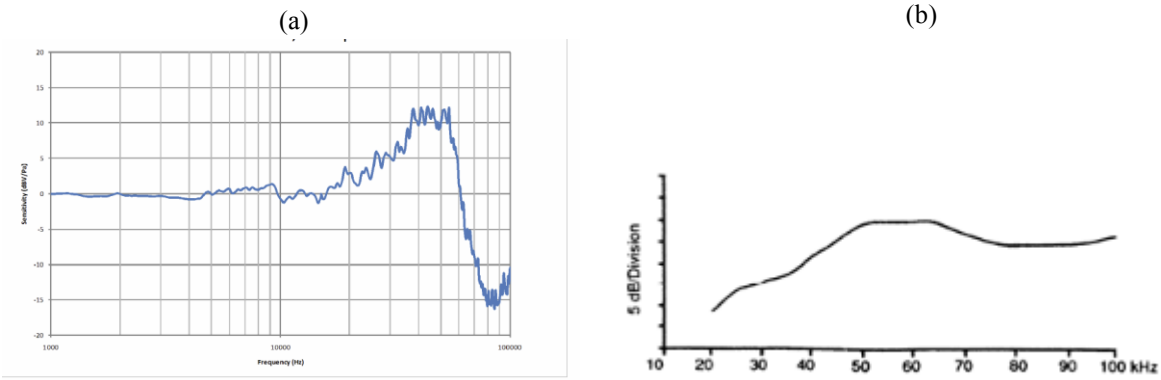


Fig 38: (a) Free field response of MEMS microphone SPM0404UD5, (b) transmitter sensitivity of Senscomp 600. Both curves are taken from the datasheet

The transducer was identified using a log-swept chirp as described in 3.4.3.1. Fig 39(a) shows the output after convolution with the filter defined by (40). The solid vertical line corresponds to the linear part of the response whereas the dashed vertical line corresponds to the quadratic term. Fig 39(b) shows the linear part of the impulse response in more detail.

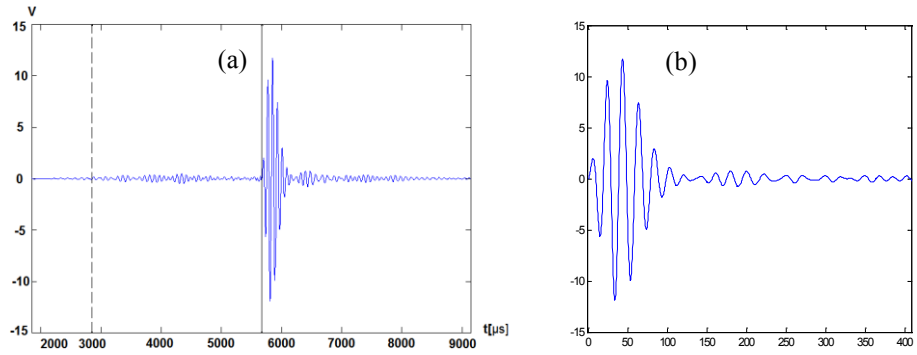


Fig 39: An exponential swept chirp as described in section 3.4.3.1 was used for system identification. The x-axis represents the time in µs. The y-axis is in volts and corresponds to the input signal of the AD converters. (a) Estimated impulse response in the time domain showing the linear and non-linear parts .The start of the non-linear part was calculated using (41) (b) Linear part of the system impulse response only.

Using the impulse response the Fourier transform is calculated and compared to the expected transfer function of the system in the frequency domain. The result is shown in Fig 40 and matches sufficiently well as it must be considered that transmitter and receiver data was taken from the datasheet.

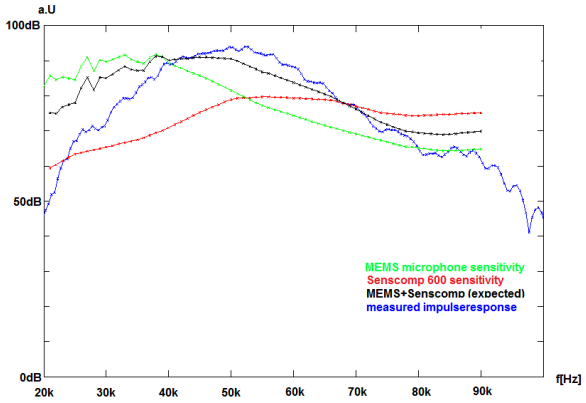


Fig 40: Estimated frequency response using measurement (blue) and calculated response from the datasheet (black). Correlation is not very good due to manual extraction from the datasheet (In fact the picture in Fig 38 had to be used). The x-axis is given in Hz and the y-axis in arbitrary units.

Using this impulse response and knowing the driving signal it is possible to predict the acoustic signal at the mainlobe using (38). By application of (36) the sound pressure can be calculated for different polar angles and distances.

3.4.3.3 Controlling the driving signal of the transducer

As the relationship between the electrical driving signal and the sound pressure at the mainlobe is not only a constant, but also depends on the impulse response of the transducer, it can be beneficial to apply a deconvolution process. Considering only the linear case for now, the output signal of the transducer can be written according to (46) where n(t) is noise.

$$y(t) = g_1(t) * x(t) + n(t) \tag{46}$$

While the most simplistic approach would be to directly calculate x(t) by calculation of the Fourier transform of the desired output signal Y(f) and the known frequency response $G_1(f)$, and successively calculating the inverse Fourier transform of $Y(f)/G_1(f)$, this gives bad results. The reason is that $|G_1(f)|$ can be small for some frequencies strongly amplifying the noise. A much better approach is using Wiener deconvolution as shown in (47). The term S(f) is the

mean power spectral density of the yet unknown input signal X(f). The output F(f) is the resulting filter and can be used for calculating an estimate of the proper driving signal X(f) for a desired output signal Y(f) according to (48). As X(f) is a-priori unknown it can only be estimated where in our case the power spectral density of the unprocessed chirp was used. N(f) is assumed to be white Gaussian noise.

$$F(f) = \frac{G_1^*(f)S(f)}{|G_1(f)|^2 S(f) + N(f)} \tag{47}$$

$$X(f) = F(f) \cdot Y(f) \qquad x(t) = \mathcal{F}^{-1}\{X(f)\} \tag{48}$$

An example is given in Fig 41 where (a) shows a linear frequency modulated chirp and (b) shows the driving signal after applying the Wiener deconvolution using the previously measured impulse response $g_1(t)$. The expected output signal is shown in Fig 41(c). A real measurement was performed with the driving signal of Fig 41(b). The result in Fig 42 show excellent agreement for correlation with Fig 41(c). Correlation is better than 88% for all channels which implies that the difference between the used MEMS microphones is negligible.

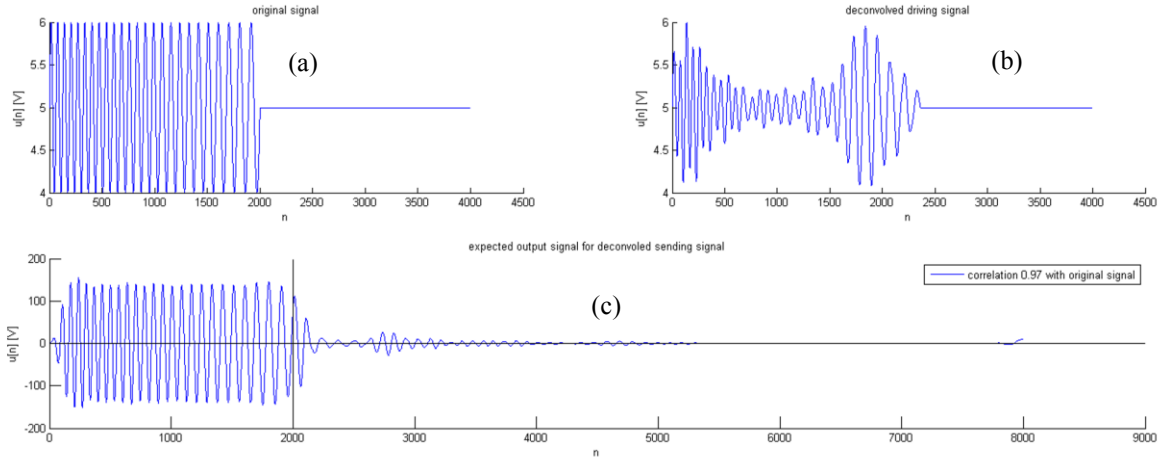


Fig 41: (a) Linear frequency modulated chirp (35 kHz-65 kHz) (b) Driving signal after applying Wiener deconvolution. (c) Expected output signal.

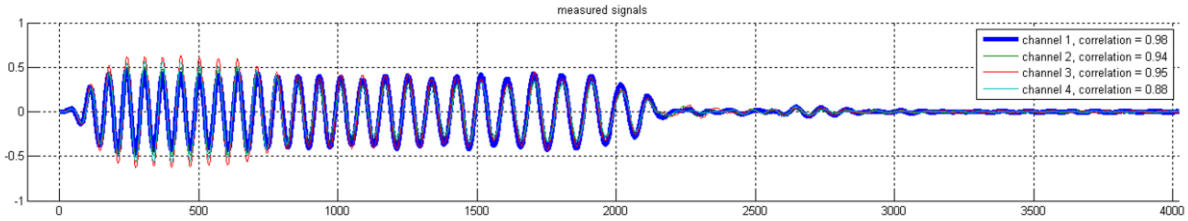


Fig 42: Measured output signal in all channels and correlation with predicted output signal of Fig 41(c). Excellent agreement was obtained.

The results of the previous sections can be summarized as following. Using the method of system impulse response identification and by applying Wiener deconvolution the sound field can be controlled at the mainlobe. By applying spatial modelling the sound field can be predicted at arbitrary polar angles and distances.

3.4.4 Other Suitable Technologies for Sound Generation

As pulse compression is employed within this work it is desirable to apply transducers with a large bandwidth as the width of the pulse is approximately 1/B (See 4.2). Suitable transducer technologies for this includes piezo composites due to their larger bandwidth and improved matching on air compared to standard piezo ceramic discs. For example the 1-3 piezo composite from Smart Materials supports a bandwidth of up to 25% of the center frequency with resonance frequencies from 40kHz to 2MHz [Are01]. The basic construction of such a composite is shown in Fig 43 where small PZT fibers of 105µm are arranged, filled, cut, plated and finally polarized.

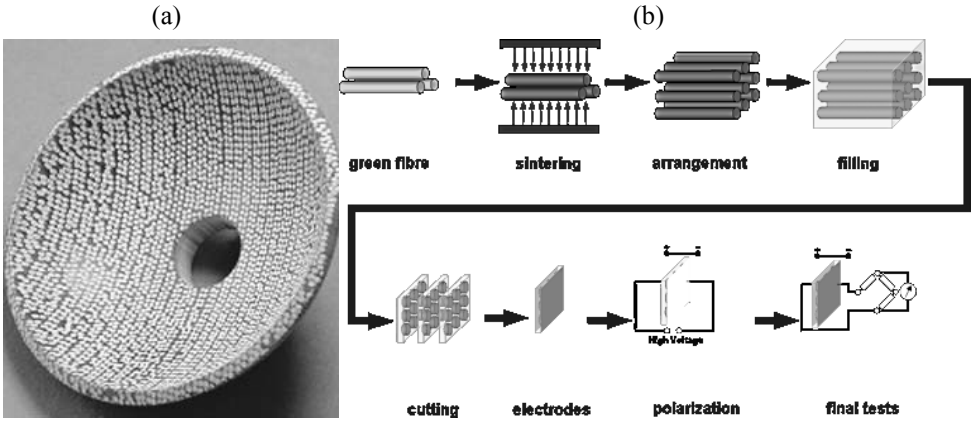


Fig 43: (a) 1-3 piezo composite transducer, (b) Manufacturing of 1-3 Piezo composites using arrange & fill method. Images taken from [Sma05]

A second type of transducers which is promising for this application is ElectroMechanical Film (EMFi). EMFi is a thin electret film with an internal cellular structure. During manufacturing the film is polarized within a very strong electrical field resulting in a permanent inherent charge [Paa00]. Benefits of this type of transducer are its low acoustic impedance of $2.6 \cdot 10^4 kg/(m^2 s)$ providing good matching to air, as well as a strong quasi piezoelectric effect with a d_{33} of 130-450 pC/N which is more than a magnitude larger than in other ferroelectric films[Jim08].

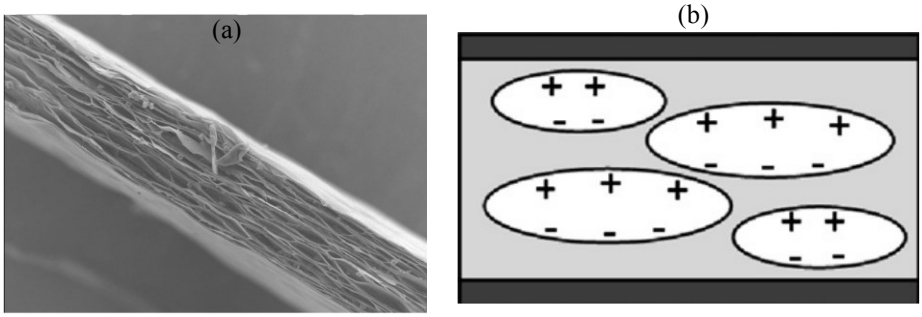


Fig 44: (a) Electron microscope scan of EMFi film with 70µm thickness. (b) Internal structure showing voids and oriented dipoles. Images taken from [Paa00]

Another benefit of the EMFi material is that different shapes of sensors can easily be constructed. EMFi films for such a purpose are readily available as screen printable sheets. Using such films focusing sound sources can easily be constructed as shown in Fig 45(a) [Hoi02]. The respective sound pressure at a distance of 6cm along a line within the focal plane is shown in Fig 45(b). Enhanced lateral resolution is achievable with such type of a transducer due to the large bandwidth of more than 300kHz on air.

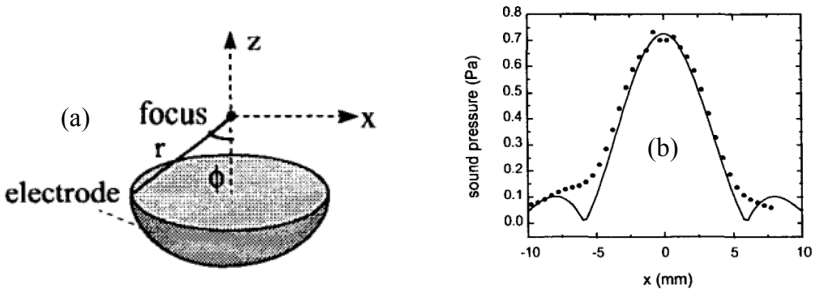


Fig 45: (a) Focusing sound device for ultrasound in air built from EMFi films. (b) Sound pressure at 6cm distance along a line on the focal plane. Images taken from [Hoi02]

Another suitable and well investigated transducer is based on the thermoacoustic effect. Despite other sound generation technologies (electrostatic, dynamic, etc.), where sound is produced by compressing/expanding the medium adiabatically, this type of speaker does not perform any mechanical-work. The speaker generates heat which in turn leads to a variation in pressure and results in a sound wave [Das13]. For a good efficiency it is important that the heat capacity of the transducer is as low as possible and the conductor is as thin as possible. Suitable technologies are indium tin oxide (ITO) based films on glass.

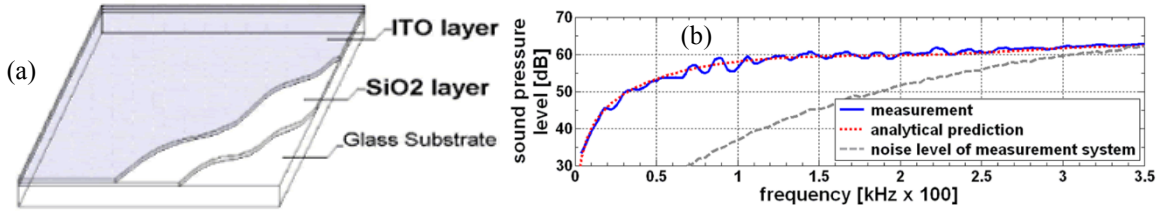


Fig 46: (a) ITO based film on glass. (b) 300nm ITO film on quartz glass. Sample size 9 x 14mm measured at 60mm distance [Das13].

The author of this work has also investigated the use of ITO based films and EMFi based films for sound generation. While application of the EMFi films and the thermoacoustic transducer are viable due to the increased bandwidth and smaller wavelength, allowing even better axial and lateral resolution, the obtainable sound pressure is rather limited. The Senscomp 600 electrostatic transducer can easily produce a sound pressure level exceeding 110dB SPL at 1m distance. This allows pulse/echo measurements on air for distances up to 100m. Another issue with EMFi films is the maximal permissible temperature of 50°C [Jim08]. The piezo 1-3 composite, while theoretically suitable, is quite expensive at small quantities starting from a few hundred Euro whereas a single electrostatic speaker is available for less than 20€. Due to these reasons the prototype design of the compact 3D sensor was realized with the electrostatic speaker.

3.5 Microphones

While the presented electrostatic transducer provides sufficient bandwidth, its lateral dimensions and strong radiation pattern are not suitable for application as a receiver. The reason for this, as will be described in more detail in chapter 5, is that multiple receivers are required for 3D localization of the reflection point. As this information is solely contained in the Time of Arrivals (ToA), any directional characteristic of the receivers further complicates the correlative signal processing. As a side note it shall be mentioned that the Ultrasonics working group at the Institute of Electrodynamics, Microwave and Circuit Engineering has presented another interesting concept for a compact 3D sensor where the radiation pattern of the transmitting device was used to obtain bearing information [Kan11]. Although basic sensor performance could be shown in a practical test setup the overall accuracy of the sensor was rather limited – especially for the detection of the azimuth.

For the compact 3D sensor presented in this work Micro-Electrical-Mechanical Systems (MEMS) microphones are chosen. These microphones are low in cost, size and have a large bandwidth. Basic construction of a MEMS microphone is shown in Fig 47. The basic operating principle is similar to a condenser microphone. Incident sound waves result in a pressure difference on a flexible membrane, which in turn is displaced from its static position. This change in capacitance is sensed, amplified and converted to an analog or digital output signal. The capacitor, as shown in Fig 47, consists of a conductive back-plate with holes and the moveable and conductive membrane. An air ventilation hole is used for pressure equalization.

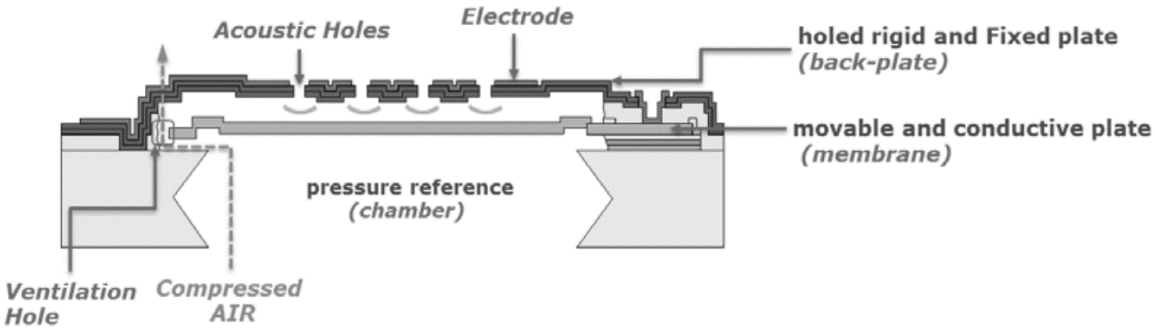


Fig 47: Basic construction of a MEMS microphone (Image taken from [ST14]

For the 3D compact sensor the Knowles SPM0404UD5 ultrasonic sensor is used. Its frequency response is shown in Fig 48. The sensitivity of the microphone is -42dBV (re 1V/Pa) with a noise floor of -150dBV/√Hz (re 1V/Pa). The acoustic overload point is 115dB SPL for a THD < 10%.

As multiple MEMS microphones are employed it is desirable that all of these microphones respond identical to an incident sound wave. For this to be the case, mounting and relative positioning of the microphones on the sensor need to be symmetrical. Furthermore, the microphones are required to have identical frequency responses and directivity patterns. To verify this a dedicated test setup as shown in Fig 49 was prepared. The sensor with the microphones can be tilted and rotated while keeping its center position. Therefore, the incident sound wave does not change (as it is a property of the transmitting device). Signals received at different tilt- and rotational angles of the sensor where then compared with the initial reference signal. The results are acceptable and the normalized correlation amplitude was 0.94 on average with respect to the reference. Results for different orientations of the sensor are shown in Fig 50.

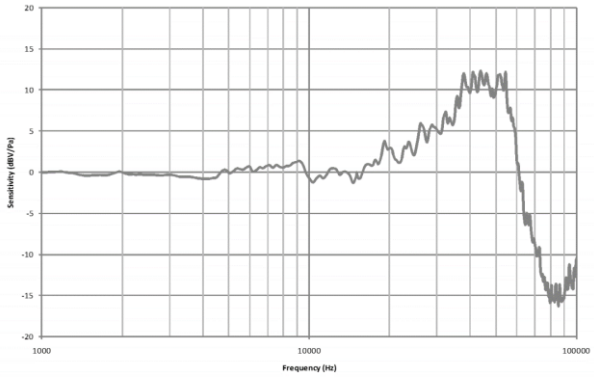


Fig 48: Freefield microphone response relative to 1kHz for Knowles SPM0404UD5 MEMS microphone (Image taken from datasheet)

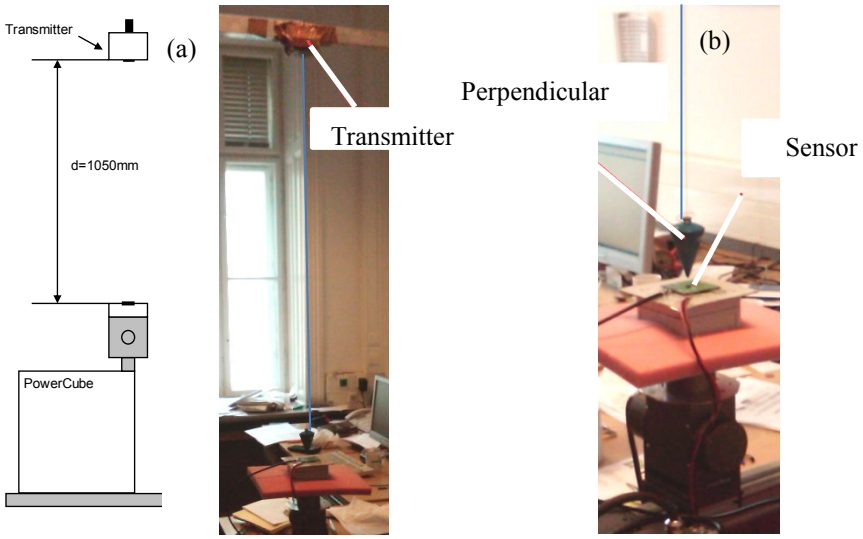


Fig 49: (a) Dedicated test setup with a transmitter mounted on a rod 1,05m above the sensor. (b) The sensor is mounted on a fixture which can be tilted, rotated and moved along an axis. Using a perpendicular the sensor was initially placed exactly along the acoustic axis of the transducer. Measurements have been performed using different tilt and rotational settings where the received signals have been compared to the initial received signal at the origin. Again the results are satistfactory with a mean cross correlation maxima up to 0.94 for tilt settings as high as 20°. For tilt settings as high as 40° mean correlation maxima drop to 0.89.

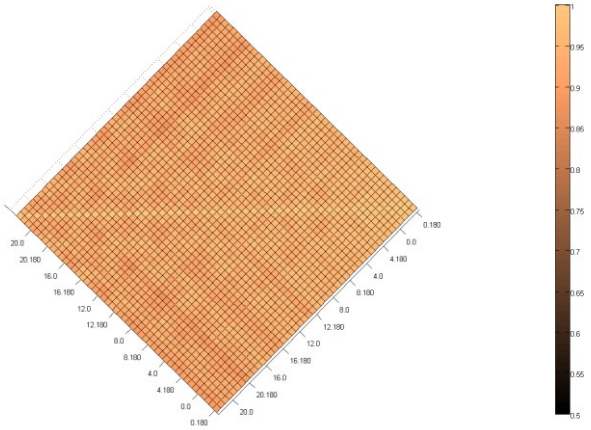


Fig 50: Binary cross correlation maxima for a single microphone for different tilt- and rotational angles of the sensor. The axis labels include the tilt-angle as a first number and the rotational angle as a second number, E.g. 20.0 corresponds to 20° tilt and 0° rotation. Minimum correlation was 0.88 whereas average was 0.94.

4 Signal Processing

4.1 Matched Filtering

Detection and time delay estimation of an a-priori known signal embedded in noise is a common problem encountered in RADAR applications. Basic design goals are maximizing the peak signal-to-noise ratio and minimizing the range estimation uncertainty.

A class of linear filters which optimally fulfills this task is a matched filter. The basic block diagram is shown in Fig 51 where a known input signal s(t), received at time Δ, and noise n(t) is present at the input of a receiver. A linear filter h(t) shall be used to detect the presence and estimate the time delay Δ under the requirements stated above.

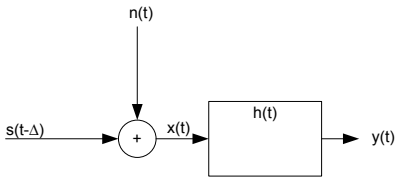


Fig 51: Block diagram showing a received time delayed input signal s(t), additive white Gaussian noise n(t) and a linear filter h(t) used for signal detection and time delay estimation.

Assume n(t) is white with a power spectral density (PSD) of $N_0/2$. Let $H(f) = \mathcal{F}\{h(t)\}$. The output noise power spectral density is given by $N_0/2 \cdot |H(f)|^2$. The average noise output power is calculated as

$$\frac{N_0}{2} \int_{-\infty}^{+\infty} |H(f)|^2 \, df \tag{49}$$

Let $S(f) \cdot e^{-2\pi f \Delta}$ be the Fourier transform of the signal s(t) delayed by Δ. The output signal of the filtered input signal at time instant τ can then be calculated by the inverse Fourier transform (50).

$$y(\tau) = \int_{-\infty}^{+\infty} S(f) e^{-2\pi f \Delta} H(f) e^{2\pi f \tau} \, df \tag{50}$$

The peak signal power to noise ratio SNR_{Peak}, at the specific time instant Δ, is then calculated as:

$$SNR_{Peak} = \frac{[y(\Delta)]^2}{\frac{N_0}{2} \int_{-\infty}^{+\infty} |H(f)|^2 \, df} = \frac{\left[\int_{-\infty}^{+\infty} S(f) H(f) \, df\right]^2}{\frac{N_0}{2} \int_{-\infty}^{+\infty} |H(f)|^2 \, df} \tag{51}$$

The matched filter for s(t) is defined as time reversed complex conjugate of the input signal, i.e. $s^*(-t)$. Therefore, its Fourier transform is simply the complex conjugate of S(f) by application of the two Fourier correspondances $\mathcal{F}\{s(-t)\} = S(-f)$ and $\mathcal{F}\{s^*(t)\} = S^*(-f)$.

$$p = \frac{\left[\int_{-\infty}^{+\infty} S(f) S^*(f) \, df\right]^2}{\frac{N_0}{2} \int_{-\infty}^{+\infty} |S^*(f)|^2 \, df} \tag{52}$$

Using the fact that $|S^*(f)|^2 = |S(f)|^2 = S(f) \cdot S^*(f) = E$, where E is the energy of the input signal, the result is identical to [Tur60] with.

$$\mathrm{SNR_{Peak}} = \frac{2\mathrm{E}}{\mathrm{N_0}} \tag{53}$$

This result can be interpreted as following – Independent of the original shape of the signal *s*(*t*), the total energy of the signal defines the peak signal power to noise ratio. For example if only limited output power is available at the transmitter the same peak signal power to noise ratio can be achieved by using a longer transmitting signal.

Following [Tur60] one can also compare the gain in the peak signal power to noise ratio. As shown in (53) $\mathrm{SNR_{Peak}}$ is the peak signal to noise ratio at the output. It is also possible to calculate the peak signal to noise ratio $\mathrm{SNR_{PeakInput}}$ at the input. If input power is equally distributed over time for the input signal s(t) then the peak power can be calculated from the energy and the duration T as E/T. The relevant noise present at the input is calculated from the equivalent noise bandwidth B of the input signal S(f). The peak signal to noise ratio $\mathrm{SNR_{PeakInput}}$ therefore is:

$$\mathrm{SNR_{PeakInput}} = \frac{\mathrm{E/T}}{\mathrm{N_0 B}} \tag{54}$$

Taking the ratio of $\mathrm{SNR_{Peak}}$ and $\mathrm{SNR_{PeakInput}}$ the well-known relationship (55) is obtained. Or in other words – At the output of the matched filter detection can be improved by using either a longer duration of the signal or larger bandwidth.

$$\mathrm{SNR_{PeakInput}}/\mathrm{SNR_{Peak}} = 2\mathrm{BT} \tag{55}$$

Maximizing the peak signal power to noise ratio optimizes the detection problem – i.e. the answer to the question whether the sending signal was present or not. For optimum range resolution we would also like that the output of the filter is maximal at time instant Δ but minimal elsewhere. Considering the output of the filter, and ignoring any part of the output signal due to noise for now, the output of the matched filter equals the autocorrelation function of the sending signal s(t) time-delayed by Δ.

$$\mathrm{y}(\tau) = \int_{-\infty}^{+\infty} \mathrm{S(f)S^*(f)e^{-2\pi f\Delta}e^{2\pi f\tau}\,df} = \mathrm{r_{ss}}(\tau - \Delta) \tag{56}$$

This is graphically shown in Fig 52 for the autocorrelation function and output. While for detecting the presence of a signal only the factor E is important, for proper time delay estimation we require that α is as small as possible. A very important result given in [Tur60] is that the factor α·β is approximately constant and in the order of unity, i.e. α·β ≈ c = constant where β equals B, the bandwidth. For a small value of α a large value of β is therefore required.

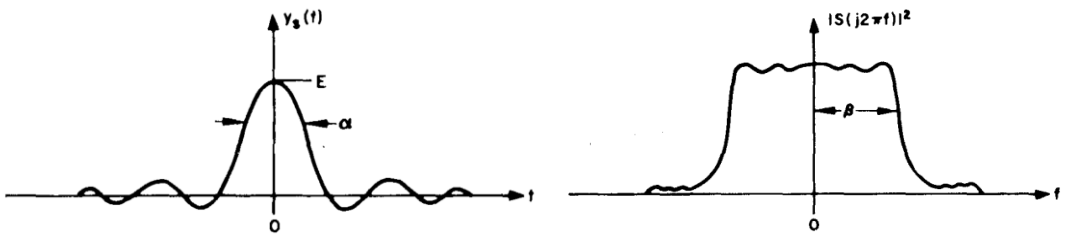


Fig 52: Output signal after matched filtering]. The width α of the output signal solely depends on the bandwidth β whereas the maximum peak E solely depends on the total signal energy. Image taken from [Tur60].

Mathematically a matched filter is identical to pulse compression techniques where the received signal is correlated with the known sending signal.

4.2 Linear Chirp

A specific class of signals used in RADAR applications is a linear frequency modulated chirp where the analytical signal is given in (57) with rect(x) = 1 if |x| < ½ or 0 otherwise.

$$s(t) = \mathrm{rect}\left(\frac{t}{T}\right) e^{2\pi j(f_0 t + kt^2/2)} \tag{57}$$

The instantaneous frequency of the signal defined in (57) is calculated as

$$\omega(t) = \frac{d\varphi(t)}{dt} = 2\pi(f_0 + kt) \tag{58}$$

Over the interval T the frequency therefore changes from $f_0 - k \cdot T/2$ to $f_0 + k \cdot T/2$. The total frequency swept is $k \cdot T = B$. The so called dispersion factor D of a linear frequency modulated chirp is calculated according to (59). This factor has a unique interpretation as the width of the collapsed pulse after matched filtering is approximately 1/B. The duration of the signal before pulse compression was T and therefore the ratio of T and 1/B equals D.

$$D = TB \tag{59}$$

An simulation of a linear frequency modulated chirp with typical parameters used within this work is shown in Fig 53(a). The amplitude shall be 1, the center frequency f_0 is 50kHz the bandwidth of the chirp is 30kHz and the total time duration of the chirp is 500µs. Furthermore bandlimited white noise within the frequency range of the chirp is present. The estimated SNR at the input is 15.7dB (theoretical is 16.2dB). At the output of the matched filter shown in Fig 53(b) the estimated SNR is 31.2dB resulting in a total peak SNR gain of 15.5dB. This is in close agreement with the theoretical gain of 14.8dB calculated using (55).

