Götz, Georg; Martinez Ornelas, Abraham; Schlecht, Sebastian; Pulkki, Ville

Autonomous robot twin system for room acoustic measurements

Published in:
Journal of the Audio Engineering Society

DOI:
10.17743/jaes.2021.0002

Published: 01/04/2021

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Götz, G., Martinez Ornelas, A., Schlecht, S., & Pulkki, V. (2021). Autonomous robot twin system for room acoustic measurements. *Journal of the Audio Engineering Society*, *69*(4), 261-272.
https://doi.org/10.17743/jaes.2021.0002

Autonomous robot twin system for room acoustic measurements

Georg Götz*, Abraham Martinez Ornelas, Sebastian J. Schlecht, and Ville Pulkki

Aalto Acoustics Lab, Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland

Whilst room acoustic measurements can accurately capture the sound field of real rooms, they are usually time consuming and tedious if many positions need to be measured. Therefore, this contribution presents the Autonomous Robot Twin System for Room Acoustic Measurements (ARTSRAM) to autonomously capture large sets of room impulse responses with variable sound source and receiver positions. The proposed implementation of the system consists of two robots, one of which is equipped with a loudspeaker, while the other one is equipped with a microphone array. Each robot contains collision sensors, thus enabling it to move autonomously within the room. The robots move according to a random walk procedure to ensure a big variability between measured positions. A tracking system provides position data matching the respective measurements. After outlining the robot system, this paper presents a validation, in which anechoic responses of the robots are presented and the movement paths resulting from the random walk procedure are investigated. Additionally, the quality of the obtained room impulse responses is demonstrated with a sound field visualization. In summary, the evaluation of the robot system indicates that large sets of diverse and high-quality room impulse responses can be captured with the system in an automated way. Such large sets of measurements will benefit research in the fields of room acoustics and acoustic virtual reality.

0. Introduction

Various physical phenomena interact when sound propagates inside a room. Consequently, the resulting sound field and its spatial variations can be fairly complex [1, 2]. Between any two positions, the combined effects on sound waves can be summarized in terms of a transfer-function. Its time-domain representation is called room-impulse response (RIR).

Knowledge about the sound field in rooms is important for various applications, including acoustic planning in architectural acoustics [3–7], source localization, enhancement, and separation algorithms [8–10], active room compensation systems [11], or six-degrees-of-freedom (6DoF) audio rendering for computer games and virtual / augmented reality (VR/AR) [3, 12]. In general, the sound field inside a room can be reconstructed over large areas if the spatial Nyquist theorem is fulfilled [13], i.e., if the sound field is sampled with a sufficient quantity of measurement points. The theoretically required number of measurements can become large, especially if the sound field consists of high frequencies [13]. Various approaches exist to reduce the required number of measurements with technical means [14–16]. For 6DoF audio rendering, inaccuracies of human hearing can be exploited to reduce the number of required measurements while ensuring a plausible listening experience [17].

RIRs are usually obtained by computing simulations or by conducting room acoustic measurements. Although room simulation algorithms have been extensively studied, current approaches still lack the accuracy and perceptual quality that can be achieved with room acoustic measurements [18]. However, room acoustic measurements require considerable measurement effort, thus making them impractical for collecting large amounts of RIRs.

In principle, room acoustic measurements do not necessarily need to be conducted manually by an experimenter. Instead, they could be automated by using robots to save valuable hours of work. Previous studies demonstrate that room acoustic measurements can be automated with robot systems [5–7, 9, 10, 19]. However, some questions still remain open regarding the flexibility and versatility of the robots, the collection of position-dependent RIRs at room scale, and the system's autonomy. In this paper, we want to address these questions by proposing a new robot system for RIR measurements. We believe that this is a good opportunity to share design decisions and insights we gained while developing the system. This might help other researchers to speed up the process of building similar systems.

*Correspondence should be addressed to: georg.gotz@aalto.fi

The main objective of this project was to develop a flexible, versatile, and cost-efficient RIR measurement system, utilizing off-the-shelf components. The resulting system should be capable of measuring large quantities of position-dependent RIRs at room-scale in various types of rooms. Therefore, we introduce the Autonomous Robot Twin System for Room Acoustic Measurements (ARTSRAM). The ARTSRAM consists of two robots, one of which is equipped with a loudspeaker, thus acting as a sound source. The other robot is equipped with a first-order microphone array to capture RIRs. Both robots are equipped with collision sensors, thus allowing them to move autonomously within a room without requiring any prior mapping of the room or external control by an experimenter. The robots are moving according to a random walk algorithm [20, 21]. A tracking system provides position data of both robots matching the respective measurements. Investigations of the random walk implementation in a test scenario and simulations indicate a sufficiently uniform coverage of the room. Furthermore, measurements acquired with the ARTSRAM provide reasonable room acoustic information that are exemplarily used in this paper for calculating a sound field visualization of a shoebox-shaped room.

The remainder of this paper is organized as follows. Section 1 reviews prior work on robot systems for room acoustic measurements. Section 2 clarifies the design principles behind the ARTSRAM and highlights how they differ from previous studies. Section 3 introduces the robot system and describes its setup as well as all of the used components. In Section 4, the system is validated from four perspectives. Firstly, anechoic measurements of the utilized mobile loudspeaker are presented. Secondly, they are compared with anechoic measurements of the entire robot system to determine the sound field alterations caused by the robot fixture. Thirdly, we evaluate the proposed random walk procedure in a test measurement and compare the results with simulations. Finally, we use a set of ARTSRAM measurements and demonstrate their validity by calculating a sound field visualization for the measured room. Section 5 discusses the results and Section 6 concludes the paper.

1. Prior Work

Robots have already been used in previous studies for conducting various room acoustic measurement tasks. For example, Witew et al. [5] built a truss-based apparatus to move 32 microphones with an arbitrarily fine resolution along a 5.3 m $\times$ 8 m measurement area. Using this device, they measured large sets of RIRs in an automated way to visualize the sound fields of a concert hall and demonstrate the scattering effect caused by the auditorium seats. Xiang et al. [6] used a step-motor-driven mechanism to move microphones along a grid inside a scale model. The motivation behind their work was to investigate the relationship between receiver positions, aperture size, and the sound decay of coupled rooms. Čmejla et al. [9] used a robot arm at a fixed position inside a 6 m $\times$ 6 m room to move a loudspeaker within a 46 cm $\times$ 36 cm $\times$ 32 cm cube. They measured RIRs with six static microphone arrays, while moving the loudspeaker inside the cube with a uniform horizontal and vertical resolution of 2 cm and 4 cm, respectively. The measurements were conducted in a laboratory room with variable reverberation times to build up a database of RIRs that can be used to evaluate source enhancement, localization, and separation algorithms. Uehara et al. [7] proposed to use measurements conducted by a robot for visualizing the distribution of room acoustic parameters. Their developed robot is operated with a remote control. It carries a cardioid microphone to measure RIRs at various positions in a room, while the sound source is kept stationary. Other recent work by Feng et al. [19] used mobile robots for conducting in-situ surface impedance measurements of several material samples inside a room. After a preliminary decomposition of the scene and a selection of all desired measurement positions, the robots work autonomously. Lastly, Le Roux et al. [10] conceptualized a robot system for collecting datasets of speech utterances inside rooms with various acoustic environments. Their robot was mainly conceptualized for collecting datasets of speech, but they hypothesized that their robot system could also be used for acquiring RIRs in the future.

