

Article

Comparison of Direct Intersection and Sonogram Methods for Acoustic Indoor Localization of Persons

Dominik Jan Schott ^{1,*}, Addythia Saphala ², Georg Fischer ³, Wenxin Xiong ⁴, Andrea Gabrielli ¹,
Joan Bordoy ⁴, Fabian Höflinger ¹, Kai Fischer ³, Christian Schindelhauer ⁴ and Stefan Johann Rupitsch ¹

¹ Department of Microsystems Engineering, University of Freiburg, 79110 Freiburg, Germany; and.gabrielli@gmail.com (A.G.); fabian.hoeflinger@imtek.uni-freiburg.de (F.H.); stefan.rupitsch@imtek.uni-freiburg.de (S.J.R.)

² Department of Medical Informatics, Biometry and Epidemiology, Friedrich-Alexander-Universität (FAU), 91052 Erlangen, Germany; addythia.saphala@fau.de

³ Fraunhofer Institute for Highspeed Dynamics, Ernst-Mach-Institute (EMI), 79588 Efringen-Kirchen, Germany; georg.fischer@emi.fraunhofer.de (G.F.); kai.fischer@emi.fraunhofer.de (K.F.)

⁴ Department of Computer Science, University of Freiburg, 79110 Freiburg, Germany; xiongw@informatik.uni-freiburg.de (W.X.); bordoy@informatik.uni-freiburg.de (J.B.); schindel@informatik.uni-freiburg.de (C.S.)

* Correspondence: dominik.jan.schott@imtek.uni-freiburg.de; Tel.: +49-761-203-7185



Citation: Schott, D.J.; Saphala, A.; Fischer, G.; Xiong, W.; Gabrielli, A.; Bordoy, J.; Höflinger, F.; Fischer, K.; Schindelhauer, C.; Rupitsch, S.J. Comparison of Direct Intersection and Sonogram Methods for Acoustic Indoor Localization of Persons. *Sensors* **2021**, *21*, 4465. <https://doi.org/10.3390/s21134465>

Academic Editors: Riccardo Carotenuto, Massimo Merenda and Demetrio Iero

Received: 1 June 2021
Accepted: 21 June 2021
Published: 29 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: We discuss two methods to detect the presence and location of a person in an acoustically small-scale room and compare the performances for a simulated person in distances between 1 and 2 m. The first method is Direct Intersection, which determines a coordinate point based on the intersection of spheroids defined by observed distances of high-intensity reverberations. The second method, Sonogram analysis, overlays all channels' room impulse responses to generate an intensity map for the observed environment. We demonstrate that the former method has lower computational complexity that almost halves the execution time in the best observed case, but about 7 times slower in the worst case compared to the Sonogram method while using 2.4 times less memory. Both approaches yield similar mean absolute localization errors between 0.3 and 0.9 m. The Direct Intersection method performs more precise in the best case, while the Sonogram method performs more robustly.

Keywords: presence detection; passive localization; room impulse response; acoustic localization; indoor localization

1. Introduction

Acoustic localization systems can provide, partly due to the comparably slower wave propagation, a high accuracy indoors similar to radio-based solutions, which are not covered by ubiquitous satellite signals of Global Navigation Satellite Systems (GNSS) [1–3]. For some applications, it may not be desirable to equip persons or objects with additional hardware as trackers due to inconvenience and privacy reasons. Previously, we reported coarsely about indoor localization by Direct Intersection in [4]. In this work, we report in detail on two algorithms for this application and their performances. The proposed system is categorized as a passive localization system [5] and is implemented solely with commercial off-the-shelf (COTS) hardware components.

Echolocation, such as the method used by bats to locate their prey, is a phenomenon where the reflected sound waves are used to determine the location of objects or surfaces that reflect the sound waves due to a change in acoustic impedance. This concept has been extensively used for various investigations in the physics and engineering fields, such as sound navigation and ranging (Sonar) [6,7] and even using only a single transducer for transmission and reception [8].

We draw the approach from bats, which can perceive the incoming reflected wave's direction due to its precise awareness of head angle, body motion, and timing. While the exhaustive echolocation method of bats is not completely understood, one of the more obvious aspects is the back-scattered signals' difference of arrival in time between left and right ears, which can be used to calculate the incoming sound wave's direction [9]. This approach differs from approaches that more generally detect changes in the systems response of a medium, where the responses act like fingerprints. However, in application, insignificant changes in a room may lead to distortions in the response. This makes a better knowledge of the specific room necessary. In contrast, determining times-of-arrival of back-scattered waves is less dependent on the complete impulse response; we therefore chose this approach. We investigate two different algorithms based on the time difference of arrival of the first-order reflection to interpret the returned signals in a small office room of approximately $3\text{ m} \times 4\text{ m} \times 3\text{ m}$ similar to [10], which are characteristic for the strong multipath fading effects that partially overlap and interfere with the line-of-sight reverberations [11]. The signal frequency employed in our experiment is significantly higher than the Schroeder frequency; therefore, we can assume the sound wave behaves much like rays of light [12]. The physiological structure and the shape of the binaural hearing conformation of bats, together with the natural and instinctive ability to perform head movements to eliminate ambiguities, enhances the echolocation and therefore guarantees excellent objects spatial localization [13]. Our system setup is a fixed structure, and we compensate the adaptive bats head movements by adding two additional microphones to the system. Furthermore, we raise the question of the performance of two approaches and compare the memory consumption and execution time.

The detection of more than one person or object is not investigated in this work.

2. Related Work

Indoor presence detection may be achieved through a variety of different technologies and techniques. For one, radio-frequency (RF)-based approaches have been implemented. In general, these may be classified into two different employed techniques: received signal strength indicator (RSSI)- and radio detection and ranging (Radar)-based approaches. The former offers low-complexity systems with cheap hardware [14,15], whereas with the latter one, higher accuracy may be achieved [16]. The other main concept employed in indoor presence detection is using ultrasonic waves, which are applied in active trackers indoors [17,18] and even underwater [19,20]. An entirely passive approach, as in [21], generally analyzes audible frequencies, which can include speech and potentially violate privacy regulations, similar to vision-based approaches. Acoustic solutions, which operate close to or in the audible range, can be perceived by persons and animals alike, which may cause irritation and in the worst case harm [22]. Therefore, special care has to be invested in designing acoustic location systems. While radio-based solutions are less critical in this concern, due to the fact that most organisms lack sensitivity to radio frequency signals, the frequency allocation is much more restrictive due to licensing and regulations. While LIDAR systems are highly accurate, but comparably costly, other light-based systems have gathered interest again, due to their high accuracy potential, with low systems costs and power consumption [23].

2.1. RF-RSSI

Mrazovac et al. [24] track the RSSI between stationary ZigBee communication nodes, detecting changes to infer a presence from it. In the context of home automation, this work is used to switch on and off home appliances. Seshadri et al. [15], Kosba et al. [14], Gunasagaran et al. [25], and Retscher and Leb [26] analyze different signal strength features for usability of detection and identification using standard Wi-Fi hardware. Kaltiokallio and Bocca [27] reduce the power consumption of the detection system by distributed RSSI processing.

This technique was then improved by Yigitler et al. [28], who built a radio tomographic map of the indoor area. The difference from the previously sampled map of RSSI values is the notification of a presence or occupancy. This general concept is known in the field of indoor localization as fingerprinting. Hillyard et al. [29] utilize these concepts to detect border crossings.

2.2. RF-Radar

Suijker et al. [30] present a 24 GHz FMCW (Frequency-Modulated Continuous-Wave) Radar system to detect indoor presence and to be used for intelligent LED lighting systems. An interferometry approach is implemented by Wang et al. [16] for precise human tracking in an indoor environment. Another promising approach in the RF domain is, instead of using a time-reversal approach (as Radar does), deriving properties of the medium (and contained, noncooperative objects) by means of wave front shaping as proposed by del Hougne et al. [31,32]. This approach would also in principle be conceivable in the acoustic wave domain.

2.3. Ultrasonic Presence Detection and Localization

A direct approach to provide room-level tracking is presented by Hnat et al. [33]. Ultrasonic range finders are mounted above doorways to track people passing beneath. More precise localization can be achieved by using ultrasonic arrays as proposed by Caicedo and Pandharipande [9,34]. The arrays' signals can be used to obtain the range and direction-of-arrival (DoA) estimates. The system is used for energy-efficient lighting systems. Pandharipande and Caicedo [7] enhanced this approach to track users by probing and calculating the position via the time difference of arrival (TDoA). Prior to that, Nishida et al. [35] proposed a system consisting of 18 ultrasonic transmitters and 32 receivers, embedded in the ceiling of a room with the aim to track elderly people and prevent them from experiencing accidents. A time-of-flight (ToF) approach was proposed by Bordoy et al. [36], who used a static co-located speaker-microphone pair to estimate human body and wall reflections. Ultrasonic range sensing may be combined with infrared technology, as has been done by Mokhtari et al. [37], to increase the energy efficiency. In lower frequency regimes, the resonance modes of a room start to dominate the measured signals. This fact may be used to deduce source locations as proposed by Nowakowski et al. [38] (cf. [39,40]).

2.4. Ultrasonic Indoor Mapping

Indoor mapping and indoor presence detection are two views of the same problem. In both instances, one tries to estimate the range and direction for a geometrical interpretation. Ribeiro et al. [41] employ a microphone array co-located to a loudspeaker to record the room impulse response (RIR). The multiple reflections can be estimated from this RIR with the use of l_1 -regularization and least-squares (LS) minimization, and a room geometry can be inferred, achieving a range resolution of about 1 m. A random and sparse array of receivers is proposed by Steckel et al. [42] for an indoor Sonar system. In addition to that, the authors use wideband emission techniques to derive accurate three-dimensional (3D) location estimates. This system is then enhanced with an emitter array to improve the signal-to-noise-ratio (SNR) [43]. Another approach, implementing a binaural Sonar sensor, is proposed by Rajai et al. [44]. A sensor was used to detect the wall within a working distance of one meter. In a recent work by Zhou et al. [45], it is shown that a single smartphone with the help of a gyroscope and an accelerometer can be used to derive indoor maps by acoustic probing. Bordoy et al. [46] use an implicit mapping to enhance the performance of acoustic indoor localization by estimating walls and defining virtual receivers as a result of the signals' reflections.