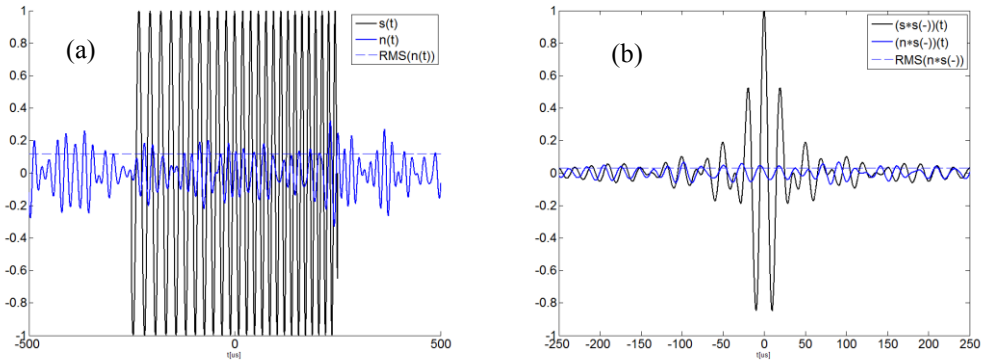


Fig 53: (a) Signals at the input of the receiver showing bandlimited Gaussian white noise within the frequency range of interest and the linear frequency modulated chirp. (b) Output signal after matched filtering showing a peak SNR increase of about 15.5dB compared to the input.

In general the output of the matched filter can directly be calculated by convolution of the input signal s(t) with $s^*(-t)$. The output is calculated as

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} \mathrm{rect}\left(\frac{\tau}{T}\right) e^{2\pi j\left(f_0\tau + \frac{k\tau^2}{2}\right)} \mathrm{rect}\left(\frac{\tau - t}{T}\right) e^{-2\pi j\left(f_0(\tau - t) + \frac{k(\tau - t)^2}{2}\right)} d\tau \\ y(t) &= e^{2\pi j\left(tf_0 - \frac{kt^2}{2}\right)} \int_{-\infty}^{\infty} \mathrm{rect}\left(\frac{\tau}{T}\right) \mathrm{rect}\left(\frac{\tau - t}{T}\right) e^{2\pi jkt\tau} d\tau \end{aligned} \tag{60}$$

Following the calculation shown in [Kla60] the evaluation of the integral yields

$$y(t) = \frac{1}{\pi kt} e^{2\pi j f_0 t} \sin\left(\pi(ktT - kt^2)\right), 0 \le t \le T \tag{61}$$

The analytical form of the output spectra is more complex to calculate but for larger values of D approaches a rectangular shape. This can be seen by rearranging equation (61) and interpreting f_0 as the carrier of the signal.

$$y(t) = e^{2\pi j f_0 t}\left[T\frac{\sin\left(\pi(ktT - kt^2)\right)}{\pi ktT}\right], 0 \le t \le T \tag{62}$$

By introducing a normalized time variable $\tau = t \cdot B$, where B is the bandwidth, and using the fact that $B = k \cdot T$ and (59) we get

$$y(\tau) = Te^{2\pi j f_0/B\,\tau}\frac{\sin\left(\pi|\tau| - \pi\frac{\tau^2}{D}\right)}{\pi|\tau|}, |t| \le T \tag{63}$$

For larger values of D the envelope of the function equals a sinc function in the time domain which corresponds to a rectangular window in the frequency domain. This becomes clear by taking the well-known Fourier correspondence

$$\mathcal{F}\left\{\frac{\sin(at)}{at}\right\} = \frac{\pi}{|a|}\mathrm{rect}\left(\frac{\pi}{a}f\right) \tag{64}$$

Hence the Fourier transform of (63) simply equals rect(f) in the normalized frequency domain. Therefore, the total bandwidth for larger values of D approaches B. The output spectrum, as well as the most important parameters of the chirp, within the resulting waveforms are shown in Fig 54. It can be seen that the output spectrum is centered on the carrier with approximate bandwidth B (only valid for larger values of D). Within the time domain response it can be seen that the chirp 3dB bandwidth is approximately 1/B. The peak SNR gain has already been shown in Fig 53.

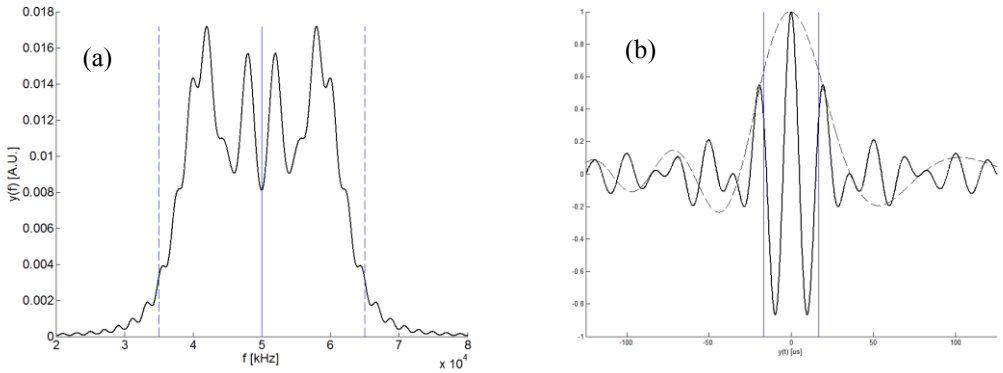


Fig 54: (a) Spectrum at output of matched filter for linear frequency modulated chirp. The center frequency of the chirp is 50kHz where the bandwidth is approx. 30kHz. (b) Autocorrelation function of a linear frequency modulated chirp. The 3dB bandwidth is approximately 1/B.

Another important parameter is target separation, i.e. the ability to separate two nearby targets from each other. For larger values of D the envelope of the autocorrelation function reduces to the sinc function. The minimum spacing, such that the two peaks of the envelope are 3dB apart, is 1.53/B as shown in Fig 55 .

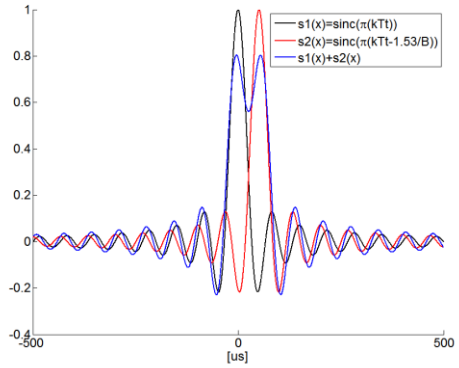


Fig 55: Target separation capability using matched filtering with a linear frequency modulated chirp. A spacing of 1.53/B is sufficient where B is the bandwidth of the chirp.

At a speed of sound of 343m/s this corresponds to a distance resolution of 1.75cm. As the measurement is performed in a pulse/echo configuration the range separation capability with these parameters is half of it, i.e. about 9mm.

4.3 Limits on the uncertainty of range estimation

Let $x[n] = s[n;\theta] + w[n]$, $0 \le n \le N-1$ be the received signal where $w[n]$ is Gaussian white noise and s is the signal for which the unknown parameter θ shall be estimated. In this context the unknown parameter corresponds to the time delay of the transmitted signal. A minimum variance for any unbiased estimator can be found by calculation of the Cramer-Rao lower bound as in (65) [Kay93].

$$\mathrm{Var}(\hat{\theta}) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1}\left(\frac{\partial s[n;\theta]}{\partial\theta}\right)^2} \tag{65}$$

In our case let $s[n;\theta]$ be the real part of (57) where δ is the sampling interval and $\theta = \tau$, the unknown time-shift. Therefore s(t) can be written according to:

$$s(t) = \mathrm{rect}\left(\frac{t-\tau}{T}\right)\cos(2\pi f_0(t-\tau) + \pi k(t-\tau)^2) \text{ with } t = n\delta \tag{66}$$

Using the initial equation $x[n] = s[n;\theta] + w[n]$ and by substituting $s[n\delta]$ (66) for $s[n;\theta]$ as well as using the fact that s(t) is zero outside $|t-\tau| > T/2$, the result (67) is obtained.

$$x[n] = \cos(2\pi f_0(n\delta-\tau) + \pi k(n\delta-\tau)^2) + w[n]\,, |n\delta-\tau| \le \frac{T}{2} \tag{67}$$

Calculation of the partial derivative with respect to the unknown time delay τ yields

$$\frac{\partial s[n;\tau]}{\partial\tau} = -\sin(2\pi f_0(n\delta-\tau) + 2k(n\delta-\tau)^2)\left(2\pi f_0 + 4k(n\delta-\tau)\right), |n\delta-\tau| \le \frac{T}{2} \tag{68}$$

Assume N is sufficiently large, such that the received signal is fully contained within the observed interval, (68) can be directly applied to calculate the Cramer-Rao lower bound. Limits on the uncertainty of range estimation using the linear frequency modulated chirp and the chirp parameters of the previous section (f_0 = 50kHz, B = 30kHz, T = 500µs) and different SNRs are given in Fig 56. This theoretical result was compared with a simulation of the matched filter within MATLAB. Using the output of the matched filter the maximum peak was determined within the signal which was taken as an estimate for the time delay τ. Due to the discrete time resolution parabolic interpolation is a well-known technique to improve the accuracy of the result. The formula used for interpolation is given in (69). The benefit can be clearly seen in

Fig 56 where, without interpolation, at higher SNR the accuracy of the result is limited by the round-off error.

$$\hat{n} = n_0 + \frac{1}{2} \cdot \frac{y[n_0 - 1] - y[n_0 + 1]}{y[n_0 - 1] - 2y[n_0] + y[n_0 + 1]} \tag{69}$$

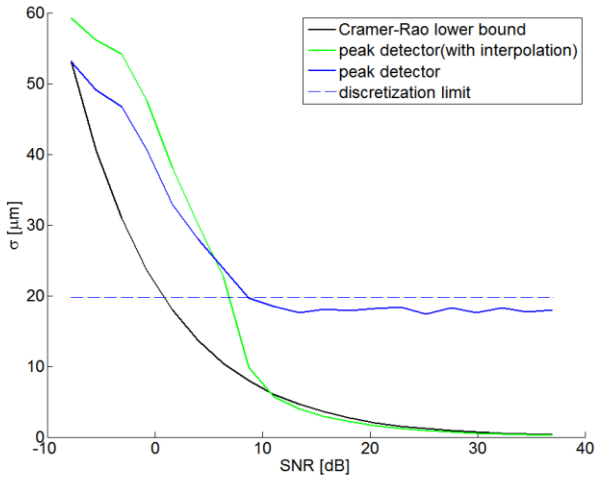


Fig 56: Cramer-Rao lower bound for range uncertainty for a linear frequency modulated chirp with B = 30kHz, f_0 = 50kHz, T = 500µs and δ = 1/5MS. If a peak detector with interpolation is used the performance is almost identical to the Cramer-Rao lower bound for an SNR better than 10dB. If the SNR is lower interpolation performs worse as only the neighboring samples are used for interpolation which amplifies noise. If a peak detector without interpolation is used the performance is mostly limited by the discrete sampling.

4.4 1-Bit binary correlation for efficient signal processing

Reduction of the computational complexity and cost of hardware implementation is beneficial for the deployment of multiple low cost sensors. The discrete implementation of a matched filter as described within 4.1 is given in (70) where x[n] is the input signal of the matched filter and y[n] is the output signal. The signal h[m] is the matched filter and by definition equals $s^*[-n]$, where s is the sending signal. If L is the length of the sending signal then $s^*[-n] = h[n]$ is non zero within the discrete interval [–(L-1),0].

$$y[n] = (h * x)[n] = \sum_{m=-\infty}^{\infty} h[m]x[n-m] = \sum_{m=-L+1}^{0} h[m]x[n-m] \tag{70}$$

Let N be the length of the recorded input signal x, then for large values of N, approximately $N \cdot L^2$ multiplications and additions are required. At high sample rates, large bit depths and multiple channels, this introduces a high computational complexity requiring at least a DSP or FPGA for real-time processing.

Application of 1-bit correlation proves to be an efficient solution to this problem. Instead of the multiply and accumulate operation, which is typical for the FIR filter structure shown in (70), the multiplication can be reduced to a logical AND operation. Furthermore, depending on the bit-width of the hardware, multiple samples can be processed at once. An excellent introduction and reference to 1-bit correlation and their properties is given in [Elm95]. Therefore, within this section only the most important properties relevant to this work are repeated. Applying the identity between cross correlation and convolution, i.e. $s^*(-t) * x$ equals the cross correlation for s and x, (70) can be rewritten as:

$$y[n] = \sum_{m=-\infty}^{\infty} s[m]x[m+n] = \sum_{m=0}^{L-1} s[m]x[m+n] \tag{71}$$

The 1-bit representation of an input signal x[n] is defined as the sign of the input signal, i.e.

$$D\{x[n]\} = \begin{cases} 1 \text{ if } x[n] \geq 0 \\ -1 \text{ otherwise} \end{cases} \tag{72}$$

Furthermore we can define the AND operation as

$$AND(x[n], y[n]) = \begin{cases} 1 \text{ if } x[n] = 1 \text{ and } y[n] = 1 \\ -1 \text{ otherwise} \end{cases} \tag{73}$$

Applying (70), (71), (72) and (73) and assuming real valued signals only yields the definition of the 1-bit normalized cross correlation for two input signals x[n] and s[n] as

$$y[n] = CC_{x,s}[n] = \frac{1}{L} \sum_{m=0}^{L-1} AND(x[m], s[m+n]) \tag{74}$$

Fig 57(a) compares the auto correlation properties of the linear frequency modulated chirp of Fig 53 with the discrete implementation of (74). For presentation purposes both functions have been normalized to 1. With exception of some minor differences the auto correlation functions are very similar. Fig 57 (b) shows the main difference as the 1-bit quantization at the input raises the noise floor considerably.

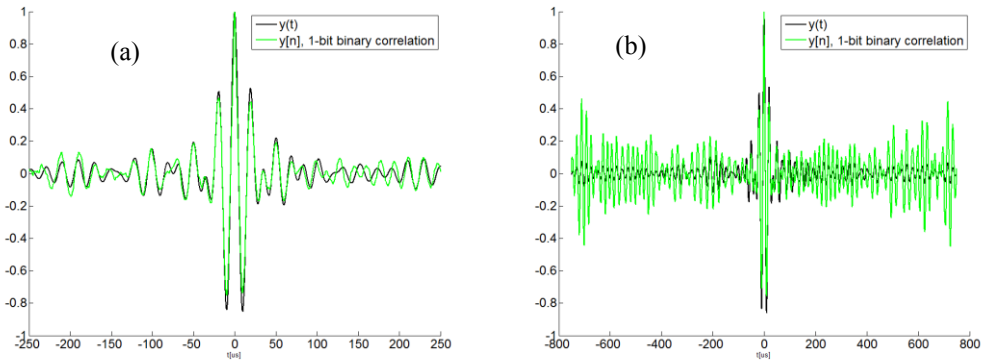


Fig 57: (a) compares the auto correlation function of the linear frequency modulated chirp for the analog case and in case 1-bit binary correlation is applied. The envelope of the AKF is nearly identical. (b) shows the rise of the noise floor due to the 1-bit binary quantization.

As a result of the increased noise floor, due to the 1-bit quantization, the SNR is degraded. This negatively affects the ranging uncertainty. Only at an input SNR of > 30dB the performance loss is less than 3dB as shown in Fig 58. Parabolic interpolation for improved peak determination was not delivering reliable results. For this reason it was excluded.

Despite the loss in performance 1-Bit binary correlation was applied throughout the work for the basic signal processing as it is favorable for practical sensor construction. Furthermore, as the 1-bit correlation ignores the amplitude of the echo, this process is essentially object independent, e.g. weak or strong reflectors are treated equally which was considered as a desired design property of the sensor. In addition the low complexity and low cost of such an implementation keeps system costs at a minimum.

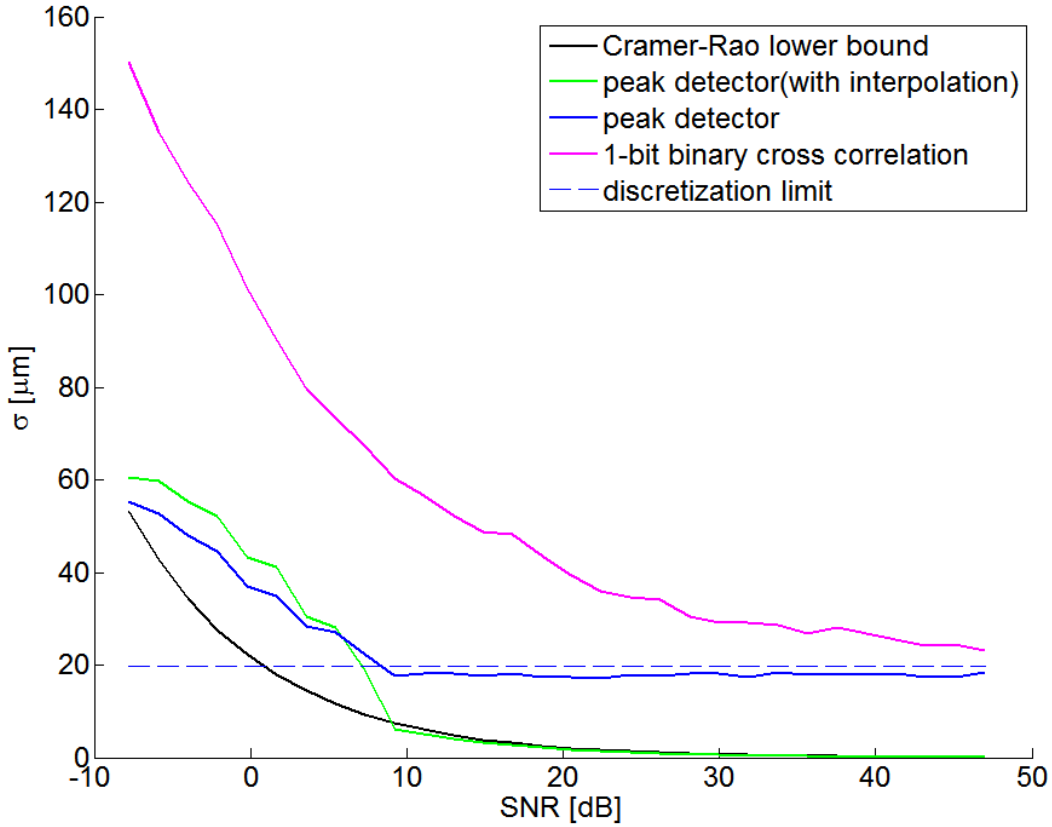


Fig 58: Comparison of ranging uncertainty with Cramer-Rao lower bound for a linear frequency modulated chirp. No interpolation used. Used sample rate = 5MS/s.

4.5 Choice of sample rate

It can be seen from Fig 58 that the sample rate introduces a lower bound on the uncertainty of the ToA measurements. Calculating this ranging uncertainty by assuming uniform and random placement of reflectors yields $\delta/\sqrt{12}$ where δ is the sampling interval. Results for different sample rates are given in Tab. 3. The expected uncertainty for the ToA is the combined uncertainty of the distance estimation algorithm using matched filtering, which is limited by the Cramer-Rao lower bound, and by the discrete sampling. Comparing this with the measurement results obtained by [Kan11] in Fig 59 (a) shows no correlation – in fact the estimated uncertainty was much larger in all cases for the measurements.

Ignoring this result for a moment and investigating the uncertainty of the time-difference-of-arrivals (TDoA) yields an interesting observation. As the receiving channels and the absolute values of the ToA are uncorrelated the combined uncertainty for the time difference of arrival (TDoA) is calculated as $\sqrt{\delta^2/12 + \delta^2/12} = \delta/\sqrt{6}$. It can be seen in that in this case there is a good correspondence for the sample rates 1MS/s (408ns ≈ 440ns) and 2MS/s (204ns ≈ 230ns) using the results of Fig 59(b). At higher sample rates the limitations by the chirp and the noise at the input need to be considered as well. The limitation by the Cramer Rao bound can be estimated from Fig 58. At 5MS the uncertainty for an SNR of ≈ 7.78dB (1-bit binary correlation) is approximately 9.3µm. The combined uncertainty due to the discretization and the Cramer-Rao Bound yields a total uncertainty of 31µm corresponding to 90ns which is in moderate agreement with 130ns.

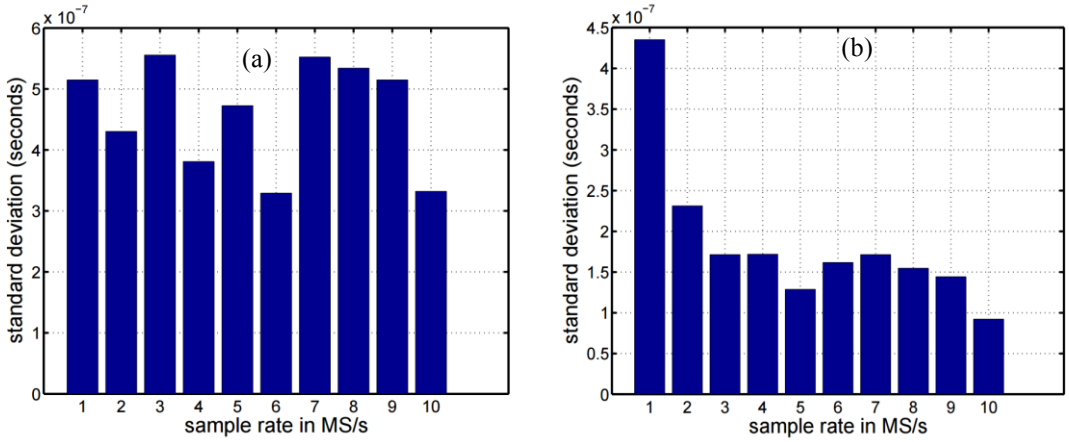


Fig 59: (a) Uncertainty in ToA and (b) TDoA measurements for a linear frequency modulated chirp using 1-bit binary correlation. Images taken from [Kan11] where the measurement was performed at 1.5m distance for a cylindrical reflector.

Sample Rate	δ/√12 (ToA)	δ/√6 (TDoA)	δ/√6·c
1MS/s	289ns	408ns	140µm
2MS/s	144ns	204ns	70µm
5MS/s	58ns	82ns	28µm
10MS/s	29ns	40ns	14µm

Tab. 3: Limitations in ranging uncertainty due to discrete sampling (assuming no interpolation)

As a summary the following conclusions can be drawn:

- For ToA measurements the uncertainty is neither due to signal processing nor to the uncertainty introduced by the discrete sampling. Therefore, it can be concluded that any uncertainty is due to disturbances during transmission, e.g. slightly varying speed of sounds due to temperature gradients, air movement, etc. The expected uncertainty for practical measurements is around 500ns, which accounts for ~170µm.
- For TDoA measurements it is important to use a high sample rate (or a suitable interpolation algorithm). The expected uncertainty for practical measurements is around 130ns, which accounts for 50 µm.

5 Compact 3D Sensor

Applying the results of the previous chapters a 3D compact sensor is presented herein. Signal processing applies 1-Bit binary cross correlation method in combination with a linear frequency modulated chirp (B = 30kHz, f_0 = 50kHz, T = 500µs) as described in section 4.2 and 4.4. As a transmitter a Senscomp 600 electrostatic open face transducer is used. This transmitter itself is fully described by the electromechanical system identification process introduced in section 3.4.2 and 3.4.3. Spatial modelling of the radiated sound wave is performed according to section 3.4.1 resulting in good correlation maxima even for echoes outside of the main lobe of transmission. As already presented briefly at the end of the state of the art in section 2.2 key driving factors for the sensor design are.

a) The sensor should allow localization of 3D reflections points with high accuracy, i.e. sub-millimetre and sub-degree.
b) The sensor should allow a fast measurement rate while at the same time obtain a large amount of information within a single shot, i.e. is all possible reflections points enclosed within the echo shall be identified. Combination of both properties is essential as the prior one enables identification of moving targets whereas the second property allows analysis of complex scenarios.
c) The sensor should be of general applicability and not restricted to a certain test setup. This implies that the sensor should especially be object independent, i.e. no matter if the type of reflector is a plane, sphere or any other kind of object the sensor should be able to correctly identify the reflection point on the surface of the object.
d) The measurements returned by the sensor should have high quality and should contain a low number of outliers. Outliers for ultrasound measurements in air are typically due to degraded SNR, reflections from complex objects or objects with discontinuities, and echo overlapping.

5.1 Basic Sensor Construction

The sensor design shown in Fig 60 is similar to [Kan11] and consists of a centered electrostatic transducer and four (or more) microphones mounted $2 \cdot d$ = 80mm apart. The pair of microphones m_1/m_2 is aligned on the x-axis and a second pair m_3/m_4 is aligned on the y-axis of the sensor. For the transmitter a Senscomp 600 open face electrostatic transducer was used which provides sufficiently large bandwidth and good acoustic coupling on air. The microphones used are Knowles SPM0404UD5 microphones which omni-directional directivity and a sensitivity of -42dBV (re 1V/Pa).

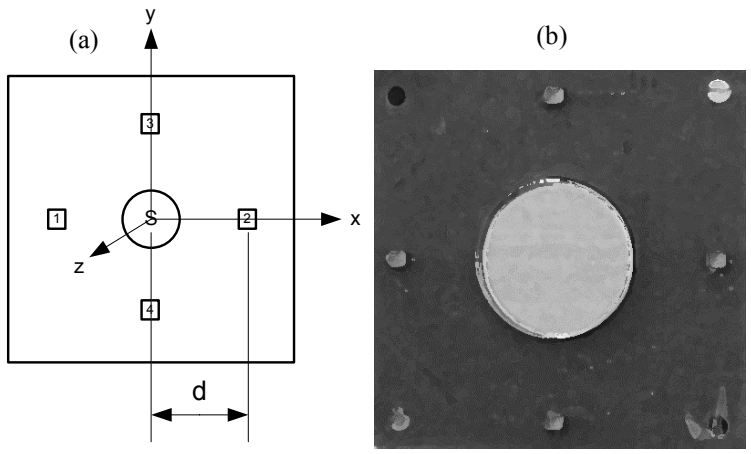


Fig 60: (a) 3D compact sensor with centered electrostatic transducer and four microphones where one pair m_1/m_2 is aligned along the x-axis and the second pair m_3/m_4 is aligned along the y-axis. (b) 3D compact sensor built from the Senscomp 600 open-face electrostatic transducer and four MEMS microphones.

5.2 Elementary localization of 3D reflection points

For now, it is assumed for the sake of simplicity, that only a single reflector is present in the vicinity of the sensor. At time $t = 0$ the sensor emits a linear frequency modulated chirp using the electrostatic transducer. After some time $t = 2 \cdot r/c$, where r is the distance between the transmitter and the obstacle, and c the speed of sound, the sound wave is reflected and eventually received by the four microphones m_1, m_2, m_3 and m_4. After matched filtering four Time-of-Arrivals (ToA) t_1, t_2, t_3 and t_4 are obtained. Calculation of the 3D reflection point is straightforward as it can be reduced to two 2D-problems yielding a simple analytical solution as already shown in [Kan11].

This is a desirable property because an analytical solution for 3D localization using trilateration is already quite complex [Man96]. Furthermore, as three ToAs are the minimum for 3D localization, no redundancy is available in the information, preventing reliable localization in real-world applications. As will be shown in 5.4 reliable localization mostly boils down to solving the echo correspondence problem, i.e. given a set of ToAs, corresponding ToAs belonging to the same object/reflector have to be identified. Typical issues encountered herein are ambiguities due to reflectors at close proximity, or due to object discontinuities on reflectors. Furthermore, the usage of multiple templates for correlation also requires a suitable selection process.

5.2.1 Formula for 3D localization

The formulas for 3D localization are best derived by using two 2D projections where each projection always includes one of the microphone pairs m_1/m_2 or m_3/m_4, the sender s and the reflection point **p**. Calculation is illustrated in Fig 61 where the x-coordinate is obtained from the two ToAs t_1, t_2 (a) whereas y is calculated using t_3 and t_4 as shown in (b). It shall be noted that the vertical axis does not correspond to the z-axis of the sensor. Therefore, z needs to be calculated afterwards.

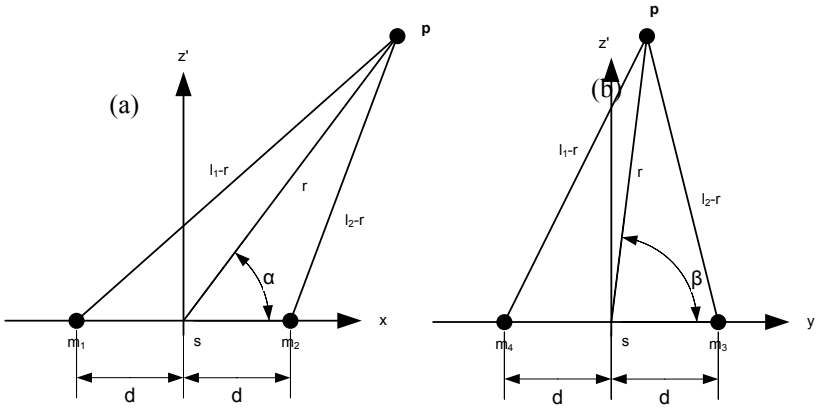


Fig 61: Calculation of the 3D reflection point is performed using two 2D projections where each projection includes the reflection point **p**, the sender and the respective microphone pair. (a) is used for calculation of the x-coordinate using microphone pair m_1/m_2. (b) is used for calculation of the y coordinate using pairs m_3/m_4. Note that the z-axis within the projection is not the same z-axis as within Fig 60. It is defined as being perpendicular to the x (or y) axis and is required to lie on the plane defined by the reflector **p** and the microphone pair m_1/m_2 (or m_3/m_4)

By application of the law of cosines x and y are calculated according to (75) and (76). The variables l_1, l_2, l_3 and l_4 are calculated from the speed of sound and the respective ToAs. The constant d is the distance of the microphones to the origin.

$$x = r\cos\alpha\,, \cos\alpha = \frac{(l_2 - r)^2 - d^2 - r^2}{-2dr}\,, r_{12} = \frac{l_1^2 + l_2^2 - 2d^2}{2(l_1 + l_2)} \tag{75}$$

$$y = r\cos\beta, \cos\beta = \frac{(l_3 - r)^2 - d^2 - r^2}{-2dr}, r_{34} = \frac{l_3^2 + l_4^2 - 2d^2}{2(l_3 + l_4)} \tag{76}$$

Finally, the z coordinate can be calculated according to (77)

$$z = \sqrt{r^2 - x^2 - y^2} \tag{77}$$

5.2.2 Simplified formula for 3D localization

Using a limiting argument for the distance r the equations (75), (76) and (77) can be further simplified as given in (78).

$$\begin{aligned} x &= r \cdot \cos\alpha, \qquad \cos\alpha = \frac{l_1 - l_2}{2d} \\ y &= r \cdot \cos\beta, \qquad \cos\beta = \frac{l_4 - l_3}{2d} \\ r &= \frac{(l_1 + l_2 + l_3 + l_4)}{8} \end{aligned} \tag{78}$$

5.2.3 Localization in spherical coordinates

A representation of the Cartesian coordinates (x,y,z) in spherical coordinates (r,θ,φ), where θ is the inclination and φ is the azimuth, is easily obtained by application of (79).

$$\begin{aligned} r &= (r_{12} + r_{34})/2 \\ \theta &= \cos^{-1}\left(\frac{z}{r}\right), \varphi = \tan^{-1}\left(\frac{y}{x}\right) \end{aligned} \tag{79}$$

5.3 Uncertainty of localization

The calculation of uncertainty in this section closely follows the methodology outlined in the Guide to the Uncertainty of Measurement (GUM) [GUM08]. In general, a measurement includes imperfections referred to as error. Although error is an idealized concept as the value of the measurand is in general not known [GUM08] it is still useful to gather a basic understanding of the system. Error consists of two components - a systematic and a random component. While the random component of the error can often be reduced by averaging (in case the mean is zero), reduction of the systematic components requires different means. In the case of the compact 3D sensor the following error components can be identified

- Uncertainty in localization due to error propagation of correlated and uncorrelated ToA noise. Uncorrelated noise is due to independent electronic channels, self-noise of the ultrasonic microphones, and Brownian motion of the air. Correlated noise, which is common to all receivers, is due to disturbances during the propagation of the acoustic wave, i.e. varying speed of sound, wind, etc.
- A systematic error due to the chosen calculation method, which assumes a single reflection point for all microphone/transmitter pairs, whereas in practice this requirement is only fulfilled on a point reflector. This error is bounded in the case of convex reflectors and further described in 5.3.3.

As recommended by the GUM systematic errors should be compensated (See [GUM08], 6.3.1) as far as possible. In this case the resulting uncertainty is only due to the ToA uncertainties. The corresponding results for the standard and simplified formulas are given in 5.3.1.

Compensation of systematic errors, as in the case of the compact 3D sensor, is only possible after the type of reflector is identified or known by a-priori information. Compensation might

therefore not be appropriate for all applications. In this case the concept F2.4.5 (Uncertainty when corrections from a calibration curve are not applied) of the [GUM08] shall be followed.