2. Design Principles

Although robot systems have previously been used for room acoustic measurements, none of them has met our requirements. In this project, our goal was to develop a system that is flexible regarding different rooms, measures RIRs at room scale, doesn't require prior path and grid planning, mainly uses off-the-shelf components, and works without external supervision. In this section, we want to highlight why these design principles were important factors during the development of the ARTSRAM.

A desirable property of an RIR measurement robot would be, that it can be used in many different rooms and in an ad hoc way, without setting up large amounts of additional hardware or cables. The installation time of the robot should be low to leverage the full potential of automating the measurements. This is especially important if the system is supposed to be used for different rooms. A measurement system that takes a long time to set up might easily end up being more inefficient than manual measurements by an experimenter when many rooms are measured.

Furthermore, it is important to think about whether the system is supposed to be used on a microscopic or a macroscopic scale. When it is required to cover only a small measurement area, a simple robot arm or turntable might be sufficient. In contrast, RIR measurement over large areas of the room would require robots that are capable of covering bigger distances.

Another important factor that needs to be considered in large-scale measurements is the autonomy of the robot system. In grid-based approaches, robots execute measurements along pre-defined grid points in a room. Such systems can work nicely in regular geometries, but setting up grids and corresponding movement paths for irregular room shapes or rooms with obstacles may be time-

*JAES Vol. 69, No. 4, April 2021 — this is the author's accepted manuscript.
Published version: https://www.aes.org/e-lib/browse.cfm?elib=21033*

consuming. Moreover, if the robots are supposed to move automatically, they require some mechanism that enables them to accurately move to a specific grid point and verify their position. A precise positioning of the robots can be realized using position-tracking solutions or computer-vision-based approaches with visual markers. Both alternatives require considerable implementation efforts, especially if the robots are supposed to be used in different rooms. Additionally, accurate positioning may be disturbed by non-ideal lighting conditions or electromagnetic interference with the tracking system.

In contrast, measurement procedures based on random walks do not require prior path and grid planning. A random walk generates a path by making random steps. Each random step involves moving into a random direction with a random step size [20]. Random walks can also generate paths on pre-defined graphs. In this case, each step consists of randomly choosing a neighbouring node [21]. The theory of random walks is well researched. For scenarios with limited complexity, quantitative measures, such as the time until a node is first visited *(access time)*, the time until the walk returns to its origin *(return time)*, or the time until all nodes were visited *(cover time)*, can be derived analytically [20–22]. For example, finite graphs consisting of N nodes have a cover time with a lower bound of $N \log(N)$ [22]. In our use case, such a lower bound translates into the intuitive assumption that RIR measurement robots moving according to a random walk will require more steps to reach a certain coverage than grid-based robots. This drawback is balanced by the reduced path planning time and the increased flexibility of the random walk procedure.

The aforementioned design principles also impose certain limitations on the choice of components. For example, the desired mobility of the robots reduces the alternatives for loudspeakers and microphones, because weight and cabling need to be considered. Furthermore, the autonomy of robots and their versatility with respect to room types requires a robot system with collision sensors or a certain awareness of its surrounding. Our design choices will therefore be highlighted in the following section.

3. Robot-based measurement system

In this section, we describe the novel Autonomous Robot Twin System for Room Acoustic Measurements (ARTSRAM). The ARTSRAM is capable of measuring room impulse responses (RIRs) with variable sound source and receiver positions. It consists of two independent robots that are able to move freely in a room. Both robots are equipped with collision sensors, thus allowing them to explore the room autonomously. The measurements of RIRs are complemented with corresponding position information of the robots.

3.1. Measurement system overview

The aim of this section is to outline an overview of the ARTSRAM and the communication between its components. More detailed elaborations on the single components will follow in the subsequent sections.

Figure 1 illustrates the overall structure of the robot system. The base for each robot is an iRobot Create 2 Roomba robot, which is controlled by a Raspberry Pi single-board computer. HTC Vive trackers of the second generation are used for tracking the position of the robots. Although it would be possible to mount several microphone arrays and loudspeakers on each of the robots to enable multi-way RIR measurements, we chose to clearly distinguish between a source and a receiver robot in this paper. The source robot uses a Minirig MRBT-2 portable loudspeaker to play back excitation signals that are recorded by the receiver robot with a Zoom H3-VR first-order microphone array. The resulting RIRs are stored on a microSD card inserted into the receiver robot's Raspberry Pi.

The entire measurement procedure is implemented in Python. It is controlled by a main measurement script running on a separate measurement laptop. The main measurement script sends commands over a TCP network socket to the Raspberry Pis of the source and receiver robot. Subsequently, the corresponding server scripts running on the Raspberry Pis handle the commands. For example, the server scripts can trigger robot movements or an RIR measurement. The HTC Vive trackers directly communicate with the measurement laptop over another wireless connection, thus allowing the main measurement script to immediately access and store the positions of both robots.

The Python source code of the ARTSRAM is publicly available, c.f. Section 8.

3.2. Robot components in detail

This section provides more detailed descriptions of the single components of the ARTSRAM. Additionally, we aim to explain our motivation behind choosing the components.

3.2.1. Robot base

The iRobot Create 2 Roomba robot is used as a base for the measurement robots. The robot has a circular shape with a diameter of approximately 35 cm and a height of approximately 10 cm. It is equipped with collision and step sensors.

The Roomba robots are connected over serial port to USB connectors with the Raspberry Pis, thus enabling the server scripts running on the Raspberry Pis to trigger movements of the robots or access the robot's sensor information. Several high-level APIs are available for this purpose, such as the Python package `irobot` [23], which was also used in this project. Due to the Roomba's way of construction, the robots can either spin around their center axis or move along straight or curved paths on floor level. This means, the robots can effectively move with three degrees-of-freedom. Collision sensor information is used to prevent the robots from hitting obstacles while exploring the room. The vacuum functionality of the robots is disabled throughout the whole measurement session to ensure a silent measurement environment.