2.5. Algorithms

The first set of methods, which are broadly applied are triangulation algorithms as described by Kundu [47]. In this work we focus on two Maximum-Likelihood approaches,

similar to the one proposed by Liu et al. [48]. The first one, Direct Intersection (DI), uses a Look-up-Table (LUT) and spheres inferred from the sensors delay measurements with error margin [49], while the other one, the Sonogram method, populates a 3D intensity map with probabilities to find likely positions of the asset. Since the approaches of the two methods are different, it is likely to expect different outcomes in accuracy, precision, computational complexity, and memory requirements.

3. System Overview

The system consists of a single acoustic transmitter, a multi-channel receiver, a power distribution board, and a central computer to analyze the recorded signals. Four microphones are placed equidistantly around the speaker and connected to the receiver board. The set-up is shown in Figure 1 as it was used for the experiment reported below.

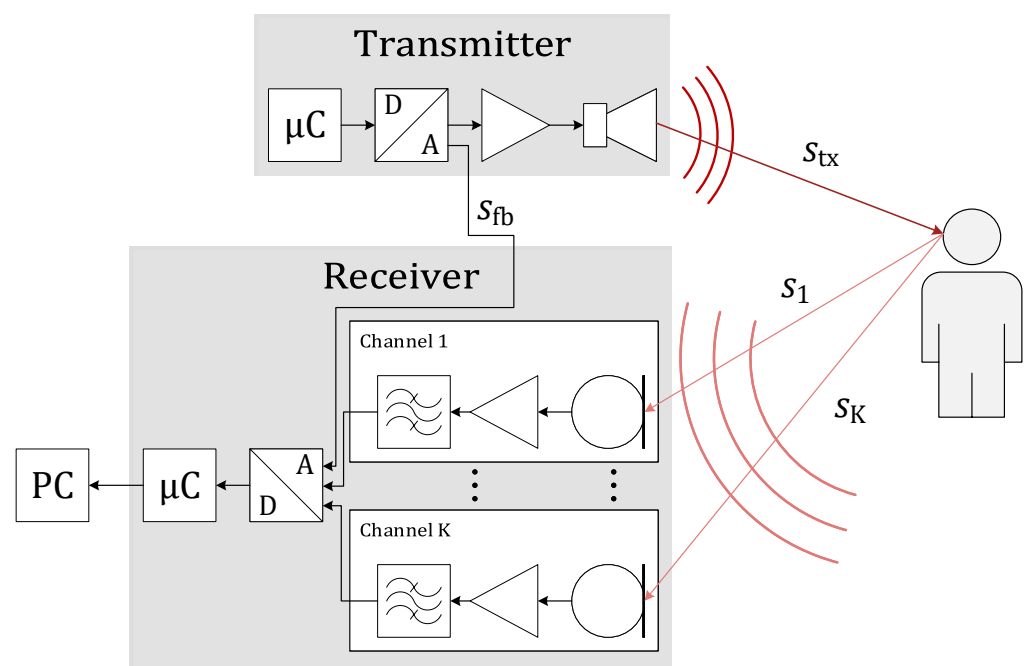


Figure 1. Schematic representation of the system.

3.1. Signal Waveform

Due to their auto-correlation properties and the ability to maximize the Signal-to-Noise-Ratio (SNR) without increasing acoustic amplitude, swept-frequency cosine, i.e., frequency modulated chirp signals, perfectly fit our case-study [50]. Auto-correlated frequency-modulated chirps are able to provide compressed pulses at the correlator output, whose width in time space is defined as follows [51]:

$$P_w = \frac{2}{B}. \quad (1)$$

The frequency-modulated signal employed in our experiments, $x_{Tx}(t)$, is mathematically defined as follows:

$$s_{tx}(t) = \begin{cases} A \cos(2\pi\phi(t)), & \text{for } 0 \leq t \leq T_s \\ 0, & \text{otherwise} \end{cases}, \text{ with} \quad (2)$$

$$\phi(t) = \frac{f_{\text{end}} - f_{\text{start}}}{2T_s} t^2 + f_{\text{start}} t, \quad (3)$$

where A denotes the signal amplitude, f_{start} is the start frequency, f_{end} the end frequency, $B = f_{\text{end}} - f_{\text{start}}$ the frequency bandwidth, T_s is the pulse duration, and $\phi(t)$ the instantaneous phase. The chirp instantaneous frequency is defined as follows:

$$f(t) = f_{\text{start}} + \frac{f_{\text{end}} - f_{\text{start}}}{T_s} t, \quad 0 \leq t \leq T_s. \quad (4)$$

Taking into account the hardware characteristics of our setup, we selected a linear up-chirp pulse with amplitude $A = 1$, $T_s = 5$ ms, $f_{\text{start}} = 16$ kHz, and $f_{\text{end}} = 22$ kHz, which result in a time-bandwidth product of $T B = 30$. The frequency response of a chirp signal directly depends on the Time-Bandwidth ($T B$) product. For chirps with $T B \geq 100$, the pulse frequency response is almost rectangular [52]. However, due to the hardware limitation of our setup, which do not allow a high ($T B$) product, the frequency response will be characterized by ripples. In order to mitigate the spectrum disturbances, we consider a window in the time domain the transmitted chirp pulse with a raised cosine window [52]. The frequency band, chirp length, and shaping window were chosen to minimize the system affecting persons and animals in hearing range. We implemented chirps, due to their property of spreading the signals energy over time compared to a single pulse to limit the maximal amplitude and resulting harmonics. While young and highly audio-sensitive people can in principle hear these frequencies, the short signal length of 5 ms compared to the repetition interval of 1000 ms further reduces the occupation of the low ultrasonic channel. Generally speaking, higher amplitudes and lower frequencies potentially increase the operation range of the system, but this comes at a health risk for humans and animals, which we seek to avoid.

3.2. Hardware Overview

To obtain 3D coordinates with static arrangement, a four-element microphone array is sampled, as well as a feedback signal. This array records the incoming echo wave with different time of arrival, depending on the incoming signal direction. Since unsuitable hardware can affect the system's performance [53], both the microphones and speaker were tested for correct signal generation and reception in an anechoic box.

3.3. Data Acquisition

Each microphone's signal was preconditioned before the digitization by the multi-channel analog-to-digital converter, which was chosen to provide each channel with the identical sample-and-hold trigger flank before conversion. Each frame consists of the signal from each microphone and a feedback, which is recorded as an additional input to estimate and mitigate playback jitter. The first layer of digital signal processing is to compress the signal, extracting the reverberated acoustic amplitude over time and removing the empty room impulse response (RIR).

3.3.1. Channel Phase Synchronization

Initially, we calculate the convolution of the feedback channel signal s_{fb} with our known reference signal s_{ref} in its analytic form to obtain the RIR and retrieve the time of transmission from the compressed signal y_{fb} , as shown in Equation (5), where j denotes the imaginary unit.

$$y_{\text{fb}} = |(s_{\text{fb}} \otimes s_{\text{ref}}) + j \cdot \mathcal{H}(s_{\text{fb}} \otimes s_{\text{ref}})| \quad (5)$$

This compressed analytic form y_{fb} of the feedback signal s_{fb} (see Figure 1) ideally holds only a single pulse from the transmitted signal, if the output stage is impedance matched. Searching for the global maximum returns both time of transmission and the output amplitude.

$$a_{\text{out}} = \max_{t \rightarrow t_0} y_{\text{fb}}(t) \quad (6)$$

In the following, we refer to the start time of a transmission as t_0 , all other channels' time scales are regarded relative to t_0 . Therefore, the signals of the microphone channels

are truncated to remove information prior to the transmission. The ring-down of small office rooms is in the order of 100 ms, so the repetition interval of consecutive transmissions is chosen accordingly to be larger. This prevents leakage of late echos into the following interval, which would result in peaks being recorded after the following interval's line-of-sight. The remaining signal frames from all microphones are compressed with the same approach as the feedback channel, shown in Equations (5) and (6), to extract each channel's compressed analytic signal y_i and line-of-sight detection time t_i .

3.3.2. Baseline Removal

In the following, we refer to the acoustic channel response after the line-of-sight as the echo profile. An example of such echo profiles is shown in Figure 2. While the line-of-sight signal ideally provides the fastest and strongest response, large hard surfaces, like desks, walls, and floors return high amplitudes, which are orders of magnitude above a person's echo. For a linear and stable channel, we can reduce this interference from the environment by subtracting the empty room echo profile from each measurement, following the approach of [54]. This profile loses its validity if the temperature changes, the air is moving, or objects in the room are moved, e.g., an office chair is slightly displaced. A dynamic approach to create the empty room profile is updating an estimation, when no change is observed for an extended time or alternatively using a very low-weight exponential filter to update the room estimation. In this work, the empty office room was sounded N times directly before each test and averaged into an empty room echo profile \bar{y}_i° for each channel i as denoted in Equation (7), to assure unchanged conditions and reduce the complexity of the measurements. The removal itself is then, as mentioned above, the subtraction of the baseline from each measurement, as in Equation (8), under the assumption of coherence.

$$\bar{y}_i^\circ = \text{mean}(y_i^\circ) \quad (7)$$

$$\tilde{y}_i = y_i - \bar{y}_i^\circ \quad (8)$$

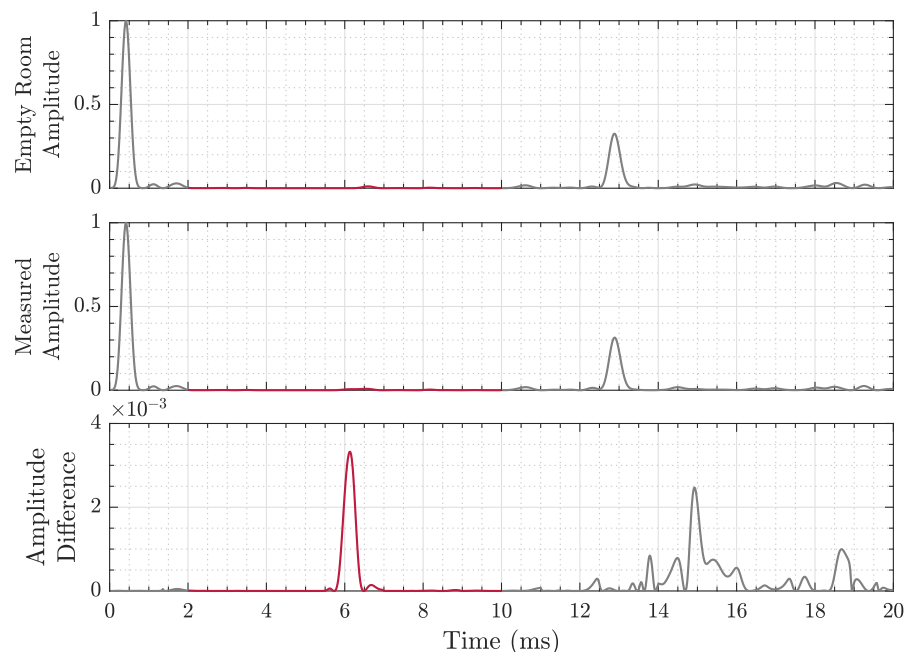


Figure 2. Exemplary magnitude plot of the compressed analytic signal, i.e., RIR, with (top) the baseline drawn from an previous recording of the empty room, (middle) the room with a person in it, and (bottom) the difference of the two above. The red highlighted line in the center marks the area of interest due to geometric constraints. Note the changed scale of the ordinate in the bottom plot.