5.3.1 Uncertainty of localization for standard formulas

Let $\mathbf{r}(t_1,t_2,t_3,t_4,c)$ be the function defined according to (80) where c is the speed of sound and $t_1,\ldots,t_4$ are the respective ToAs. The variables x, y, z and r are calculated according to (75), (76) and (77).

$$\mathbf{r}(t_1,t_2,t_3,t_4) = \begin{pmatrix} r \\ \theta \\ \varphi \end{pmatrix} = \begin{pmatrix} \sqrt{x^2+y^2+z^2} \\ \cos^{-1}\left(\frac{z}{r}\right) \\ \tan^{-1}\left(\frac{y}{x}\right) \end{pmatrix}, l_i = c \cdot t_i \tag{80}$$

Let **J** (81) be the Jacobian of (80) and I_x be the (symmetric) noise covariance matrix as shown in (82). As all receivers are identical and symmetrically arranged, a single ToA uncertainty is sufficient for modelling, where the noise is split into a correlated and uncorrelated part between the receivers. The values of the correlation coefficient δ can range between 0, i.e. all the noise is uncorrelated, to +1, where in this case noise is common to all receivers.

$$\mathbf{J} = \begin{pmatrix} \frac{\partial r}{\partial t_1} & \frac{\partial r}{\partial t_2} & \frac{\partial r}{\partial t_3} & \frac{\partial r}{\partial t_4} \\ \frac{\partial \theta}{\partial t_1} & \frac{\partial \theta}{\partial t_2} & \frac{\partial \theta}{\partial t_3} & \frac{\partial \theta}{\partial t_4} \\ \frac{\partial \varphi}{\partial t_1} & \frac{\partial \varphi}{\partial t_2} & \frac{\partial \varphi}{\partial t_3} & \frac{\partial \varphi}{\partial t_4} \end{pmatrix} \tag{81}$$

$$\mathbf{I_x} = \begin{pmatrix} \sigma_{t1}^2 & \sigma_{t1}\sigma_{t2} & \sigma_{t1}\sigma_{t3} & \sigma_{t1}\sigma_{t4} \\ . & \sigma_{t2}^2 & \sigma_{t2}\sigma_{t3} & \sigma_{t2}\sigma_{t4} \\ . & . & \sigma_{t3}^2 & \sigma_{t3}\sigma_{t4} \\ . & . & . & \sigma_{t4}^2 \end{pmatrix} = \begin{pmatrix} \sigma_t^2 & \delta\sigma_t^2 & \delta\sigma_t^2 & \delta\sigma_t^2 \\ . & \sigma_t^2 & \delta\sigma_t^2 & \delta\sigma_t^2 \\ . & . & \sigma_t^2 & \delta\sigma_t^2 \\ . & . & . & \sigma_t^2 \end{pmatrix} \tag{82}$$

Using (81) and (82) the combined uncertainty is calculated according to (83). Evaluating this formula reveals some interesting results.

$$\mathbf{U} = \mathbf{J} \cdot \mathbf{I_x} \cdot \mathbf{J^T} \tag{83}$$

First of all it can be seen in Fig 62(a) and Fig 63(a) that the ToA uncertainty directly influences the ranging. In case of uncorrelated noise an averaging effect, inherent to the computation formulas, reduces the uncertainty by a factor of two, which is in good correspondence with the intuitive reduction of the standard deviation by the square root of the number of receivers, i.e. √4 = 2. Furthermore, as we have pulse/echo measurement, the distance error is half the time of flight error times the speed of sound. This results in the factor $\sigma_t \cdot c/4 = 85.75\mu m/\mu s$. Or in other words – An uncorrelated ToA uncertainty of 1µs results in a distance uncertainty of 85.75µm. For the determination of polar and azimuth angles it can be seen that polar angle estimation uncertainty increases with higher polar angles whereas azimuth estimation uncertainty is highest on the acoustic axis of the sensor. The average uncertainty for the polar angle is 0.7°/µs whereas for the azimuth it is 0.6°/µs.

For correlated noise the ranging uncertainty is not reduced as averaging does not help in reduction of correlated noise. This result can be seen in Fig 63(a) where the ranging uncertainty is

increased to $\sigma_t \cdot c/2 = 171.5$ µm/µs. On the other hand side the uncertainty for polar angle and azimuth estimation is not strongly affected by correlated noise as it is nearly completely canceled. On average the noise suppression is in the range of ~100dB compared to the case with uncorrelated noise. This inherent property of the sensor becomes more obvious if the simplified formulas presented in (78) are inspected. Due to the ToA differences any noise for angle estimation is almost perfectly cancelled.

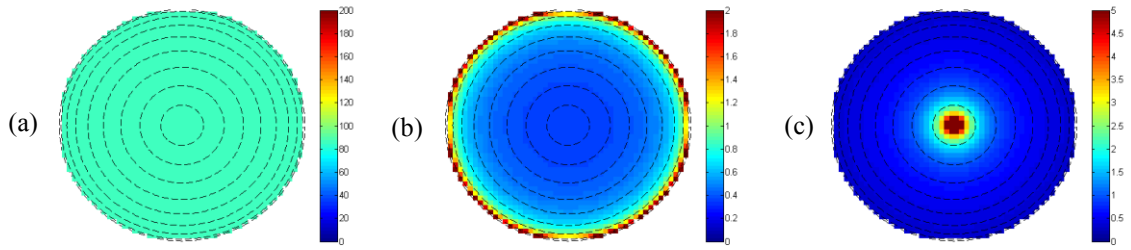


Fig 62: Performance figures for uncorrelated noise (δ=0) where the center corresponds to 0° polar angle. Each circle accounts for 10° degree covering the half-space in front of the sensor. (a) Distance estimation uncertainty σ_r in µm/µs. (b) Polar angle estimation uncertainty in °/µs. (c) Azimuth estimation uncertainty in °/µs.

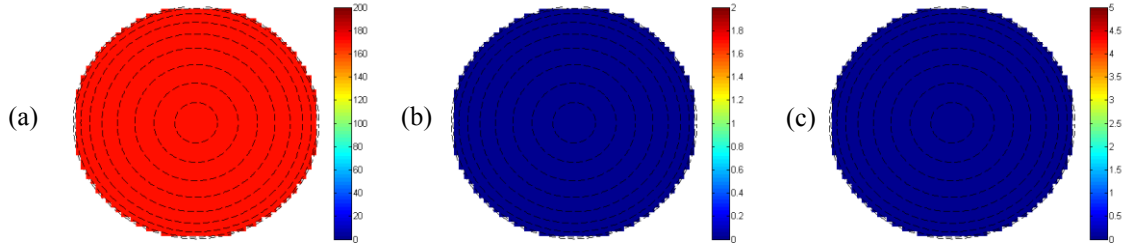


Fig 63: Performance figures for correlated noise (δ=1) where the center corresponds to 0° polar angle. Each circle accounts for 10° degree covering the half-space in front of the sensor. (a) Distance estimation uncertainty σ_r in µm/µs. (b) Polar angle estimation uncertainty in °/µs. (c) Azimuth estimation uncertainty in °/µs. Correlated noise therefore does not impact bearing estimation although ranging estimation is affected most.

5.3.2 Uncertainty of localization for simplified formulas

Performance of the simplified formulas with respect to ToA uncertainty was compared to the standard formulas. The difference in performance is negligible. Therefore, no results are presented within this section and the reader is referred to Fig 62 and Fig 63. The inherent reduction of correlated noise is more obvious in the simplified formulas shown in (78) as the determination of the angles is only defined by time differences.

5.3.3 Object and distance dependent systematic error

The formulas presented in (75), (76) and (77) were derived by assuming a point reflector. In practice this is not the fact as the reflection points for each microphone and sender are different, depending on the type of reflector and its orientation. Minimal and maximal error bounds have been calculated for different types of reflectors and orientations of the sensor. Error in this case has three different possible interpretations – the first type of error is the distance between the desired reflection point $\mathbf{p}_s$ and the calculated one $\mathbf{p}_c$, i.e. $\|\mathbf{p}_c - \mathbf{p}_s\|$. The second type of error is the length difference between the reflection points $\|\mathbf{p}_c\| - \|\mathbf{p}_s\|$, whereas the third type of error is the enclosed angle α_{err} between the two vectors. These concepts are illustrated in Fig 64(a) showing the object dependent reflection points whereas (b) shows the different types of errors.

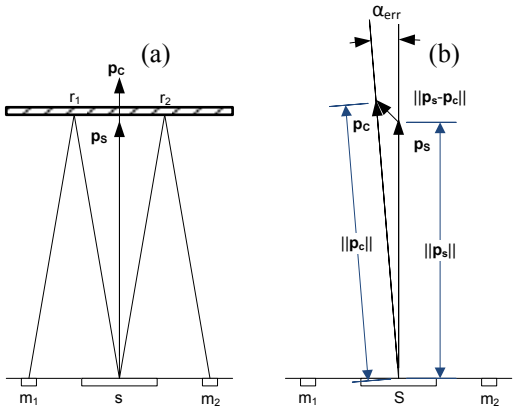


Fig 64: (a) Depending on the shape and orientation of the reflector, reflection points are not necessarily identical resulting in an error between the calculated result and the theoretical reflection point. (b) Shows the different type of errors where the first error $\|\mathbf{p_c}-\mathbf{p_s}\|$ is the norm of the distance, the second type is the difference $\|\mathbf{p_c}\|-\|\mathbf{p_s}\|$ of the lengths of the vectors, and the third is $\alpha_{err} = \cos^{-1}[\mathbf{p_c} \cdot \mathbf{p_s}/(\|\mathbf{p_c}\|\cdot\|\mathbf{p_s}\|)]$

Simulation results for a plane reflector are given in Fig 65. It is interesting to see that the simplified formula delivers better results than the standard formula. Comparing these results to [Kan11, Fig 6.1.4] it can be seen in Fig 65(a) that the error at 0° polar angle is identical to his results. In general the error for the standard algorithm is higher as [Kan11] did not account for all possible orientations of the reflector. In addition it can be seen that the error is dependent on the polar angle θ and is modulated by the azimuth angle φ.

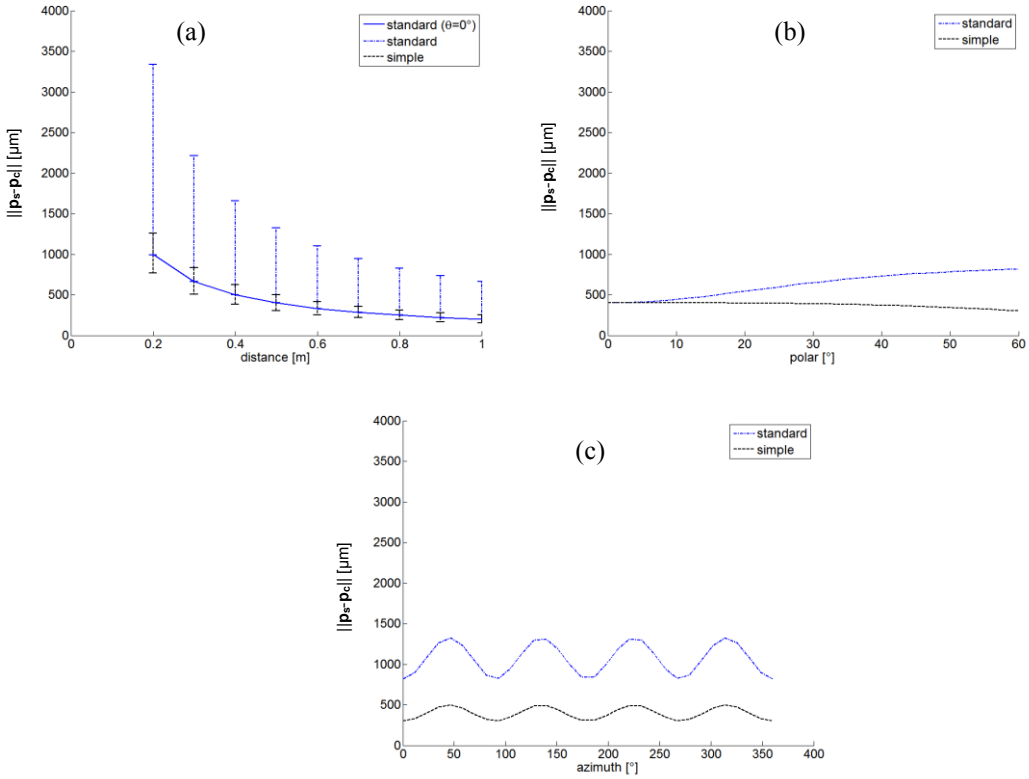


Fig 65: Distance error shown for a plane reflector where the sensor orientation was varied between 0°-360° around its z-axis and tilted by 0° - 60°. (a) Minimal and maximal error between theoretical and calculated reflection point. (b) Dependence of the error on the polar angle θ for φ = 0° and r = 0.5m. (c) Dependence of the error on the azimuth angle φ for θ = 60° and r = 0.5m.

The main error source for the standard algorithm is due to the larger direction estimation error as can be seen in Fig 66(b). The difference in length between the theoretical point and the calculated point has equal magnitude but opposite sign for both algorithms as shown in Fig 66(a).

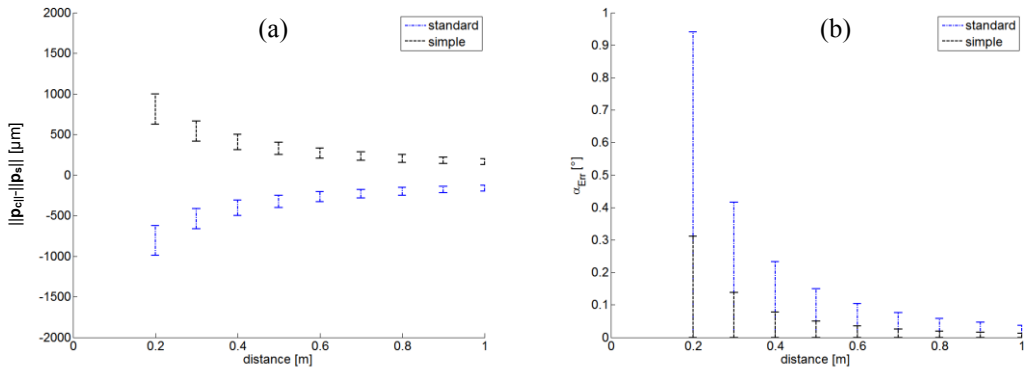


Fig 66: (a) Minimal and maximal error between the lengths of the position vector of the theoretical reflection point $\mathbf{p_s}$ and the calculated reflection point $\mathbf{p_c}$. (b) Angle α_{err} in ° enclosed between the vectors of the theoretical reflection point $\mathbf{p_s}$ and the calculated reflection point $\mathbf{p_c}$.

In case of a point reflector the performance of the algorithms change their roles. The standard algorithm introduces no additional error as it was developed for this case. The simplified algorithm introduces errors slightly larger than the standard algorithm for a plane. The results are shown in Fig 67.

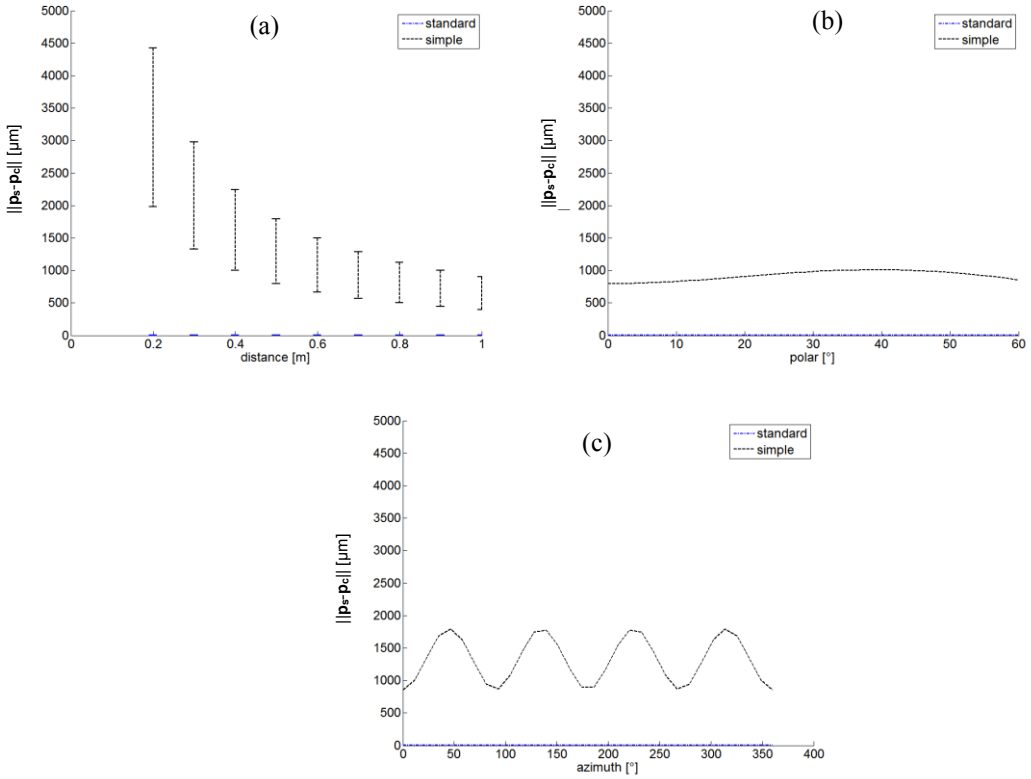


Fig 67: Distance error shown for a point reflector where the sensor orientation was varied between 0°-360° around its z-axis and tilted by 0°-60°. (a) Minimal and maximal error between theoretical and calculated reflection point. (b) Dependence of the error on the polar angle for φ = 0° and r = 0.5m. (c) Dependence of the error on the azimuth angle for θ = 60° and r = 0.5m.

It shall be noted, that although these calculations are deterministic, the object type and orientation in general is not known a-priori. In this case the combined uncertainty for a measurement of a reflector at a given position **p** is the uncertainty of the localization plus the uncertainty due

to the unknown object type. Therefore, it is interesting to compare which of these two uncertainties contributes most. Using the results of Fig 65(a) and the simple algorithm the expected error at 0.5m is approximately 500µm. At 1m this error reduces to 250 µm.

The part of the combined uncertainty due to ToA uncertainty can be estimated from the results of Fig 59. At 5MS/s it can be seen that the uncertainty of a TDoA measurement is approximately 130 ns whereas the uncertainty for a ToF measurement is 470ns. Let n_a and n_b be noise components within the signal for a pair of microphones, e.g. the pair m_1/m_2 or m_3/m_4. Then each noise component at the receiver consists of an uncorrelated part n_a'/ n_b' and a correlated part n_c. From the measurements it is known that

$$\begin{aligned} &n_a = n_a' + n_c, \quad n_b = n_b' + n_c \\ &u^2(n_a) = (470\text{ns})^2 = u^2(n_b) \\ &u^2(n_a - n_b) = (130\text{ns})^2 \end{aligned} \tag{84}$$

As n_a' and n_b' are uncorrelated (84) reduces to

$$\begin{aligned} &u^2(n_a - n_b) = u^2(n_a') + u^2(n_b') = (130\text{ns})^2 \\ &u^2(n_a') = u^2(n_b') = (130\text{ns})^2/2 \rightarrow u(n_a') = u(n_b') = 92\text{ns} \end{aligned} \tag{85}$$

Using (85) the uncertainty due to n_a' and n_b' can be calculated as

$$u^2(n_c) = u^2(n_a) - \frac{(130ns)^2}{2} \rightarrow u(n_c) = 452ns \tag{86}$$

Using the results of Fig 62 and Fig 63 the distance uncertainty is calculated approximately as

$$\begin{aligned} 452ns \cdot 170\frac{\mu m}{\mu s} &= 76.8\mu m\ (correlated) \\ 92ns \cdot 100\frac{\mu m}{\mu s} &= 9\mu m\ (uncorrelated) \end{aligned} \tag{87}$$

Comparing the results of (87) to the uncertainty of 500µm (0.5m) or 250µm (1m), due to systematic errors by unknown object dependence, the noise impact is remarkably low. Therefore, unless the systematic error in the final results is not accounted for, there is little benefit in reducing ToA uncertainty. Finally it shall be noted that for practical applications it is desirable to apply the simplified equations given in (78) as plane type reflectors are more often encountered.

5.4 Practical localization of 3D reflection points

Localization from ToA measurements have already been introduced in 5.2. Despite the simplicity of the proposed formulas herein the problem in practice is far more difficult. The amount of information contained in a single pulse/echo measurement using an ultrasonic sensor is limited by specular reflection and the field of the view of the sensor. The former one cannot be controlled, as it is a property of the environment. For obtaining a maximum of information it is desirable to maximize the field of view of the sensor. Assuming for now that multiple reflectors are present in the environment it is quite likely that overlapping of the echoes at the receiver occurs. This is especially true as only the distance to the sensor matters, e.g. it can be the case that an echo from the floor at 90° incident angle overlaps with an echo from an object at approximately the same distance in front of the sensor. Therefore, the signal processing techniques employed have to be specialized for ensuring good target separation capabilities. Low ToA uncertainty is also required and combination of these two requirements justifies the application of pulse compression techniques, where in this case a linear-frequency-modulated chirp and binary correlation, as introduced 4.4, is used. Still one problem remains – choosing a sensor with good coupling on air and large bandwidth leaves us with a chicken and egg dilemma. Due to the

spatial extent of the transmitter, closely resembling a piston, the sensor exhibits a strong directional characteristic. This in turn leads to reduced correlation outside the mainlobe, as echoes are heavily changed in phase and amplitude, reducing correlation by more than 50% [Wal12b]. This issue was counteracted by using multiple correlation templates which can either be measured or spatially predicted as introduced in 3.4.1.

Considering the previous paragraph practical localization for the sensor presented in Fig 60 can now be described. Let r_i, with $1 \leq i \leq 4$, be the 1-bit quantized received signals for the four microphone channels. Let tp_j, $1 \leq j \leq M$, be a set of spatially predicted or measured templates used within the binary cross-correlation introduced in (74). Within the 4 times M signals peaks are identified. The minimum temporal spacing, allowing proper target separation of two overlapping chirps, is 1.53/B where B is the bandwidth of the chirp signal, as shown in Fig 55(b). A value of 2/B was chosen, corresponding to approx. ~1.1cm distance for a pulse/echo measurement at c = 343m/s. A suitable method for implementation within MATLAB is the findpeaks method which allows specifying both properties - the minimal temporal spacing between adjacent peaks and a lower limit for the amplitude. After this identification process 4 x M sets $T_{ij} = \{(t_{ij1},c_{ij1}), \ldots, (t_{ijNij},c_{ijNij})\}$ are obtained where N_{ij} is the number of peaks found within the microphone channel i using template j. Each element corresponds to a ToA t_{ijk} and an associated correlation peak c_{ijk} where $1 \leq k \leq N_{ij}$. For localization subsets have to be chosen, where a suitable subset is of the form (88), i.e. it contains exactly one ToA for each microphone and it is built from the same template. An example is given in Fig 68 where three reflectors in the field of view of the sensor result in multiple possible subsets.

$$\begin{aligned} &\{(t_{1ji_1}, c_{1ji_1}), (t_{2ji_2}, c_{2ji_2}), (t_{3ji_3}, c_{3ji_3}), (t_{4ji_4}, c_{4ji_4})\}, \\ &1 \leq j \leq M \text{ and } 1 \leq i_1 \leq N_{1j}, 1 \leq i_2 \leq N_{2j}, 1 \leq i_3 \leq N_{3j}, 1 \leq i_4 \leq N_{4j} \end{aligned} \tag{88}$$

While it is obvious, that localization requires exactly one ToF for each microphone, it is not intuitively clear why we require elements within a subset to be built using the same template. Or in other words – should different templates be used which individually maximize the correlation amplitude or should the (small) performance loss in the correlation peak be accepted and only a single template be used. Inspection of the formulas for localization presented within (75) to (78) reveals that even a small error in the time differences can result in a huge error for the calculated direction of the reflector. As measured templates are manually extracted during a calibration process there is an inherent error in the selection of the "cutting window". If ToAs are built using such templates, this error will propagate into the final result, reducing overall accuracy. For measured templates the decision for using only a single template is therefore easily argued. Predicted templates on the other hand do not have this systematic error. Comparing both options in practical measurements it was observed that in some cases the usage of different templates resulted in a time difference error between the ToAs, which in turn resulted in a huge localization error. For this reason usage of different templates was again restricted to different subsets whereas within a single subset all ToAs have to be built using the same template. It shall be noted that the performance loss for correlation is very low (a few percent only – depending on the number of templates). The reason for this is, that due to the compact construction, reflection points associated with different microphones are in close proximity. Therefore, the incident, and in turn reflected wave, is approximately equal.

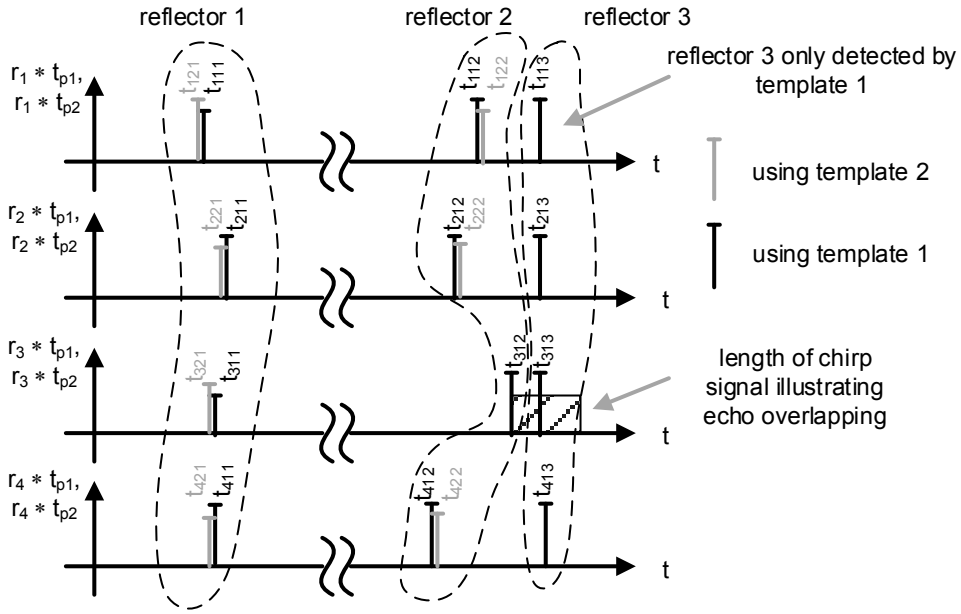


Fig 68: Simple example assuming three reflectors in the field of view of the sensor. Performing binary correlation with two templates yields multiple ToAs for each channel, where the height corresponds to the correlation amplitude. Prior to localization, subsets of ToAs have to be built, where each subset includes exactly four ToAs from each channel and all ToAs correspond to the same template.

Referring back to (88) it is in general not a-priori clear which subsets can be built from the different ToAs. For example the subset (t_{112}, t_{212}, t_{313}, t_{412}) in Fig 68 is not a valid subset as it combines information from reflector 2 and reflector 3. Using such a set of ToAs would result in an incorrectly located point – also called outlier. Furthermore, due to the usage of multiple templates, it can be the case that a reflector is detected more than once. For example reflector 1 and reflector 2 in the example of Fig 68. In this case a proper choice has to be taken. The problem of selecting a proper subset is called *echo correspondence problem* and is within the scope of the next section.

5.5 Echo correspondence problem

Given multiple ToAs the *echo correspondence problem* is defined as choosing a proper subset of these ToAs, where a subset belongs to exactly one reflector and contains exactly one ToA for each microphone channel. For the sake of simplicity we will now write $\{(t_1,c_1),(t_2,c_2),(t_3,c_3),(t_4,c_4)\}$ for such a set and assume it was chosen from a sound combination as described in (88).

Elementary for the selection of such a subset are physical sensor constraints. The time difference between any pair of ToAs cannot be larger than the spacing of the respective microphone pair divided by the speed of sound. Let d_{ij} be the Euclidian distance between the microphone M_i and M_j. Then a subset is valid if and only if

$$\left|t_i - t_j\right| c \leq d_{ij} \quad 1 \leq i,j \leq 4, i \neq j \tag{89}$$

Depending on the number of templates used it might be the case that the same reflector is detected multiple times by different templates. The reason for this is that correlation between neighboring templates is quite high as the change in amplitude/phase of the transmitted signal is not abrupt but more or less continuous. To achieve high accuracy it is essential that a proper choice is taken, as correlation with a wrong template can result in time-shifts in the order of a few wavelengths, i.e. millimeters to centimeters using distances as a unit of measurement. An example is shown in Fig 68 where reflector 1 is identified by two possible subsets $\{t_{111}, t_{211}, t_{311}, t_{411}\}$ and $\{t_{121}, t_{221}, t_{321}, t_{421}\}$. Two arbitrary subsets $\{(t_1,c_1),(t_2,c_2),(t_3,c_3),(t_4,c_4)\}$ and

$\{(t_1',c_1'),(t_2',c_2'),(t_3',c_3'),(t_4',c_4')\}$ are considered to be close if the time differences for each channel are smaller than the separation bound 2/B.

$$|t_i - t_i'| < \frac{2}{B} \text{ for all } 1 \leq i \leq 4 \tag{90}$$

The choice of 2/B, which equals 2/30kHz = 67µs = 2.2cm/c, was chosen as ToAs with a larger spacing could be due to echo overlapping (multiple reflectors, multipath, etc.), which can be separated using the linear frequency modulated chirp and pulse compression. An example is shown in Fig 69(a) where a single reflector was detected by multiple templates. The heights of the dots marked with "x" correspond to the correlation quality, where different colors correspond to the respective microphone channels. Furthermore it can be seen that correlation with different templates would yield different ToAs.

This problem is resolved by the introduction of a combined selection criterion where only the solution which maximizes this criterion is kept for further processing. The selection criterion m_c as defined in (91) is based on the product of all correlation peaks $m_{peaks} = c_1 \cdot c_2 \cdot c_3 \cdot c_4$, the matching to time-difference of arrivals m_{tdoa} (see 5.5.1), the *mid-point criterion* m_{mid} (see 5.3.3), and the *combined correlation quality* m_{comb} (see 5.5.2). Hence the combined criterion is simply the product of all the individual criteria. All individual criteria shall be normalized between 0 and 1 where any potential weighting is applied therein.

$$m_c = m_{peaks} \cdot m_{TDoA} \cdot m_{comb} \cdot m_{mid} \tag{91}$$

Fig 69(b) shows the chosen subset after applying the combined selection criterion m_c. The result is very satisfactory.

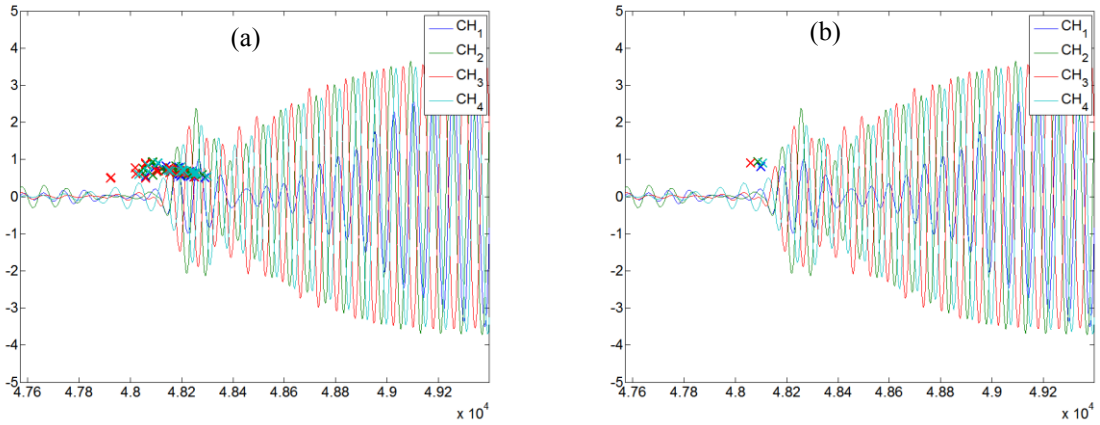


Fig 69: (a) Set of potential *close* echo subsets. Among these subsets a given subset needs to be selected. (b) Chosen subset selected using a criterion calculated from the mid-point criterion, correlation peaks, time-difference matching and combined quality.

5.5.1 Matching to time difference of arrivals (m_{TDoA})

The ToAs within a candidate for a potential subset have been obtained by correlation with numerous measured and/or predicted templates. Still it can occur that the received waveform does not perfectly match with any of these templates, for example due to echo overlapping, common mode disturbances, or due to a distortion of the reflected waveform by the reflector itself. While there is nothing that can be done for the incorrect detection of the starting time of the signal it is still possible to calculate the relative time differences of the echo without the usage of such templates. This is beneficial as any error within the time differences can result in a very large distance error for the calculated reflection point, whereas a common mode error in ranging is

bounded by its own magnitude. The method applied herein is a comparison of the time differences from the initial ToA measurement with the time differences obtained by cross-correlation of a part of the received signals.

Let $\{(t_1,c_1),(t_2,c_2),(t_3,c_3),(t_4,c_4)\}$ be an arbitrary selected subset. Then we define the start time and end times for the signal where we want to calculate the cross correlation according to (92) where O is a small offset in samples and L is the length of chirp in samples. In our case O equals 100 samples and L equals 2000, corresponding to a chirp of a length of 500µs at a sample rate of 4MS/s.

$$\begin{aligned} t_s &= \min(t_1, t_2, t_3, t_4) - O \\ t_e &= \max(t_1, t_2, t_3, t_4) + L + 0 \end{aligned} \tag{92}$$

Calculating the binary cross correlation between the channels returns the relative shifts between the waveforms. Taking channel 1 as reference three time differences, $t_{21'}$, $t_{31'}$, $t_{41'}$ are obtained. Using the ToAs t_1, t_2, t_3 and t_4 determined by correlation with the templates it is also possible to calculate $t_{21}=t_2-t_1$, etc. Based on this an error can be defined according to (93). This process is illustrated in Fig 70.