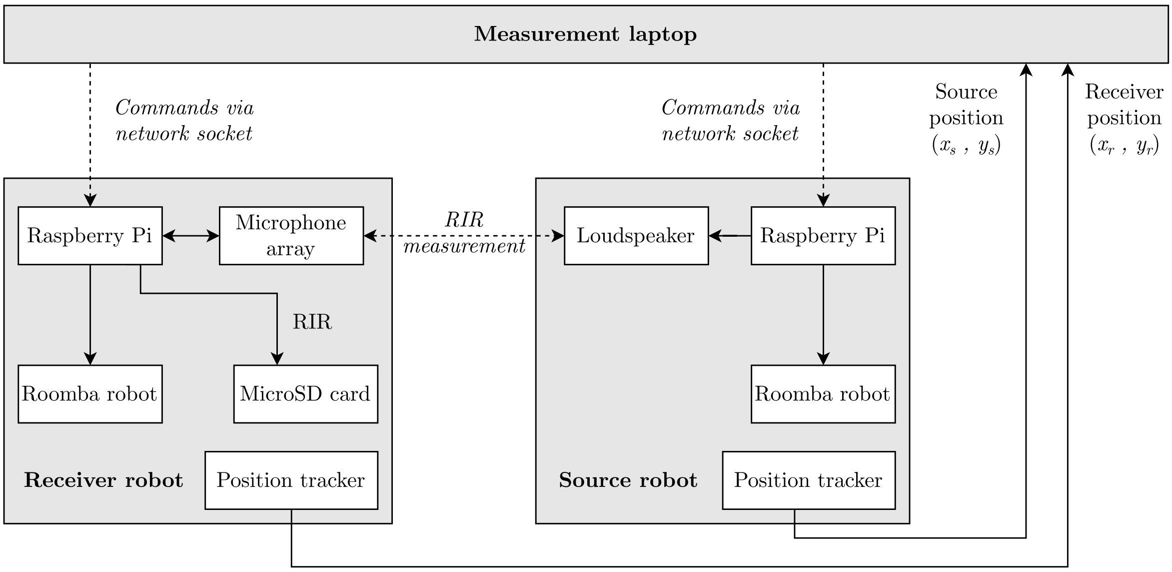


Fig. 1: Block diagram of the proposed implementation of the Autonomous Robot Twin System for Room Acoustic Measurements (ARTSRAM). The ARTSRAM can be used to autonomously measure large sets of room impulse responses (RIRs) for various source-to-receiver combinations. The system can easily be extended with additional loudspeakers and microphone arrays.

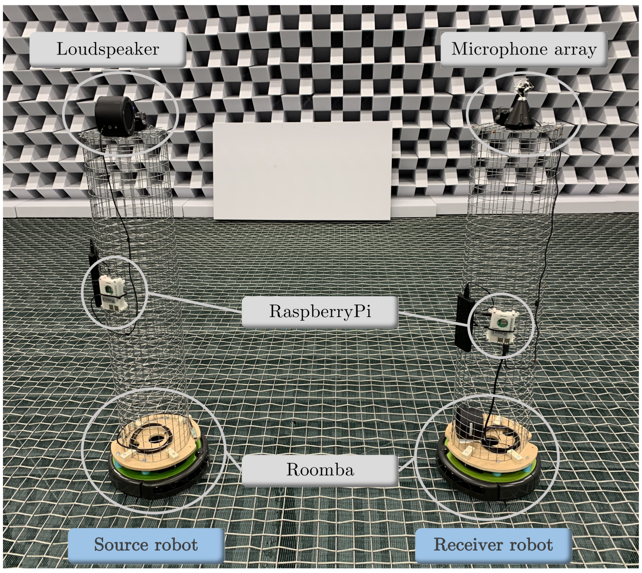


Fig. 2: The proposed robot system inside an anechoic chamber.

3.2.2. Additional Fixture

An additional fixture is mounted onto both of the robots to attach the microphone array and loudspeaker at a height of approximately 1.3 m. A picture of the whole apparatus is shown in Figure 2. A circular wooden plate was attached onto the Roomba robot. Additionally, a chicken fence was formed into a cylinder and fixed on the wooden plate. The construction allows a partial disassembly of the robot system for easier transportation. Similar to the mesh floor in an anechoic chamber, a chicken fence has minimal influence on the sound field around the robots. A detailed evaluation of the sound field alterations by the robot fixtures is presented in Section 4.2.

In principle, the currently used fixture could be extended with additional platforms at various heights inside the chicken fence to mount further loudspeakers or microphone arrays. This would allow capturing different RIRs or transfer-functions, for example from floor to ear height. Such a transfer-function could be useful for the auralization of footsteps in acoustic virtual reality scenarios.

3.2.3. Position Tracking

HTC Vive trackers were used to track the position of the robots. They are an affordable and off-the-shelf position tracking solution based on measuring synchronized light sweeps emitted by the corresponding HTC Vive lighthouses. The trackers have been used by various scientific projects and have a reportedly high accuracy and precision that can reach the millimetre range [24, 25]. The measurement script accesses the position data over Valve's OpenVR SDK and its respective Python bindings [26]. Two HTC Vive lighthouses are required for a maximum measurement area of 5 m by 5 m. However, the measurement area could potentially be extended up to 10 m by 10 m with two additional lighthouses.

3.2.4. Microphone array

In order to obtain spatial information about the sound field, the measurement system requires a microphone array consisting of multiple microphone capsules. Consequently, an RIR is measured for every microphone capsule of the

*JAES Vol. 69, No. 4, April 2021 — this is the author's accepted manuscript.
Published version: https://www.aes.org/e-lib/browse.cfm?elib=21033*

array. This allows transforming the measured RIRs to Ambisonic streams, which encode the sound field around the microphone array with a given resolution that depends on the number of capsules in the array [27, 28]. The resolution can be described by the order of the Ambisonic stream.

For the proposed measurement system, we use the Zoom H3-VR, which is a microphone array that records first-order Ambisonic streams. It can be connected to the Raspberry Pis via USB. In principle, higher-order microphone arrays that capture the sound field with a higher resolution than the HR-VR are available on the market. However, such microphone arrays are usually either unwieldy or hardly portable due to their cabling, thus making them unsuitable for the ARTSRAM.

3.3. Room impulse response measurement procedure

Exponential sine sweeps [29] are used for measuring RIRs with the ARTSRAM. After every measurement, the two robots move one after another to new positions in the room. The robots move according to a random walk procedure, in which every robot first spins for a random time span and subsequently drives straight forward for another randomly selected time period. Both time spans should be bounded, such that too short or too long movements are avoided. Time spans corresponding to spins between 15° and 180° and straight movements between 15 cm and half of the length of the longest room dimension prove successful during our tests. If a robot detects an obstacle during the random walk, it will stop its movement and drive backwards for 1 s to 3 s. As we will demonstrate in Section 4.3, good coverage of a measurement area and high variability between the measured positions can be ensured by following this random walk procedure. Furthermore, it does not require any additional path or grid planning before the measurement. A video demonstration of the measurement procedure can be found on the paper's companion page, c.f. Section 8.

4. Validation

In this section, the proposed ARTSRAM is validated from four perspectives. Firstly, anechoic measurements of the utilized mobile loudspeaker are presented. Secondly, they are compared with anechoic measurements of the entire robot system to determine the sound field alterations caused by the robot fixture. Thirdly, the proposed random walk procedure is evaluated with a test measurement and simulations. Lastly, we use ARTSRAM measurements and demonstrate their validity by calculating a sound field visualization for a shoebox-shaped room.

4.1. Loudspeaker measurement

Ideally, a loudspeaker for RIR measurements should be omnidirectional in order to excite the room as evenly as possible [30]. Additionally, its magnitude response should be flat to ensure that all frequencies are excited equally. The Minirig MRBT-2 loudspeaker is a consumer device

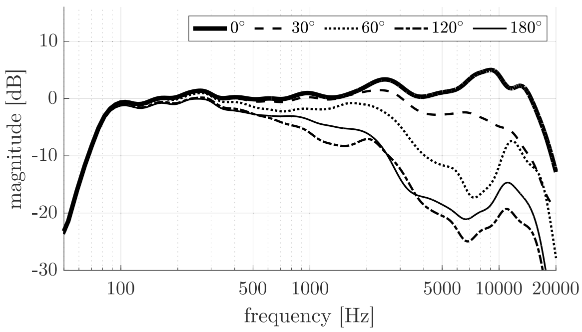


Fig. 3: Magnitude response of Minirig MRBT-2 loudspeaker, measured from five different azimuth angles. The measurements were conducted in an anechoic chamber. The responses are smoothed in 1/3 octave bands.