3.3.3. Time-Gating

For our approach we assume some features of the person, such as being closer to the observing system compared to the distant environment objects, like chairs, tables and monitors, while another area of reverberations is in the close lateral vicinity of the system, consisting, e.g., of lamps and the ceiling. This is exploited by introducing a time gate, which only allows for non-zero values in the interval of interest as in Equation (9) (also compare Figure 2).

$$\tilde{y}_{\text{tg},i} = \begin{cases} \tilde{y}_i, & \text{for } t_{\min} < t < t_{\max} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

Another assumption is that of a small reverberation area on the person. We assume the points of observation from each microphone to be sufficiently close on a person to overlap. The latter assumption introduces an error, which limits the precision of the system in the order of 10 cm [55], which we deem sufficient for presence detection, as a person's dimension is considerably larger in all directions. This estimation is based on the approximate size of a person's skull and its curvature with respect to the distance to the microphones and their spacing. The closer the microphones and the further the distance between head and device, the more the reflection points will approach each other. If we regard a simplified 2D projection, where a person with a spherical head of radius $r_H \approx 10$ cm moves in the y-plane only, the position of a reflection point $R = (x_R, z_R)$ on the head can be calculated by

$$\begin{aligned} x_R &= x_C - r_H \sin(\alpha_R), \quad \text{and} \\ z_R &= z_C - r_H \cos(\alpha_R), \end{aligned} \quad (10)$$

where x_C and z_C are the lateral and vertical center coordinates of the head and α_R is the reflection angle. The latter is calculated through

$$\alpha_R = \tan^{-1} \frac{x_C + \frac{d_M}{2}}{z_C}, \quad (11)$$

with the distance d_M between the microphone and sender. The origin is set as the speaker position. By geometric addition, the distance between two such reflection points can be calculated and reach the maximum value if the head moves towards the center. In this case, the reflection points would be on the opposing sides of the head and result in a mismatch of $2r_H$. The other extreme is laterally moving to an infinite distance, which increases the magnitude of x_C , while the distance between microphone and speaker stays constant; therefore, the reflection points converge to a single point of reflection. In this work, the distance between head center and speaker remained above 120 cm, with a projected error distance of about 1.3 cm.

3.3.4. Echo Profile

During the experiment, the reflected signals from the floor, walls, tables, and chairs have a very high amplitude. This interference can lead to masking the echo from the target object. To reduce the effect of the interference, the empty room profile is used to subtract the target impulse response from the input impulse response. If we define the reflection from objects other than the target object as noise, we can increase the signal-to-noise ratio with this method. The empty room impulse response is also called empty room echo profile in this work. In Figure 2, the upper plot is the empty room impulse response, where the experiment room is cleared of most clutter. The middle plot is the room with single static object as target, shown in Figure 3. The lower plot shows the result of subtraction between the second and first plot, and the scale is adjusted for clarity.

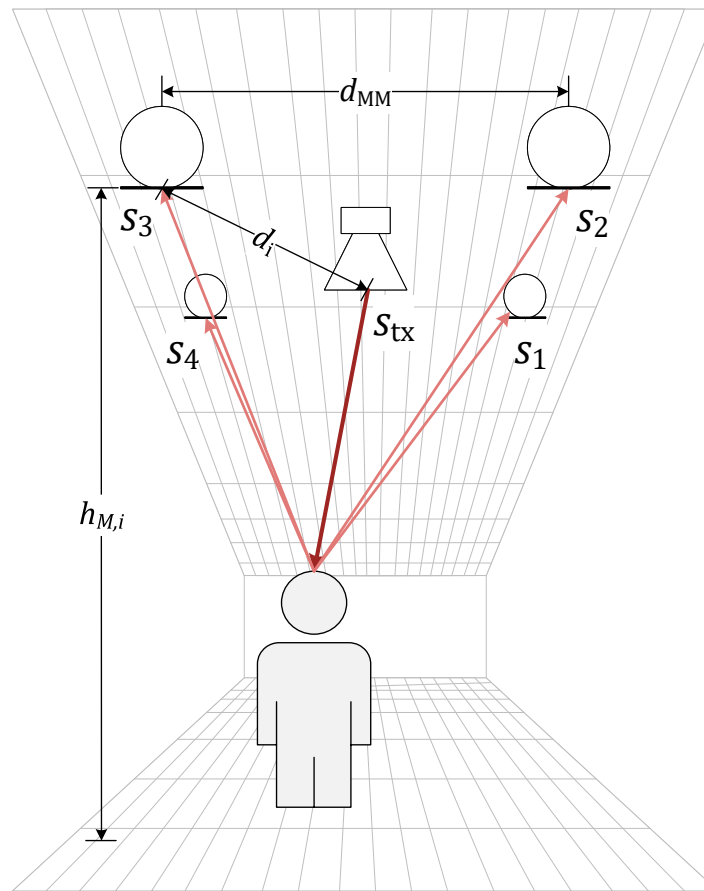


Figure 3. Experimental setup for $K = 4$ receivers spaced by $d_{MM} \approx 0.2$ m. The transmitted signal s_{tx} is observed as reflected signals s_i by the system located near the ceiling of the room.

3.3.5. Distance Maps

Look-up tables are calculated before the experiment to estimate the travel distance of a signal from the speaker to each microphone under the assumption of a direct reverberation from a point at position \vec{x} in the room and linear beam-like signal propagation. This grid is formed by setting the center speaker as origin and spanning up a 3-dimensional Cartesian coordinate system of points \vec{x} through the room in discrete steps. We limit the grid to the intervals X_1 to X_3 in steps of 1 cm to decrease the calculational effort and multipath content under the prior knowledge of the rooms geometry as follows:

$$\begin{aligned} \vec{x} &= (x_1, x_2, x_3) \in \mathbb{X}, \text{ where} \\ \mathbb{X} &= \{X_1 \times X_2 \times X_3\} \subset \mathbb{R}^3. \end{aligned} \quad (12)$$

The look-up table approach serves to minimize the processing time during execution. The distance maps provide pointers to convert from binary sampling points to distance points. Each sub-matrix contains the sum of distance between each point in the room to the corresponding i th microphone at the position $\vec{x}_{M,i}$ and to the speaker at position \vec{x}_S , which cover the flight path of the echoes, as in Equation (13):

$$M_i(\vec{x}) = \|\vec{x} - \vec{x}_S\| + \|\vec{x}_{M,i} - \vec{x}\|. \quad (13)$$

Therefore, the resultant entries in matrices M depend on the geometric arrangement of speaker and microphones, and the matrix size corresponds to the area of detection, as in Equation (12).

3.4. Data Processing

3.4.1. Direct Intersection

The main assumption for this approach (Algorithm 1) is that the highest signal peak in the observation window of each channel indicates the position of interest, as visualized in Figure 2. Each channels' peak index defines the radius r_i of a sphere around each microphone, which is contained in the point cloud L_i . While ideally those spheres overlap in exactly the point of reverberation, in practical application, where noise, interference, and jitters are present, this is not the case. To compensate this error, we pad the sphere by Δr additional points in the radius until all spheres overlap and the unity of valid estimation points U_L is not empty. The sphere radius widening Δr can be used as an indication of each measurement's quality, as a low error case will require little to no padding, while in high-error cases, the required padding will be large. Another approach is to use a fixed and small padding, which will ensure only measurements of high quality to be successful, but will fail for high error scenarios.

Algorithm 1: Direct Intersection Estimation [56,57].

Input : \tilde{y}_{tg} , observed intensity data frames,
 M_i , distance maps,
 K , number of channels,
 Δr_{max} , maximum radius spreading distance.

Output: \vec{x}_{est} , estimated 3D-position.

```

begin
   $\Delta r \leftarrow 0$  // initial estimation tolerance
   $N_{OL} \leftarrow 0$  // number of overlapping points
  for  $i = 1$  to  $K$  do
     $r_i \leftarrow \max_n \tilde{y}_{tg}$  // get index of peak
     $R_i \leftarrow \{r_i\}$ 
  while ( $N_{OL} = 0$ ) & ( $\Delta r < \Delta r_{max}$ ) do
     $\Delta r++$ 
    for  $i = 1$  to  $K$  do
       $R_i \leftarrow \{r_i - \Delta r, R_i, r_i + \Delta r\}$  // recursively add width
       $L_i \leftarrow \text{isMember}(M_i, R_i)$  // select points by radius [56]
     $U_L \leftarrow (L_1 \& \dots \& L_K)$ 
     $\vec{x}_{OL} \leftarrow \text{ind2sub}(\text{size}(U_L), \text{find}(U_L))$  // wrap into 3D coordinates [57]
     $N_{OL} \leftarrow \min(\text{length}(\vec{x}_{OL}))$ 
   $\vec{x}_{est} = \text{mean}(\vec{x}_{OL})$ 

```

3.4.2. Sonogram

The Sonogram approach (Algorithm 2) leverages available memory and processing power to build a 3D intensity map. This approach utilizes the entire echo profile difference shown in Figure 2 (bottom) and maps them into the 3D distance map explained in Section 3.3.5, with the assumption that the highest peak corresponds to the source of reverberation. The multiplication of impulse amplitude that corresponds to the same coordinates is used as an indication of possible reverberation source. Therefore, the maximum result would have the highest likelihood of being the reverberation source location.

Algorithm 2: Sonogram Estimation [58,59].