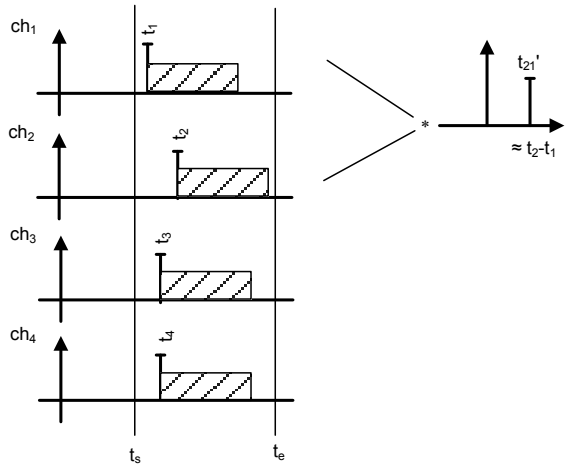


Fig 70: Using a window of the received signal defined by the starting time t_s and the end time t_e the relative shift between the signals can be obtained by cross correlation. These time differences can be compared to the time differences calculated from the ToA measurement. For a valid result it is required that the time differences are approximately equal. Errors are most often due to distorted echoes or incorrect matching with imperfect templates.

A combined criterion m_{TDoA} can be defined according to (94) which ensures that the value ranges between 0 and 1. The factor k_{TDoA} is a value between 0 and 1. A suitable choice is 0.1 in the case of our system which was determined empirically.

$$\mathrm{err}_{\mathrm{TDOA}} = \sum_{i=2}^{4} |t'_{i1} - (t_i - t_1)| \tag{93}$$

$$m_{\mathrm{TDoA}} = e^{-\mathrm{err}_{\mathrm{TDOA}} \cdot k_{\mathrm{TDOA}}} \tag{94}$$

5.5.2 Combined correlation quality

Due to the compact design of the sensor reflection points are in close proximity. Therefore, it is a sound assumption that the received echoes for each channel are very similar. Using this fact a combined correlation quality for the received input signals r_i can be defined according to (95) where L is the length of the chirp in samples. Note that this is an extension of the binary correlation introduced in (70) to multiple channels. Furthermore, due to the non-transitive behavior of correlation, it is also not the same as the product of the individual correlations obtained using

the reference templates, i.e. it is not true that if template tp_1 and received signal r_1 show a correlation peak of c_1, and tp_1 and r_2 show a correlation peak of c_2, that r_1 and r_2 show a correlation peak of $c_1 \cdot c_2$.

$$m_{comb} = \frac{1}{L}\sum_{i=0}^{L-1} \mathrm{AND}(r_1[t_1 + i], r_2[t_2 + i], r_3[t_3 + i], r_4[t_4 + i]) \tag{95}$$

5.5.3 The midpoint criterion

In case of reflectors at close distance, or in case of object discontinuities, the criteria defined above are not sufficient. This is illustrated by two examples. Fig 72 shows two reflectors at approximately the same distance, where it is assumed that the time distances are so small that the basic geometric criterion defined in (90) is not sufficient for resolving the ambiguity. The second example shown in Fig 72 is a reflector with a discontinuity, where one reflection point lies on a different part of the object. Without taking any countermeasures simply calculated reflection points would result in numerous outliers within the data. In this section we will introduce a simple geometric criterion which resolves these issues.

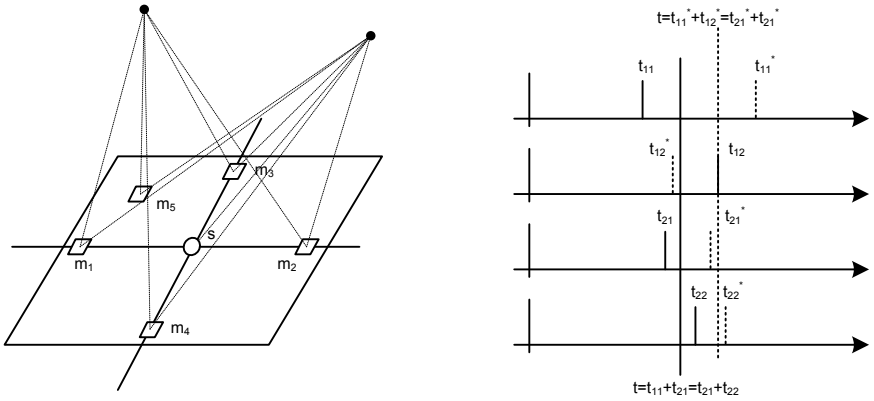


Fig 71: (a) Two reflectors at approximately the same distance results in echoes at close temporal proximity for each microphone. (b) Reflection from objects 1 are shown as vertical solid lines, whereas echoes from object 2 have dashed vertical lines. If separation by physical sensor constraints is not possible, up to eight different combinations of ToAs are possible, where only two combinations yield a valid result.

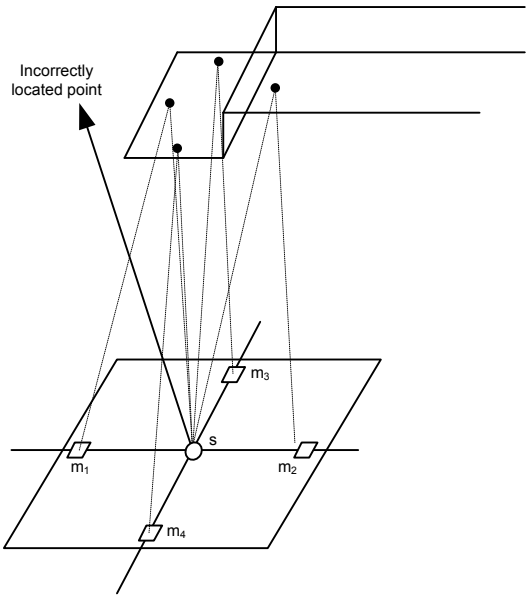


Fig 72: Using the midpoint criterion allows identification of an incorrectly calculated reflection if one pair of ToAs is enlarged by the discontinuity.

It can be shown [Wal15b] that for plane and convex reflectors, reflection points belonging to the same object, are approximately on a plane. Note that this also applies to a sphere, although this virtual plane would intersect with the sphere. In case of the compact 3D sensor presented herein this fact (96) can be used to define a criterion according to (97). It can be interpreted as following: The mean distance for the ToFs for each microphone pair is approximately equal, where a bound of 2mm at a distance of 50cm suffices to account for any object dependent deviations.

$$\frac{d_1 + d_2}{2} \approx \frac{d_3 + d_4}{2}, d_i = t_i \cdot c \tag{96}$$

Based on that the criterion is defined as following

$$k = \left| \frac{d_1 + d_2}{2} - \frac{d_3 + d_4}{2} \right|$$
$$m_{mid} = \begin{cases} 1 & \text{if } k \leq 2mm \\ 1 - - \frac{k - 2mm}{8mm} & \text{if } 2mm \leq k \leq 10mm \\ 0 & \text{otherwise} \end{cases} \tag{97}$$

The values given are appropriate for the sensor with 80mm spacing. The gradual reduction from one to zero accounts for any issues with particular close reflectors which might otherwise be ignored.

Using the criterion defined within (97) Fig 71Fig 72 can be reevaluated. For example, the solution $\{t_1,t_2,t_3',t_4\}$ can be easily discarded as the mean value for t_3' and t_4 does not equal the mean of t_1 and t_2. While there exist cases where this criterion is not sufficient it greatly reduces the number of combinations. It shall be noted that different, alternate, methods are available for solving this task. For example other authors have addressed this problem by applying statistical methods which maximizes the likelihood of observations [Per92].

Application of the midpoint criterion also helps resolving the potential outlier resulting from the scenario in Fig 72. In this case one reflection point was on a different plane disturbing a single ToA. If the disturbance is smaller than the maximal allowed time difference between ToAs due to microphone spacing (89), calculation will yield an incorrect reflection point. The absolute error in Cartesian coordinates is practically unbounded which becomes clear from (78) and (79), as the direction of the calculated reflection point is solely due to time differences. Again this is resolved by this criterion as only the mean of the pair m_1/m_2 is affected where the mean of m_3/m_4 stays the same. A close look reveals a potential issue of this criterion and why it is only required but not sufficient. Due to symmetry there exist cases where both pairs are affected equally and which cannot be detected by this criterion. Resolving of this requires additional redundancy or a second measurement using a different orientation of the sensor.

5.6 Example for 3D echo localization

The performance of the proposed compact 3D sensor is demonstrated by analyzing a complex scene shown in Fig 73, where, without appropriate measures, the resulting data would contain numerous outliers making it extraordinary hard to interpret or use the resulting information. The complexity is mostly due to the large number of reflectors including a cylinder (a), two vertical planes (b) and (c) at approximately the same distance and a tilted plane (d) with a small discontinuity. To analyze the scene the sensor is moved from x = -20cm to x = 50cm performing 61 measurements in total. The expected trace of reflection points due to the law of reflection is shown as a solid blue trace within Fig 73.

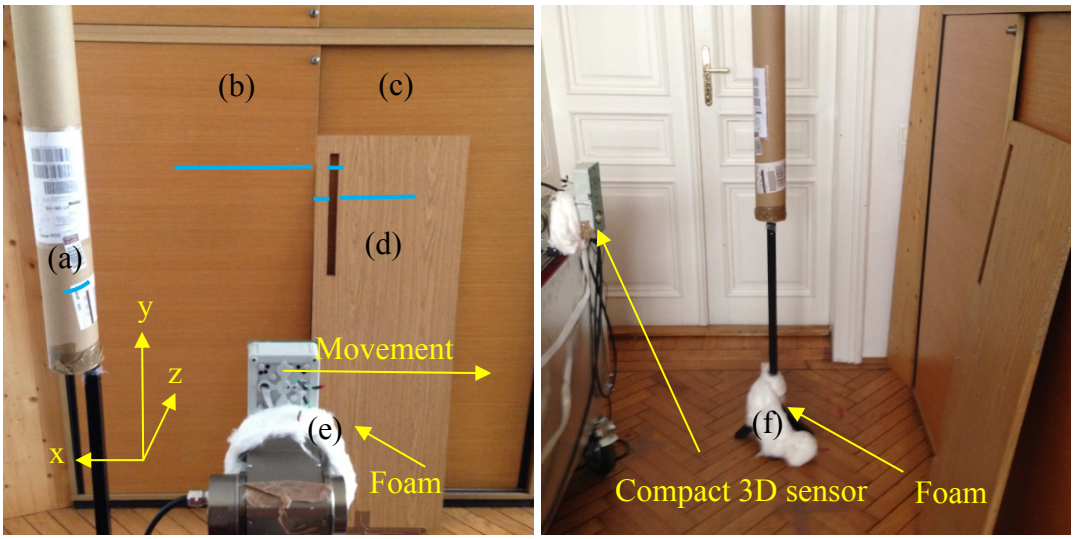


Fig 73: Complex scene including a cylindrical reflector (a), two plane reflectors (b) and (c) parallel to the sensor surface, and a titled plane (d). In addition the tilted plane (d) includes a small hole. During movement of the sensor reflection points move from plane (b) to plane (c), and as these planes are at close proximity, incorrect reflection points would result without further measures. The same applies for the tilted plane (d). A sound absorbing foam was added to avoid reflections rom the back side of the sensor (e) and from the rod (f).

The aim of this example is to show the benefits and the high quality of data obtained by the proposed method. Analyzing the scene without any further measures, i.e. by applying binary correlation using multiple templates and choosing results only based on the correlation maxima returns non satisfactory results. This is shown in Fig 74 where the respective data is marked with the same labels as in Fig 73. Labels including a number are due to multipath propagation where the wave was reflected from the object back to the sensor, from the sensor back to the object and then finally back to the sender. All the other data within Fig 74 could not be identified although a-priori knowledge of the scene was available (as we manually marked the elements in this example). The huge number of outliers can be clearly seen.

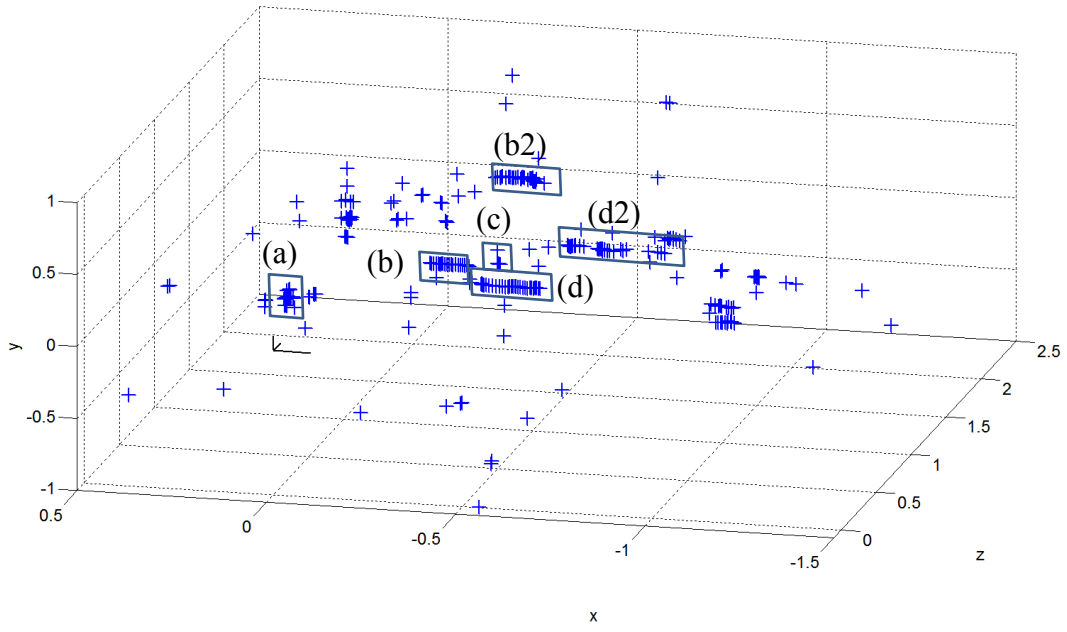


Fig 74: 3D reflection points obtained from 61 measurements where the sensor was moved from x = -0.2m to x = 0.5m. The only applied criterion was that all ToAs should have a peak with correlation amplitude > 0.5. In case of close subsets the subset which maximizes the product of all correlation amplitudes was chosen. Points not enclosed within a solid, labeled rectangle are all either outliers and/or unidentified objects.

A very basic method which is often employed is application of an energy threshold on the received echoes. This is shown in Fig 75 where it can be seen that the performance improvement is marginal – in fact important and correct data, like the cylinder (a), previously present is lost while outliers still exist. Inspection of the corresponding ToA data for the outliers marked as (e) yields an interesting insight to this issue. The ToA data for one such reflection point is shown in Tab. 4 and is a good example for the echo correspondence problem. What happened is that ToAs from different objects, in this case from the planes (b) and (d), have been used for calculation. It shall be noted that neither the basic geometric criterion presented in (89) nor the energy threshold could have resolved that issue.

ToAs	Object (b)	Object (d)	Object (e)
$t_1 \cdot c$	2.0637 [m]	2.0001 [m]	2.0001 [m]
$t_2 \cdot c$	2.0633 [m]	1.9997 [m]	2.0633 [m]
$t_3 \cdot c$	2.0628 [m]	2.0031 [m]	2.0628 [m]
$t_4 \cdot c$	2.0641 [m]	1.9960 [m]	2.0641 [m]
Midpoint error	0.05 [mm]	0.35 [mm]	-31.75 [mm]

Tab. 4: Calculation of a reflection point using the ToA data for object (e) will yield in an incorrect reflection point. Therefore, it should be avoided. As all ToAs are from planes at approximately the same distance the energy of the reflected wave is high in all cases and can, therefore, not be used to resolve this ambiguity.

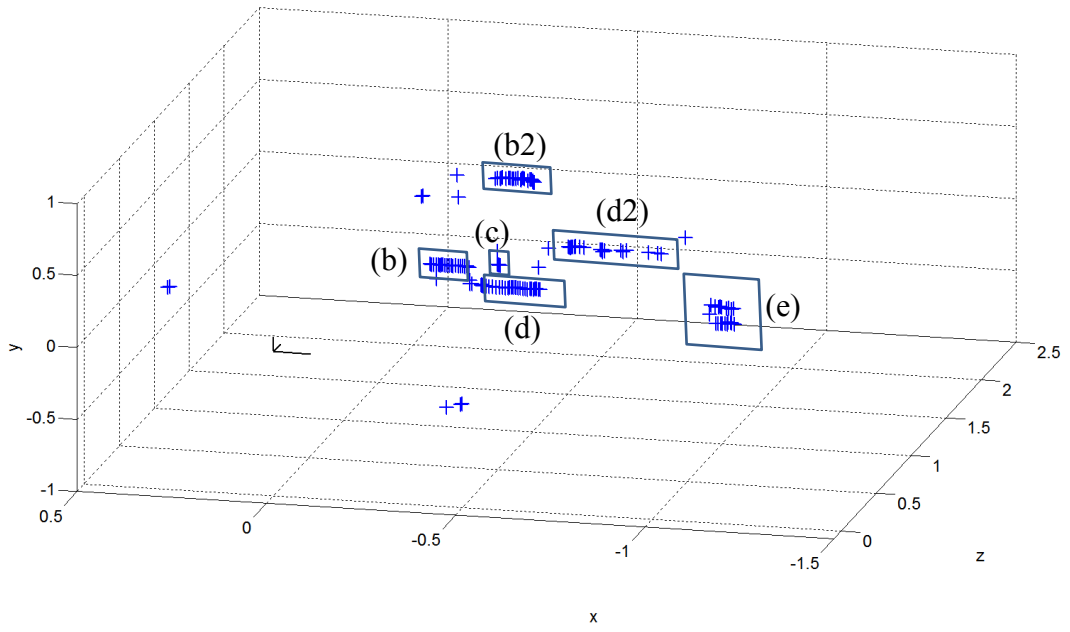


Fig 75: Applying an energy threshold (around 25% of the energy contained in an echo for a plane reflector at 1m) does improve quality but at a high price. First of all important data is lost, like the cylindrical reflector. Furthermore, outliers are still present, for example the group marked (e)

Such types of problems are typical for real world scenarios containing multiple and close reflectors. The additional redundancy introduced by the compact 3D sensor, employing four microphones, can be used to improve reliability of localization. The result after application of the midpoint criterion and by using only a marginal energy limit (0.5% percent of the energy of the reflected wave for a wall at 1m distance) is shown in Fig 76. Most importantly the incorrect group (e) no longer exists while important data is preserved.

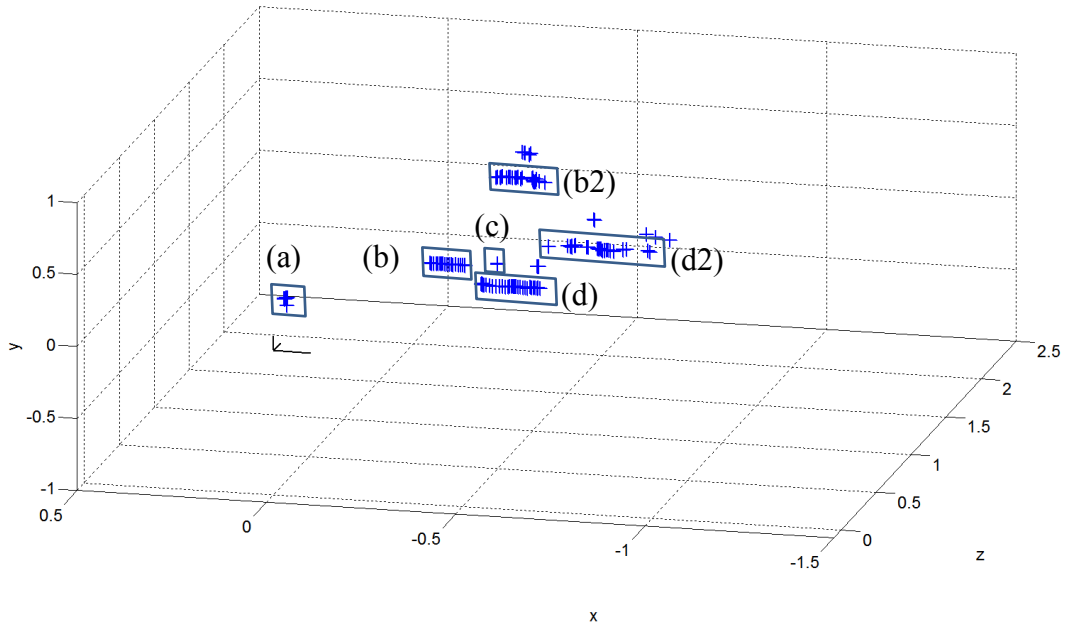


Fig 76: Scene analyzed after application of the midpoint criterion. Outliers are strongly reduced while important information is still present in the analyzed data.

5.7 Benefits of additional receivers

It was already mentioned in 5.5.3, that there exist cases when the simple criterion does not identify a potential issue due to reflection points on different objects. This is shown in Fig 77 where for each pair m_1/m_2 and m_3/m_4 one reflection point is on a different object. If the change in distance for both microphone pairs is equal, the midpoint criterion is still full-filled, while the calculated reflection point is wrong. It is possible to resolve this by an additional receiver, where a calculated reflection point is valid if and only if the length of the calculated position vector is approximately equal to half the ToF times the speed of sound. If a 5^{th} receiver is not present this issue can also be resolved in a second pass by rotation of the sensor.

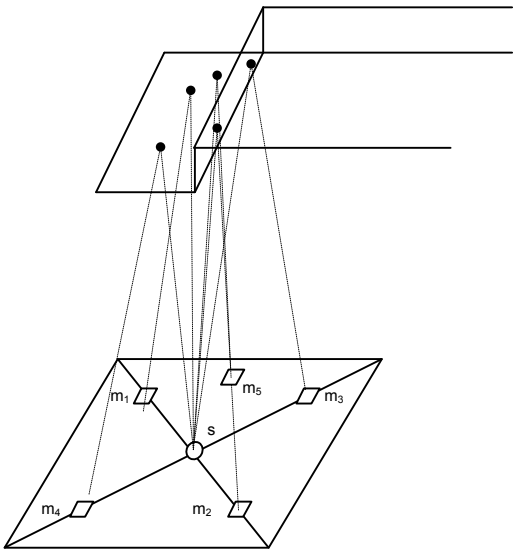


Fig 77: Resolving ambiguity in cases where ToFs for pairs of microphones are affected symmetrically.

The example scene presented in Fig 73 was analyzed again employing a 5th receiver. The results are shown in Fig 78. Not a single outlier is present and no manual post processing was performed on the data.

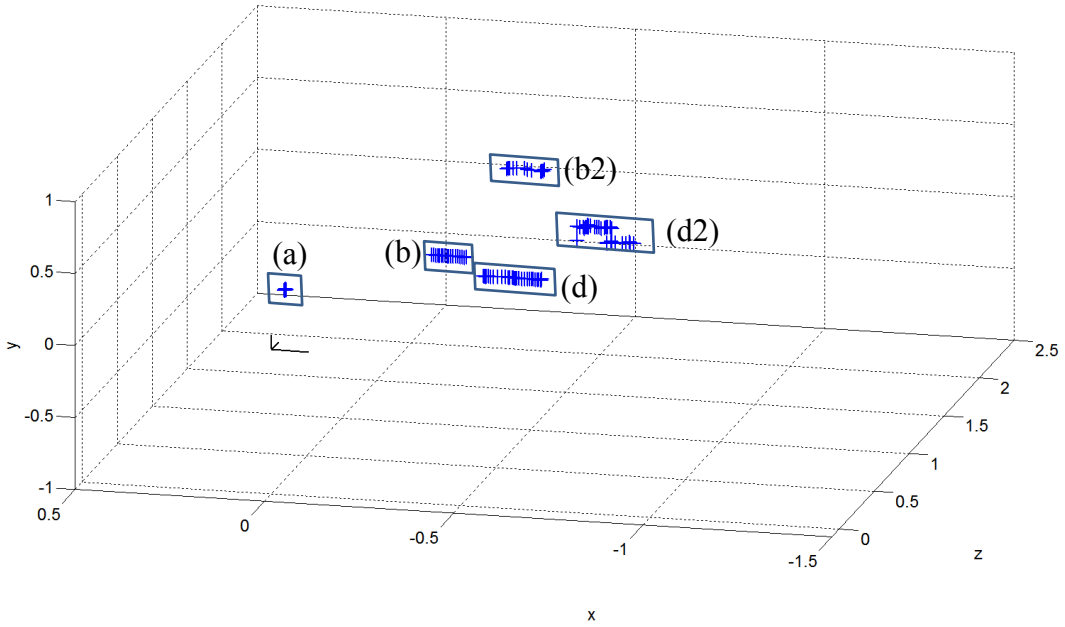


Fig 78: Benefits of using a 5-th receiver for analyzing the scene. Not a single outlier is present while the cylinder and the planes have been correctly identified.

6 Applications

The compact 3D sensor described in 5 is well suited for applications where the parameters to be derived can be obtained by specular reflections on the surface of one or more objects. Two common examples are presented herein where focus is given on the methods solving such an application by using the proposed sensor. Furthermore, performance is verified for each application in a practical test setup. The first application, *scene analysis*, assumes an a-priori unknown environment where such a sensor is used to obtain information about the objects therein. In the second application, *object classification*, unknown parameters of a-priori known type of objects shall be estimated. What will be shown is that the main difference between these two applications is the realization of a required global coordinate system and whether the sensor has to be moved to obtain sufficient information or not. In general applications can be classified using the following three categories.

6.1 Application Categories

Besides of a basic application as a static sensor, which makes only sense in completely well-defined measurement tasks, movements of sensor and/or objects is a requirement for getting a sufficient number of data. Three alternatives are feasible:

- Moving sensor with well-defined absolute knowledge of position and orientation of the sensor coordinate system within a room coordinate system. In this case object parameters can be measured with respect to their absolute position in the environment. *Map building* and *scene analysis* is an example for this type of application.
- Objects are moving and the sensor system remains on its static position. In this case, object parameters can be identified with respect to any static coordinate system, which can also be the coordinate system of the sensor. A typical example of this case is measuring parameters of objects placed on a conveyor, where actual conveyor position, direction and speed of movement are known. This application is called *object classification*.
- Sensor moving relative to objects with known features. An example of this application is using the sensor, mounted on an industrial robot, where the sensor is guiding the robot for reaching precise positions and orientation with respect to a well-known object.

Especially scene analysis is often dealing with recognition of unknown objects. In this case object types are determined by checking whether sections of measurement results originate from different elementary object types, called object hypotheses. Each of these object hypotheses is characterized by a set of parameters, minimally defined by a sufficient number of measurement values. If fitting the hypothesis, theses measurement values are eliminated from further post-processing. Hypotheses are applied in order of increasing complexity of the object types [Sch05].

6.2 Moving Sensor with Global Reference System (scene analysis)

In the application of scene analysis the 3D compact sensor is operated within an a-priori unknown environment. Given a well-defined world coordinate system it is assumed that the location and orientation of the sensor within this coordinate system can be determined. This can be done for example by using a suitable localization system like LOSNUS as described in 7.2, by using odometry, or by employing fix-point navigation. This is illustrated in Fig 79 where a reflection point **p'** is measured within the sensor coordinate system. Using the localization system, the position **r** and orientation **R** of the sensor coordinate system, relative to the world coordinate system, is obtained. It is then possible to calculate the position **p** of the reflector in world coordinates as $(\mathbf{R}^{-1}\mathbf{p'}) - \mathbf{r}$. From now on it shall be assumed that all measurements are given in world coordinates unless otherwise stated.

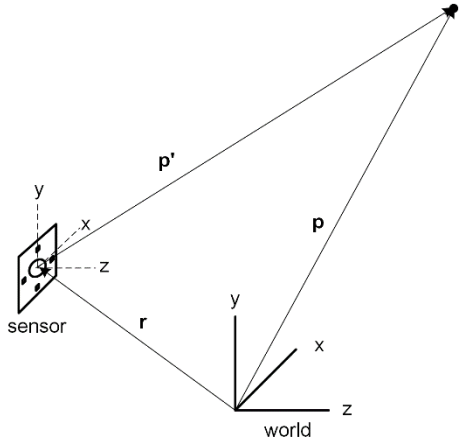


Fig 79: Transformation of a measurement **p'** in sensor coordinates to world coordinates. For this the orientation **R** and position **r** of the sensor has to be known..

Using multiple measurements the environment is analyzed and partitioned into basic geometric objects, where planes and cylinders have been considered within this work. As introduction on this topic an example is given in Fig 80(a), where the result after measurement and post processing is shown in Fig 80(b). The scene consists of two plane reflectors (A) and (C) and a cylinder (B). For the planes four parameters, the distance and normal vector to the origin, are determined. For the cylinder five parameters are necessary including radius and the main axis of the cylinder.

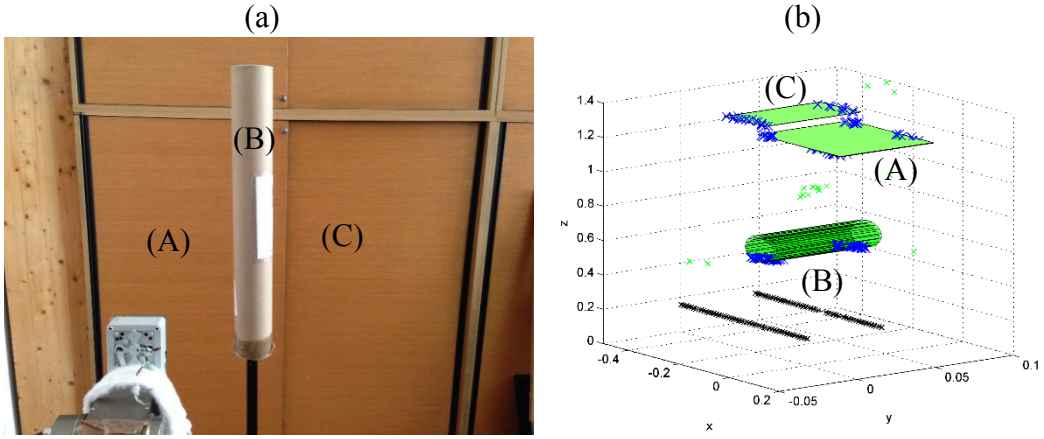


Fig 80: (a) Example scenery including two plane reflectors (A) and (C) and a cylindrical reflector (B). (b) Analyzed scene where identified objects are shown in solid green color. Measurement data is shown as blue crosses.

The following sections will explain in rigorous detail how planes (6.2.1) and cylinders (6.2.2) can be determined if only a set of 3D reflection points obtained by the compact 3D sensor is given. Using these two methods, section (6.2.3) applies them to the example of Fig 80(a), also covering practical considerations for scene analysis. The last part 6.2.4 focuses on some special issues which include multipath propagation and bending of acoustic waves around objects.

6.2.1 Plane reflectors

A plane can be fully described in its Hessian normal form by a direction vector **n** and the distance **a** to the origin [Gel89]. An example is given in Fig 81 where the sensor initially measures **p'** in its local coordinate system. Application of the 3D law of reflection (14), and by assuming a compact 3D sensor, it follows that the direction of the reflected wave and the surface normal **n** are equal. Therefore, using the reflection point **p** and the position of the sensor **r**, the surface normal **n** can be calculated according to (98).

$$\mathbf{n} = \mathbf{r} - \mathbf{p} \tag{98}$$

It can be shown that for any point **p** on a plane (99) holds [Gel89]. As **n** was already determined using (98), and **p** was assumed to lie on a plane, a single measurement is sufficient to determine all parameters. Despite the simplicity of the method described above, application in this form to a real world scenario is not yet possible. Firstly, any uncertainty in the measurements is neglected. Secondly, reflection points are assumed to lie on a plane, i.e. a-priori knowledge of the environment was required, which has to be avoided.

$$\mathbf{n} \cdot \mathbf{p} = -\mathrm{a} \tag{99}$$

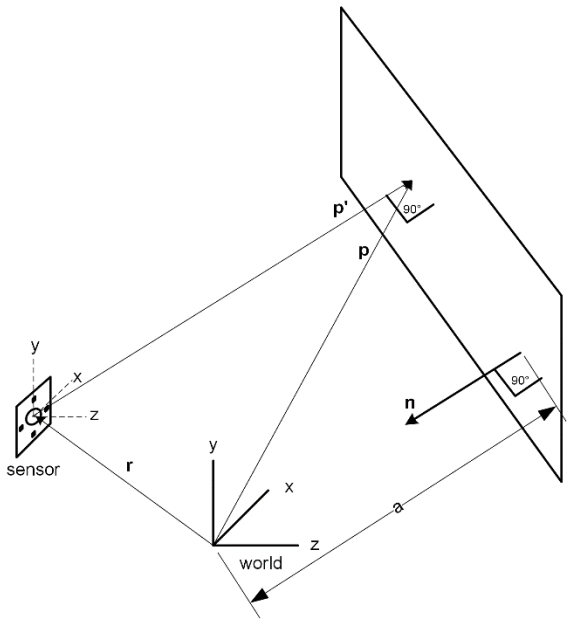


Fig 81: Plane reflector with normal vector **n** at distance a from the origin. Using a single measurement only the distance a of the plane to the origin and its surface normal can be estimated.