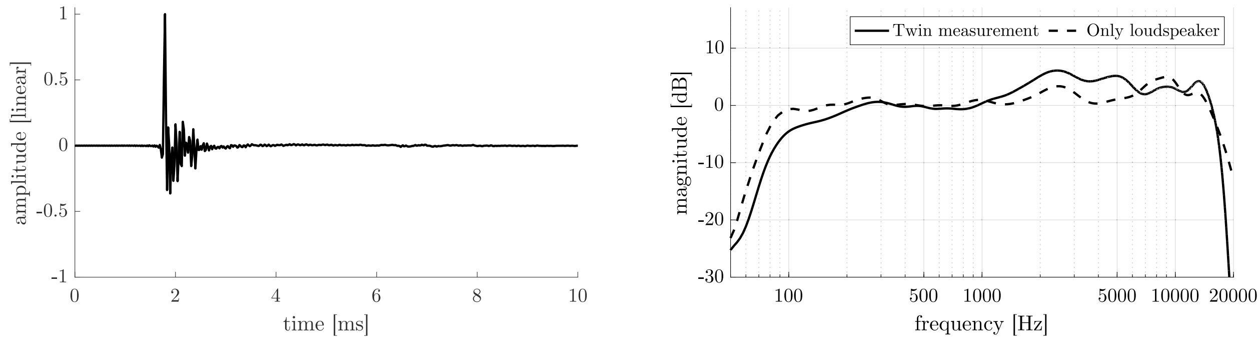
and therefore not primarily designed to conduct RIR measurements. For this reason, the loudspeaker was measured in an anechoic chamber to evaluate its suitability for the measurement system.

Figure 3 shows 1/3-octave-band-smoothed magnitude responses, measured with a 1/2" GRAS 46AF free-field microphone for five different azimuth angles. Measurements along the median plane of the loudspeaker would lead to similar responses, because the loudspeaker has a circular shape (c.f. Figure 2). The plot illustrates that the magnitude response is reasonably flat for azimuth angles of 0° and 30°. In contrast, the measured response at 60°, 120°, and 180° exhibit strong attenuation of frequencies above 1 kHz due to shadowing effects of the loudspeaker enclosure.

4.2. Anechoic robot twin measurement

Every RIR measurement setup requires at least one loudspeaker and one microphone. Adding additional components to the RIR measurement setup always comes at the cost of potentially disturbing the sound field that is supposed to be measured. In order to determine whether the Roomba robot and the additional fixture disturb the sound field, we conducted an RIR measurement with the entire system inside an anechoic chamber. In other words, we measured the anechoic response from the source robot to the receiver robot. Every disturbance caused by the robot fixtures or the utilized hardware would appear in this measurement as additional reflections in the time domain or as deviations from the loudspeaker response in the frequency domain. In the following, we will refer to this measurement as the *twin measurement*.

Figure 4a depicts the RIR that was obtained during the twin measurement. A clear direct sound peak is visible. Additionally, the response exhibits smaller ripples during the millisecond following the direct sound peak. These ripples are most likely caused by the loudspeaker itself, because they also appear in the loudspeaker's impulse response. The robots were placed 1 m apart and the distance from the microphone or loudspeaker to the wooden plate is approximately 1.1 m. Therefore, potential reflec-



(a) Impulse response (b) Magnitude response

Fig. 4: Responses from the twin measurement. During the twin measurement, the response from the source robot to the receiver robot was measured inside an anechoic chamber. The magnitude response is smoothed in 1/3 octave bands.

tions caused by the wooden plate of the fixture would be visible after approximately 4 ms. However, the plot shows only minor ripples at these time instances. The impulse response after 10 ms is not depicted in the plot, because it remains close to zero.

Figure 4b depicts the magnitude response of the twin measurement. Only minor deviations from the anechoic loudspeaker response can be observed. The influence of the additional fixture is maximal between 2 kHz to 6 kHz with deviations of up to 4.3 dB between the loudspeaker response and the twin measurement. Overall, these results imply that the Roomba robot, the utilized hardware, and the additional fixture do not evoke any major disturbances of the sound field that is supposed to be measured.

Number of steps	Percent visited	
	Source	Receiver
150	56.0 ± 3.1	56.2 ± 3.4
300	80.1 ± 2.7	80.5 ± 3.0
600	95.6 ± 1.7	95.7 ± 1.7
1000	99.2 ± 0.7	99.2 ± 0.7
5000	99.9 ± 0.1	99.9 ± 0.1

Table 1: Amount of cells visited by the robots during the proposed random walk procedure. The values are calculated from simulations of 100 runs over a measurement area of $6\,\mathrm{m} \times 4\,\mathrm{m}$. A uniform cell size of $40\,\mathrm{cm} \times 40\,\mathrm{cm}$ is assumed. All results are given as mean $\pm$ standard deviation calculated over the 100 runs.

4.3. Random walk test

As outlined in Section 3.3, the robots move according to a random walk procedure. The motivation behind this was to ensure a big variability between measured positions without requiring a manually defined measurement grid. The literature on random walks is extensive and scenarios with limited complexity can be described analytically [20–22]. However, our proposed random walk exhibits a higher complexity, because the measurement area is limited by walls and collisions with obstacles trigger interruptions to the standard walk. Therefore, we want to avoid an overly-mathematical derivation of a measurement point distribution at this point and present a more empirical evaluation instead. More precisely, in this section, we present results from a test measurement and from simulations of the random walk procedure. This evaluation should help to balance the pros and cons of our chosen approach compared to a grid-based measurement procedure.

A test measurement session was conducted in an exemplary room to evaluate the movements of the robots. Inside this room, a rectangular measurement area of approximately $6\,\mathrm{m}\times4\,\mathrm{m}$ was delimited with wooden planks on the floor. The session comprised 340 individual RIR measurements and took 4 hours to complete. During the random walk, the time spans for spinning and moving straight were randomly sampled from the intervals $\{0.3\,\mathrm{s} \leq t_{\mathrm{spin}} \leq 4\,\mathrm{s}\}$ and $\{1\,\mathrm{s} \leq t_{\mathrm{move\_straight}} \leq 20\,\mathrm{s}\}$ respectively. These time intervals approximately correspond to spins between 15° and 180° and to straight movements between 0.15 m and 3 m. Therefore, the robots could potentially move very accurately, but they were also able to traverse half of the longest measurement area dimension within one random walk step.

Figure 5 depicts the resulting movement paths of both robots. Initially, both robots were placed as close to each other as possible. From this position, they began their random walk and moved into different directions. The plots suggest that the robots cover the room fairly equally. No part of the room was entirely skipped. Only some portions close to the measurement area's borders were not reached by the random walk.

We ran additional simulations to investigate the random walk's coverage of the measurement area for different numbers of steps. During the simulations, the number of measurement positions or random walk steps was varied between 150 and 5000. A random walk of 150 steps means that each of the robots moves 150 times. For every number of steps, 100 runs of the proposed random walk procedure were simulated with the same parameters and the same measurement area dimensions as in the exemplary measurement.