Input : \tilde{y}_{tg} , observed intensity data frames,
 M_i , distance maps,
 K , number of channels,
 Δr_{max} , maximum radius spreading distance.

Output: \vec{x}_{est} , estimated 3D-position.

begin

```

 $\tilde{y}_+ \leftarrow \{\tilde{y}_{\text{tg}} > 0\}$  // remove negative intensities
forall  $\vec{x}$  do
   $I(\vec{x}) \leftarrow \prod_{i=1}^K \tilde{y}_{+,i}(M_i(\vec{x}))$ 
   $A_{x1} \leftarrow \max_{x_1 \in X_1}(I)$  // 2D matrix
   $\vec{a}_{x2} \leftarrow \max_{x_2 \in X_2}(A_{x1})$  // 1D vector
   $x_{3,\text{est}} \leftarrow \max_{x_3 \in X_3}(\text{smooth}(\vec{a}_{x2}))$  // scalar, moving average smoothed [58]
   $x_{2,\text{est}} \leftarrow \text{find}(\vec{a}_{x2}, x_{3,\text{est}})$  // select first matching value [59]
   $x_{1,\text{est}} \leftarrow \text{find}(A_{x1}, \{x_{2,\text{est}}, x_{3,\text{est}}\})$ 
   $\vec{x}_{\text{est}} \leftarrow (x_{1,\text{est}}, x_{2,\text{est}}, x_{3,\text{est}})$ 

```

4. Experiments

4.1. Set-Up

In the experiment, we use a mock-up representing a person's head as the experiment target. The hard and smooth surface of the object is intentional for the sake of usability and to remove unintended movements from our measurements at this early stage. In the set-up shown in Figure 3, the central speaker emits the well-known signal s_{tx} , and the reflected echoes from the target s_1 to s_4 are recorded by the microphone array around the speaker. The depiction in Figure 3 is exaggerated for clarity.

Table 1 shows the spherical coordinates, i.e., radial distance r , azimuth angle θ , and elevation angle ϕ of the target inside the room, with the center of the device as the reference point. The device is positioned on the ceiling, oriented downward. For each position, we measure the distance for the assumed acoustic path with a laser distance meter Leica DISTO™ D3a BT for reference. As mentioned above, the coordinate system's point of origin is set to the center of the device, the x-axis is set perpendicular to the entrance door's wall, and increasing towards the right, the y-axis is parallel to the line of sight from the door and increasing towards the rear end of the room, and the z-axis is zero in the plane of the device (upper ceiling lamp level) and decreasing towards the floor. The two-dimensional depictions are shown in Cartesian coordinates to provide clarity, while the detection results are done in spherical coordinates.

Table 1. Reference Positions.

| Position | r (m) | θ (°) | ϕ (°) |
|----------|---------|--------------|------------|
| ① | 1.58 | 77 | 59 |
| ② | 1.70 | −92 | 57 |
| ③ | 1.23 | −35 | 54 |
| ④ | 1.26 | 169 | 54 |

4.2. Results

4.2.1. Room Properties and Impulse Response

In preparation for the later experiments, we sounded the room 100 times as described in Section 3.3.2 to record the baseline profiles shown in Figures 4 and 5. This recordings were taken one time and served as a reference for all later experiment runs. During the recordings, the room was left closed and undisturbed.

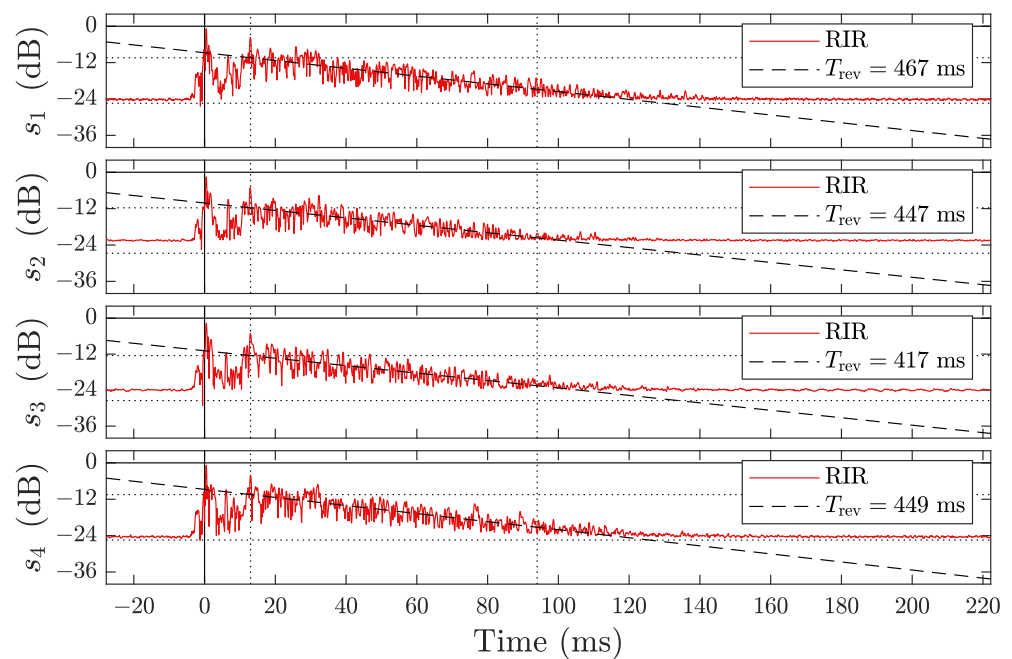


Figure 4. Empty room's impulse response magnitude of a linear chirp ($T_s = 5$ ms, 16 to 22 kHz) in logarithmic scale for all 4 channels s_1 to s_4 . The red line indicates the mean response over 100 measurements, with a linear fit indicated by a black dashed line in the interval between 13 to 94 ms (dotted vertical lines) to approximate the reverberation time constant T_{rev} of the room, given in the legend of each channel's subplot. The upper horizontal dotted line indicate the fit's level at $t = 13$ ms, while the lower indicates an additional drop by -20 dB.

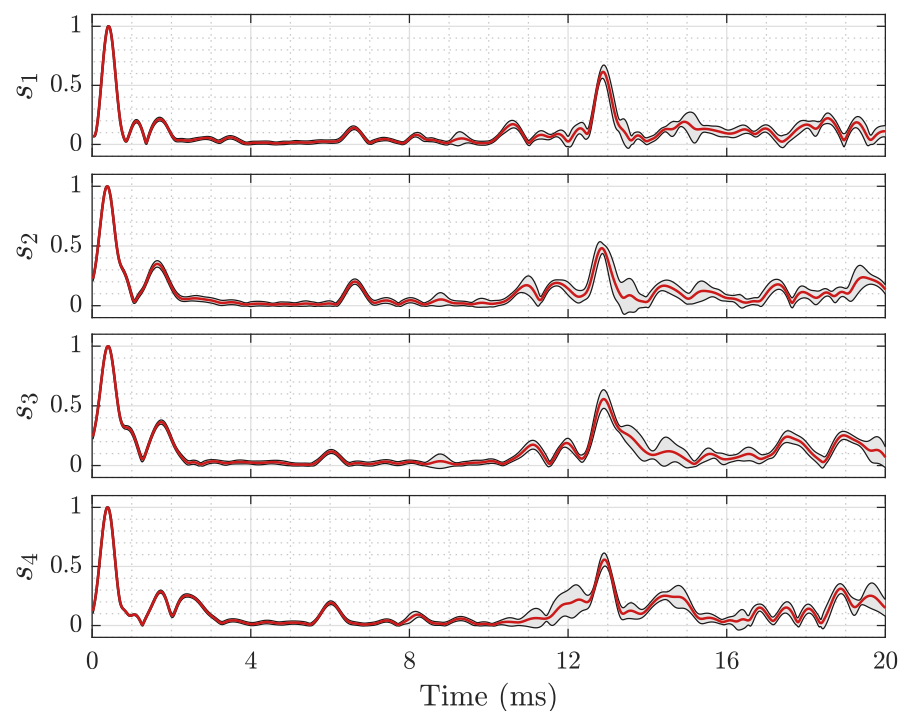


Figure 5. First 20 ms of the empty room's amplitude response for all 4 channels s_1 to s_4 . The red line indicates the mean response over 100 measurements, the grey envelope the $\pm 3\sigma$ region. The first peak marks the line-of-sight arrival time and is used for time synchronization.

The room exhibits a different room response for each microphone, as illustrated in Figure 4. We divide the response into four parts: line-of-sight, free space transition, first

order echoes, and higher order echoes, i.e., coda [60]. The signal remains in the room for more than 100 ms, before it drops below the noise floor. The definition of the reverberation time from Sabine requires a drop of the sound levels below -60 dB [61,62], for which the low signal-to-noise ratio of less than 24 dB does not suffice. Therefore, we adapted a fractional model and extrapolated the reverberation from a drop of 20 dB. The resulting mean reverberation time of the room is approximately $\bar{T}_{\text{rev}} \approx 445$ ms, which corresponds to a dampening factor $\delta \approx 15.5 \text{ s}^{-1}$ and a Schroeder frequency of approximately $f_{\text{sch}} \approx 230$ Hz, which is far below the transmission band. In this work, we focus on the response in the parts-free space transition and first-order echoes to estimate a person's position. A close-up of the first three parts of the room response is shown in Figure 5.

The recordings still show significant variances in each channel at varying positions, e.g., in the uppermost subplot of Figure 5 from 15 to 16 ms. Below 8 ms, these intervals with increased variances do not occur, indicating a stable channel. The signals' interval close to zero contains strong wall and ceiling echos. Note the very strong reverberation peak at 12.5 to 13.5 ms that is caused by the floor. As our area of interest does not fall within this distance, we omit it for analysis as well. Hence, the time-gate limits as introduced in Section 3.3.3 are $t_{\text{min}} = 3$ ms and $t_{\text{max}} = 8$ ms.