To reduce uncertainty of the estimated parameters it is beneficial to apply averaging. If N measurements are available an unbiased estimator for the parameters is given in (101), where we will from now on write $\mathbf{d_i}$ for the surface normal at the reflection point $\mathbf{p_i}$ to avoid clutter in the presentation. This vector $\mathbf{d_i}$ will be called *direction vector* in future as it also corresponds to the direction of the reflected wave.

$$\mathbf{d_i} = \frac{(\mathbf{r_i} - \mathbf{p_i})}{\|\mathbf{r_i} - \mathbf{p_i}\|} \tag{100}$$

$$\mathbf{n} \approx \frac{1}{N}\sum_{i=1}^{N}\mathbf{d_i} \quad a \approx -\frac{1}{N}\sum_{i=1}^{N}\mathbf{p_i}\cdot\mathbf{n} \tag{101}$$

In the next step the requirement for a single plane reflector is dropped. Let $\mathbf{p_i}$ and $\mathbf{r_i}$ be a set of measurements where $1 \le i \le N$. First, the direction vector $\mathbf{d_i}$ is calculated for all measurements. These direction vectors are grouped into M different groups $G_k = \{\mathbf{d_{k1}}, \mathbf{d_{k2}}, \dots, \mathbf{d_{kL_k}}\}$ where L_k is the number of elements within the group G_k, $1 \le i \le M$. The set of all groups $\{G_1,\dots,G_M\}$ is the smallest set, where for each group the condition $\cos^{-1}(\mathbf{d_{ki}}\cdot\mathbf{d_{kj}}) < C_d$ holds, for $1 \le i,j \le L_k$. C_d is a limiting angle and depends on the uncertainty of the direction estimation. Suitable values are in the range of 1°. A group G_k can, therefore, be interpreted as following: All elements within a group G_k have a similar surface normal vector, but are not necessarily elements of a plane or from a plane at the same distance.

For each group an average normal vector $\mathbf{n_k}$ can be calculated similar to (101). By projection of the corresponding reflection points the groups $A_k = \{a_{k1}, a_{k2}, \dots, a_{kL_k}\}$ are be calculated. Using these groups A_k, again the minimal number of groups is identified, where two elements belong to the same group if and only if $|a_{ki}-a_{kj}| < C_0$. A suitable distance bound C_0 is in the range of a few millimeters. A group is considered to be a plane if such a group contains a minimal number of elements, and, in addition, if the area occupied by the elements within the group is not zero. Such elements are removed from further post processing as the elements are most likely on a plane.

6.2.2 Cylindrical reflectors

An infinite cylinder in 3D can be uniquely described by five parameters. The main axis of the cylinder has four degree of freedom. This can be argued as following: Without loss of generality we choose two coordinate planes, e.g. XY and XZ. Two points on these planes are sufficient for defining a line in 3D. The fifth parameter is the radius of the cylinder. For presentation purposes we will use a simplified representation of the cylinder, keeping in mind that it is always possible to reduce the number of parameters to five. Such a unique and minimal representation is required for checking if two solutions are equal.

A more intuitive representation is show in Fig 82. The cylinder is described by a direction vector **n**, which defines a plane through the origin and is parallel to the main axis of the cylinder.

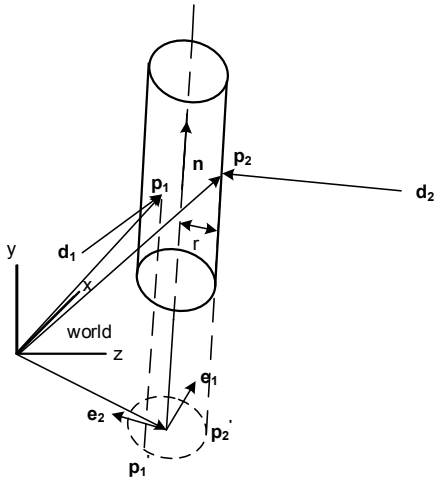


Fig 82: A cylinder in 3D described by a direction vector **n**. Using the direction vector **n** a plane through the origin as well as two unit vectors $\mathbf{e_1}$ and $\mathbf{e_2}$ can be defined. The projection of any point on the cylinder surface onto that plane lies on a circle. This fact can be exploited to verify if the object in question is a cylinder and, furthermore, can also be used to estimate the radius r and the center **c**. Estimation of the direction vector is possible from two measurements only as the cross product of $\mathbf{d_1}$ and $\mathbf{d_2}$ always points into the direction of the cylinder.

Using $\mathbf{n}$ we can define two perpendicular unit vectors $\mathbf{e_1}$ and $\mathbf{e_2}$, which together with the vector $\mathbf{n}$, define a system of coordinates. The benefit of this coordinate system is that any projection of a point on the cylinder surface is on a circle, which in turn can be used for estimation of the center and the radius. Furthermore, using the fact that the direction of the reflected wave corresponds to the surface normal on the object, reveals that the vector $\mathbf{n}$ can be estimated by the cross product of $\mathbf{d_1}$ and $\mathbf{d_2}$, i.e. from only two measurements on the surface of the cylinder.

This already outlines the main idea behind the algorithm. Using N measurements $\mathbf{p_1},\ldots,\mathbf{p_N}$ the corresponding direction vectors $\mathbf{d_1},\ldots,\mathbf{d_N}$ are obtained. Using $\mathbf{d_1},\ldots,\mathbf{d_N}$ an estimate for the direction vector $\mathbf{n}$ of the main axis of the cylinder is computed by application of (102). Enforcing a positive z coordinate is required to avoid ambiguity, which would otherwise occur if the order of the vectors within the cross product would be exchanged.

$$\mathbf{n_{ij}} = \begin{cases} \dfrac{\mathbf{d_i}\times\mathbf{d_j}}{\|\mathbf{d_i}\times\mathbf{d_j}\|} & \text{if the z component is positive} \\ \dfrac{-\mathbf{d_i}\times\mathbf{d_j}}{\|\mathbf{d_i}\times\mathbf{d_j}\|} & \text{otherwise} \end{cases} \qquad (102)$$

$$\mathbf{n} = (n_x \quad n_y \quad n_z)^T = \binom{N}{2}^{-1} \sum \mathbf{n_{ij}}$$

The unit vectors $\mathbf{e_1}$ and $\mathbf{e_2}$, which both lie on the plane through the origin are calculated according to (103). By construction both vectors are perpendicular to each other and to the normal vector $\mathbf{n}$.

$$\mathbf{e_{11}} = (n_y \quad -n_x \quad 0)^T \quad \mathbf{e_{12}} = (0 \quad n_z \quad -n_y)^T$$
$$\mathbf{e_1} = \frac{\mathbf{e_{11}} + \mathbf{e_{12}}}{\|\mathbf{e_{11}} + \mathbf{e_{12}}\|} \quad \mathbf{e_2} = \mathbf{n}\times\mathbf{e_1} \qquad (103)$$

Using the vectors $\mathbf{e_1}$ and $\mathbf{e_2}$ the projection of every point $\mathbf{p_1},\ldots,\mathbf{p_N}$ on the surface of the cylinder is calculated using (104). This equals a reduction to 2D, as a point on a plane only has two degrees of freedom.

$$\mathbf{pp_i} = (\mathbf{ppx_i}, \mathbf{ppy_i})^T = (\mathbf{p_i}\cdot\mathbf{e_1}, \mathbf{p_i}\cdot\mathbf{e_2})^T \qquad (104)$$

Similar to (105) the projections for the respective sensor positions are calculated as

$$\mathbf{rp_i} = (\mathbf{rpx_i}, \mathbf{rpy_i})^T = (\mathbf{r_i}\cdot\mathbf{e_1}, \mathbf{r_i}\cdot\mathbf{e_2})^T \qquad (105)$$

An example showing the estimation of the radius and center is given in Fig 83. Each projected pair of sensor position $\mathbf{rp_i}$ and reflection point $\mathbf{pp_i}$ together define a line in 2D space. It is then possible to calculate the intersection point $\mathbf{c_{ij}}$ of these lines according to (106).

$$cx_{ij} = \frac{(ppx_i * rpy_i - ppy_i * rpx_i)\cdot(ppx_j - rpx_j) - (ppx_i - rpx_i)\cdot(ppx_j * rpy_j - ppy_j * rpx_j)}{(ppx_i - rpx_i)\cdot(ppy_j - rpy_j) - (ppy_i - rpy_i)\cdot(ppx_j - rpx_j)}$$
$$cy_{ij} = \frac{(ppx_i * rpy_i - ppy_i * rpx_i)\cdot(ppy_j - rpy_j) - (ppy_i - rpy_i)\cdot(ppx_j * rpy_j - ppy_j * rrx_j)}{(ppx_i - rpx_i)\cdot(ppy_j - rpy_j) - (ppy_i - rpy_i)\cdot(ppx_j - rrx_j)} \qquad (106)$$

Using $\mathbf{c_{ij}}$ the center $\mathbf{c}$ can be estimated from the average according to (107). The individual radii r_i are calculated as $r_i = \|\mathbf{pp_i} - \mathbf{c}\|$, where again for the final estimate the average is used.

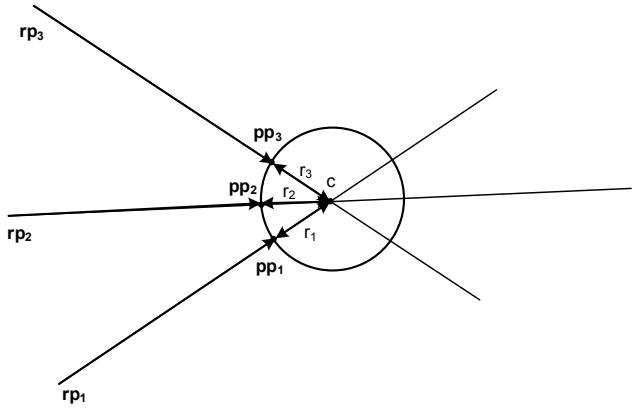


Fig 83: Projected robot positions $\mathbf{r_{pi}}$ and reflection points $\mathbf{p_{pi}}$. The lines defined by the points $\mathbf{p_{pi}}/\mathbf{r_{pi}}$ all intersect at the center **c** of the circle. The center is estimated by calculating the intersection and by averaging. Once the center **c** is known the radius can easily be estimated by the average distance of $\mathbf{pp_i}$ - **c**.

$$\mathbf{c} = \binom{N}{2}^{-1} \sum \mathbf{c_{ij}} \tag{107}$$

$$r = \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{pp_i} - \mathbf{c}\| \tag{108}$$

Using the calculated values of **c**, r and **n** the 5-parameter representation of the cylinder can easily be obtained as well.

While above description is sufficient for a single cylinder and the assumption that all points are on that cylinder, it cannot be used in this form in practice. Investigation of (102) reveals that any uncertainty in the measurements is greatly amplified if the enclosed angle between the two direction vectors is small. This is intuitively obvious as otherwise we could estimate the cylinder from multiple measurements at the same position. Therefore, for any measurement used within the calculation, we require that (109) holds, where C_1 has the meaning of an angle. A suitable choice is 15°.

$$\|\mathbf{d_i} \times \mathbf{d_j}\| = \|\mathbf{n_{ij}}\| > \sin(C_1) \tag{109}$$

As in general multiple cylinders can be present in the environment, the fact that the cross products of direction vectors from the same cylinder point into the direction of the cylinder axis is used. Using this fact M groups G_k, $1 \leq k \leq M$ of size M_k are built according to (110) - or in other words – if the dot product of two vectors is close to 1 then both vectors are approximately parallel. In the context of the direction vectors it has the meaning of pointing into the same direction. The constant C_2 again has the meaning of an angle and 5° was used within this work.

$$\begin{gathered} \mathbf{G_k} = \left\{\mathbf{n_{i_1 j_1}}, \ldots, \mathbf{n_{i_{M_k} j_{M_k}}}\right\} \text{ where for all } \mathbf{n_{i_a j_a}}, \mathbf{n_{i_b j_b}}, 1 \leq a, b \leq M_k \\ \mathbf{n_{i_a j_a}} \cdot \mathbf{n_{i_b j_b}} > \cos(C_2) \end{gathered} \tag{110}$$

For each such group the mean normal vector **n** is calculated identically to the second part of (102). Using the definition of the normal vectors (103) the projection for each reflection point and sensor position is calculated. For example for $\mathbf{n_{i1j1}}$ the projection $\mathbf{pp_{i1}}/\mathbf{rp_{i1}}$ as well as $\mathbf{pp_{j1}}/\mathbf{rp_{j1}}$. In the next step, by application of (106), intersections $\mathbf{c_{ij}}$ are calculated, where in addition, criterion (111) is applied. While this criterion is not required if only a single cylinder is present, it is useful if some other points are present as well. It basically ensures that the intersection point, i.e. a point on the center axis of the cylinder, must be further away from the sensor as the reflection point on the surface of the cylinder.

$$\|\mathbf{pp_i} - \mathbf{rp_i}\| < \|\mathbf{c_{ij}} - \mathbf{rp_i}\| \text{ and } \|\mathbf{pp_j} - \mathbf{rp_j}\| < \|\mathbf{c_{ij}} - \mathbf{r_{pj}}\| \quad (111)$$

As we have dropped the requirement that only a single cylinder is present, the center and radius cannot be directly calculated from the 2D image. By rasterizing the 2D plane with suitable resolution, e.g. 1cm, a histogram can be built. In case of a cylinder multiple points will accumulate at the same position, the center, whereas other objects do not. Next a threshold is applied for the number of elements within a bin and suitable candidates for estimating the center of a cylinder are obtained. The center is then calculated by averaging. Again the radius is estimated by the average distance of the reflection point $\mathbf{rp_i}$ to that center point. The calculation is similar to (107) and (108) with the exception that points are selected according to the procedure described above.

6.2.3 Example for scene analysis

Measurements have been performed in an office room with a size of 500x300x330cm as shown in Fig 84. A cylinder (b) with a diameter of 10.5cm was mounted on a rod and placed in front of a cabinet. For the experiment the sensor (four microphone configuration) was mounted on a linear belt (Schunk PowerCube™ PLB 070) which allows movement along the x-axis. 60 measurements at two vertical positions along the x-axis have been obtained as indicated by the two arrows in Fig 84. What is also special in this test setup is, that the two doors (a) and (c) of the cabinet are not at the same depth, although at close proximity. Therefore, the problem encountered herein therefore is similar to Fig 72 where reflection points are on different objects, possibly resulting in outliers.

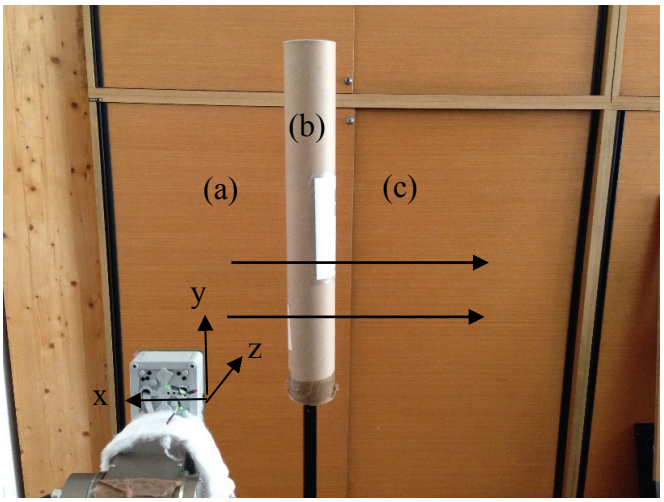


Fig 84: Measurement of the scene containing two planes (a) and (c) and a cylinder (b). Two traces of measurement points at two vertical positions were obtained where each traces consists of 60 measurements in total.

In the first step 3D reflection points are obtained within the local sensor coordinate system and subsequently transformed into the world coordinate system. The result is shown in Fig 85, where the high quality of the data can be observed. Reflection points of all objects in Fig 84 are present, whereas only a small number of outlier exist. Outliers, within this context, correspond to detected echoes which cannot directly be attributed to a physical reflector. The two black traces at the bottom part of Fig 84 correspond to the sensor movement. Therefore, all objects are observed from two different vertical positions. Mirror objects are present as well, which are due to multipath propagation. Objects (a2) and (c2) exist because the reflected wave from the wall is reflected once more by the flat sensor surface and subsequently by the wall itself. Therefore, the calculated distance corresponds approximately to twice the distance from the respective objects (a) and (c). The object (d) is more complicated and exists due to reflection on the back-side of the cylinder. As the reflection occurs on a different location it was decided to label it (d) instead of (b2).

Using this data the algorithms described within 6.2.1 and 6.2.2 are applied. No manual post processing has to be performed and the required time for analyzing the full scene is approx. 15 seconds. The result is shown in Fig 86. For presentation purposes the objects (a2), (c2) and (d) are not shown in the final result although they have been detected as well. The respective parameters of the planes (a) and (c) are given in Tab. 5. It shall be noted that the distance (b2) and (c2) are not exactly twice the distance of (b) and (c) due to the choice of the world coordinate system and the fact that the sensor was mounted 0.15m off the X/Y plane, i.e. at z = 0.15m. Parameters for the cylinder are given in Tab. 6.

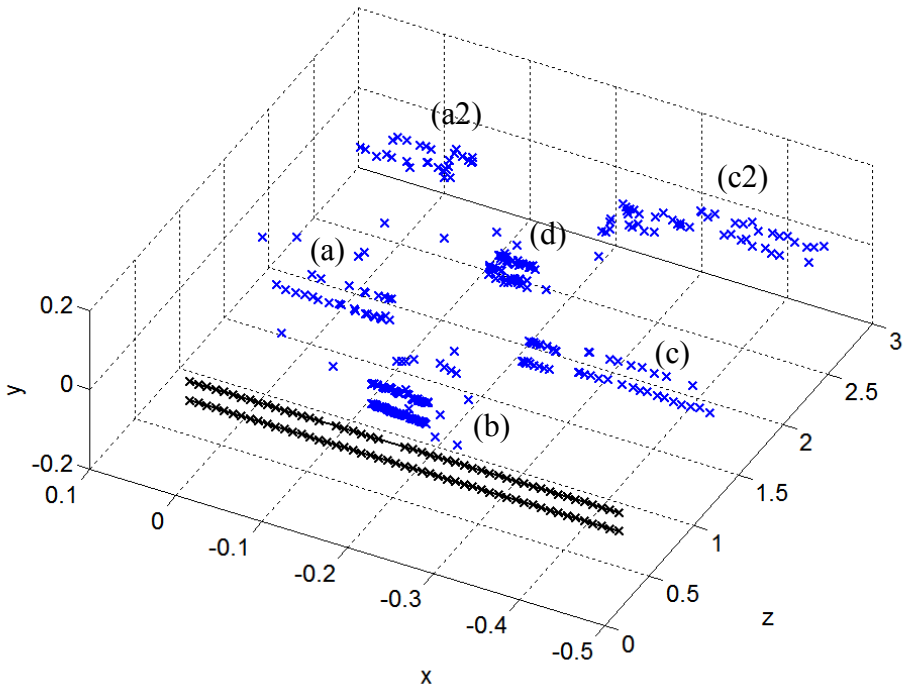


Fig 85: The compact 3D sensor was moved along the solid black traces where for each trace 60 measurements have been taken. The cylinder (b) and both planes (a) and (c) are both present, where on the other side only a very small amount of outliers exist. Items marked with a number are due to multipath propagation. The used settings were 50% for individual correlation quality and a mid-point error limit of 2mm.

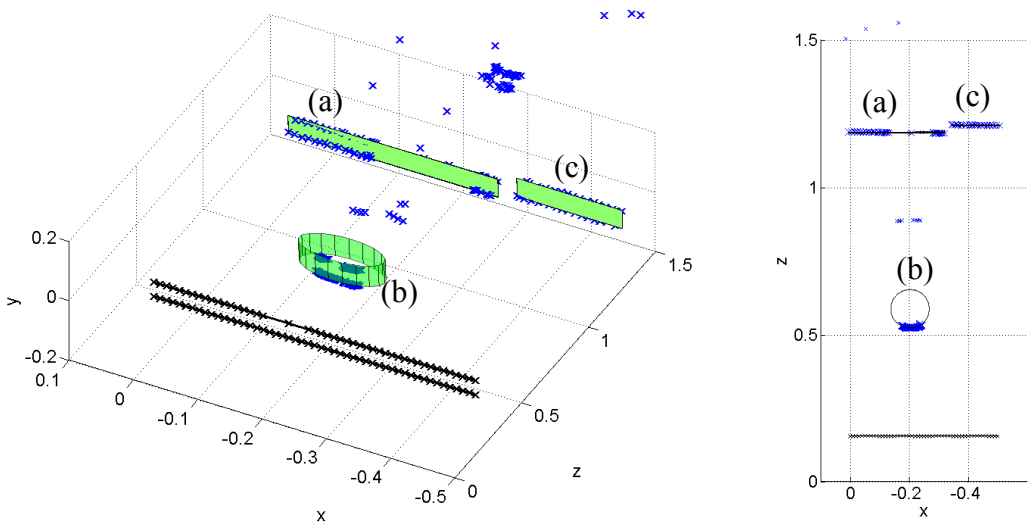


Fig 86: Analyzed scene showing the two planes (a) and (c) and the cylinder (b).

Normal vector	Distance	# Elements	Label
$(0.003,0.027,1.000)^T$ [m]	1.185 [m]	48	(a)
$(0.003,0.027,1.000)^T$ [m]	1.209 [m]	26	(c)
$(0.003,0.043,0.999)^T$ [m]	2.217 [m]	46	(a2)
$(0.000,0.066,0.998)^T$ [m]	2.264 [m]	29	(c2)

Tab. 5: Planes described in Hessian normal form by a normal vector and the distance to the origin. The number of elements used within the identification is show in the third column.

Direction	Offset	Radius	# Elements	Label
$(0.003,0.043,0.999)^T$ [m]	$(-0.203,0.017,0.587)^T$ [m]	6.6 [cm]	85	(b)

Tab. 6: Cylinder described by an offset from the origin and a direction vector.

The remainder of this section explains the application of the algorithms described within 6.2.1 and 6.2.2 for the identification of planes and cylinders. In this example multiple planes are present with similar direction vectors and are, therefore, within the same group. By application of (101) the distances of the planes to the origin are calculated for all elements within the group. Fig 87 shows the histogram of this data where the planes (a), (c), (a2), and (c2) are readily identified. In addition some elements from the cylinder surface (b) are also present. The reason for this is that if the sensor is in front of the cylinder, it can (not yet) be distinguished if the data belongs to a plane. Finally, the surface area spawned by the elements for each plane is calculated and, if above a certain threshold, the data is removed from further analysis. Two important aspects shall be mentioned herein. First, while the resolution of the histogram is finite, the calculated distance of the plane is not limited in accuracy because the distance is calculated from the average of the elements belonging to this bin. Secondly, usage of a threshold for the surface area is recommended because it removes planes from further processing while it does not remove important data like (b). In case of (b) the surface area is very small as the direction vector is only parallel to the normal vector of a plane at a specific position.

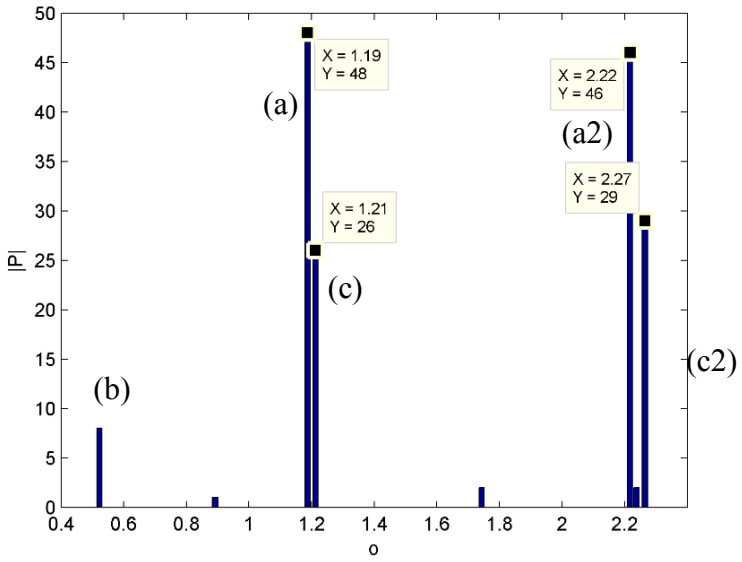


Fig 87: Histogram showing the distances to the origin for all elements having a direction vector similar to $(0, 0, 1)^T$ [m], i.e. corresponding to the planes (a) and (c). The plane normal and distance are accurately calculated from the elements within a histogram, therefore, neither resolution nor accuracy is limited by this process. The histogram is more or less only an efficient and quick method for grouping the data.

After identification of the planes the leftover points are used within the cylinder identification. In the first step the cross products of the remaining direction vectors are calculated, which in turn are used for estimation of the main axis of the cylinder. Using this calculated vector **n** a plane through the origin is defined and by application of (101), (102) and (103) the respective projections of the sensor positions $\mathbf{rp}_i$ and the reflection points $\mathbf{pp}_i$ are calculated. This is shown in Fig 88 where the blue crosses correspond to the sensor position, the red crosses to reflection points on the surface of the cylinder, and the green dots to the calculated intersections. It shall be noted that some other data, not necessarily belonging to the cylinder, might also be present as the cross product of their respective direction vectors could be parallel to the cylinder axis, for example the elements towards the lower right within Fig 88. Using the fact that only for a cylinder the intersection points accumulate, this ambiguity can be resolved by applying a suitable threshold within a 2D histogram as shown in Fig 89. The threshold is chosen based on the average noise floor and on a suitable high safety margin.

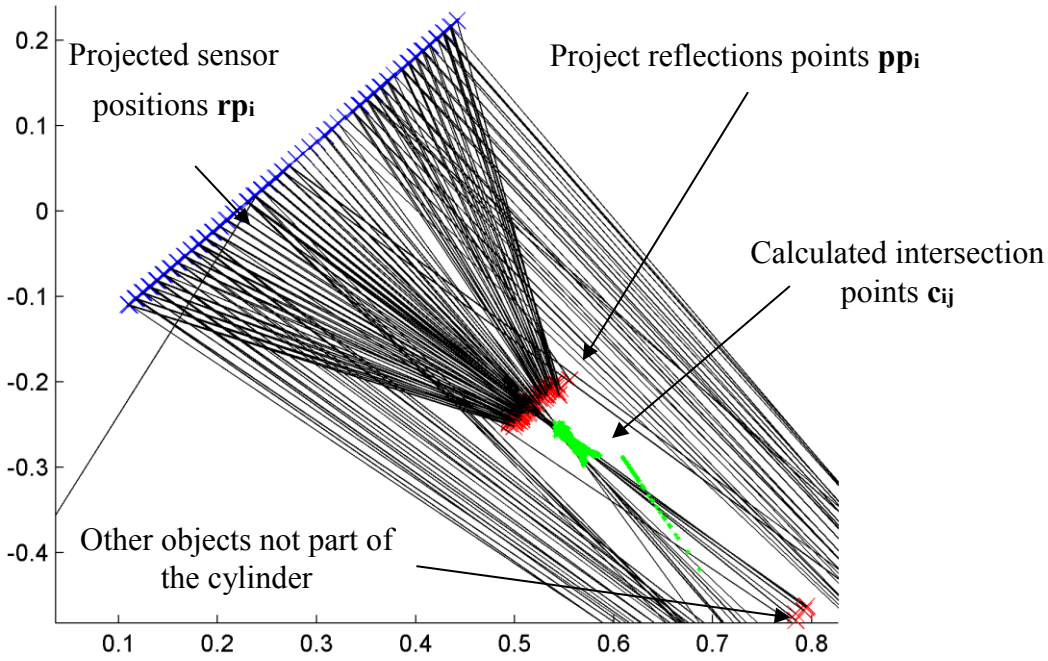


Fig 88: Projected reflection points $\mathbf{pp}_i$ and sensor positions $\mathbf{rp}_i$. The 2D plane used for projection is defined by a plane through the origin with a normal vector identical to the main axis of the cylinder. The main axis of the cylinder is calculated from the cross products of the direction vectors $\mathbf{d}_i$.

Once the candidates have been identified the center of the cylinder is calculated by averaging, whereas the radius is calculated as the average from the distances $\mathbf{pp}_i$ to the center c. Finally, using the projection vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ the result is transformed back into the world coordinate system completing the identification. It shall be noted that these image processing algorithms are only suitable as all reflection points are specular and, therefore, contain additional information about the object. Using diffuse reflection the proposed algorithms would not work. This is also the reason why existing algorithms for image processing are not easily applied or are too complex.

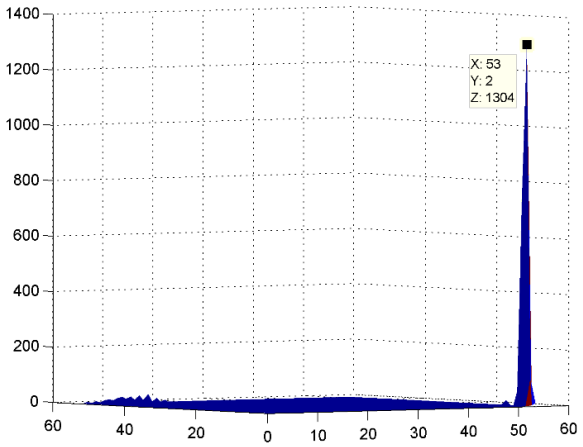


Fig 89: Histogram calculated from the data of Fig 88. At the center of the cylinder multiple points accumulate. These points are then used for estimation of the radius and the center of the cylinder.

6.2.4 Practical Problems

6.2.4.1 Diffraction of Sound-Waves

As simple as the scene presented in Fig 84 might seem it contains some interesting problems worth discussion. As introduced in 0 sound waves are not only reflected by objects but also can bend "around" objects. This phenomenon is called diffraction and allows sound waves to propagate into the shadow region. This can result in echoes which normally not exist assuming straight wave propagation only. While the amplitude can be calculated by application of the formulas presented in 0 this is, in general, not required for our purposes. The sensor, by design, ignores any amplitude information in the echo with the exception of an integral energy indicator. The important factor is the total time of flight. Let q be any point in the shadow region of an object and **p** be the position of a sensor. The total time of flight from **p** to **q** can then be estimated by using a string pulled taut and dividing the length by the speed of sound. This is illustrated in Fig 90.

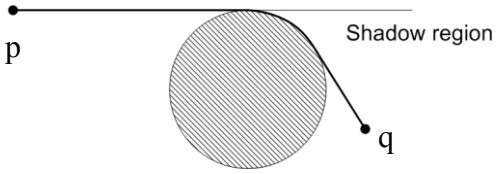


Fig 90: A plane wave propagates from left to right through point **p** towards the point **q** in the shadow region of a cylinder. The time of flight can be calculated by assuming a string pulled taut and dividing the total length by the speed of sound. This is identical to Fermat's principle of least time.

A test setup was developed for investigating this effect on the localization accuracy of the compact 3D sensor. It is shown in Fig 91(a) where a cylinder of 84mm diameter was placed in front of a wall. The sensor was mounted on a linear belt (Schunk PowerCube™ PLB 070). Movement was performed in steps of 1mm starting from x=0mm to x=300mm yielding 301 measurements in total. The axis of movement is approximately parallel to the door of the cabinet. Therefore, if the sensor is left or right to the cylinder, reflection points on the wall closely follow the sensor movement. When the sensor is in front of the cylinder the direct path to the wall is blocked for some microphones. Therefore, some microphones no longer directly get a wall echo but instead the diffracted and reflected wave. This artificially enlarges the time of flight of some ToAs and might lead to an incorrectly calculated direction of the reflection point. This is illustrated in Fig 91(b).

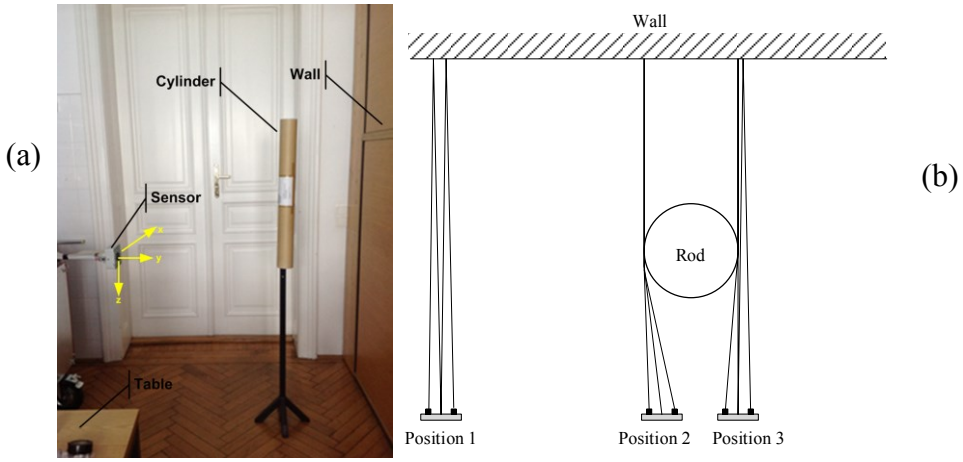


Fig 91: (a) Example scene for verifying the impact of diffraction. (b) Depending on the position of the sensor none, some or all ToAs are affected by diffraction.