Table 1 summarizes the calculated statistics from the simulations. To quantify the coverage of the random walk,

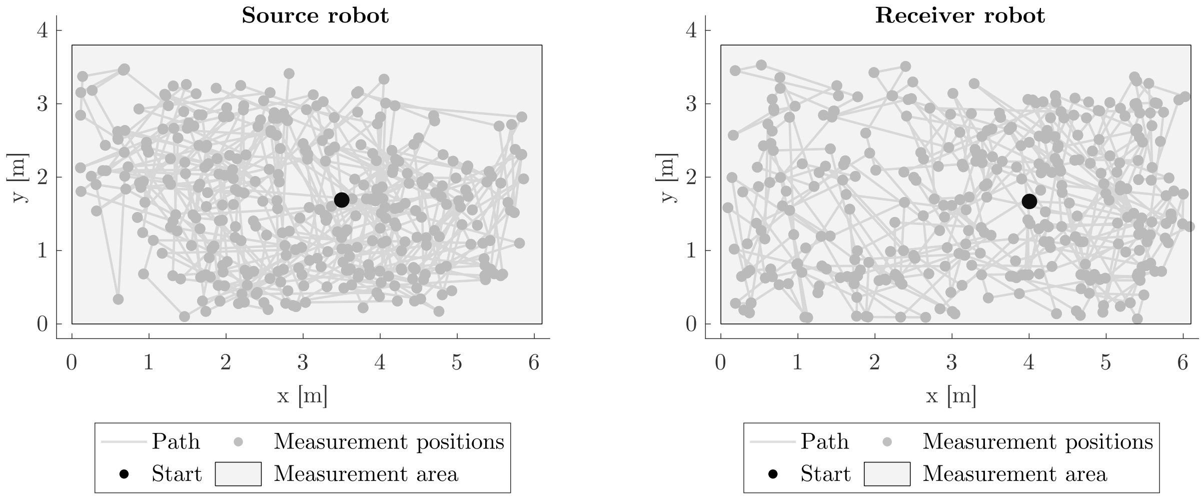


Fig. 5: Robot paths measured during a measurement session with 340 measurements. Both robots followed the proposed random walk procedure.

the measurement area was divided into 150 uniform cells of 40 cm × 40 cm. The "percent visited" value quantifies how many of the 150 cells were visited by each of the robots after the random walk with a certain number of steps is completed. A cell counts as visited, if a measurement was conducted in it. Differences between runs are accounted for by calculating the mean and standard deviation over all 100 simulated runs. The randomization of the measurement procedure means that more measurements need to be conducted than would be required if the approach was grid-based. In fact, the random walk only visits about half of the cells during 150 steps, whereas a grid-based procedure would visit all of them. However, the coverage gets significantly better already for double the number of steps. After 600 steps, more than 95 % of the cells were visited at least once by both the source and the receiver.

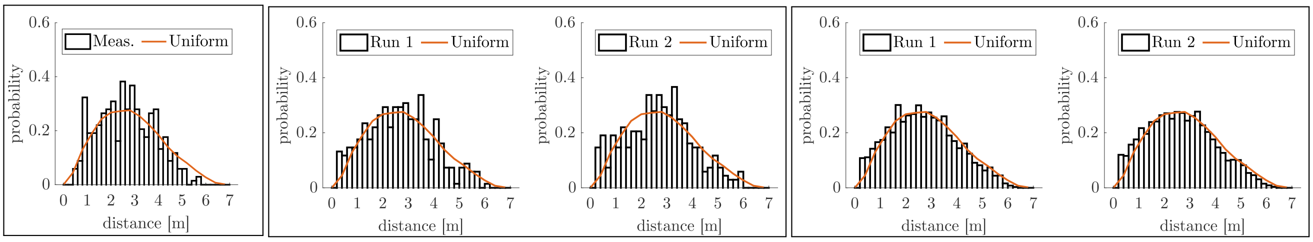
Additionally, we used the data of the test measurement and the simulations to evaluate the random walk procedure with respect to the achievable variability of measurements. For this reason, we investigated the distribution of measured source-to-receiver configurations. Figures 6a - 6c depict probability density functions (PDFs) of observed source-to-receiver distances for the measurement, simulations with 340 steps and simulations with 5000 steps, respectively. Although they show entries for a big range of different distances, the distributions are considerably non-uniform. An exceptionally high number of measurements was conducted with intermediate source-to-receiver distances, whereas fewer measurements were conducted with very small or very large distances. One reason for this is that the robots cannot get closer to each other than approximately 0.35 m because of their physical extent. Furthermore, big distances between the robots occur less frequently due to the spatial limitations of the measurement area. Interestingly, the resulting distributions are similar to the distribution that would be observable for a uniform measurement grid. To demonstrate this, PDFs of a uniform grid are included as a reference in Figure 6.

Another important observation is that random walks with fewer steps exhibited higher inter-run variability of the distributions than random walks with more steps. To illustrate this, Figure 6 shows PDFs of multiple runs. While the PDFs vary considerably among different runs of the 340 step walk, the PDFs of individual runs coincide better for a random walk with 5000 steps. This has practical implications, because it means that the measurement position distributions for runs with few steps are less predictable and reproducible. In other words, the distribution of our measurement, which is just one exemplary run of a 340 step random walk, could look somewhat different if we repeated it. For a higher number of steps, we could in contrast be more confident that the distribution converges to some average distribution.

4.4. Sound field visualization

The following section evaluates the validity of ARTSRAM-based RIR measurements by using them to visualize a sound field inside a room. A set of 330 RIRs was measured with the ARTSRAM inside the same shoebox-shaped room that was used during the evaluation of the random walk in Section 4.3. The room has a size of 7.85 m × 5.35 m × 3.15 m and the measurement area was again restricted with wooden planks on the floor. During the measurement session, the source robot remained static, while the receiver robot was exploring the room according to the proposed random walk procedure. The measurement session took 3 hours. It is 1 hour shorter than the session in the previous section, because only one of the robots moved.

Figure 7 shows multiple snapshots of the sound field recorded in the room. The visualization is based on a directional analysis of the sound energy propagating through the room. At every receiver position, direction-of-arrivals (DOAs) were calculated for multiple time windows along the entire RIR. The DOA estimation was based on the sound intensity vector calculated from



(a) Measurement with 340 steps

(b) Simulations with 340 steps per walk

(c) Simulations with 5000 steps per walk

Fig. 6: Probability density functions (PDFs) of source-to-receiver distance. The plots compare PDFs resulting from the proposed random walk with the PDF of a simulated uniform grid.

Ambisonic streams [31] and was implemented using the SDM toolbox [32].

For every time window, the DOA information can be used together with the time index of the window to extrapolate a position in space where the corresponding sound energy originated from. In a simplified way, this can be done by following a ray that points from the receiver position towards the DOA with a length of $l = c(nT + \frac{T}{2})$, where n is the window number, T is the window length, and c is the speed of sound. For the early part of the RIR, this will result in rays that point from the receiver towards the sound source and its image-sources. In principle, this is reciprocal to calculating energy responses based on the image-source method. At the global observation time $t = 0$, i.e., at the emission time of the pulse from the source, the approach returns energy contributions clustered at the primary source and its image-sources. With increasing t, sound energy propagates along the calculated rays. Consequently, it is possible to visualize the temporal progression of the sound field as shown in Figure 7.