If we transfer the room dimensions into the wavelength space, hence

$$\Lambda = \frac{l}{\lambda_g} = \frac{\ell f_g}{c}, \quad (14)$$

with c as the speed of sound and l the room dimensions in the respective Cartesian direction, we can draw an estimator from [63] for the number of modes below the reference frequency f_g as

$$N_{\text{mode}} = \frac{4\pi}{3}(\Lambda_x \Lambda_y \Lambda_z) + \frac{\pi}{2}(\Lambda_x \Lambda_y + \Lambda_y \Lambda_z + \Lambda_z \Lambda_x) + \frac{1}{2}(\Lambda_x + \Lambda_y + \Lambda_z). \quad (15)$$

This lets us calculate approximately 15×10^6 modes below 16 kHz and 40×10^6 modes below 22 kHz, which leaves about 25×10^6 modes in the sounding spectrum in-between. If we regard the number of eigenfrequencies below the Schroeder frequency, Equation (15) yields $N_{\text{sch}} \approx 73$ modes that strongly influence the sound characteristics of the room [64].

4.2.2. Direct Intersection

The localization by Direct Intersection from all 100 runs is shown for each of the four reference positions in Figure 6. While the statistical evaluation is performed in spherical coordinates due to the geometric construction during the estimation, this overview plots, as well as those for the Sonogram localization are drawn in Cartesian coordinates that allow for easier verification and intuitive interpretation. The lateral spread of the estimation point cloud in Figure 6 ① is misleading as the points are situated on a sphere around the origin. The projected lateral extent is almost entirely due to the angular errors.

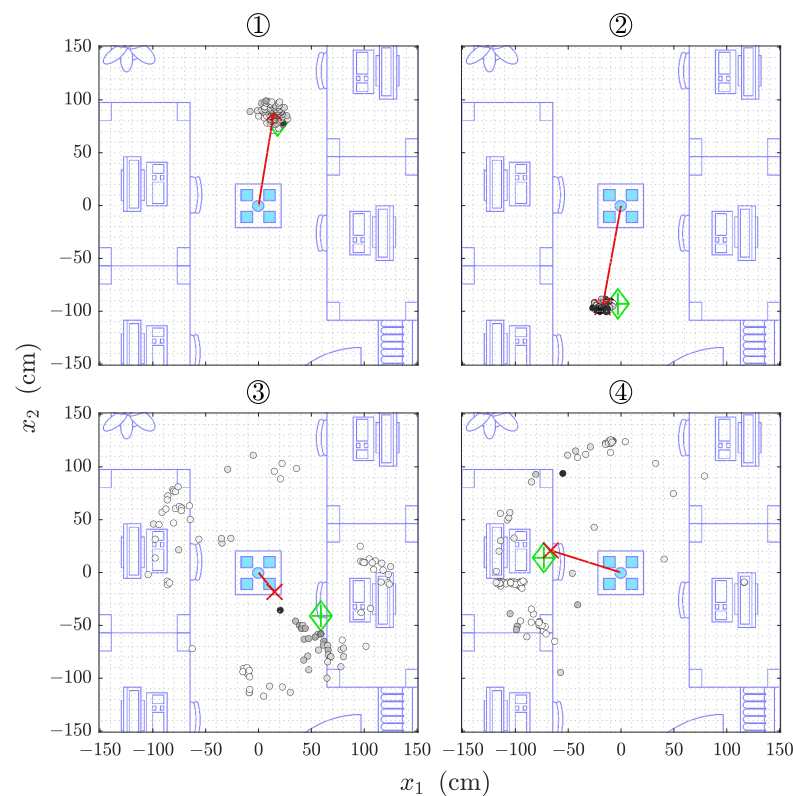


Figure 6. 2D projection of 100 estimations of 3D positions ① to ④ by Direct Intersection. The single estimations are indicated by the black circled markers, the red cross marks the Cartesian averaged position and is highlighted by the red line to the origin, and the green diamond indicates the reference position. The points' infill is proportional to the observed intensity relative to the radius spreading (darker is higher).

Positions ① and ② show a distance estimation deviation of $\sigma_r \approx 10$ cm, as well as azimuth and elevation angle errors of $\sigma_\theta \approx \sigma_{CE} < 5^\circ$ for both Direct Intersection and Sonogram localization (compare Tables 2 and 3). For positions ③ and ④, which are situated closer to the desks, the deviation increases to almost 40 cm in distance and almost arbitrary azimuth angles with $\sigma_\theta \approx 120^\circ$ and more, but a far less affected elevation angle estimation with a $\sigma_\theta < 10^\circ$. The deviations are calculated around the mean estimator for each value. For simplicity of interpretation, the mean error for each dimension is shown in Section 4.2.3.

Table 2. Direct Intersection Estimated Positions.

| Position | r (m) | θ ($^\circ$) | ϕ ($^\circ$) |
|----------|-----------------|-----------------------|---------------------|
| ① | 1.83 ± 0.14 | 81 ± 4 | 61 ± 1 |
| ② | 2.01 ± 0.11 | -100 ± 3 | 61 ± 1 |
| ③ | 1.92 ± 0.37 | 4 ± 96 | 59 ± 4 |
| ④ | 2.12 ± 0.25 | -58 ± 135 | 60 ± 3 |

Table 3. Sonogram Estimated Positions.

| Position | r (m) | θ ($^\circ$) | ϕ ($^\circ$) |
|----------|-----------------|-----------------------|---------------------|
| ① | 1.85 ± 0.10 | 80 ± 4 | 58 ± 2 |
| ② | 2.03 ± 0.11 | -100 ± 3 | 60 ± 2 |
| ③ | 1.77 ± 0.26 | -41 ± 69 | 47 ± 7 |
| ④ | 1.96 ± 0.34 | 31 ± 119 | 51 ± 9 |

The error distributions for each dimension are shown in Figure 7, where each column depicts one of the spherical dimensions (radius, azimuth angle, and elevation angle), while each row represents the results from the reference position indicated to the left of the plot. For the first two positions, the distributions are almost unimodal, but for the latter two, this does not hold true, making the mean value and standard deviation unsuitable estimators.

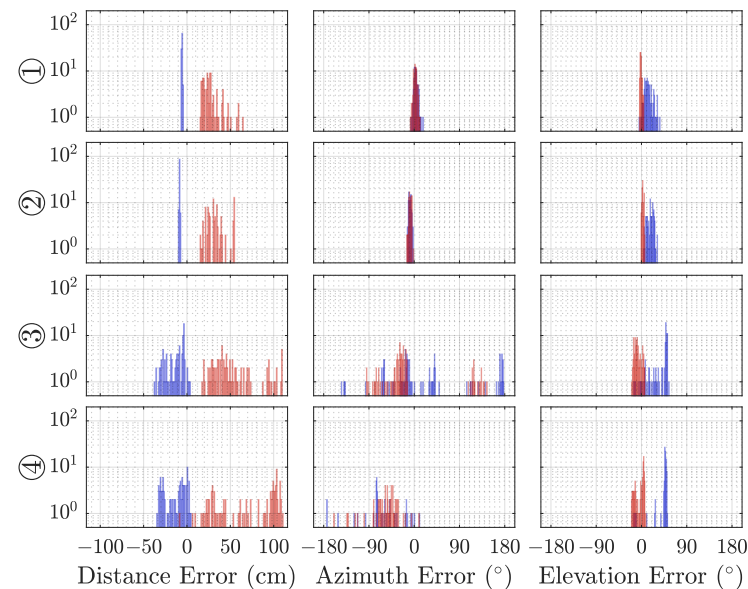


Figure 7. Histograms of the error in estimation compared to the reference over 100 localization repetitions at each position by DI (blue) and Sonogram (red) estimation. Each row depicts the 3 degrees of freedom for each position.

The distribution of the error in the absolute distance between the estimated positions and reference positions (see Figure 8) is likewise a few dozen centimeters for the first two cases, but around 1 m for the latter two. If we recall the reference positions from Table 1, the true distances are between 1 and 2 m, which puts the error in the same order as the expected value.

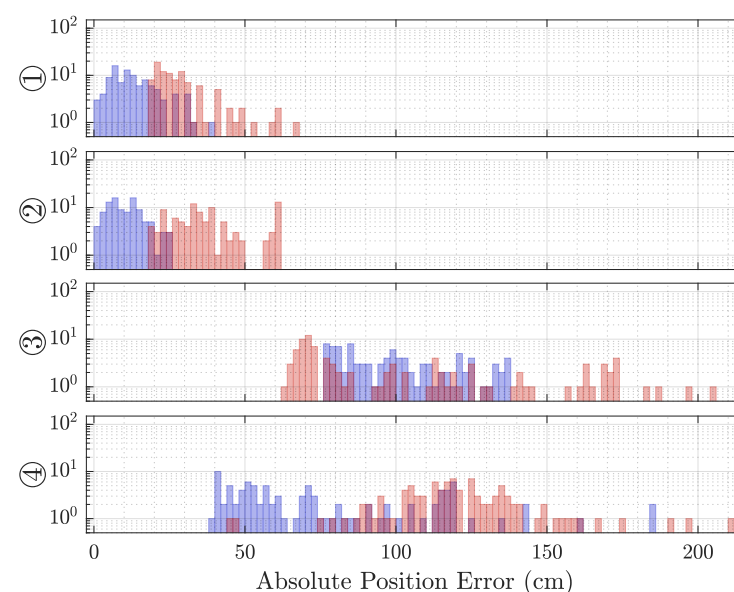


Figure 8. Histograms of the absolute distance error in estimation compared to the reference over 100 localization repetitions at each position by DI (blue) and Sonogram (red). Each row depicts the 3 degrees of freedom for each position.

The Direct Intersection method allows for an investigation into the time variance of the detected maximum peak, which is depicted in Figure 9. In the first two cases, we observe unimodal distributions of around 10 samples in width, while the latter cases show detected peaks all over the interval.

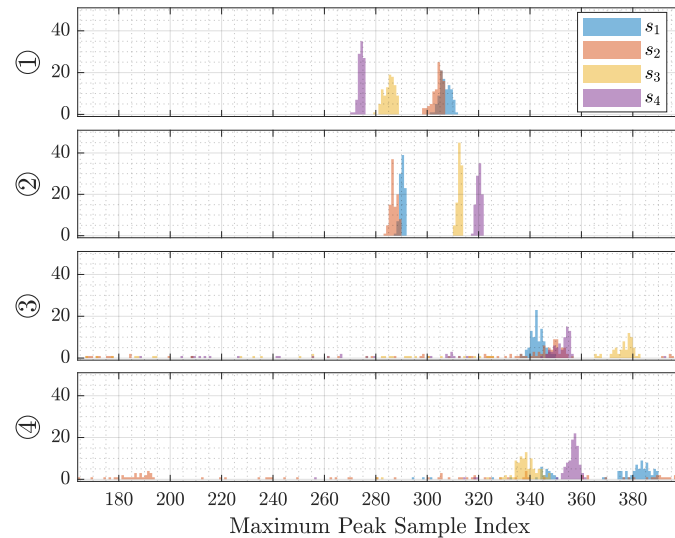


Figure 9. Histograms of the highest peak position of each microphone’s channel over 100 localization repetitions at each position by DI. Each row depicts the 3 degrees of freedom for each position.