The respective ToAs for the measurement are shown in Fig 92 (a) for the microphones m_1 and m_2. While the sensor moves from left to right first the ToA for microphone m_1 is enlarged by at most 5mm, corresponding to approx. 15µs, followed by m_2. Fig 92 (b) shows the energy of the reflected wave in arbitrary units. In case of diffraction the energy decreases extremely quickly due to the short wavelength. The reader is referred to Fig 17(a) to Fig 17(c) for comparison. Despite that the pulse compression techniques correctly identify the ToAs.

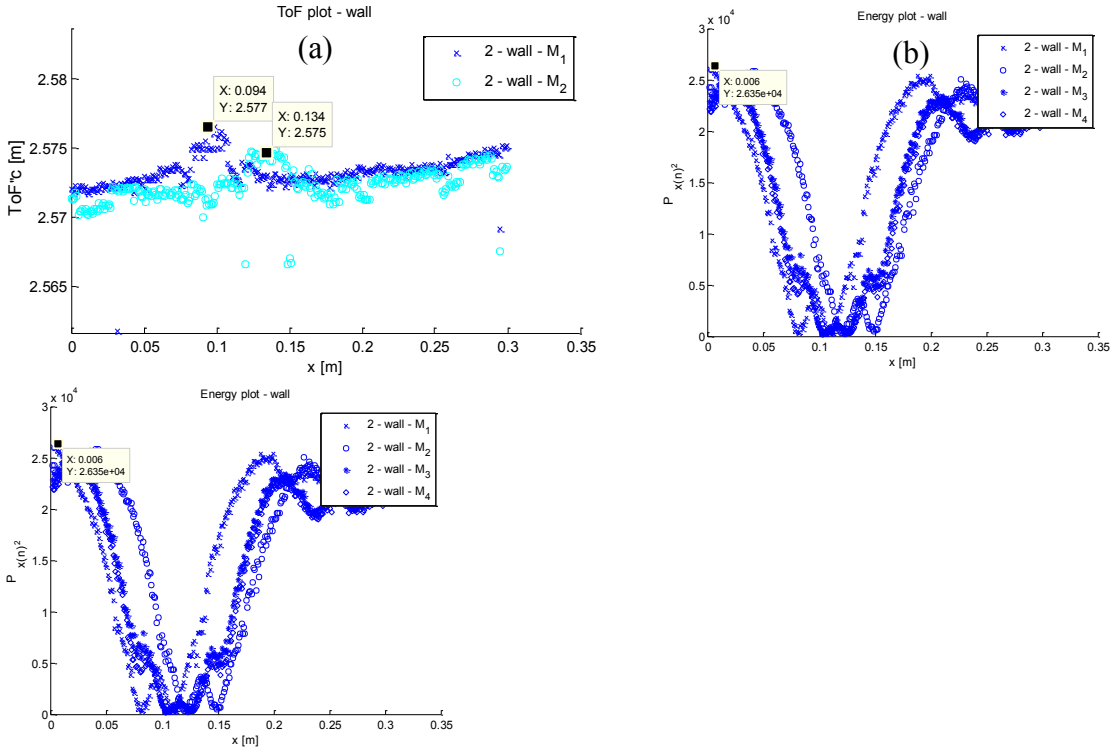


Fig 92: (a) ToA for microphone pairs m_1/m_2 for the wall. The ToA is enlarged due to bending of the acoustic waves around the cylinder. From x = 0m to x = 0.3m the ToAs increase slightly as the sensor is not moving perfectly parallel to the wall. (b) Shows the energy of reflected echoes. If the sensor is in front of the cylinder the energy of the received signal is strongly reduced.

Using these ToAs within the localization provides an interesting result. Despite that ToAs were available in front of the cylinder no reflection points have been calculated behind the cylinder. This is shown in Fig 93.

As the result was very astonishing a simulation with MATLAB was created modelling diffraction around objects. For the sake of simplicity the wall as well as the rod are assumed to extend

infinitely on the vertical axis. Excellent agreement between simulation and measurement was obtained and the result is shown in Fig 94 which shall be compared to the data of Fig 92 (a). Using this simulated data for the calculation of the reflection points confirms the results of Fig 93 as shown in Fig 95. Only a few reflection points are located behind the cylinder where most are either diverted to the left or right. The reason is, that if only a single ToA is affected, the calculated direction of the echo changes. When both microphones are affected equally the reflection point is located behind the cylinder, e.g. at x = -0.13m.

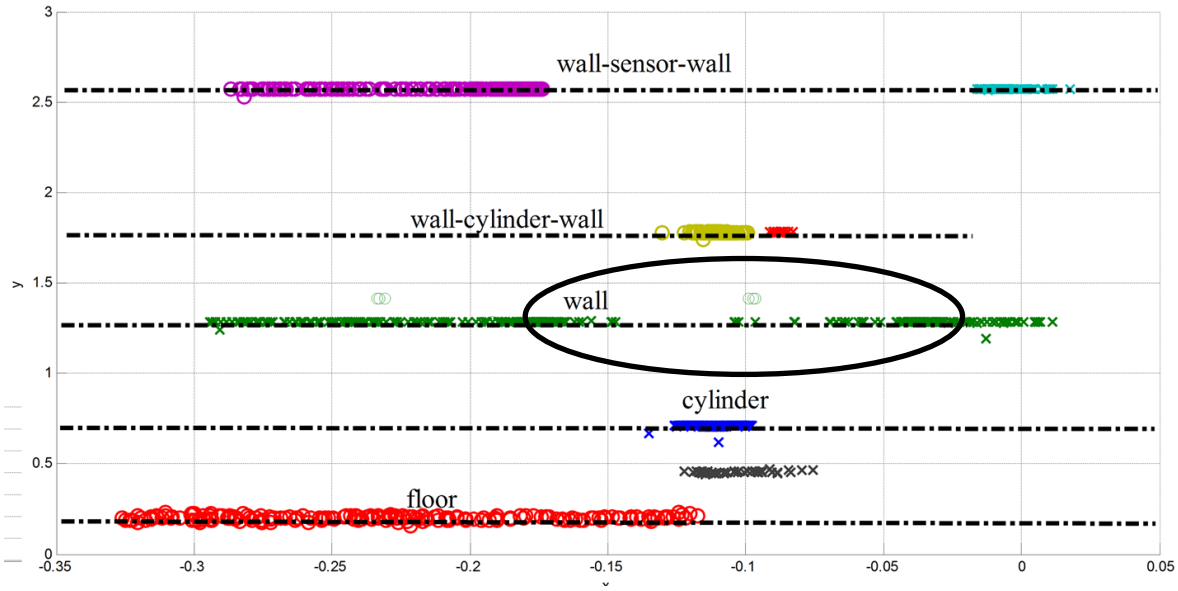


Fig 93: Calculated 3D reflection points using the ToA data of Fig 92. In addition to the echoes from the wall and cylinder, multiple objects due to multipath are present as well. They are similar to the ones already described within 6.2.3 and are not of interest within this context. The interesting part is marked within a black ellipse. Despite ToA data was obtained for the wall no reflection points are present between -0.15m and -0.07cm.

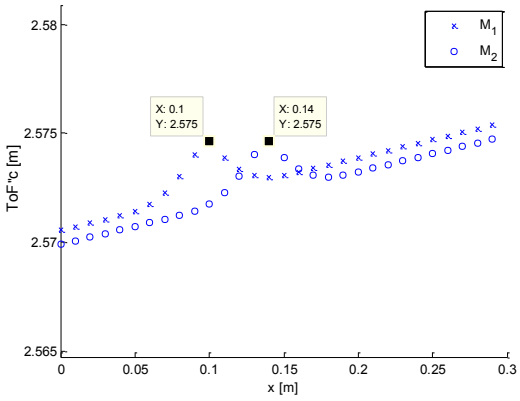


Fig 94: Simulation of diffraction were acoustic waves bend around a cylinder artificially enlarging the ToAs.

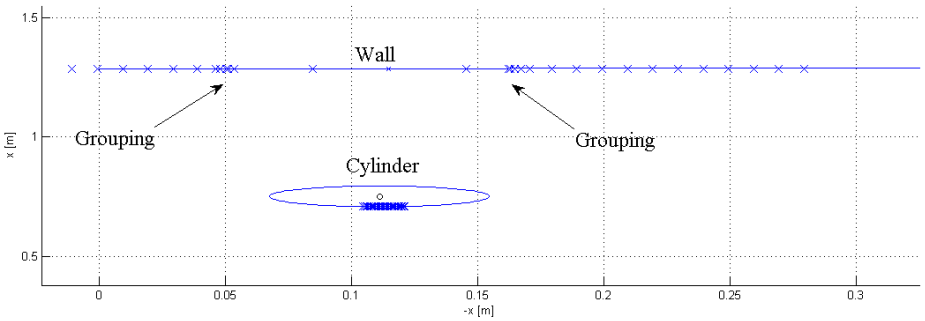


Fig 95: Calculated reflection points using the simulated data of Fig 94. As not all ToAs are affected equally the direction of echo is either diverted to the left and/or right. Therefore multiple reflection points are at close proximity.

While it might be argued that the calculated reflection point is correct the direction vector is not. Therefore, using such data for scene analysis has to be avoided. Two methods are suitable for this: Application of the midpoint criterion can remove some cases as only a single ToA is affected. Furthermore, an energy threshold is also an efficient and suitable method for this due to the huge difference in energy (approx. 30dB).

6.2.4.2 Multipath Propagation

Multipath propagation can easily occur in more complex environments. To investigate the effect on scene analysis a dedicated test setup was created. Care was taken to ensure that the test setup, demonstrating multipath issues, is realistic to stress the importance of dealing with this subject during scene analysis. Furthermore, it will be shown that applying an energy threshold is not a suitable method for multipath identification, as depending on the transducer and its directivity pattern, the multipath signal can have more bigger energy.

Fig 96 shows the test setup used. Two rods have been placed in front of a cabinet (C) at approximately the same distance to the sensor, whereas the rod (A) on the left hand side is closer towards the acoustical axis of the sensor. The rod (B) is located on the right side.

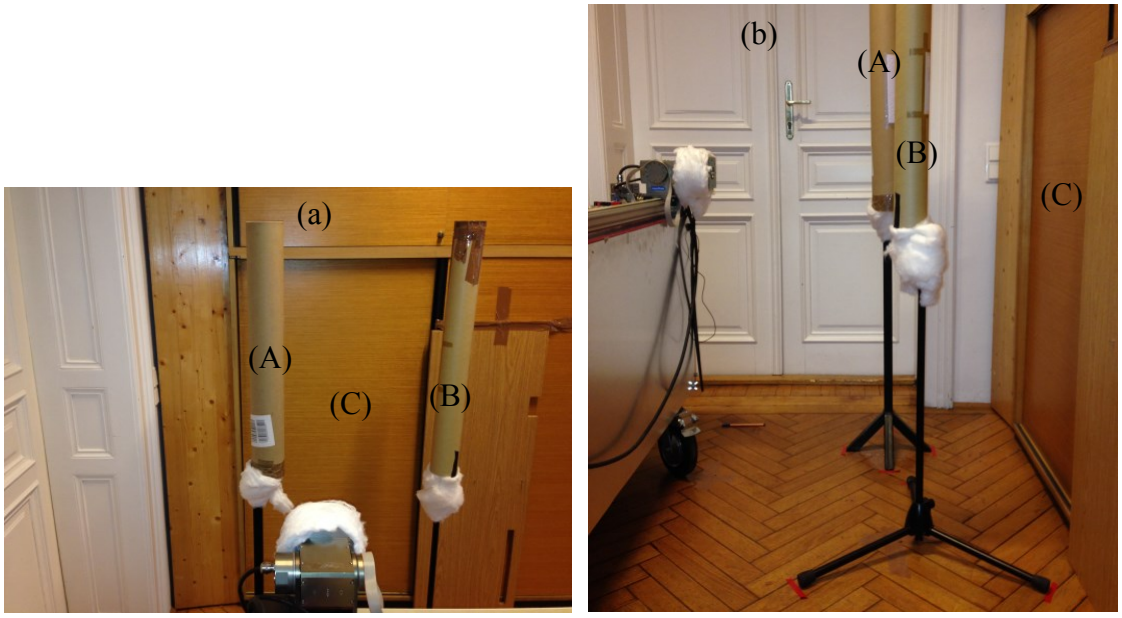


Fig 96: (a) Front view of test setup showing two rods located to the left and right of the sensor. (b) View from the side. Both rods are at approximately the same distance although one is a bit closer to the sensor.

The result is shown in Fig 97 where the corresponding data from the reflectors is given in Tab. 7. While the calculated reflections points (A), (B) and (C) from the two rods and the cabinet are easily explained, the echo (B1) is not.

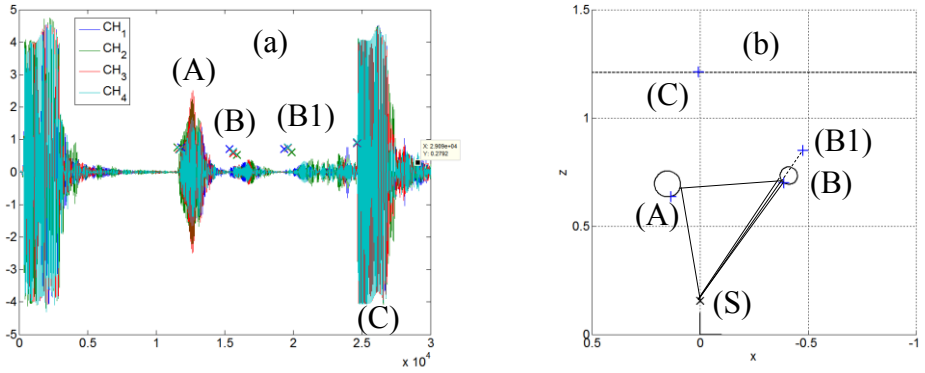


Fig 97: (a) Analog received signals for the individual channels M_1-M_4 shown in different colors (For presentation purposes the analog signal is shown). The ToAs for each identified reflector are marked with an "x". (b) Shows the calculated reflection points. The two rods and the cabinet are indicated by two circles and a horizontal dashed line. The result (B1) is due to multipath propagation.

Looking at (B1) in Fig 97(b) it can be seen that the direction of the echo is similar to the direction of the echo (B), as indicated by the line intersecting the sensor position (S) at $(0,0,0.155)^T$ [m]. Manual inspection of the scene identifies a possible multipath from the sensor (S) over the rod (A) towards rod (B) and eventually back to the sender. The law of reflection is easily fulfilled for this path due to the round rods which scatter the wave into all directions.

Reflection point	Direction Vector	Energy	Label
$(0.135, 0.016, 0.637)^T$ [m]	$(0.269, 0.031, 0.963)^T$ [m]	11.2%	(A)
$(-0.387, -0.003, 0.701)^T$ [m]	$(-0.579, -0.004, 0.817)^T$ [m]	0.18%	(B)
$(-0.475, 0.003, 0.8484)^T$ [m]	$(-0.565, 0.003, 0.825)^T$ [m]	0.25%	(B1)
$(0.007, 0.022, 1.211)^T$ [m]	$(0.006, 0.020, 0.999)^T$ [m]	100%	(C)

Tab. 7: Calculated reflection points using the ToA data within Fig 97(a). What is interesting to see is that the energy level of the multipath signal (B1) is stronger than the directly reflected echo from the rod (B). The reason for this is the directivity pattern of the electrostatic transducer.

To strengthen the assumptions the approximate distances are compared. For the multipath from the two rods the total distance-of-flight is ||A-S|| + ||B-A|| + ||B-S||. This accounts for a total amount of 1.696m. This is in excellent correspondence with 2·||B1-S|| = 1.681m. The small difference is due to the fact that the reflection points on the two rods are not identical as the law of reflection is fulfilled on different points on the surface. For testing purposes the multipath between rods (A) and (B) was covered with a sound absorbing foam as shown in Fig 98(a). It is easily identified that within Fig 98(b) the direct reflection is still present whereas the multipath is no longer detected.

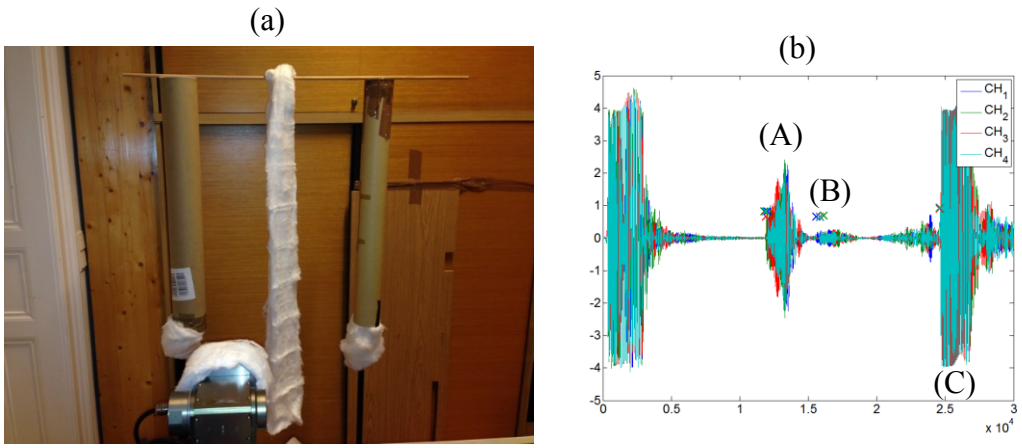


Fig 98: (a) For testing the multi-path was blocked by the insertion of a small piece of sound absorbing foam. (b) The multi-path reflection (B1) is no longer present.

The issue of incorrectly calculated reflection points due to multi-path propagation can be dealt with by post-processing the result from scene analysis using a ray-tracing approach. This can reveal potential multipaths which can then be used to explain some more of the previously obtained ToA data. It shall especially be noted that an energy limit is not a sound approach, as due to the directional characteristic of the transmitter, it can be the case that the reflected echo from the multipath has more energy than the direct echo. This is also true for the data shown in Tab. 7 where the energy of echo (B1) is greater than (B).

6.3 Moving Objects with Common Sensor Reference System

Classifying objects according to their shape, size, orientation and location is a common problem often encountered in industrial applications. This chapter focuses on a specific subclass of this problem where objects are in movement and speed and direction of the movement are known or can be measured. Furthermore, it is required that a (simplified) geometrical-acoustic model of the object is available. By employing an array of compact 3D ultrasonic sensors the proposed system can correctly determine the model parameters and orientation.

Although the actual implementation depends very much on the type of object the general methodology is best demonstrated by an example. First, some suggestions on proper placement of the sensors, depending on the application, are given. Based on the example a basic algorithm is derived which can estimate all unknowns, but ignores measurement uncertainty and any outliers. The chapter concludes with an improved and more robust algorithm which is verified by a physical measurement.

6.3.1 Example for object classification by a static sensor array

The example application consists of a sound-hard rigid box moving on a conveyor belt from right to left with constant velocity $\mathbf{v}_{s2}$ as shown in Fig 99 . The unknown parameters are width, length, height and orientation – all of which should be estimated during object movement.

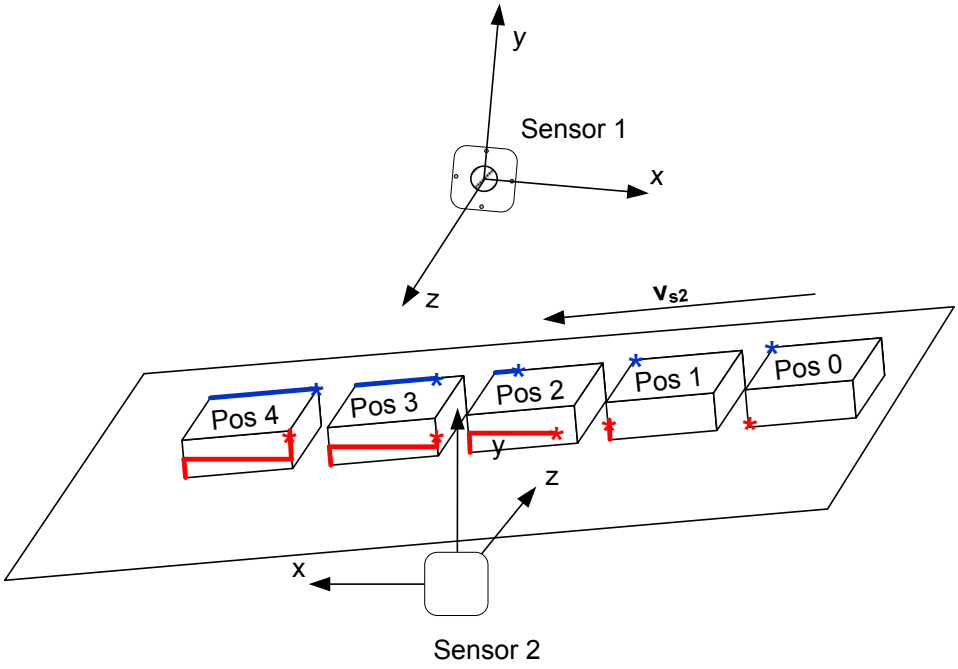


Fig 99: A sound-hard rigid box moves with constant velocity $\mathbf{v}_{s2}$ from right to left. While the object is moving an array of two sensors performs multiple measurements. The trace of reflection points for sensor 1 is shown in blue whereas sensor 2 is shown in red. Using these measurements the unknown parameters length, width, height and orientation of the box are estimated.

In the first step the location of the sensors needs to be fixed where the following general guidelines can be observed.

- If no object is present the sensors should be in the field of view of each other. This allows an array of sensors to self-establish a common reference frame. If this is not possible at least an active or passive reflector shall be present in the field of view of both sensors. The calibration process is detailed in chapter 7.1.
- Sensors shall be placed such that the information obtained from the ToA data is independent. For example a non-suitable placement would put both sensors at the same

side along each other resulting in practically identical information. In the scope of the example application of Fig 99 it is obvious that both sensors should be placed opposite of each other easily allowing determination of the width.

- ToA data can be obtained by different physical wave propagation means. The easiest information to use is specular reflection on a flat surface. This type of information provides good SNR. Furthermore, the surface normal can often be used to estimate the orientation of the object.
- On sharp edges and/or corners the wave is partially reflected and diffracted. Reflection on such objects, although limited in amplitude, is useful for determination of object boundaries. In case of our example shown in Fig 99 the height and length of the box is estimated by sensor 2 using reflection from the top edge of the box. Diffraction, e.g. bending of the acoustic wave around the object, also occurs for any signal emitted by sensor 1 and eventually received by sensor 2 and vice-versa. Due to the high frequency the SNR obtainable by diffracted sound waves is relatively small as can be seen from the example in 6.2.4.1. It is not employed in this example.
- Although not used within this application it might be useful to apply multi-path propagation. For example a passive reflector could be mounted above the linear belt. If the type of the passive reflector is known and its position/orientation is determined within the calibration space it can be possible to obtain further information on the object.

Application of the guidelines and corresponding to practical verification suggested that a proper placement is as following: Sensor 2 is placed such that at least some reflection points are on the front face of the box. This information will be used for determination of the orientation of the box and, in combination with the information of sensor 1, for width estimation. Sensor 1 is placed such that all echoes are due to edge and/or corner reflections. This information is used to estimate the length and height of the box. Within the next section the basic identification algorithm is presented.

6.3.1.1 Basic identification algorithm for sound-hard rigid box moving on a conveyor belt

Without loss of generality all measurements are assumed to be available in the reference frame of sensor 2. A suitable method for establishing a common reference frame and transformation of measurements is described within chapter 7.1. While the box is moving from right to left, as shown in Fig 99, the box comes into the vicinity of either sensor 1 or sensor 2. This time instant is called t_0 and, in combination with the velocity, is used to account for the sensor movement. Let $\mathbf{p_{ij}}$ be an arbitrary measurement of sensor i at time instant t_{ij}, where $t_{ij} > t_0$ and $1 \leq j \leq M_i$. M_i is the total number of measurements for sensor i. Furthermore, let $\mathbf{v_{s2}}$ be the direction of movement for objects on the belt in the reference frame of sensor 2. A measurement is then transformed to time instant t_0 according to (112)

$$\mathbf{pc_{ij}} = \mathbf{p_{ij}} - \mathbf{v_{S2}}\left(t_{ij} - t_0\right) \quad t_{ij} > t_0, 1 \leq j \leq M_i, i \in \{1,2\} \tag{112}$$

The sensor positons have to be transformed as well. Otherwise, if the position of the sensor would remain the same, the direction of the echo would be (incorrectly) altered. Therefore, the sensor position $\mathbf{s_i}$ is transformed as:

$$\mathbf{sc_{ij}} = \mathbf{s_i} - \mathbf{v_{S2}}\left(t_{ij} - t_0\right) \quad t_{ij} > t_0, 1 \leq j \leq M_i, i \in \{1,2\} \tag{113}$$

Similar to chapter 6.2.1 the direction of the echo $\mathbf{dirc_{ij}}$ is calculated from the reflection point and the sensor position. It shall be noted that by (114) the dot product of two direction vectors corresponds to the cosine of the enclosed angle. This fact will be used later on.

$$\mathbf{dirc_{ij}} = \frac{\mathbf{pc_{ij}} - \mathbf{sc_{ij}}}{\left\|\mathbf{pc_{ij}} - \mathbf{sc_{ij}}\right\|} \tag{114}$$

Fig 100 illustrates the application of (112) - (114) to a set of measured data. The transformed sensor positions are shown as solid black circles whereas measurements for sensor 2 are shown in blue and measurements for sensor 1 in red. It can be observed that echoes for sensor 2 are initially due to edge reflection on the left side of the box ($\mathbf{p_{21}}$, $\mathbf{p_{22}}$). Next some measurements are obtained on the front surface of the box ($\mathbf{p_{23}}$, $\mathbf{p_{24}}$ and $\mathbf{p_{25}}$). Eventually some data might be obtained by edge reflection on the right side of the box (not present in this example). For sensor 1 data is initially due to reflection on the corner ($\mathbf{p_{11}}$), then due to edge reflection ($\mathbf{p_{12}}$ and $\mathbf{p_{13}}$) and finally again due to reflection on the right corner ($\mathbf{p_{14}}$ and $\mathbf{p_{15}}$).

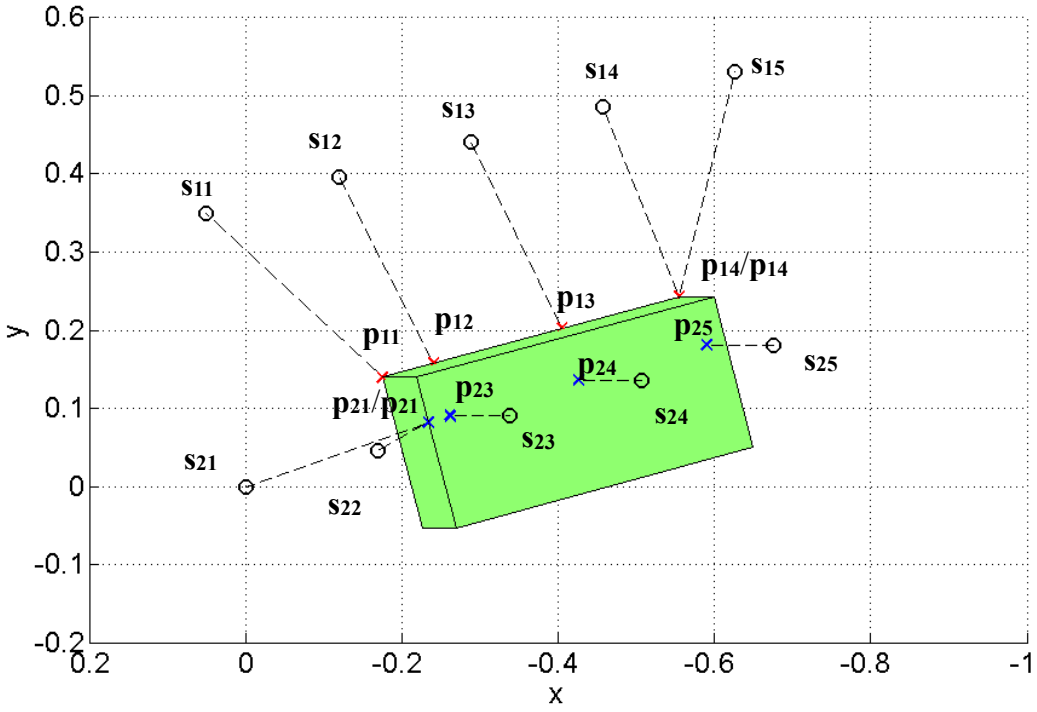


Fig 100: Simulated example showing the transformed ToA data using (112) - (114) within a common reference frame. Reflection points for sensor 2 are shown in blue whereas reflection points for sensor 1 are shown in red.

To simplify calculations the model presented in Fig 101 will be used. The box is described by the edge points $\mathbf{p_1}$-$\mathbf{p_4}$ and $\mathbf{p_{1L}}$-$\mathbf{p_{4L}}$. The length, width and height as well as the orientation can be directly calculated from these data.

Using the reflection points of sensor 2 it can be assumed that at least some of them are on the front side surface of the box, e.g. $\mathbf{p_{23}}$, $\mathbf{p_{24}}$ and $\mathbf{p_{25}}$ in the example of Fig 100. Due to the law of reflection all of them will exhibit a similar direction vector. Using an angular threshold C groups G_k of size O_k can be built by application of (115).

$$G_k = \left\{ \mathbf{pc_{2k_1}}, \ldots, \mathbf{pc_{2k_{O_k}}} \middle| \cos^{-1}\left(\mathbf{dirc_{2k_i}} \cdot \mathbf{dirc_{2k_j}}\right) < C \right\} \tag{115}$$

Or in other words – Two elements belong to the same group if their angular difference of the direction vector is less than the threshold C. Among the groups built let G_f be the group with the largest cardinality. This group is also most likely the group which contains all echoes from the front side surface of the box, because the other echoes are due to edge reflection, and therefore quickly change their incident direction. Using the elements of the group G_f it is then possible to estimate the normal vector $\mathbf{n_1}$ using (116).