The plots in Figure 7 are based on 330 first-order Ambisonic RIRs measured at the receiver positions depicted in Figure 7f. The grey dots in the figure correspond to sound energy contributions at the positions in the room, which were calculated with the previously outlined procedure. They propagate along rays between the (image-) sources and the receiver, thus rendering the different snapshots over time. The superposition of these energy contributions at one time instance yields the depicted wave fronts, which are indicated by the dashed auxiliary lines. Dots outside the room are contributions in the mirror rooms. They are depicted to illustrate the reflection formation process via image-sources and the notional continuation of the wave propagation into the mirror rooms. An animation of these sound field plots is available on the companion page of this paper, c.f. Section 8.

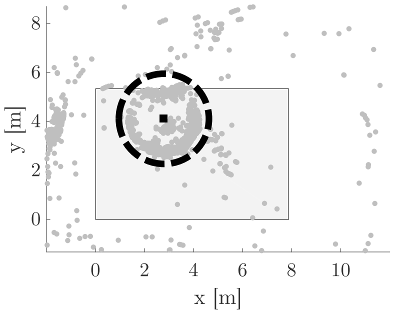
Although the measurement grid is non-uniform, the measurements reproduce many of the acoustic phenomena that can be expected from a theoretical point of view. Firstly, Figure 7a shows the direct sound as a spherical wave front propagating from the sound source into the room. Secondly, distinct wave fronts originating from first order reflections at the walls of the room can be observed in Figures 7b – 7d. Lastly, the visualization in Figure 7e shows a second-order reflection that is caused by the first-order reflection depicted in Figure 7d. For clarity reasons, the second-order reflection was plotted while it is still in the mirror room, i.e., before the reflection happened.

5. Discussion

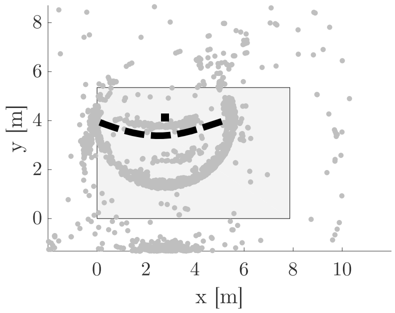
The aim of this paper was to develop a flexible, versatile, scalable, and autonomous robot system for capturing large sets of RIRs with variable sound source and receiver positions. During the development process, the design principles highlighted in Section 2 played a major role in choosing the components of the system. Keeping this in mind, the results of the previous validation section should be interpreted accordingly.

The anechoic loudspeaker measurement outlined in Section 4.1 illustrates that the Minirig loudspeaker deviates considerably from an ideal omni-directional loudspeaker. Nevertheless, its frequency response is sufficiently flat for sound radiating towards directions between 0° and 30° off the loudspeaker's main axis. This applies for directions along the horizontal and median plane in the same way, because the loudspeaker is circular. Although the Minirig loudspeaker is certainly not ideal for RIR measurements, we assume that it is still a good choice, because comparable mobile loudspeakers with similar dimensions will most likely not perform significantly better. Omnidirectional loudspeakers [33] or variable directivity higher-order Ambisonics speakers [34, 35] can be heavy and usually require a large amount of cabling and additional hardware. Obviously, this means that such loudspeakers cannot be used in a system like ours, for which mobility has been a crucial design principle. Additionally, the loudspeaker directivity of the Minirig is probably less problematic for the use case of acoustic virtual reality rendering, because usual sources in such scenes (human voice, many instruments, gunshots or other FX) exhibit similarly directive sound radiation patterns [36, 37]. In this case, it would be necessary to measure multiple source orientations for every measurement position. Such an extension of the random walk procedure would be very time-efficient, because spins of the robots can be performed quickly.

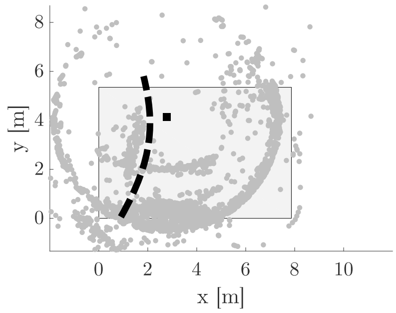
Section 4.2 explored how the sound field is influenced by the fixture that is mounted on the robots. The results suggest that the fixture does not disturb the sound field considerably. To the greatest extent, reflections can be avoided



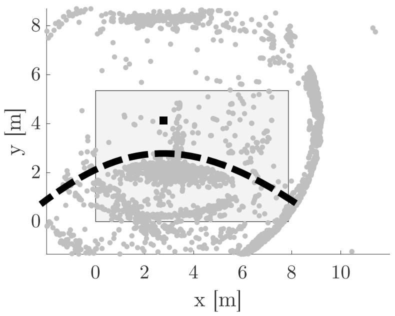
(a) t = 4.17 ms (direct sound)



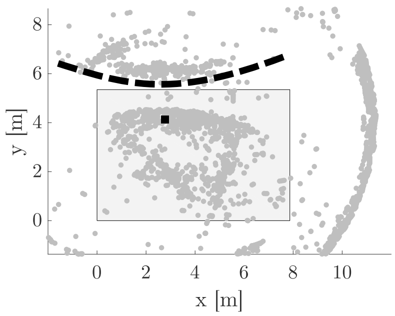
(b) t = 8.33 ms (1st order reflection at long wall with y=5.3)



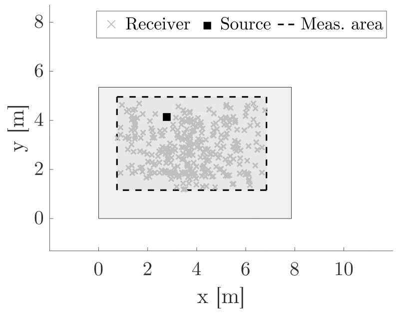
(c) t = 13.54 ms (1st order reflection at short wall with x=0)



(d) t = 18.75 ms (1st order reflection at long wall with y=0)



(e) t = 25 ms (2nd order reflection at long wall with y=5.3)



(f) Source and receiver positions

Fig. 7: Visualization of the sound energy (grey dots) inside a room, based on 330 ARTSRAM measurements. The measurement procedure followed the proposed random walk. The static sound source is depicted as the black square. The bold dashed auxiliary lines indicate observable wave fronts, which are classified in the respective subcaptions of the plots.

by placing the loudspeaker and microphone on a mesh construction, such as the proposed chicken fence. Only minor disturbances are caused by the robot base itself.

The results of Section 4.3 indicate that a high coverage of the room and a big variability in measured positions can be achieved with the proposed random walk procedure. As a consequence, large and diverse sets of RIRs can be captured, without requiring the ARTSRAM to measure positions along a predefined grid. In comparison to a grid-based approach, the random walk procedure requires a higher number of measurement steps to reach a comparable coverage. Approximately four times more measurements were required for the simulated scenario of Section 4.3. However, grid-based measurement procedures are less flexible, because they require additional grid and path planning as well as positioning efforts. We believe that the increased amount of measurements in the random walk are more tolerable than the additional efforts of grid-based approaches in the long run. Once the random walk measurement is running, additional measurements can be easily accommodated. Furthermore, additional measurements are less problematic because a session continues autonomously without any need for supervision and could potentially run over night or during weekends.