4.2.3. Sonogram

The Sonogram localization on the same data as before in Section 4.2.2 is shown in Figure 10 for all four cases. The lateral distribution of the estimated locations is not following the spherical shape as closely as is the case for those by Direct Intersection estimations (compare, e.g., Figure 6 ①).

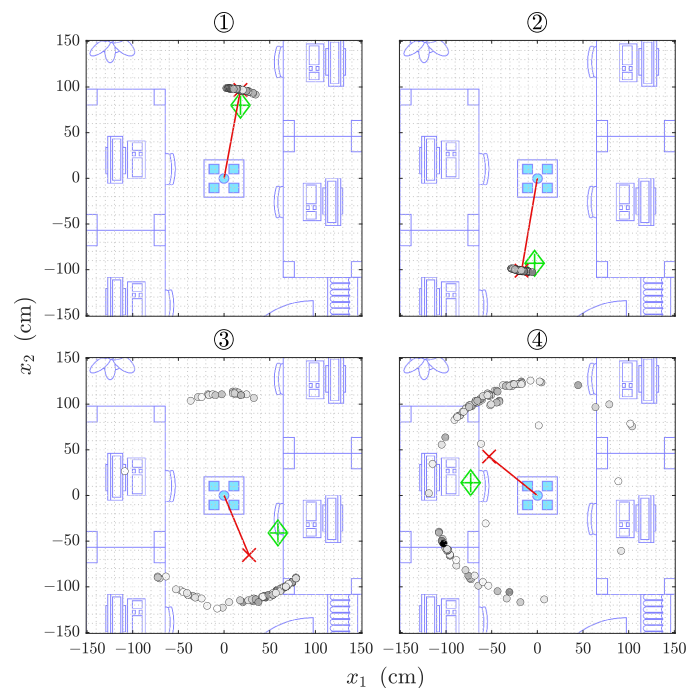


Figure 10. The same position estimation plot as in Figure 6 for positions ① to ④ but by Sonogram. The reference position is given by the green diamond, the averaged estimation by the red cross, and each circle represents a single estimated position. The circles’ infill is proportional to the observed intensity.

Similar to before, the method performs well in the cases ① and ②, exhibiting small deviations (see Table 3), but far less precise with the largest deviation increase in the azimuth angle as well. The corresponding mean errors to the reference positions are listed in Table 4.

Table 4. Mean Error for Direct Intersection and Sonogram.

| Position | Direct Intersection | | | Sonogram | | |
|----------|---------------------|--------------|------------|----------|--------------|------------|
| | r (m) | θ (°) | ϕ (°) | r (m) | θ (°) | ϕ (°) |
| ① | 0.25 | 3 | 2 | 0.27 | 2 | 1 |
| ② | 0.31 | 8 | 4 | 0.34 | 8 | 3 |
| ③ | 0.69 | 39 | 5 | 0.53 | 6 | 7 |
| ④ | 0.87 | 47 | 6 | 0.70 | 138 | 3 |

The cases ③ and ④ display two larger clusters of estimated positions, which leads to the bimodal error distributions in Figure 7.

The absolute error is similarly distributed around lower values for the former two cases and widely spread for the two latter cases (see Figure 8). Note that the error distribution plots for the Sonogram are of slightly different horizontal scale, as no errors below 20 cm were observed, while the observed maximal error exceeds 200 cm.

Lastly, the performance of both algorithms with regard to execution time is listed in Table 5 and mean required memory in Table 6. The distribution of those measures is shown in Figures 11 and 12. The Direct Intersection method requires roughly $2.4\times$ less memory than the Sonogram localization. With a best-case mean execution time of 0.66 s, the former algorithm is almost $1.7\times$ faster than the best case mean of the latter method, while the worst-case mean—almost unchanged for the Sonogram approach—is with a factor of 7.1 for the Direct Intersection by far slower than the worst case mean execution time of the Sonogram method.

Table 5. Runtime Performance: Time.

| Position | Direct Intersection Time (s) | Sonogram Time (s) |
|----------|------------------------------|-------------------|
| ① | 0.94 ± 0.17 | 1.14 ± 0.07 |
| ② | 0.66 ± 0.13 | 1.20 ± 0.02 |
| ③ | 6.38 ± 6.60 | 1.10 ± 0.01 |
| ④ | 8.58 ± 7.12 | 1.10 ± 0.01 |

Table 6. Runtime Performance: Memory.

| Direct Intersection Memory ($\times 10^8$ bit) | Sonogram Memory ($\times 10^8$ bit) |
|---|--------------------------------------|
| 1.600 ± 0.004 | 3.840 ± 0.002 |

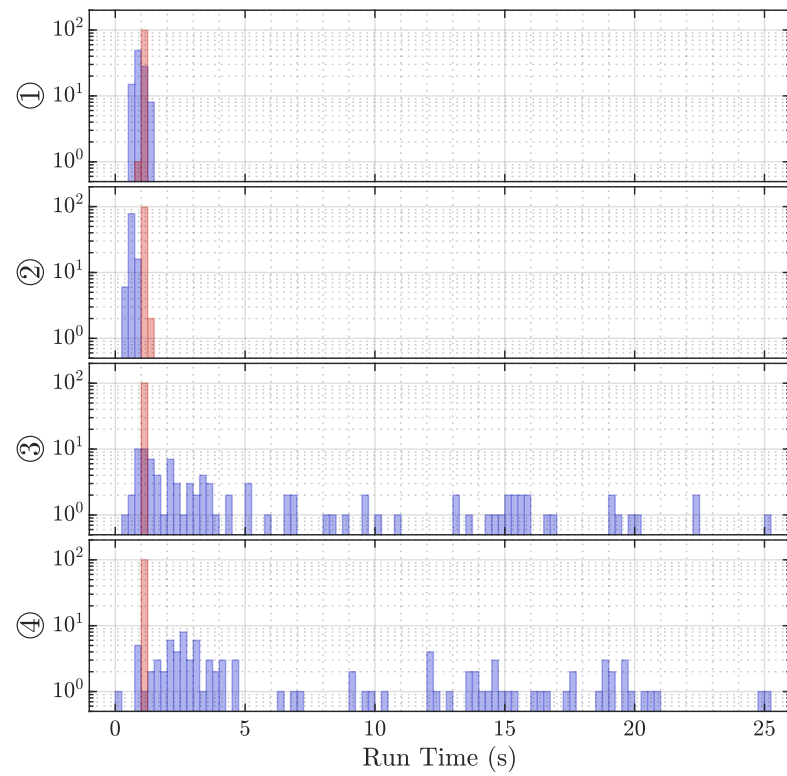


Figure 11. Histograms of the execution time of 100 localization repetitions at each position by DI (blue) and Sonogram (red).

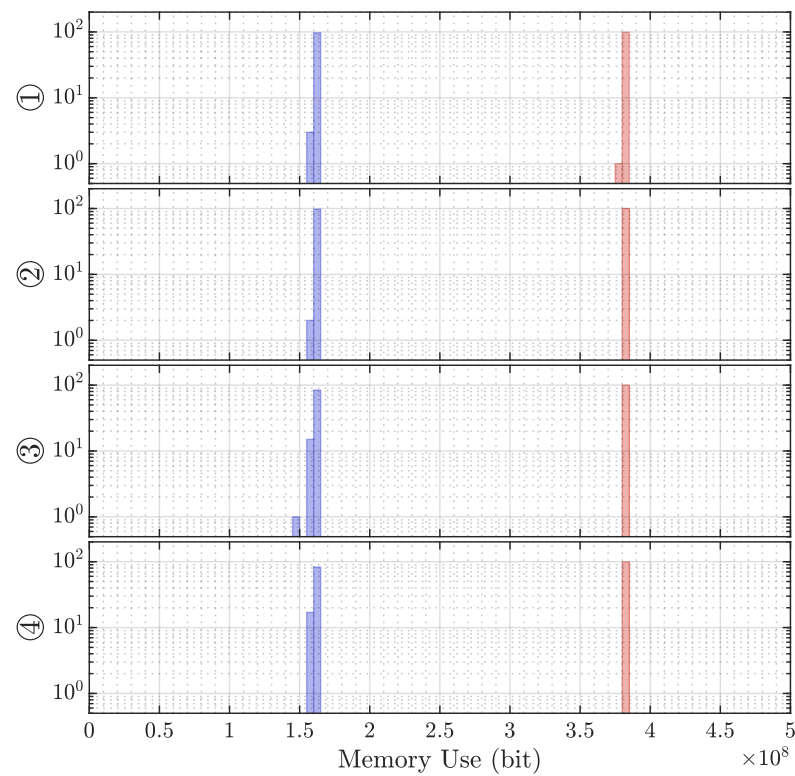


Figure 12. Histograms of the memory allocation during 100 localization repetitions at each position by DI (blue) and Sonogram (red).

The Direct Intersection execution time varies strongly, as we observe it anywhere between 0.25 and 25.0 s; thus, without further limitations, it does not allow for a well-confined prediction of the localization algorithm's execution time.

5. Discussion

5.1. Localization

The localization methods discussed in Section 4 are based on the time of arrival of the line-of-sight reflection from the target. This is possible because the frequency-modulated signal in our experiments is significantly higher than the Schroeder frequency of the room. The Direct Intersection method provides throughout all cases distance estimations that are too short, while the Sonogram-based localization returns distance estimations that are longer than the reference (compare Figure 7). Regarding the absolute error distribution, we observe that the Direct Intersection method performs more accurately, especially in the better cases ① and ②, as well more precise in the first three of the four observed cases, as drawn from Figure 8. The possible cause of the degradation of both methods performance for cases ③ and ④ is in the peak detection algorithm, as Figure 9 shows a wide error range of detected possible peaks. While this was observed specifically for the Direct Intersection method, this also implies the low signal-to-noise ratio of the underlying echo profile, and consequently also affects the Sonogram estimation. Interestingly, the lower estimation errors for cases ① and ② implicate a better performance for the larger distances than the closer ones, which is counter-intuitive from a power perspective, but if we recall the empty room impulse responses shown in Figure 5, where noise is included as the curves' variance, and compare it to the magnitudes of a person's signal in Figure 2, the difference in magnitude is in the same order. For higher distances, the variance increases, as fluctuations in the speed of sound cause phase distortions, but for lower distances, interference effects dominate. The frequency band of the chirp between 16 and 22 kHz sets the wavelength range to approximately 2.2 to 1.6 cm, which is close to the distance between reflection points on a person's head, as shown above in Section 3.3.3. Proximity to objects increases interference as well, which explains the lower performance in the closer positions ③ and ④, where the projected distance onto the sensor system's aperture between the person and the wall, screen, and desk is reduced. If we regard the error distributions of each position in Figure 7 again, the angles and distances roughly fit non-line-of-sight paths, especially for the Sonogram method.