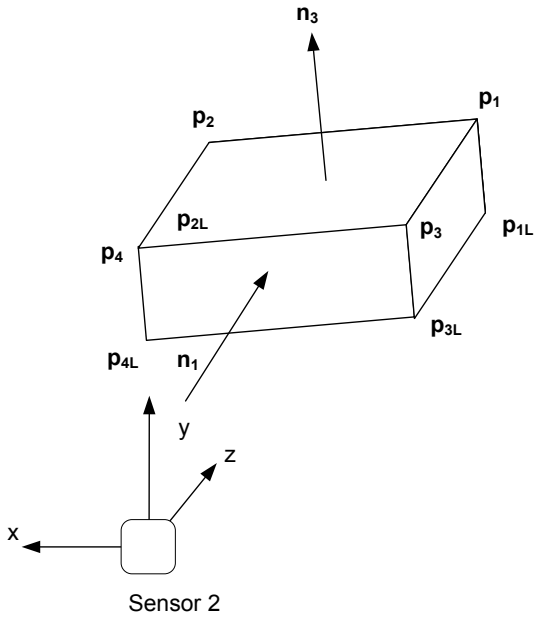


Fig 101: Model of a box described by the corner points $\mathbf{p_1}$, $\mathbf{p_2}$, $\mathbf{p_3}$ and $\mathbf{p_4}$ as well as $\mathbf{p_{1L}}$, $\mathbf{p_{2L}}$, $\mathbf{p_{3L}}$ and $\mathbf{p_{4L}}$.

$$\mathbf{n_1} = \frac{1}{O_k} \sum_{i=1}^{O_k} \mathbf{dirc_{2k_i}} \tag{116}$$

Using the Hessian Normal form the distance to the origin is estimated as

$$d_1 = \frac{1}{O_k} \sum_{i=1}^{O_k} \mathbf{pc_{2k_i}} \cdot \mathbf{n_1} \tag{117}$$

By model assumption $\mathbf{n_2}$ is identical to $\mathbf{n_1}$. As all measurements, ignoring any outliers for now, are on the line defined by the point $\mathbf{p_1}$ and $\mathbf{p_2}$, these measurements are also on the plane with normal vector $\mathbf{n_2}$. Therefore, the distance d_2 to the origin can be calculated as:

$$d_2 = \frac{1}{M_2} \sum_{i=1}^{M_2} \mathbf{pc_{1k_i}} \cdot \mathbf{n_1} \tag{118}$$

An estimator for the width of the box is given by

$$\text{width} = d_2 - d_1 \tag{119}$$

To estimate the length and orientation of the box principal component analysis (PCA) is used. Using the measurements of sensor 1 the first principal component is obtained. This component points into the direction of the vector defined by the (unknown) endpoints $\mathbf{p_1}$ and $\mathbf{p_2}$. Intuitively this is clear because it is also the direction of the greatest variance. Let $\mathbf{u}$ be the eigenvector

with the largest eigenvalue obtained by PCA. We then apply the PCA for dimensionality reduction using the eigenvector $\mathbf{u}$. That is for each measurement $\mathbf{pc_{1i}}$, $1 \leq i \leq M_1$, z_{1i} is calculated according to:

$$z_{1i} = \mathbf{u} \cdot (\mathbf{pc_{1i}} - \overline{\mathbf{pc_1}}) \tag{120}$$

An estimator for the length of the box is then given by

$$\text{length} = \max(z_1) - \min(z_1) \tag{121}$$

The two edge points $\mathbf{p_1}$ and $\mathbf{p_2}$ can be reconstructed as:

$$\mathbf{p_1} = \mathbf{u}\min(z_1) + \overline{\mathbf{pc_1}} \quad \mathbf{p_2} = \mathbf{u}\max(z_1) + \overline{\mathbf{pc_2}} \tag{122}$$

The points $\mathbf{p_3}$ and $\mathbf{p_4}$ are calculated using the plane normal $\mathbf{n_1} = \mathbf{n_2}$ and the previously calculated points $\mathbf{p_1}$ and $\mathbf{p_2}$.

$$\mathbf{p_3} = \mathbf{p_1} - \mathbf{n_1} \cdot \text{width} \quad \mathbf{p_4} = \mathbf{p_2} - \mathbf{n_1} \cdot \text{width} \tag{123}$$

The height of the box cannot be directly measured, but it is a sound assumption that the bottom and top surface of the box are parallel to the belt, i.e. it is lying on the belt. The normal vector $\mathbf{n_3}$ of the box, which by the previous assumption is also identical to the belt surface normal, is given by:

$$\mathbf{n_3} = \mathbf{u} \times \mathbf{n_1} \tag{124}$$

Let $\mathbf{b}$ be an arbitrary point on the surface of the belt. It shall be noted that such a point is readily obtained using the setup presented in Fig 99 because such a point $\mathbf{b}$ fulfills the law of reflection for a wave originating from sensor 2, being reflected by the belt at position $\mathbf{b}$, and then eventually received by sensor 1. As the position of the receiving microphones and the sound emitting device is known this point can be calculated – although not as easily as in case of the compact 3D sensor. Using this point $\mathbf{b}$ and the normal vector $\mathbf{n_3}$ the belt surface can be described by a plane in Hessian normal with a distance d_3 calculated according to (125)

$$d_3 = -\mathbf{n_3} \cdot \mathbf{b} \tag{125}$$

In the next step the distance of the points $\mathbf{p_1} - \mathbf{p_4}$ to the belt surface is calculated using (126). The right part of the first equation is the formula for calculating the normal distance of a point to a plane and is also an estimator for the height.

$$\begin{aligned} \|\mathbf{p_i} - \mathbf{p_{iL}}\| &= \text{height} = \mathbf{n_3} \cdot \mathbf{p_i} + d_3 \\ \mathbf{p_{iL}} &= \mathbf{p_i} - \|\mathbf{p_i} - \mathbf{p_{iL}}\|\mathbf{n_3} \end{aligned} \tag{126}$$

Fig 102 shows the execution of above algorithm in case of no outliers and noise. The sole input to the algorithm is the measurement data, the velocity of the belt and an arbitrary point on the surface of the belt (if the height shall be estimated as well). Outputs of the algorithm are all parameters of the box including height, width, length and orientation. What has not been dealt with up to now is measurement uncertainty and outliers. This is within the scope of the next section.

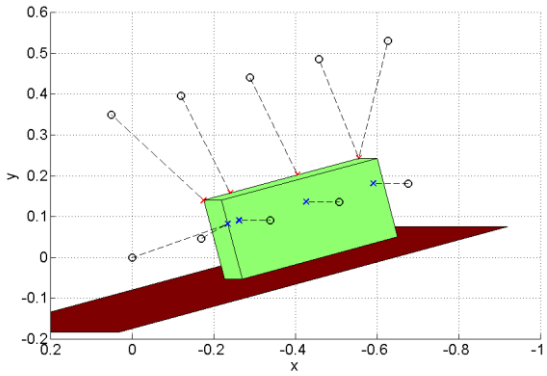


Fig 102: Application of the algorithm described within 6.3.1.1 to measurement data obtained by a box moving from left to right in front of the sensor array with a known velocity. As no measurement uncertainty was applied in this example all parameters have been determined with zero error.

6.3.1.2 Robust algorithm for sound-hard rigid box moving on a conveyor belt

The algorithm presented in 6.3.1.1 did not account for measurement uncertainty and outliers. Potential sources of uncertainty in the context of this discussion include the uncertainty of the 3D compact sensor (See chapter 5.3), the uncertainty of the calibration, the number of outliers in a measurement and the total number of measurements available. The minimum number of measurements required for identification is three: a single measurement of sensor 2 on the front side surface as well as a single measurement for each corner from sensor 1. While theoretically feasible, in practice this is not sufficient, as it cannot be assured, that measurements are taken at a specific point on the box surface.

Investigation of the algorithm described in 6.3.1.1 reveals the following two problems:

- The estimation of the length and points $\mathbf{p_1}$ and $\mathbf{p_2}$ uses a minimum and maximum operation. Therefore, no benefit is taken from additional data.
- Principal component analysis, as applied in this context, is not resilient to outliers as the main target is to estimate the direction of greatest variance.

Therefore, the following modification is applied to the previously described algorithm. Let P be the probability that a measurement taken is an outlier and M_1 be the total number of measurements from sensor 1. As all measurements taken by sensor 1 should be on a straight line (by model assumption) we expect that the PCA has a very strong first principal component and the other components are very weak. Or in other words – the variance of the data is focused exactly along one direction. This fact is exploited by building random subsets of size $K = \lceil (1-2\cdot P)\cdot M_1 \rceil$ for which the PCA is computed. From this subset the one is chosen which minimizes the ratio between the largest eigenvalue and the sum of the remaining two eigenvalues. Let $\mathbf{u}$ be the corresponding eigenvector for the largest eigenvalue.

As an additional filtering step the image is reconstructed and the difference to the input data is calculated. Using the chosen eigenvector $\mathbf{u}$, all measurements are projected into the 1-dimensional space using (127).

$$y_{1i} = \mathbf{u} \cdot (\mathbf{pc_{1i}} - \overline{\mathbf{pc_1}}), 1 \leq i \leq M_1 \tag{127}$$

By reconstruction the data points $\mathbf{pr_{1i}}$ from its projection y_{1i} the distance to the original data points is calculated and compared to a threshold. A point is kept for further processing if the error err_{1i} is less than a chosen bound (1cm in our example)

$$\mathbf{pr_{1i}} = \mathbf{u} \cdot y_{1i} + \overline{\mathbf{pc_1}}, 1 \leq j \leq M$$
$$\mathrm{err}_{1i} = \|\mathbf{pr_{1i}} - \mathbf{pc_{1i}}\| \tag{128}$$

As the extents of the box are estimated from the furthest left and right data points another estimator is required. Initially min/max was used but this is not a robust estimator as the performance gets worse with increasing number of data. For solving this problem, first the points on the corners are identified. This can be done easily as the measured direction vector and the direction of the eigenvector **u** are orthogonal. A point, belongs to a corner if evaluation of (129) yields a value greater than C, where C is an angular constant in radians. In the simulation it was chosen as 5°.

$$\left|\cos^{-1}\left(\mathbf{u} \cdot \mathbf{dir_{j_1}}\right) - \pi/2\right| > \frac{C}{180}\pi \tag{129}$$

Using this set, it is partitioned into two distinct sets, using a k-Means algorithm where k equals 2. The points $\mathbf{p_1}$ and $\mathbf{p_2}$ of equation (122) are then determined by using the median from the two groups.

For estimation of d_1, the distance of the front plane, it is sufficient to sort the calculated projections of equation (117) and remove some elements from the top and beginning. The number of elements to be removed shall be proportional to the outlier probability P. The same algorithm shall be applied for d_2.

6.3.1.3 Performance investigation of robust-algorithm with respect to measurement uncertainty and outliers

For investigation on the effect of measurement uncertainty and outliers a simulation framework was created. The simulation framework can introduce uncertainty to the 3D localization, movement of the object on the conveyor and on the initial calibration. Furthermore, outliers can be added to the measurements. Presentation within this section is limited to two special cases. In the first step it was checked if the improved algorithm makes use of additional measurement data in the sense that the uncertainty of the estimated parameters is reduced. The second test verifies if the algorithm can correctly handle a specific amount of outliers.

The first simulation added random Gaussian noise to the 3D coordinates where the standard deviation was set to 1cm. Verification was performed by comparison with the model, where the reported performance indicator is the estimated error bound for 95% of all measurements. For the origin of the box the Euclidian distance was used as an error measure. The error in estimating the surface normal $\mathbf{n_1} = \mathbf{n_2}$ and $\mathbf{n_3}$ of the box is given in degrees. Tab. 8 shows the result of the simulation where N is the number of measurement points. It can be seen that the uncertainty is reduced by additional data.

	N=5	N=20
95% error bound for height	1.56 [cm]	0.81 [cm]
95% error bound for length	2.6 [cm]	2.07 [cm]
95% error bound for width	2.25 [cm]	1.2 [cm]
95% error bound for origin	2.77 [cm]	1.8 [cm]
95% error bound for $\mathbf{N_1}$ [°]	3.52 [°]	2.02 [°]
95% error bound for $\mathbf{N_3}$ [°]	2.55 [°]	1.59 [°]

Tab. 8: Error bound for object parameters with 95% confidence interval. Contrary to the simplified algorithm the robust algorithm makes proper use of additional data and uncertainty is reduced by additional data. 3D Location uncertainty was simulated as 1cm (1-σ).

The second simulation adds approx. 10% outliers to the measurement data. Fig 103 shows such an example where the outliers are marked within a circle. The performance of the algorithm is not affected by a moderate number of outliers.

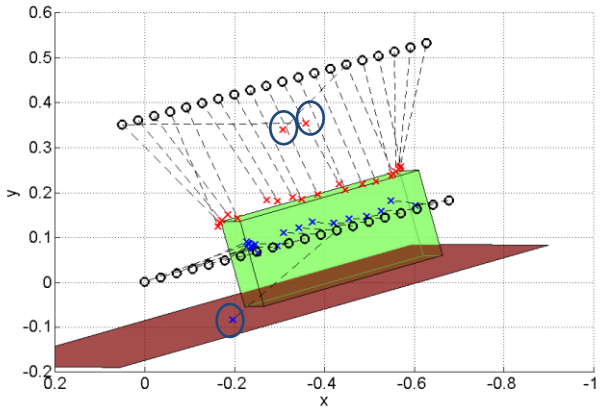


Fig 103: Simulation including 3D location uncertainty and approx. 10% outliers in the measurement data. The robust algorithm performance is not affected by the outliers (marked with blue circles).

6.3.1.4 Practical verification of robust algorithm

The experiment was performed in a standard office room. A small box with dimensions 38cm x 16cm x 18cm was mounted on a linear belt. Using the linear belt the box was moved from the origin to x = 80cm in steps of 4cm. For the measurement two compact 3D sensors have been used. The initial calibration was performed using an active device mounted on a rigid plate. The complete test setup is shown in Fig 104.

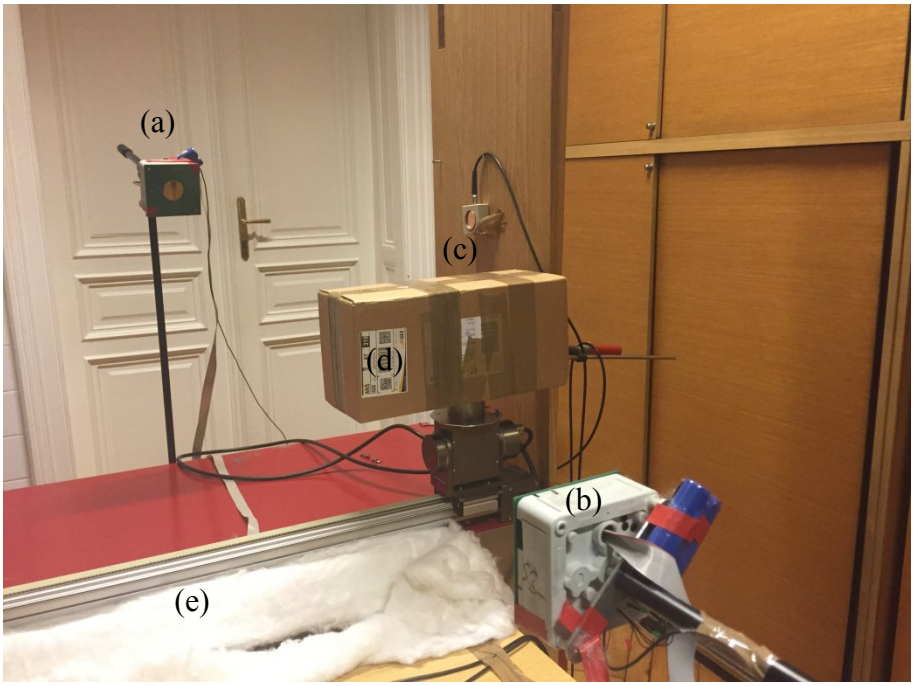


Fig 104: Test setup showing two 3D compact sensors where sensor 1 is labeled as (a) and sensor 2 is labeled as (b). (c) is an active calibration device further explained in section 7.1. The box (d) is mounted on a linear belt (e).

Executing of the algorithm yields the results shown in Fig 105. It can be seen that at a distance of approximately 10cm it is possible for sensor 1 (red points) to obtain echoes from the corner. Reflections from sensor 2 (blue trace) are not limited by amplitude as echoes from the plane

are very strong. The length of the box was estimated with 36,2cm which is in reasonable correspondence with the physical extents of 38cm. The width of the box was estimated with 16,2cm which shows excellent correspondence with the physical size of 16cm.

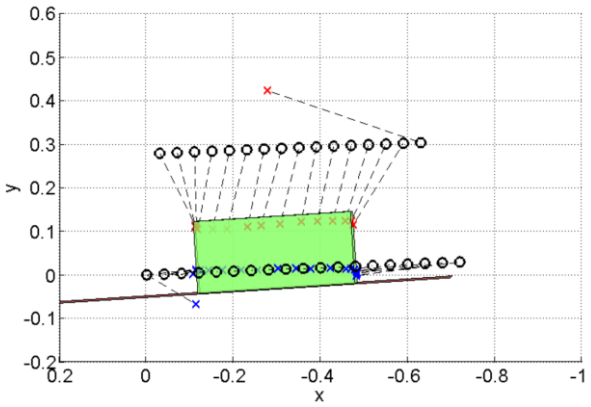


Fig 105: Identified box shown in green where measurements for sensor 2 are shown in blue and measurements for sensor 1 in red.

As the surface of the conveyor belt is not well defined in this setup the height measurement was ignored and is only drawn for presentation purposes. Orientation shows good correspondence although this could not be verified due to the lack of a reference.

7 Calibration

Referring to the application categories introduced in section 6.1 all applications, with the exception of tracking an object, require some form of initial system calibration. In case of a moving sensor, i.e. for *map building* and *scene analysis*, measurements need to be transformed into a non-moving reference coordinate system. For this task the position and orientation of the compact 3D sensor needs to be determined. A suitable indoor positioning system (IPS) is presented in 7.2. It does not only support localization but also includes methods for fully automatic self-calibration with well-defined uncertainty. Another potential solution is the use of acoustic retro-reflectors introduced in 7.3 which can be used as fix-points for navigation. In the case of moving objects, e.g. for *object classification*, and if multiple static 3D compact sensors are used, it is also possible to automatically establish a common reference frame among them. A suitable method for this task is introduced in 7.1.

7.1 Self-Calibration of an array of sensors

Some applications require the use of multiple compact 3D sensors to obtain sufficient amount of information. For example, the application given in 6.3.1 uses two compact 3D sensors to determine the parameters of a moving box on a conveyor belt. One issue typically encountered within such a task is that measurements from different sensors are only available in their local coordinate system, i.e. the data cannot be used as a whole. Within this context the problem of "self-calibration" shall, therefore, be defined as obtaining a common reference coordinate system to which all measurement data can be related to.

The presented method herein requires at least two compact 3D sensors, one passive and/or active reflection device and three measurements. After the calibration process measurements can be transformed either into the local coordinate system of sensor 1 or sensor 2. A suitable setup is shown in Fig 106. Care has to be taken that there exists a line-of-sight (LOS) path between sensor 1 and sensor 2. Furthermore, in the case of a passive reflection device, a multipath shall exist. The reflection point on the passive device for sensor 1 is called $\mathbf{r}_{|1}$ and $\mathbf{r}_{|2}$ for sensor 2. If such a passive device is not available an additional active transmitter can be used as well.

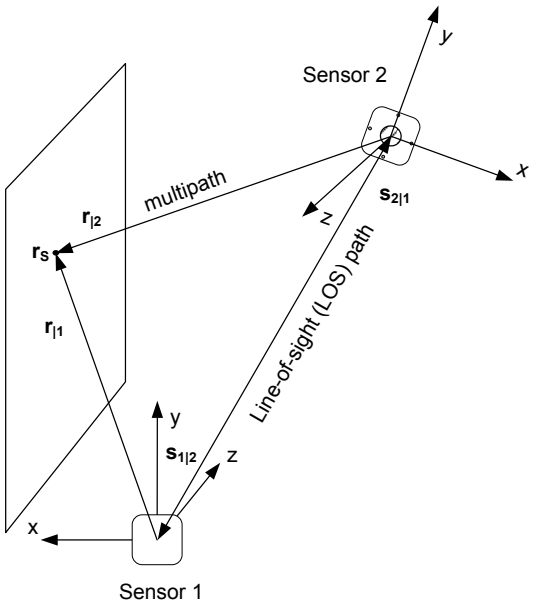


Fig 106: Suitable setup for automatic self-calibration of an array of compact 3D sensors. There are no restrictions on the orientation and position of sensor 1 and sensor 2. What is required is that there exists a line-of-sight path between sensor 1 and sensor 2 and a passive reflecting device. If no passive reflector is available an active device can be used instead (which would be placed at the position $\mathbf{r}_{|1}/\mathbf{r}_{|2}$)

The section starts with an informal description of the algorithm and later gives a mathematical more accurate description. First, sensor 1 emits an ultrasonic signal which is processed by sensor 2. From this measurement sensor 2 can obtain the vector $\mathbf{s_{1|2}}$, i.e. the position of sensor 1 expressed in the coordinates of sensor 2. As sensor 2 now knows the position of sensor 1, which is also the sound emitting device, it can calculate the position of the reflection point $\mathbf{r_{|2}}$. The same process is repeated with sensor 2 firing, yielding $\mathbf{s_{2|1}}$ and $\mathbf{r_{|1}}$. Using these measurements a rotation matrix $\mathbf{R_{1\to2}}$ is determined corresponding to the relative orientation between both sensors. A measurement $\mathbf{p_{|1}}$ in the local coordinate system of sensor 1 can then be transformed into the coordinate system of sensor 2 by application of (130).

$$\mathbf{p_{|2}} = \mathbf{R_{1\to2}}\mathbf{p_{|1}} + \mathbf{s_{1|2}} \tag{130}$$

In case of an active calibration device $\mathbf{r_{|1}}$ and $\mathbf{r_{|2}}$ are directly obtained by locating the transmitter. The section now continues with a more formal description of the algorithm above. First, the 3D localization formulas presented in 5.2 have to be extended to allow localization of a potentially displaced transmitter assuming direct wave propagation.

Let $\mathbf{m_{1|i}}$, $\mathbf{m_{2|i}}$, $\mathbf{m_{3|i}}$ and $\mathbf{m_{4|i}}$ be the positions of the microphones of sensor i expressed in its local reference frame. Furthermore, let t_{1ji}, t_{2ji}, t_{3ji}, and t_{4ji} be the ToA measurements of microphones 1, 2, 3 and 4 between an arbitrary transmitter j and the sensor i. The position $\mathbf{t_{j|i}}$ of transmitter j in the reference frame of sensor i can then be determined by solving (131) for the argument $\mathbf{t}$.

$$\mathbf{t_{j|i}} = \arg\min\left(\sum_{k=1}^{4}\left(\left\|\mathbf{m_{k|i}} - \mathbf{T}\right\| - t_{kji}\right)^2\right) \tag{131}$$

Following the notation introduced previously it follows that $\mathbf{s_{1|2}} = \mathbf{t_{1|2}}$ and $\mathbf{s_{2|1}} = \mathbf{t_{2|1}}$.

For obtaining the reflection point $\mathbf{p_{j|i}}$ for multipath propagation with displaced transmitter j and sensor i (132) can be used. Similar to (131) the system of equations is solved for the unknown argument $\mathbf{p}$. Therefore $\mathbf{r_{|2}} = \mathbf{p_{1|2}}$ and $\mathbf{r_{S|1}} = \mathbf{p_{2|1}}$.

$$\mathbf{p_{j|i}} = \arg\min\left(\sum_{k=1}^{4}\left(\left\|\mathbf{m_{k|i}} - \mathbf{p}\right\| + \left\|\mathbf{t_{j|i}} - \mathbf{p}\right\| - t_{kji}\right)^2\right) \tag{132}$$

The task of finding a common reference frame can now be solved accordingly: Without loss of generality sensor 2 is used as reference frame, i.e. its origin equals the null-vectors and its respective unity vectors in x, y and z are $(1\ 0\ 0)^T$, $(0\ 1\ 0)^T$ and $(0\ 0\ 1)^T$. Let $\mathbf{p_{|1}}$ be a measurement obtained by sensor 1. Using (130) the corresponding vector $\mathbf{p_{|2}}$ in the coordinate system of sensor 2 is obtained. As $\mathbf{s_{1|2}}$ was already gathered as one of the calibration measurements the only unknown left is $\mathbf{R_{1\to2}}$. Settings $\mathbf{p_{|1}} = \mathbf{s_{2|1}}$, and using the fact that the sensor position of sensor 2 in the coordinate system of sensor 2 equals the null vector, yields:

$$\mathbf{0} = \mathbf{R_{1\to2}}\mathbf{s_{2|1}} + \mathbf{s_{1|2}} \tag{133}$$

From (133) it follows that

$$\mathbf{s_{2|1}} = -\mathbf{R_{1\to2}}^{-1}\mathbf{s_{1|2}} \tag{134}$$

Again using (130) we multiply both sides by $\mathbf{R_{1\to2}}^{-1}$ and finally substitute $\mathbf{s_{2|1}}$ of (134).

$$\begin{aligned}\mathbf{R_{1\to2}}^{-1}\mathbf{p_{S|2}} &= \mathbf{p_{S|1}} + \mathbf{R_{1\to2}}^{-1}\mathbf{s_{1|2}}\\ \mathbf{p_{S|1}} = \mathbf{R_{1\to2}}^{-1}\mathbf{p_{S|2}} - \mathbf{R_{1\to2}}^{-1}\mathbf{s_{1|2}} &= \mathbf{R_{1\to2}}^{-1}\mathbf{p_{S|2}} + \mathbf{s_{2|1}}\end{aligned} \tag{135}$$

Together equations (135) and (130) allow transformation between the two coordinate systems.

Using the measurements $\mathbf{s_{1|2}}$, $\mathbf{s_{2|1}}$, $\mathbf{r_{|2}}$ and $\mathbf{r_{|1}}$ from the calibration process the transformation matrix can be obtained by the following system of equations.

$$\begin{aligned}\mathbf{0} &= \mathbf{R_{1\to 2}}\mathbf{s_{2|1}} + \mathbf{s_{1|2}} \\ \mathbf{r_{|2}} &= \mathbf{R_{1\to 2}}\mathbf{r_{|1}} + \mathbf{s_{1|2}}\end{aligned} \tag{136}$$

(136) can be rewritten as

$$\begin{aligned}\mathbf{0} &= \left(\mathbf{0} - \mathbf{s_{1|2}}\right) - \mathbf{R_{1\to 2}}\mathbf{s_{2|1}} \\ \mathbf{0} &= \left(\mathbf{r_{|2}} - \mathbf{s_{1|2}}\right) - \mathbf{R_{1\to 2}}\mathbf{r_{|1}}\end{aligned} \tag{137}$$

(137) is now in the form suitable for a minimum least square approach and can be written as

$$\begin{aligned}&\mathbf{err}(\mathbf{R_{1\to 2}}) = \sum_{\mathbf{k=1}}^{\mathbf{2}} \|\mathbf{p_k^*} - \mathbf{R_{1\to 2}}\mathbf{p_k}\|^{\mathbf{2}} \\ &\mathbf{p_1^*} = -\mathbf{s_{1|2}}, \mathbf{p_1} = \mathbf{s_{2|1}} \\ &\mathbf{p_2^*} = \mathbf{r_{|2}} - \mathbf{s_{1|2}}, \mathbf{p_2} = \mathbf{r_{|1}}\end{aligned} \tag{138}$$

The formulation of (138) is identical to Wahba's problem [Wah65] and can be solved using singular value decomposition of matrix **B** yielding the desired matrix $\mathbf{R_{1\to 2}}$.

$$\begin{aligned}&\mathbf{B} = \sum_{\mathbf{k=1}}^{\mathbf{2}} \mathbf{P_k}^{*}\mathbf{P_k}^{\mathbf{T}} \\ &\mathbf{B} = \mathbf{USV}^{\mathbf{T}} \\ &\mathbf{M} = \begin{bmatrix}1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \det(\mathbf{U})\det(\mathbf{V})\end{bmatrix} \\ &\mathbf{R_{1\to 2}} = \mathbf{UMV}^{\mathbf{T}}\end{aligned} \tag{139}$$

It shall be noted the calibration accuracy can be improved by extending (139) to an arbitrary number of measurements.

Tab. 9 shows the values obtained for the example in Fig 104. In this case an active calibration device was used. Application of (139) yields the result shown in (140).

Point	3D coordinates
$\mathbf{s_{2|1}}$	$(-0.00, 0.28, 1.64)^T$ [m]
$\mathbf{s_{1|2}}$	$(0.02, 0.14, 1.66)^T$ [m]
$\mathbf{r_{|1}}$	$(-0.49, 0.22, 1.02)^T$ [m]
$\mathbf{r_{|2}}$	$(0.57, 0.0, 0.57)^T$ [m]

Tab. 9: Measurement data for calibration of the setup shown in Fig 104. An active calibration device was used in this case.

$$\boldsymbol{R_{1\to 2}} = \begin{bmatrix}-0.98 & 0.17 & 0.01 \\ 0.17 & 0.95 & -0.25 \\ -0.06 & -0.24 & -0.97\end{bmatrix} \tag{140}$$

7.2 Indoor Localization System with automatic self-calibration

This section briefly introduces the indoor localization system *LOSNUS* (**L**ocalization **o**f **S**ensor **N**odes by **U**ltra**s**ound) with a focus on automatic self-calibration and applicability for this work. *LOSNUS* is an indoor localization system developed by the ultrasonic working group at the Institute of Electrodynamics, Microwave and Circuit Engineering (EMCE). Besides the work on the compact 3D sensor the author of this work focused on automatic self-calibration, reliable localization and methods for assessing the uncertainty of an indoor localization system. For a more in depth discussion the interested reader is referred to [Wal13b].

Such a local position system (LPS) is especially useful for *map building* and *scene analysis* as it can determine the position and orientation of the 3D compact sensor within the room with low uncertainty. In fact only three of the four receivers available on the 3D compact sensor need to participate in a *LOSNUS* locating sequence. *LOSNUS*, by design, is optimized for locating numerous quasi-static devices. A basis setup consists of multiple active, static and synchronized beacons emitting a broadband ultrasonic wave in combination with a transmitter code as shown in Fig 107. Using a well-defined transmitting sequence these signals are received by passive nodes, which in combination with the a-priori known beacon positions, can determine their own location.

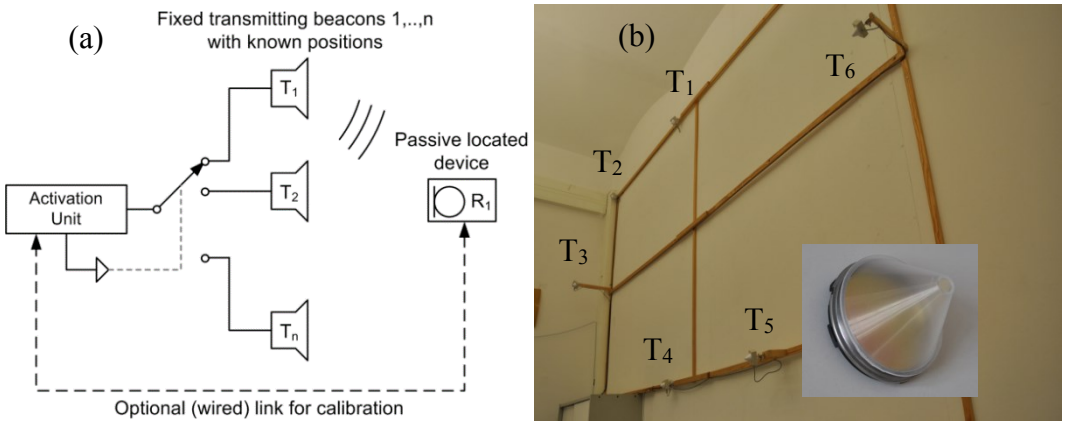


Fig 107: (a) Basic setup for the LPS *LOSNUS*. It consists of multiple static beacons which are sequentially fired using an activation unit. A passive device receiving the transmitted waveforms measures the relative time-of-arrivals, identifies the corresponding transmitters, and by using the a-priori known beacon positions determines its own position. (b) Test setup at Technische Universität Wien, EMCE, showing six static transmitters mounted on the wall. The transmitters have been modified by attaching a small cone to reduce the directional characteristic of the electrostatic transducers.

Calibration of a LPS, i.e. obtaining knowledge of the beacon positions with low uncertainty is an essential requirement if high locating accuracy is desired. For this task *LOSNUS* supports a fully automatic self-calibration which requires only four receivers, six transmitters (beacons) and a known reference distance. This calibration process as well as a method for establishing the total system uncertainty is presented within the next sections.

7.2.1 Self-calibration of indoor local positioning system *LOSNUS*

Calibration in the implemented form requires four receivers, at least six static beacons and a reference distance with known uncertainty. The output of the calibration is a well-defined reference frame in which all beacon positions are known. For calibration time-of-flight measurements are used requiring synchronization between transmitters and receivers. This requirement can be dropped during practical operation.

First, the reference frame is defined by choosing three arbitrary beacons. These three beacons, without loss of generality, are assumed to lie on the y-z plane where one beacon is used as the origin of the coordinate system, i.e. $(0\ 0\ 0)^T$. The second beacon defines the y-axis, i.e. $(0\ y_2\ 0)$

whereas the third defines the z-axis, i.e. (0 y_3 z_3). All parameters y_2, y_3, and z_3 of these beacons are unknown. The z-axis is defined by the cross-product of the x-axis and y-axis. The minimum number of transmitters and receivers can be argued as following. Let N be the number of transmitters and M be the number of receivers. Then there are a total of (N + M)·3 unknowns. Using three of the transmitters for fixing the reference frame, as mentioned before, eliminates six unknowns. On the other hand N·M ToF measurements are available. For the system of equations to be solvable we require:

$$(N + M)3 - 6 \geq \mathrm{N} \cdot \mathrm{M} \tag{141}$$

This requirement is fulfilled by exactly six beacons and four arbitrary placed receivers.