Section 4.4 demonstrated the validity of ARTSRAM-based RIRs measurements and their corresponding position data. It was exemplarily shown that such measurements can be used to visualize the sound field inside a room and reproduce well known acoustic phenomena, such as specular reflections. Another interesting perspective in this context is the question whether the sound field can theoretically be reconstructed from an irregular sampling with a distribution like in Section 4.3. The theoretical limit of sound field reconstruction is given by the Nyquist theorem [13]. If high frequency contents are supposed to be considered, the amount of required measurements may become very large. Taking into account the increased number of measurements due to the random walk procedure, the measurement time to reach a sufficient resolution may exceed what is feasible to measure with our proposed system. However, promising approaches exist for reconstructing sound fields with fewer measurements [14–16]. An irregular, as opposed to a uniform sampling, may be beneficial in terms of reconstructing the sound field because of a lower coherence between measurements [14]. Datasets obtained with the ARTSRAM might be valuable to accelerate further developments of sound field reconstruction algorithms, especially for data-heavy approaches such as [15].

6. Conclusions

In this paper, we proposed the Autonomous Robot Twin System for Room Acoustic Measurements (ARTSRAM). It consists of two autonomous robots equipped with electroacoustic transducers and position-tracking devices. The robots are programmed to perform random walks in rooms for large-scale room acoustic measurements. For example, the ARTSRAM can be used to capture large sets of RIRs in an automated way, including the corresponding position data of the sound source and receiver. Collision sensors make the robots autonomous and ensure a high flexibility and versatility, because the system can be used in various rooms without big installation efforts.

Although the components of the system were chosen according to our best knowledge, some non-ideal properties had to be accepted in order to ensure a high mobility of the robots. For example, currently the robots are equipped with a microphone array of only first order and mobile loudspeakers with a considerable directivity. However, we assume that the system's flexibility and the achievable acquisition speed of RIRs compensate for the minor shortcomings of the incorporated components. The modular structure of the robot system easily allows replacing its components as soon as better alternatives become available.

Our evaluation indicates that the proposed system is capable of measuring large, diverse, and high-quality RIR data sets in an automated way. The acoustic disturbance caused by the custom-made robot fixture was shown to be negligible. Furthermore, the implemented random-walk procedure exhibited a good coverage of an exemplary measurement area. Finally, it was shown that ARTSRAM-based RIR measurements and their corresponding position data are capable of accurately capturing acoustic phenomena, such as specular reflections. Therefore, the ARTSRAM can be seen as a valid tool to obtain position-dynamic RIR measurements for future research in the fields of room acoustics and acoustic virtual reality.

7. Acknowledgement

The authors would like to express their gratitude to Aleksi Öyry for the helpful advice and support while building the additional fixture and mounting it on the robots.

The project has received funding from from the Academy of Finland, project no. 317341, and from Nordic Sound and Music Computing Network (NordicSMC), project no. 86892.

8. Companion page and source code

A companion page for this paper can be found at:

`http://research.spa.aalto.fi/publications/papers/artsram`

It includes a video demonstration of a measurement session with the ARTSRAM and an animation of the sound field visualization depicted in Figure 7.
The source code of the ARTSRAM implementation can be found at:

`https://github.com/georg-goetz/ARTSRAM`

References

[1] Akama, T., Suzuki, H., and Omoto, A., "Distribution of selected monaural acoustical parameters in concert halls," *Appl. Acoust.*, 71(6), pp. 564–577, 2010, doi:10.1016/j.apacoust.2010.01.004.

[2] Mourjopoulos, J., "On the variation and invertibility of room impulse response functions," *J. Sound Vib.*, 102(2), pp. 217–228, 1985, ISSN 0022-460X, doi:10.1016/s0022-460x(85)80054-7.

[3] Llopis, H. S., Pind, F., and Jeong, C.-H., "Development of an auditory virtual reality system based on pre-computed B-format impulse responses for building design evaluation," *Build. Environ.*, 169, pp. 106553:1 – 106553:10, 2019, doi:10.1016/j.buildenv.2019.106553.

[4] Kuttruff, H., *Room Acoustics*, Spon Press, London, UK, 4th edition, 2000.

[5] Witew, I. B., Vorländer, M., and Xiang, N., "Sampling the sound field in auditoria using large natural-scale array measurements," *J. Acoust. Soc. Am.*, 141(3), pp. EL300–EL306, 2017, doi:10.1121/1.4978022.

[6] Xiang, N., Escolano, J., Navarro, J. M., and Jing, Y., "Investigation on the effect of aperture sizes and receiver positions in coupled rooms," *J. Acoust. Soc. Am.*, 133(6), pp. 3975–3985, 2013, doi:10.1121/1.4802740.

[7] Uehara, M., Ishikawa, N., and Okawa, S., "Visualization of Distribution of Room Acoustic Parameters by Using Mobile Robot," in *Proc. 23rd Int. Congr. Acoust. (ICA)*, pp. 4611–4616, Deutsche Gesellschaft für Akustik (DEGA), Aachen, Germany, 2019.

[8] Adavanne, S., Politis, A., Nikunen, J., and Virtanen, T., "Sound Event Localization and Detection of Overlapping Sources Using Convolutional Recurrent Neural Networks," *IEEE J. Sel. Topics Signal Process.*, 13(1), pp. 34–48, 2018, doi:10.1109/jstsp.2018.2885636.

[9] Čmejla, J., Kounovský, T., Gannot, S., Koldovský, Z., and Tandeitnik, P., "MIRaGe: Multichannel Database Of Room Impulse Responses Measured On High-Resolution Cube-Shaped Grid In Multiple Acoustic Conditions," `https://arxiv.org/abs/1907.12421`, 2019, (Last accessed: 29 October 2020).

[10] Le Roux, J., Vincent, E., Hershey, J. R., and Ellis, D. P. W., "Micbots: Collecting large realistic datasets for speech and audio research using mobile robots," in *Int. Conf. Acoust., Speech, Sig. Proc. (ICASSP)*, pp. 5635–5639, IEEE, Brisbane, Australia, 2015, doi:10.1109/icassp.2015.7179050.

[11] Cecchi, S., Carini, A., and Spors, S., "Room Response Equalization—A Review," *Appl. Sci.*, 8(1), p. 16, 2017, doi:10.3390/app8010016.

[12] Raghuvanshi, N. and Snyder, J., "Parametric directional coding for precomputed sound propagation," *ACM Trans. Graphics*, 37(4), p. 108, 2018 July, doi:10.1145/3197517.3201339.

[13] Ajdler, T., Sbaiz, L., and Vetterli, M., "The Plenacoustic Function and Its Sampling," *IEEE Trans. Signal Process.*, 54(10), pp. 3790–3804, 2006, doi:10.1109/tsp.2006.879280.

[14] Verburg, S. A. and Fernandez-Grande, E., "Reconstruction of the sound field in a room using compressive sensing," *J. Acoust. Soc. Am.*, 143(6), pp. 3770–3779, 2018, doi:10.1121/1.5042247.