5.2. Performance

The Direct Intersection method requires less than half the memory for its computations compared to the Sonogram method, as the information is very early condensed in the peak selection part of the algorithm. The index look-up is in itself a cheap operation, but due to the sphere-spreading loop to decrease the probability of the algorithm not returning any valid position at all, comes at higher execution duration. The observed worst case for Direct Intersection is with 25 s so high that no real-time tracking is possible anymore. If we look closer at Figure 6 ③, we see that the estimation point gray scale infill is proportional to the inverse spreading factor, so darker colors mean less radial spread before intersecting points could be found. The notion that including strong outliers by allowing the sphere thickness to be spread so far is not confirmed if we consider Figure 6 ④.

6. Conclusions

Both methods show mean distance estimation errors ranging between approximately 0.3 and 0.9 m for objects in distances between 1.2 and 1.7 m, with angular errors between 2° and 138° in azimuth, 1° and 7° in elevation. The Sonogram Estimation allows for analysis of room response in more detail, and the results are more accurate (i.e., average error) in three out of four observed cases, but inversely, the precision (i.e., error variance) of the Direct Intersection is higher in three of the cases. The Direct Intersection method allows for less expensive computation by reducing maximum radius spreading, while the Sonogram

method's cost can be reduced effectively by limiting the vertical search interval, e.g., to the clutter free area above the desks. For a full-range sounding of the room, we observed that the locations close to the clutter area are estimated worse regarding both accuracy and precision. For a pragmatic operation on hardware with higher memory limitations the Direct Intersection method will perform faster and with similar precision and accuracy, and can be limited in execution time by restricting the sphere radius spreading at the cost of not being able to estimate the position for several intervals. We esteem further investigation into limiting the degradation of the estimation process by single unreliable channels as most promising for improving passive acoustic indoor localization.

Author Contributions: Conceptualization, D.J.S., A.S., and F.H.; methodology, D.J.S., A.S., and J.B.; software, D.J.S. and A.S.; validation, D.J.S., A.S., G.F., W.X., A.G., and J.B.; formal analysis, D.J.S. and A.S.; investigation, D.J.S. and A.S.; resources, F.H., K.F., C.S., and S.J.R.; data curation, A.S.; writing—original draft preparation, D.J.S., A.S., G.F., W.X., A.G., and J.B.; writing—review and editing, D.J.S., G.F., W.X., and A.G.; visualization, D.J.S.; supervision, J.B. and F.H.; project administration, F.H., K.F., C.S., and S.J.R.; funding acquisition, F.H., K.F., C.S., and S.J.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the Fraunhofer Gesellschaft and the state of Baden-Württemberg in the Framework of the MERLIN project, and also the German Ministry of Education and Research (BMBF) under the grant FKZ 16ME0023K (“Intelligentes Sensorsystem zur autonomen Überwachung von Produktionsanlagen in der Industrie 4.0 - ISA4.0”).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to express their gratitude to the anonymous reviewers for many useful suggestions and support in deepening their understanding of acoustics.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|------|--|
| MDPI | Multidisciplinary Digital Publishing Institute |
| DI | Direct Intersection |
| DoA | Direction of Arrival |
| FMCW | Frequency-Modulated Continuous-Wave |
| LS | Least-Squares |
| RF | Radio-Frequency |
| RIR | Room Impulse Response |
| RSSI | Received Signal Strength Indicator |
| SONO | Sonogram |
| SNR | Signal-to-Noise Ratio |
| TDoA | Time Difference of Arrival |
| ToF | Time of Flight |

References

1. Zafari, F.; Gkelias, A.; Leung, K.K. A Survey of Indoor Localization Systems and Technologies. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 2568–2599. [[CrossRef](#)]
2. Billa, A.; Shayea, I.; Alhammedi, A.; Abdullah, Q.; Roslee, M. An Overview of Indoor Localization Technologies: Toward IoT Navigation Services. In Proceedings of the 2020 IEEE 5th International Symposium on Telecommunication Technologies (ISTT), Shah Alam, Malaysia, 9–11 November 2020; pp. 76–81. [[CrossRef](#)]
3. Obeidat, H.; Shuaieb, W.; Obeidat, O.; Abd-Alhameed, R. A Review of Indoor Localization Techniques and Wireless Technologies. *Wirel. Personal Commun.* **2021**. [[CrossRef](#)]

4. Höflinger, F.; Saphala, A.; Schott, D.J.; Reindl, L.M.; Schindelbauer, C. Passive Indoor-Localization using Echoes of Ultrasound Signals. In Proceedings of the 2019 International Conference on Advanced Information Technologies (ICAIT), Yangon, Myanmar, 6–7 November 2019; pp. 60–65. [\[CrossRef\]](#)
5. Pirzada, N.; Nayan, M.Y.; Subhan, F.; Hassan, M.F.; Khan, M.A. Comparative analysis of active and passive indoor localization systems. *AASRI Procedia* **2013**, *5*, 92–97. [\[CrossRef\]](#)
6. Caicedo, D.; Pandharipande, A. Distributed Ultrasonic Zoned Presence Sensing System. *IEEE Sens. J.* **2014**, *14*, 234–243. [\[CrossRef\]](#)
7. Pandharipande, A.; Caicedo, D. User localization using ultrasonic presence sensing systems. In Proceedings of the 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Seoul, Korea, 14–17 October 2012; pp. 3191–3196. [\[CrossRef\]](#)
8. Kim, K.; Choi, H. A New Approach to Power Efficiency Improvement of Ultrasonic Transmitters via a Dynamic Bias Technique. *Sensors* **2021**, *21*, 2795. [\[CrossRef\]](#)
9. Caicedo, D.; Pandharipande, A. Transmission slot allocation and synchronization protocol for ultrasonic sensor systems. In Proceedings of the 2013 10th IEEE International Conference on Networking, Sensing and Control (ICNSC), Evry, France, 10–12 April 2013; pp. 288–293. [\[CrossRef\]](#)
10. Carotenuto, R.; Merenda, M.; Iero, D.; Della Corte, F.G. An Indoor Ultrasonic System for Autonomous 3-D Positioning. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 2507–2518. [\[CrossRef\]](#)
11. Patwari, N.; Ash, J.; Kyperountas, S.; Hero, A.; Moses, R.; Correal, N. Locating the nodes: Cooperative localization in wireless sensor networks. *IEEE Signal Process. Mag.* **2005**, *22*, 54–69. [\[CrossRef\]](#)
12. Kuttruff, H. Geometrical room acoustics. In *Room Acoustics*, 5th ed.; Spon Press, Taylor & Francis: Abingdon-on-Thames, UK, 2009.
13. Zhang, S.; Ma, X.; Dong, Z.; Zhou, W. Ultrasonic Spatial Target Localization Using Artificial Pinnae of Brown Long-eared Bat. *bioRxiv* **2020**. [\[CrossRef\]](#)
14. Kosba, A.E.; Saeed, A.; Youssef, M. RASID: A robust WLAN device-free passive motion detection system. In Proceedings of the 2012 IEEE International Conference on Pervasive Computing and Communications, Lugano, Switzerland, 19–23 March 2012. [\[CrossRef\]](#)
15. Seshadri, V.; Zaruba, G.; Huber, M. A Bayesian sampling approach to in-door localization of wireless devices using received signal strength indication. In Proceedings of the Third IEEE International Conference on Pervasive Computing and Communications, Kauai, HI, USA, 8–12 March 2005; pp. 75–84. [\[CrossRef\]](#)
16. Wang, G.; Gu, C.; Inoue, T.; Li, C. A Hybrid FMCW-Interferometry Radar for Indoor Precise Positioning and Versatile Life Activity Monitoring. *IEEE Trans. Microw. Theory Tech.* **2014**, *62*, 2812–2822. [\[CrossRef\]](#)
17. Bordoy, J.; Schott, D.J.; Xie, J.; Bannoura, A.; Klein, P.; Striet, L.; Höflinger, F.; Häring, I.; Reindl, L.; Schindelbauer, C. Acoustic Indoor Localization Augmentation by Self-Calibration and Machine Learning. *Sensors* **2020**, *20*, 1177. [\[CrossRef\]](#)
18. Pullano, S.A.; Bianco, M.G.; Critello, D.C.; Menniti, M.; La Gatta, A.; Fiorillo, A.S. A Recursive Algorithm for Indoor Positioning Using Pulse-Echo Ultrasonic Signals. *Sensors* **2020**, *20*, 5042. [\[CrossRef\]](#)
19. Schott, D.J.; Faisal, M.; Höflinger, F.; Reindl, L.M.; Bordoy Andreú, J.; Schindelbauer, C. Underwater localization utilizing a modified acoustic indoor tracking system. In Proceedings of the 2017 IEEE 7th International Conference on Underwater System Technology: Theory and Applications (USYS), Kuala Lumpur, Malaysia, 18–20 December 2017. [\[CrossRef\]](#)
20. Chang, S.; Li, Y.; He, Y.; Wang, H. Target Localization in Underwater Acoustic Sensor Networks Using RSS Measurements. *Appl. Sci.* **2018**, *8*, 225. [\[CrossRef\]](#)
21. Cobos, M.; Antonacci, F.; Alexandridis, A.; Mouchtaris, A.; Lee, B. A Survey of Sound Source Localization Methods in Wireless Acoustic Sensor Networks. *Wirel. Commun. Mob. Comput.* **2017**, *2017*, 3956282. [\[CrossRef\]](#)
22. Hage, S.R.; Metzner, W. Potential effects of anthropogenic noise on echolocation behavior in horseshoe bats. *Commun. Integr. Biol.* **2013**, *6*, e24753. [\[CrossRef\]](#)
23. Rahman, A.B.M.M.; Li, T.; Wang, Y. Recent Advances in Indoor Localization via Visible Lights: A Survey. *Sensors* **2020**, *20*, 1382. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Mrazovac, B.; Bjelica, M.; Kukolj, D.; Todorovic, B.; Samardzija, D. A human detection method for residential smart energy systems based on Zigbee RSSI changes. *IEEE Trans. Consum. Electron.* **2012**, *58*, 819–824. [\[CrossRef\]](#)
25. Gunasagaran, R.; Kamarudin, L.M.; Zakaria, A. Wi-Fi For Indoor Device Free Passive Localization (DfPL): An Overview. *Indones. J. Electr. Eng. Inform. (IJEEI)* **2020**, *8*. [\[CrossRef\]](#)
26. Retscher, G.; Leb, A. Development of a Smartphone-Based University Library Navigation and Information Service Employing Wi-Fi Location Fingerprinting. *Sensors* **2021**, *21*, 432. [\[CrossRef\]](#)
27. Kaltiokallio, O.; Bocca, M. Real-Time Intrusion Detection and Tracking in Indoor Environment through Distributed RSSI Processing. In Proceedings of the 2011 IEEE 17th International Conference on Embedded and Real-Time Computing Systems and Applications, Toyama, Japan, 28–31 August 2011. [\[CrossRef\]](#)
28. Yigitler, H.; Jantti, R.; Kaltiokallio, O.; Patwari, N. Detector Based Radio Tomographic Imaging. *IEEE Trans. Mob. Comput.* **2018**, *17*, 58–71. [\[CrossRef\]](#)
29. Hillyard, P.; Patwari, N.; Daruki, S.; Venkatasubramanian, S. You’re crossing the line: Localizing border crossings using wireless RF links. In Proceedings of the 2015 IEEE Signal Processing and Signal Processing Education Workshop (SP/SPE), Salt Lake City, UT, USA, 9–12 August 2015. [\[CrossRef\]](#)