During the calibration phase each transmitter i fires at least once, whereas the emitted wave is received by a receiver j yielding a ToF measurement t_{ij}. This yields 24 ToF equations of the form (142) where c is the speed of sound and *l* is the length of the acoustic cone. The additional variable *l* is required as the wave does not directly propagate from the transmitter to the receiver but has to take the "detour" through the cone opening.

$$0 = \left\|\mathbf{t_i} - \mathbf{r_j}\right\| - \mathrm{t_{ij}c} - \mathrm{l} \tag{142}$$

For solving the system of equations (142) can be rewritten as a minimization problem

$$\mathrm{err} = \sum_{\mathrm{i}=1}^{6}\sum_{\mathrm{j}=1}^{4}\left(\left\|\mathbf{t_i} - \mathbf{r_j}\right\| - \mathrm{t_{ij}c} - \mathrm{l}\right)^2 \tag{143}$$

It can be shown that if a minimization algorithm solves (143) it also solves a scaled version of the same system by a factor a. This fact is used to account for the a-priori unknown speed of sound. Let *r* be the a-priori known reference distance and assume that the extents are determined during calibration by receiver $\mathbf{r_i}$ and $\mathbf{r_j}$, e.g. by mounting them on the reference distance. Then the scaling factor *a* can be calculated according to:

$$\mathrm{a} = \frac{\mathrm{r}}{\left\|\mathbf{r_i} - \mathbf{r_j}\right\|} \tag{144}$$

The final results are then calculated from the output $\mathbf{t_i}$, $\mathbf{r_j}$ and *l* according to (145).

$$\mathbf{t_i'} = \mathbf{t_i}\mathrm{a} \quad \mathbf{r_j'} = \mathbf{r_j}\mathrm{a} \quad \mathrm{l'} = \mathrm{al} \tag{145}$$

For suppression of random components averaging of ToF measurements is necessary. This process is problematic as it cannot be safely assumed that measurement conditions are constant over time. This problem can be avoided by fixing the sum of all individual distances over time to an arbitrary reference value. From this a scaling factor according to (146) can be defined which is individually applied to all measurements where $t_{ij}(k)$ is the measurement taken at the discrete time index k. It shall be noted that the final result is not affected due to the post-scaling by application of (144). An example is given in Fig 108 where the change of the scaling factor c is shown over a time of 6 hours.

$$\mathrm{c(k)} = \frac{\sum_{\mathrm{i}=1}^{6}\sum_{\mathrm{j}=1}^{4}\mathrm{t_{ij}\ (k)}}{\sum_{\mathrm{i}=1}^{6}\sum_{\mathrm{j}=1}^{4}\mathrm{t_{ij}\ (1)}} \tag{146}$$

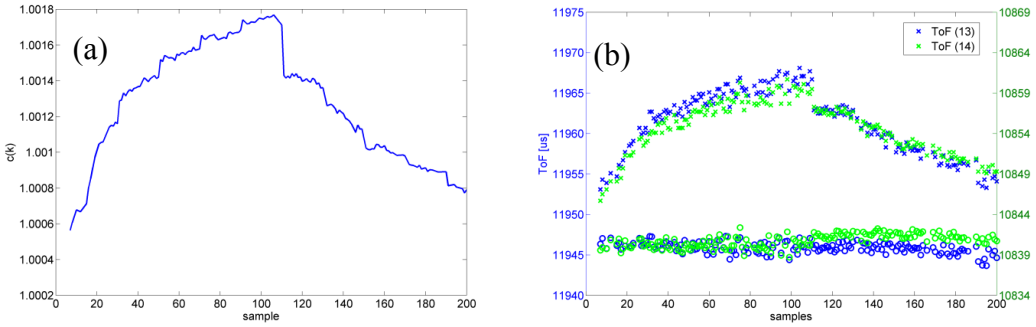


Fig 108: (a) scaling factor c(k) over time (total duration = 6 hours). (b) Temperature compensated measurements (bottom) show no drift over time whereas uncompensated measurements (top) correlate with the temperature. The total drift accounts for approx. 5mm which would introduce a severe error in the final result.

An example for a calibration result is shown in Fig 109. All the beacon coordinates and receiver positions (coordinate not shown) have been identified. The factor a, calculated according to (144), equals 1.012. The acoustic cone length l was determined as 81.2µs = 2.8cm which is in good correspondence with the physical extents of the cone.

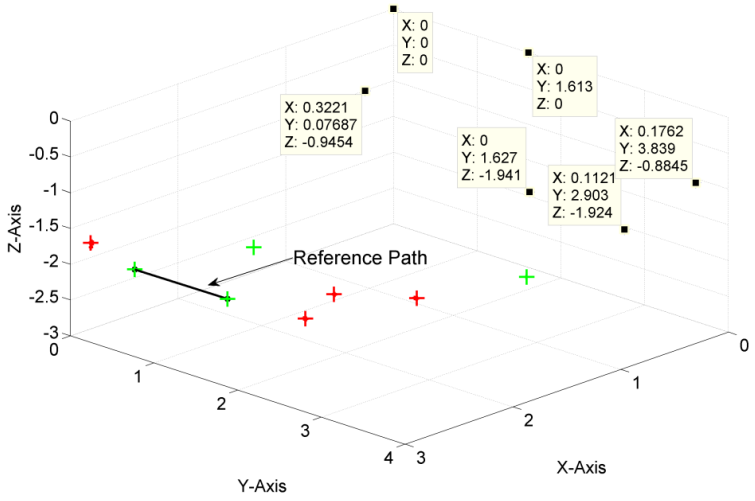


Fig 109: Automatic calibration of LPS LOSNUS for the test setup shown in Fig 107. All unknown coordinates of the beacons have been defined. The green '+' symbols indicate receivers used during calibration. The red '+' where additional receivers used for error checking but not used within the calibration process.

7.2.2 Uncertainty of localization and calibration

The accuracy of locating is based on two factors: accuracy of the system calibration and accuracy of the actual locating procedure. These two factors are completely independent being two different procedures performed at different points of time. System calibration is an operation done only once after system installation. Further on, the realized calibration accuracy is influencing any result of locating as the actual locating accuracy is an additional factor which reduces the overall accuracy.

7.2.2.1 Uncertainty in ToA and TDoA measurements

Using binary cross correlation for ToA/ToF measurements the random component was tested to be normal distributed with a standard deviation of 400ns. The random component is a complex function of the electronics noise, microphone self-noise, acoustic noise, sampling rate as well as length and bandwidth of the used chirp, which shall all be subsumed in above measure. ToF and ToA measurements are modeled according to equation (147) where in case of ToA the term t_0 described the unknown transmission time. In case of ToF the term t_0 contains any electrical and acoustic delays.

$$\mathrm{ToF/ToA} = t_t + t_0 + N_t \quad \text{with} \quad \overline{N_t} = 0, {s_{N_t}}^2 = (400\mathrm{ns})^2 \tag{147}$$

Time difference measurements are obtained by using one ToA as a reference. As measurements are uncorrelated the variances add. Note that in this case we assumed no correlated noise as compared to 4.5 as the measurements are performed at different time instants.

$$\mathrm{TDoA} = t_{t1} - t_{t2} + N_t \quad \text{with} \quad \overline{N_t} = 0, {s_{N_t}}^2 = (400\mathrm{ns})^2 + (400\mathrm{ns})^2 \tag{148}$$

7.2.2.2 Uncertainty in the calibration phase

Using the proposed method of 7.2.1 for calibration the obtained uncertainty for each calibrated transmitter position equals:

$$u(\mathbf{t_i}) \approx \|t_i\| \frac{u_r}{r} \tag{149}$$

$\mathbf{t_i}$ is the coordinate of the transmitter, u_r is the uncertainty of the reference path and *r* is the length of the reference path. The same applies for the receivers $\mathbf{r_i}$. All other systematic or random components are reduced either by the algorithm or averaging. A receiver position used within the calibration is called reference position. Having two such positions and a linear belt available allows creation of arbitrary reference positions with a known uncertainty.

7.2.2.3 Uncertainty of localization

For reporting the locating uncertainty the concept F.2.4.5 of the "Guide to the expression of uncertainty in measurement" [Gum08] is applied. In this case a known correction factor b for a systematic effect is not applied but reported in the uncertainty of the measurement. Measurements are reported according to equation (150) as three dimensional coordinates (x, y, z) including a systematic effect with unknown direction and a combined standard uncertainty, also with unknown direction. **p** is the (unknown) value of the measurand. One should note that equation (150) cannot be evaluated but is just a formal method for reporting results and should be interpreted as described herein.

$$\|\mathbf{p} - (x, y, z)^T\| = b \mp u_c \tag{150}$$

The reminder of this section focuses on the derivation of the factor b and the combined uncertainty u_c. A suitable test setup for evaluating the uncertainty consists of a linear belt with a LOSNUS node attached. The start and end positions $\mathbf{r_1}$ and $\mathbf{r_2}$ are determined during calibration with low uncertainty. Using the belt a total of M known reference positions can be obtained by moving the receiving node from $\mathbf{r_1}$ to $\mathbf{r_2}$ using M steps with step size Δ.

$$\mathbf{r_k} = \frac{\mathbf{r_2} - \mathbf{r_1}}{\|\mathbf{r_2} - \mathbf{r_1}\|} k\Delta + \mathbf{r_1} \; 1 \le k \le M \tag{151}$$

At each reference position multiple measurements are performed. Let SP_k be the set of all measurements at position k. The center of gravity for these points can be calculated according to:

$$\overline{\mathbf{p_k}} = \frac{1}{|SP_k|} \sum_{\mathbf{p} \in \mathbf{SP_k}} \mathbf{p} \tag{152}$$

The systematic deviation b_k between the center of gravity and the reference point is calculated for each position k as:

$$b_k = \|\mathbf{r_k} - \overline{\mathbf{p_k}}\| \tag{153}$$

According to F.7a of the GUM the mean correction factor *b* is calculated as

$$b = \frac{1}{M}\sum_{k=1}^{M} b_k \tag{154}$$

The variance of the correction factor b can be estimated using (155). This shall be interpreted as the uncertainty introduced by using only a single systematic factor b, although this factor in general is different for every measurement k.

$$u^2(b) = \frac{1}{M-1}\sum_{k=1}^{M}(b_k - b)^2 \tag{155}$$

Another part which has to be accounted for is the variance of b_k itself. There are two potential error sources for this: First the uncertainty of the reference positions $\mathbf{r_k}$ and secondly the uncertainty of the center of gravity due to the finite number of measurements at each position. Note that the latter one is reduced by the number of measurements taken at each position.

$$\begin{aligned} u^2(b_k) &= u^2(\mathbf{r_k}) + \frac{u^2(SP_k)}{|SP_k|} \\ \overline{u^2(b_k)} &= \frac{1}{M}\sum_{k=1}^{M} u^2(b_k) \end{aligned} \tag{156}$$

The uncertainty of the reference positions can be estimated as following. Let u_1 and u_2 be the uncertainty of the reference positions $\mathbf{r_1}$ and $\mathbf{r_2}$ estimated using (149). Then the uncertainty of an interpolated reference position $\mathbf{r_k}$ can be calculated using (157), i.e. it gradually fades from u_1 to u_2 over M steps.

$$u(\mathbf{r_k}) = \frac{u_2 - u_1}{\|R_2 - R_1\|} k\Delta + u_1 \tag{157}$$

Using F.7d of the GUM the mean variance of the individual measurements can be calculated. Note that this figure is closely related to the dilution-of-precision (DOP). The DOP is a factor, which, if multiplied by the uncertainty of a ToF measurement, gives the expected uncertainty of the located point in distance units.

$$\begin{aligned} u^2(SP_k) &= \frac{1}{|SP_k| - 1}\sum_{\mathbf{P} \in \mathbf{P_k}} \|\mathbf{p} - \overline{\mathbf{p_k}}\| \\ \overline{u^2(SP_k)} &= \frac{1}{M}\sum_{i=1}^{M} u^2(\mathbf{SP_k}) \end{aligned} \tag{158}$$

Using (155), (156) and (158) the total combined uncertainty is given as:

$$u_c^2 = u^2(\bar{b}) + \overline{u^2(b_k)} + \overline{u^2(SP_k)} \tag{159}$$

In other words: the total uncertainty is a combination of the uncertainty due to the usage of a single correction factor b, the uncertainty associated with the estimation of the correction factor, and finally the uncertainty of the measurement itself.

For the test setup shown in Fig 107 the systematic correction factor b was determined as 4.7mm whereas the total combined uncertainty was 1.3mm. This is a remarkable result for an indoor locating system as typical reported results for common ultrasonic LPS are often in the range of centimeters [Wal13b]. Furthermore, most literature dealing with GPS and LPS often defines

their own meaning of the terms “accuracy” and “precision” rendering results non-comparable between different authors. For this reason a suitable method for applying the GUM in the context of LPS was shown and applied to a practical test setup.

7.3 Fix-Point Navigation using retro-reflectors

Another potential method which allows tracking of the sensor during movement is the usage of acoustic retro-reflectors. These acoustic retro-reflectors serve as fix-points within the environment. If at least three such fix-points are available in the field-of-view of the compact 3D sensor its position and orientation relative to these points can be calculated. This makes it possible to relate measurements from different positions to each other, which is an essential requirement for scene analysis and map-building.

It shall be noted that although it is possible to use any well-known and static object for such a task the acoustic retro-reflector has two important properties making it especially suitable for this task. First of all the echo exhibits a strong amplitude due to specular reflection on a flat surface. Secondly the reflection point is well defined and does not change with the direction of the incident acoustic wave. Fig 110 shows a simulation of such a reflector. It consists of three perpendicular planes where the individual reflection points for each microphone/transmitter pair are shown. The calculated reflection point, by application of the equations given in 5.2, is exactly in the corner at position $(0\ 0.5\ 1)^T$. In addition all plausibility checks introduced in 5.5 are also satisfied for this type of reflector.

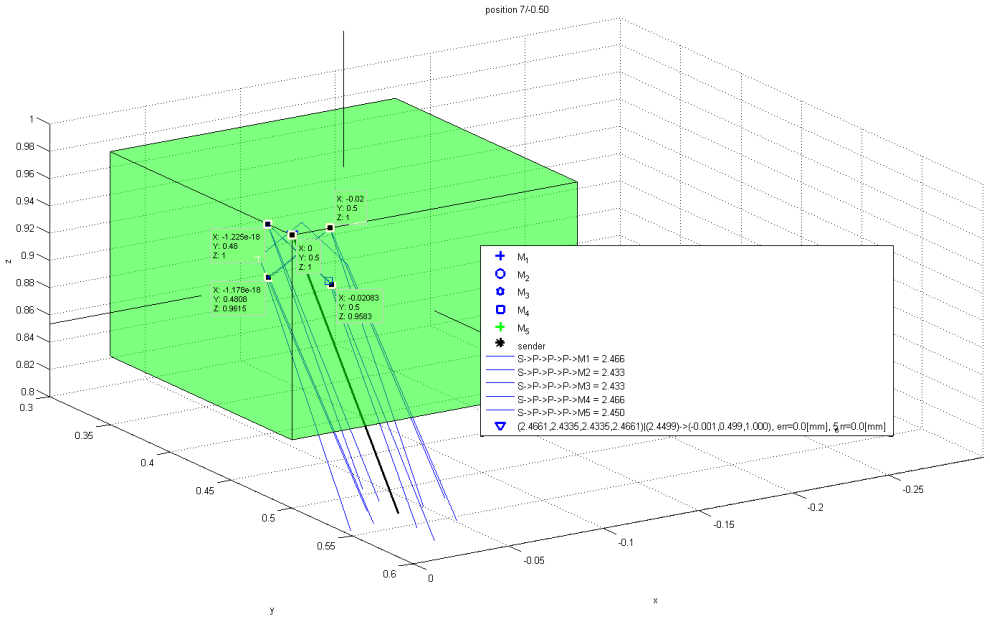


Fig 110: Simulation of an acoustic retro-reflector for the 3D compact sensor. The calculated reflection point using the equations of section 4.2 correspond to the corner of the device.

Performance of an acoustic retro-reflector was verified with a practical test setup as shown in Fig 111. All calculated reflection points are very close to each other with a spread of only 9 mm. Using multiple such retro-reflectors as fix-points within the environment is therefore a viable alternative compared to more complex solutions like using an indoor localization system.

Fig 111: Acoustic retro-reflector built from three sound-hard planes. The acoustic-retro reflector was located from multiple positions on the belt. The calculated reflection points are all very close to each other with a total point-spread of approx. 9 mm.

8 Conclusion

In this work a universal compact 3D sensor for accurate localization of 3D reflection points using Ultrasound on Air is presented. Key improvements compared to existing work in this area is the larger amount of information obtained by a single measurement without any compromise on the quality of the data. Identification of the electrostatic transducer, spatial modelling and efficient 1-bit binary pulse compression techniques ensure a wide field-of-view, low time-of-arrival uncertainty and good echo separation capabilities. The novel sensor construction in combination with suitable methods for processing the time-of-arrival (ToA) information allows object independent localization of reflection points, efficient solving of the echo correspondence problem and avoids outliers due to object discontinuities or boundaries.

8.1 Applicability of presented methods

Using such a compact 3D sensor as a building block two applications have been targeted with the proposed sensor system delivering excellent results: For scene analysis and map building the sensor is able to correctly identify complex 3D objects like planes, cylinders and others in 3D. Performance was verified in difficult environments where echo overlapping, object discontinuities, multipath propagation and diffraction effects exit. To obtain sufficient information sensor movement is a prerequisite. This in turn requires determination of the sensor position and orientation to allow fusing data from different measurement positions. Two methods have been proposed for this task: The indoor localization system LOSNUS can be statically installed within a room. After an initial fully-automatic self-calibration the system is able to locate the sensor with a total uncertainty in the millimeter range. As a second and less complex alternative navigation using fix-points with the environment have been investigated.

The second application uses an array of static compact 3D sensors suitable for solving complex object classification tasks for moving objects. In the proposed example an array of two sensors is used to classify a moving box on a conveyor belt. All parameters of the box including its orientation are determined by a few measurements. Similar to scene analysis and map building a common reference frame for the group of sensors is required. The proposed calibration method requires no human intervention and only a single passive reflector.

During the work results were presented at several international conferences as [Wal11], [Wal12a], [Wal12b], [Wal12c], [Wal13a], [Wal14a], [Wal14b], [Wal15b] and [Wal15c] and in scientific journals [Wal13b] and [Wal15a]. The author has also co-authored several other papers as [Moh12] and [Moh14].

8.2 Outlook

Despite the promising performance different areas for improving the proposed sensor system exist.

- Scene analysis and map building was verified for planes and cylinders in 3D. Extending the basic geometric shapes by additional objects would make it possible to analyze more complex environments.
- Due to the correlative signal processing multi-path echoes are easily detected by the sensor as the SNR is still sufficient. Fake objects, due to multipath, are therefore still present in the final identified scene. By post-processing using suitable raytracing algorithms it would be possible to identify some of these objects which in turn could be removed or marked.
- While the electrostatic transmitter is well suited for generation of ultrasonic sound waves some interesting other techniques for sound generation have been proposed. Es-

pecially the EMI-Film material and/or the thermoacoustic transducer would allow better control of the directivity of the transmitter as arbitrary shapes can be created with high precision.

- Right now the system uses an AD-interface mounted inside a standard PC and MATLAB for signal processing. If the sensor is applied in practical scenarios a stand-alone system would be preferred. For this a new hardware consisting of transmitter driving circuits, microphone preamplifiers and signal processing should be developed.

9 Glossary

9.1 Notation

A vector **x** is written in bold lower case letters. A matrix **X** is written in bold upper case letters. For vectors the symbol · corresponds to the dot-product whereas × corresponds to the cross product. ||**x**|| is used for the Eculidian norm of a vector **x**.

The Fourier transform of a function f(t) in the time domain is written as $\mathcal{F}\{f(t)\}$. Unless otherwise stated the unitary Fourier transform according to (160). Convolution in the time domain is indicated by a $*$ symbol.

$$\begin{aligned} \mathcal{F}\{f(t)\} = F(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(\tau)\, e^{-i\omega\tau} d\tau \\ f(t) = \mathcal{F}^{-1}\{F(\omega)\} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(\nu)\, e^{i\nu t} d\nu \end{aligned} \tag{160}$$

For an unknown parameter θ the estimator is written as $\hat{\theta}$.

9.2 Symbols

c	Speed of sound [m/s] (343 m/s at dry air, 20° Celsius)
C_v	Head capacity [kJ/(kg·K)]
f	Frequency [Hz]
H_n, H_n'	Hankel function of the first kind and its first derivate
$\boldsymbol{I_x}$	Noise covariance matrix
k	Wave number
J_n, J_n'	Bessel function of the first kind and its first derivate
$\boldsymbol{J}$	Jacobian of a multidimensional function
M	Molar mass [g/mol] (28.97 g/mol for Air)
p_0	Static pressure (101325 Pa = average static pressure at sea level)
p_{Sat}	Partial pressure of water vapour [Pa]
p	Sound pressure [Pa]
R	Molar gas constant [J/(mol·K)] (8.31446 J/(mol·K for air)
RH	Relative humidity
$\boldsymbol{v}$	Sound particle velocity [m/s]
ν_t	Thermal conductivity [W/mK]
Z	Complex acoustic impedance [Ns/m^3] (approx. 414 Ns/m^3 for Air)
α	Attenuation coefficient [Nepers/m]
γ	Adiabatic index (often assumed 1.4 for air)

η	Dynamic viscosity [Pa·s]
λ	Wavelength [m]
ρ_0	Density of air [kg/m^3] (1.2041 kg/m^3 at 20° Celsius and p_0)
Γ	Reflection factor
Φ	Acoustic velocity potential
ϑ	Temperature in degree Celsius [°]
ω	Angular frequency [rad/s]

Tab. 10: Common symbols used within the work

9.3 Abbreviations

ToF	…	Time of Flight
MEMS	…	Microelectromechanical systems
FCM	…	Fuzzy Clustering Method
RANsac	…	RANdom Sample Consensus
LOSNUS	…	Localization of Sensor Nodes by Ultra-Sound
RCD	…	Region of constant depth
EKF	…	Extended Kalman Filter
PCA	…	Principal component analysis
FPGA	…	Field programmable gate array
EMFi	…	Electromechanical Film
THD	…	Total Harmonic Distortion
RADAR	…	Radio Detection and Ranging
DOP	…	Dilution of Precision
GPS	…	Global Positioning System
LPS	…	Local Positioning System
IPS	…	Indoor Positioning System
ITO	…	Indium Tin Oxide

10 References

[Tur60] G. Turin, “An Introduction to Matched Filters”, IEEE Transactions on Information Theory, 1960, doi:10.1109/TIT.1960.1057571

[Kla60] J. R. Klauder, A. C. Price, S. Darlington; W.J. Albersheim, “The Theory and Design of Chirp Radars”, Bell System Technical Journal, Volume 4, 1960

[Sav64] A. Savitzky, M. J. E. Golay, “Smoothing and Differentiation of Data by Simplified Least Squares Procedures”, Analytical Chemistry, Volume 36, Issue 8, 1964, DOI: 10.1021/ac60214a047

[Wah65] G. Whaba, “A Least Square Estimate of Spacecraft Attitude”, SIAM, Volume 8, November 1965

[Mey79] E. Meyer, E.-G. Neumann, “Physikalische und Technische Akustik”, ISBN 978-3-528-28255-4, Braunschweig 1979, DOI: 10.1007/978-3-663-13982-9

[Mor85] Moravec, H.P., Elfes, A, "High resolution maps from wide angle sonar", IEEE International Conference on Robotics and Automation, Volume 2, March 1985, doi: 10.1109/70.88147

[Gel89] W. Gellert, S. Gottwald, Hellwich M., Kästner H., Künstner H., “VNR Concise Encyclopedia of Mathematics”, 2nd edition, New York, 1989, 978-0442205904

[Boz91] O. Bozma, R. Kuc, "Building a sonar map in a specular environment using a single mobile sensor", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 13, December 1991, doi: 10.1109/34.107000

[Leo91] J.J. Leonard, H.F. Durrant-Whyte, "Mobile robot localization by tracking geometric beacons", IEEE Transactions on Robotics and Automation, Volume 7, pp.376-382, June 1991, doi: 10.1109/ROBOT.1985.1087316

[Per92] H. Peremans, J.M.Van Campenhout, L. Levrouw, “Steps toward triaural perception”, Proceedings SPIE 1611, Sensor Fusion IV: Control Paradigms and Data Structures, 1992; doi:10.1117/12.57920.

[Kay93] S. Kay, “Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory”, April, 1993, ISBN 007-6092031871

[Zol93] M. Zollner, E. Zwicker, “Elektroakustik”, 3rd edition, ISBN 3-540-64665-5, Heidelberg 1993

[Per93a] H. Peremans, K. Audenaert, J.M. Van Campenhout, "A high-resolution sensor based on tri-aural perception," IEEE Transactions on Robotics and Automation, Volume 9, pp.36-48, Feb 1993, doi: 10.1109/70.210793

[Per93b] H. Peremans, J.M. Van Campenhout, "Tri-aural perception on a mobile robot", IEEE International Conference on Robotics and Automation, May 1993, doi: 10.1109/ROBOT.1993.291993

[Kle94] L. Kleeman, R. Kuc, "An optimal sonar array for target localization and classification," IEEE International Conference on Robotics and Automation, May 1994, doi: 10.1109/ROBOT.1994.351089

[Elm95] H. Elmer, “Improved Ultrasonic Distance Measurement in Air”, PHd thesis, Technische Universität Wien, Vienna, 2005

[Kle95] H. Akbarally, L. Kleeman, "A sonar sensor for accurate 3D target localization and classification", IEEE International Conference on Robotics and Automation”, May 1995, doi: 10.1109/ROBOT.1995.525710

[Mat95] P. Mattila, F. Tsuzuki, H. Vaataja and K. Sasaki, "Electroacoustic model for electrostatic ultrasonic transducers with V-grooved backplates," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Volume 42, Januar 1995, doi: 10.1109/58.368320

[Man96] I. E. Manolakis, “Efficient solution and performance analysis of 3-D position estimation by trilateration”, IEEE Transactions on Aerospace and Electronic Systems, 1996, doi: 10.1109/7.543845

[Bar98] B. Barshan, B. Ayrulu, "Performance comparison of four time-of-flight estimation methods for sonar signals", Electronics Letters, Volume 34, pp. 1616-1617, 6 August 1998, doi: 0.1049/el:19981127

[Far00] F. Angelo, “Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique”, 108th AES convention, February 2000

[Paa00] M. Paajanen, J. Lekkala, K. Kirjavainen, “ElectroMechanical Film (EMFi) — a new multipurpose electret material”, Sensors and Actuators A: Physical, Volume 84, August 2000, doi:10.1016/S0924-4247(99)00269-1

[Aky01] T. R Akylas, C. C. Mei, “School-wide Program on Fluid Mechanics - Modules on Waves influids”, MIT course 1.138J/2.062J

[Are01] T.E. Gómez Álvarez-Arenas, “Wide-band air-coupled ultrasonic piezoelectric transducers”, International Symposium on Piezocoposite Applications, Dresden, September 2011

[Hoi02] H. Hoislbauer, R. Schwodiauer, S. Bauer-Gogonea, S. Bauer, “Patterned piezoelectricity in charged, cellular polymers for air-borne ultrasound devices”, 11th International Symposium on Electrets, 2002, doi: 10.1109/ISE.2002.1042943

[Kle02] L. Kleeman, “On-the-fly Classifying Sonar with Accurate Range and Bearing Estimation”, IEEE International Conference on Intelligent Robots and Systems, October 2002, doi: 10.1109/IRDS.2002.1041385

[Sch05] H. Schweinzer, P. Krammer, "Scene analysis with ultrasonic sensors", 10th IEEE Conference on in Emerging Technologies and Factory Automation, Sept. 2005, doi: 10.1109/ETFA.2005.1612523

[Sma05] Smart Materials. “1-3 fiber random composite”, http://www.smart-material.com/13Crand-product-main.html, last checked on 14 November 2015.

[BK08] Bruel & Kjaer, “1/8″Pressure-field Microphone Type 4138”, Product datasheet

[Jim08] A. Jiménez, Á. Hernández, J. Ureña, M. Pérez, F. Álvarez, C. Marziani, J. García, J. Villadangos, “EMFi-based ultrasonic transducer for robotics applications”, Sensors and Actuators A: Physical, Volume 148, November 2008, doi:10.1016/j.sna.2008.07.024

[Gum08] JCGM , “Evaluation of measurement data — Guide to the expression of uncertainty in measurement”, Joint Committee for Guides in Metrology, 2008

[Bro09] D. Browne, L. Kleeman, "An advanced sonar ring design with 48 channels of continuous echo processing using matched filters", IEEE/RSJ International Conference on Intelligent Robots and Systems, October 2009, doi: 10.1109/IROS.2009.5354228

[Och09] A. Ochoa, J. Urena, A. Hernandez, M. Mazo, J.A. Jimenez, M.C. Perez, "Ultrasonic Multitransducer System for Classification and 3-D Location of Reflectors Based on

PCA", IEEE Transactions on Instrumentation and Measurement, Sept. 2009, doi: 10.1109/TIM.2009.2016820

[Ler09] R. Lerch, G. Sessler, D. Wolf, “Technische Akustik – Grundlagen und Anwendungen”, ISBN 978-3-540-23430-2, DOI 10.1007/978-3-540-49833-9, Erlangen 2009

[Nov10] A. Novak, L. Simon, F. Kadlec and P. Lotton, "Nonlinear System Identification Using Exponential Swept-Sine Signal," IEEE Transactions on Instrumentation and Measurement, Volume 59, August 2010, doi: 10.1109/TIM.2009.2031836

[Kre10] B. Kreczmer, “Objects Localization and Differentiation Using Ultrasonic Sensors”, ISBN: 978-953-7619-83-1, InTech, DOI: 10.5772/9271.

[Aar11] R. M. Aarts, A. J. E. M. Janssen, “Spatial impulse responses from a flexible baffled circular piston”, Journal of the Acoustical Society of America, Volume 129, Issue 5, DOI: 10.1121/1.3562558, 2011

[Kan11] G. Kaniak, “Improved ultrasonic environment perception in air for specular scenes”, Dissertation, Technical University of Vienna, 2010, http://katalog.ub.tuwien.ac.at/AC08417581

[Gia12] N.I. Giannoccaro, L. Spedicato, C. Di Castri, “A New Strategy for Spatial Reconstruction of Orthogonal Planes Using a Rotating Array of Ultrasonic Sensors”, IEEE Sensors Journal, vol. 12, May 2012, doi: 10.1109/JSEN.2011.2170062

[Wal11] C. Walter, G. Kaniak, H. Schweinzer, “Improvements on the Robustness of Ultrasonic Bearing Measurement Methods”, 15th International Conference on Sensors and Measurement Technology, Nürnberg, Juni 2011

[Wal12a] C. Walter, H. Schweinzer, S. Mohammad, “Methods for 3D Scene Analysis and Localization based on Ultrasound”, Informationstagung Mikroelektronik 2012, Wien, April 2012

[Wal12b] C. Walter, H. Schweinzer, “An accurate compact ultrasonic 3D sensor using broadband impulses requiring no initial calibration”, 2012 IEEE International Conference on Instrumentation and Measurement Technology Conference (I2MTC), Graz, 2012, DOI: 10.1109/I2MTC.2012.6229255

[Wal12c] C. Walter, H. Schweinzer, “Hochauflösende Laufzeitmessung zur genauen Abstandsbestimmungan Luft”, 1. Workshop des Fachausschusses Ultraschall der DEGA e.V., Messtechnische Anwendungen von Ultraschall, Juni 2012

[Moh12] M. Syafrudin, H. Schweinzer, C. Walter, “An automated calibration method for the-Local Positioning System LOSNUS”, 2012 IEEE International Conference on Instrumentation and Measurement Technology Conference (I2MTC), Graz, 2012, DOI: 10.1109/I2MTC.2012.6229256

[Das13] M. Daschewski, R. Boehm, J. Prager, M. Kreutzbruck, A. Harrer, “Physics of thermo-acoustic sound generation”, Journal of Applied Physics, Volume 114, issue 11, doi: 10.1063/1.4821121

[Gia13] N.I. Giannoccaro, L. Spedicato, “Exploratory data analysis for robot perception of room environments by means of an in-air sonar scanner”, Ultrasonics, Volume 53, August 2013, doi:10.1016/j.ultras.2013.01.015

[Ste13] J. Steckel, A. Boen, H. Peremans, "Broadband 3-D Sonar System Using a Sparse Array for Indoor Navigation", IEEE Transactions on Robotics, February 2013, doi: 10.1109/TRO.2012.2221313

[Wal13a] C. Walter, H. Schweinzer, “Locating and classifying of objects with a compact ultrasonic 3D sensor”, 2013 IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN), Montbeliard-Belfort, France, October 2013

[Wal13b] C. Walter, M. Syafrudin, H. Schweinzer, “A Self-Contained and Self-Checking LPS with High Accuracy”, ISPRS Int. J. Geo-Inf. 2013, 2, pp. 908-934, 2013, doi:10.3390/ijgi2040908

[Bab14] M. Baba, K. Ohtani, S. Komatsu, "3D shape recognition system by ultrasonic sensor array and genetic algorithms", Instrumentation and Measurement Technology Conference, May 2004, doi: 10.1109/IMTC.2004.1351468

[Moh14] M. Syafrudin, C. Walter, H. Schweinzer, “Robust locating using LPS LOSNUS under NLOS conditions”, 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN), DOI: 10.1109/IPIN.2014.7275531

[Wal14a] C. Walter, H. Schweinzer, “Locating of objects with discontinuities, boundaries and intersections using a compact ultrasonic 3D sensor”, 2014 IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN), Busan, October 2014, DOI: 10.1109/IPIN.2014.7275532

[Wal14b] C. Walter, H. Schweinzer, “Zuverlässige 3D-Ortung von Reflexionspunkten unter Verwendung eines kompakten Ultraschallsensors”, 2. Workshop des Fachausschusses Ultraschall der DEGA e.V., Messtechnische Anwendungen von Ultraschall, Juni 2014, Drübeck

[ST14] ST Microelectronics, “AN4426 – Tutorial for MEMS microphones”, January 2014

[Wal15a] C. Walter, H. Schweinzer, “Reliable recognition and locating of objects using a 3D ultrasound sensor”, Technisches Messen, vol 82, January, 2015, DOI: 10.1515/teme-2014-0033,

[Wal15b] C. Walter, H. Schweinzer, “Robuste 3D-Ortung mit Ultraschall”, Österreichische Patentanmeldung Nr. A 50415/2014.

[Wal15c] C. Walter, H. Schweinzer, “Applications of a compact ultrasonic 3D sensor for recognition of shape and orientation of objects”, 17th International Conference on Sensors and Measurement Technology, Nürnberg

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die beigefügte Dissertation selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel genutzt habe. Alle wörtlich oder inhaltlich übernommenen Stellen habe ich als solche gekennzeichnet.

Ich versichere außerdem, dass ich die beigefügte Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe und, dass diesem Promotionsverfahren keine endgültig gescheiterten Promotionsverfahren vorausgegangen sind.