30. Suijker, E.M.; Bolt, R.J.; van Wanum, M.; van Heijningen, M.; Maas, A.P.M.; van Vliet, F.E. Low cost low power 24 GHz FMCW Radar transceiver for indoor presence detection. In Proceedings of the 2014 44th European Microwave Conference, Rome, Italy, 6–9 October 2014; pp. 1758–1761. [CrossRef]
31. Del Hougne, P.; Imani, M.F.; Fink, M.; Smith, D.R.; Lerosey, G. Precise Localization of Multiple Noncooperative Objects in a Disordered Cavity by Wave Front Shaping. *Phys. Rev. Lett.* **2018**, *121*, 063901. [CrossRef]
32. Del Hougne, M.; Gigan, S.; del Hougne, P. Deeply Sub-Wavelength Localization with Reverberation-Coded-Aperture. *arXiv* **2021**, arXiv:2102.05642.
33. Hnat, T.W.; Griffiths, E.; Dawson, R.; Whitehouse, K. Doorjamb: Unobtrusive room-level tracking of people in homes using doorway sensors. In Proceedings of the The 10th ACM Conference on Embedded Network Sensor Systems (SenSys 2012), Toronto, ON, Canada, 6–9 November 2012. [CrossRef]
34. Caicedo, D.; Pandharipande, A. Ultrasonic array sensor for indoor presence detection. In Proceedings of the 2012 20th European Signal Processing Conference (EUSIPCO), Bucharest, Romania, 27–31 August 2012; pp. 175–179.
35. Nishida, Y.; Murakami, S.; Hori, T.; Mizoguchi, H. Minimally privacy-violative human location sensor by ultrasonic Radar embedded on ceiling. In Proceedings of the 2004 IEEE Sensors, Vienna, Austria, 24–27 October 2004. [CrossRef]
36. Bordoy, J.; Wendeborg, J.; Schindelbauer, C.; Reindl, L.M. Single transceiver device-free indoor localization using ultrasound body reflections and walls. In Proceedings of the 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Banff, AB, Canada, 13–16 October 2015. [CrossRef]
37. Mokhtari, G.; Zhang, Q.; Nourbakhsh, G.; Ball, S.; Karunanithi, M. BLUESOUND: A New Resident Identification Sensor—Using Ultrasound Array and BLE Technology for Smart Home Platform. *IEEE Sens. J.* **2017**, *17*, 1503–1512. [CrossRef]
38. Nowakowski, T.; de Rosny, J.; Daudet, L. Robust source localization from wavefield separation including prior information. *J. Acoust. Soc. Am.* **2017**, *141*, 2375–2386. [CrossRef] [PubMed]
39. Conti, S.G.; de Rosny, J.; Roux, P.; Demer, D.A. Characterization of scatterer motion in a reverberant medium. *J. Acoust. Soc. Am.* **2006**, *119*, 769. [CrossRef]
40. Conti, S.G.; Roux, P.; Demer, D.A.; de Rosny, J. Measurements of the total scattering and absorption cross-sections of the human body. *J. Acoust. Soc. Am.* **2003**, *114*, 2357. [CrossRef]
41. Ribeiro, F.; Florencio, D.; Ba, D.; Zhang, C. Geometrically Constrained Room Modeling With Compact Microphone Arrays. *IEEE Trans. Audio Speech Lang. Process.* **2012**, *20*, 1449–1460. [CrossRef]
42. Steckel, J.; Boen, A.; Peremans, H. Broadband 3-D Sonar System Using a Sparse Array for Indoor Navigation. *IEEE Trans. Robot.* **2013**, *29*, 161–171. [CrossRef]
43. Steckel, J. Sonar System Combining an Emitter Array With a Sparse Receiver Array for Air-Coupled Applications. *IEEE Sens. J.* **2015**, *15*, 3446–3452. [CrossRef]
44. Rajai, P.; Straeten, M.; Alirezaee, S.; Ahamed, M.J. Binaural Sonar System for Simultaneous Sensing of Distance and Direction of Extended Barriers. *IEEE Sens. J.* **2019**, *19*, 12040–12049. [CrossRef]
45. Zhou, B.; Elbadry, M.; Gao, R.; Ye, F. Towards Scalable Indoor Map Construction and Refinement using Acoustics on Smartphones. *IEEE Trans. Mob. Comput.* **2020**, *19*, 217–230. [CrossRef]
46. Bordoy, J.; Schindelbauer, C.; Hoeflinger, F.; Reindl, L.M. Exploiting Acoustic Echoes for Smartphone Localization and Microphone Self-Calibration. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 1484–1492. [CrossRef]
47. Kundu, T. Acoustic source localization. *Ultrasonics* **2014**, *54*, 25–38. [CrossRef] [PubMed]
48. Liu, C.; Wu, K.; He, T. Sensor localization with Ring Overlapping based on Comparison of Received Signal Strength Indicator. In Proceedings of the 2004 IEEE International Conference on Mobile Ad-hoc and Sensor Systems, Fort Lauderdale, FL, USA, 25–27 October 2004. [CrossRef]
49. Dmochowski, J.P.; Benesty, J.; Affes, S. A Generalized Steered Response Power Method for Computationally Viable Source Localization. *IEEE Trans. Audio Speech Lang. Process.* **2007**, *15*, 2510–2526. [CrossRef]
50. Cook, C. Pulse Compression-Key to More Efficient Radar Transmission. *Proc. IRE* **1960**, *48*, 310–316. [CrossRef]
51. Springer, A.; Gugler, W.; Huemer, M.; Reindl, L.; Ruppel, C.C.W.; Weigel, R. Spread spectrum communications using chirp signals. In Proceedings of the IEEE/AFCEA EUROCOMM 2000. Information Systems for Enhanced Public Safety and Security, Munich, Germany, 19 May 2000; pp. 166–170. [CrossRef]
52. Milewski, A.; Sedek, E.; Gawor, S. Amplitude Weighting of Linear Frequency Modulated Chirp Signals. In Proceedings of the 2007 IEEE 15th Signal Processing and Communications Applications, Eskisehir, Turkey, 11–13 June 2007; pp. 383–386. [CrossRef]
53. Carotenuto, R.; Merenda, M.; Iero, D.; G. Della Corte, F. Simulating Signal Aberration and Ranging Error for Ultrasonic Indoor Positioning. *Sensors* **2020**, *20*, 3548. [CrossRef] [PubMed]
54. Schnitzler, H.U.; Kalko, E.K.V. Echolocation by Insect-Eating Bats: We define four distinct functional groups of bats and find differences in signal structure that correlate with the typical echolocation tasks faced by each group. *BioScience* **2001**, *51*, 557–569. [CrossRef]
55. Saphala, A. Design and Implementation of Acoustic Phased Array for In-Air Presence Detection. Master’s Thesis, Faculty of Engineering, University of Freiburg, Freiburg, Germany, 2019.
56. MathWorks. ismember. Available online: <https://de.mathworks.com/help/matlab/ref/double.ismember.html> (accessed on 27 May 2021).

-
57. MathWorks. ind2sub. Available online: <https://www.mathworks.com/help/matlab/ref/ind2sub.html> (accessed on 27 May 2021).
 58. MathWorks. smooth. Available online: <https://de.mathworks.com/help/curvefit/smooth.html> (accessed on 27 May 2021).
 59. MathWorks. find. Available online: <https://de.mathworks.com/help/matlab/ref/find.html> (accessed on 27 May 2021).
 60. Dylan Mikesell, T.; van Wijk, K.; Blum, T.E.; Snieder, R.; Sato, H. Analyzing the coda from correlating scattered surface waves. *J. Acoust. Soc. Am.* **2012**, *131*, EL275–EL281. [[CrossRef](#)] [[PubMed](#)]
 61. Kuttruff, H. Decaying modes, reverberation. In *Room Acoustics*, 5th ed.; Spon Press, Taylor & Francis: Abingdon-on-Thames, UK, 2009.
 62. Lerch, R.; Sessler, G.M.; Wolf, D. Statistische Raumakustik. In *Technische Akustik*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2009. [[CrossRef](#)]
 63. Lerch, R.; Sessler, G.M.; Wolf, D. Wellentheoretische Raumakustik. In *Technische Akustik*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2009. [[CrossRef](#)]
 64. Kuttruff, H. Steady-state sound field. In *Room Acoustics*, 5th ed.; Spon Press Taylor & Francis: Abingdon-on-Thames, UK, 2009.