[15] Lluís, F., Martínez-Nuevo, P., Møller, M. B., and Shepstone, S. E., "Sound field reconstruction in rooms: Inpainting meets super-resolution," *J. Acoust. Soc. Am.*, 148(2), pp. 649–659, 2020, doi:10.1121/10.0001687.

[16] Antonello, N., Sena, E. D., Moonen, M., Naylor, P. A., and van Waterschoot, T., "Room Impulse Response Interpolation Using a Sparse Spatio-Temporal Representation of the Sound Field," *IEEE/ACM Trans. Audio, Speech, Language Process.*, 25(10), pp. 1929–1941, 2017 July, doi:10.1109/taslp.2017.2730284.

[17] Werner, S., Klein, F., and Götz, G., "Investigation on spatial auditory perception using non-uniform spatial distribution of binaural room impulse responses," in *Proc. 5th Int. Conf. Spatial Audio (ICSA)*, pp. 137–144, Verband Deutscher Tonmeister (VDT), Ilmenau, Germany, 2019, doi:10.22032/dbt.39936.

[18] Brinkmann, F., Aspöck, L., Ackermann, D., Lepa, S., Vorländer, M., and Weinzierl, S., "A round robin on room acoustical simulation and auralization," *J. Acoust. Soc. Am.*, 145(4), pp. 2746–2760, 2019 April, doi:10.1121/1.5096178.

[19] Feng, Y., Khodayi-mehr, R., Kantaros, Y., Calkins, L., and Zavlanos, M. M., "Active Acoustic Impedance Mapping Using Mobile Robots," in *IEEE Conf. Decision Control (CDC)*, pp. 3910–3915, IEEE, Miami Beach, USA, 2018, doi:10.1109/cdc.2018.8618924.

[20] Hughes, B. D., *Random Walks and Random Environments - Volume 1: Random Walks*, Oxford University Press, New York, NY, USA, 1995.

[21] Lovász, L., "Random Walks on Graphs: A Survey," in D. Miklos, V. T. Sos, and T. Szony, editors, *Combinatorics, Paul Erdős is Eighty*, number 2 in Bolyai Society Mathematical Studies, pp. 1–46, János Bolyai Mathematical Society, Budapest, Hungary, 1993.

[22] Aldous, D. J., "Lower bounds for covering times for reversible Markov chains and random walks on graphs," *J. Theor. Probab.*, 2(1), pp. 91–100, 1989, doi:10.1007/bf01048272.

[23] Witherwax, M., "iRobot - A Python implementation of the iRobot Open Interface," `https://github.com/julianpistorius/irobot`, 2016, (Last accessed: 15 September 2020).

[24] Borges, M., Symington, A., Coltin, B., Smith, T., and Ventura, R., "HTC Vive: Analysis and Accuracy Improvement," in *Int. Conf. Intelligent Robots Systems (IROS)*, pp. 2610–2615, IEEE/RSJ, Madrid, Spain, 2018, doi:10.1109/iros.2018.8593707.

[25] Niehorster, D. C., Li, L., and Lappe, M., "The Accuracy and Precision of Position and Orientation Tracking in the HTC Vive Virtual Reality System for Scientific Research," *i-Perception*, 8(3), pp. 1–23, 2017 May-June, doi:10.1177/2041669517708205.

[26] Bruns, C., "pyOpenVR - Unofficial python bindings for Valve's OpenVR virtual reality SDK," `https://github.com/cmbruns/pyopenvr`, 2016, (Last accessed: 15 September 2020).

[27] Rafaely, B., *Fundamentals of Spherical Array Processing*, number 8 in Springer Topics in Signal Processing, Springer, Berlin, Heidelberg; Germany, 2015, doi:10.1007/978-3-662-45664-4.

[28] Zotter, F. and Frank, M., *Ambisonics - A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, number 19 in Springer Topics in Signal Processing, Springer Nature, Cham, Switzerland, 2019, doi:10.1007/978-3-030-17207-7.

[29] Farina, A., "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," in *108th Conv. Audio Eng. Soc.*, pp. 1–23, AES, Paris, France, 2000.

[30] ISO 3382-2, "Acoustics - Measurement of room acoustic parameters - Part 2: Reverberation time in ordinary rooms," Standard, International Organization for Standardization (ISO), Geneva, Switzerland, 2008.

[31] Merimaa, J. and Pulkki, V., "Spatial Impulse Response Rendering I: Analysis and Synthesis," *J. Audio Eng. Soc.*, 53(12), pp. 1115–1127, 2005 December.

[32] Tervo, S., "SDM Toolbox," `https://www.mathworks.com/matlabcentral/fileexchange/56663-sdm-toolbox`, 2018, (Last accessed: 15 September 2020).

[33] Leishman, T. W., Rollins, S., and Smith, H. M., "An experimental evaluation of regular polyhedron loudspeakers as omnidirectional sources of sound," *J. Acoust. Soc. Am.*, 120(3), pp. 1411–1422, 2006, doi:10.1121/1.2221552.

[34] Farina, A. and Chiesi, L., "A novel 32-speakers spherical source," in *140th Conv. Audio Eng. Soc.*, pp. 1–6, AES, Paris, France, 2016.

[35] Avizienis, R., Freed, A., Kassakian, P., and Wessel, D., "A Compact 120 Independent Element Spherical Loudspeaker Array with Programmable Radiation Patterns," in *120th Conv. Audio Eng. Soc.*, pp. 1–7, AES, Paris, France, 2006.

[36] Monson, B. B., Hunter, E. J., and Story, B. H., "Horizontal directivity of low- and high-frequency energy in speech and singing," *J. Acoust. Soc. Am.*, 132(1), pp. 433–441, 2012 July, doi:10.1121/1.4725963.

[37] Pätynen, J. and Lokki, T., "Directivities of Symphony Orchestra Instruments," *Acta Acust. united Ac.*, 96(1), pp. 138–167, 2010 January/February, doi:10.3813/AAA.918265.

THE AUTHORS

Georg Götz

Sebastian J. Schlecht

Georg Götz is a doctoral candidate at the Acoustics Lab, Department of Signal Processing and Acoustics, of Aalto University, Finland. He received his B. Sc. (2016) and M. Sc. (2019) degrees in Media Technology from Ilmenau University of Technology, specializing in spatial audio and psychoacoustics. He was the deputy chair of the AES Student Section Ilmenau from 2013 to 2016.

The research of his doctoral thesis is about machine-learning-based virtual acoustics rendering. His other research interests include room acoustics, psychoacoustics, and VR/AR technology.

●

Abraham Martinez Ornelas received his M. Sc. from Aalto University.

●

Sebastian J. Schlecht is a Professor of Practice for Sound in Virtual Reality at the Acoustics Lab, Department of Signal Processing and Acoustics and Media Labs, Department of Media, of Aalto University, Finland. He received the Diploma in Applied Mathematics from the University of Trier, Germany in 2010, and an M.Sc. degree in Digital Music Processing from School of Electronic Engineering and Computer Science at Queen Mary University of London, U.K. in 2011. In 2017, he received a Doctoral degree at the International Audio Laboratories Erlangen, Germany, on the topic of artificial spatial reverberation and reverberation enhancement systems. From 2012 on, Dr. Schlecht was also external research and development consultant and lead developer of the 3D Reverb algorithm at the Fraunhofer IIS, Erlangen, Germany.

His research interests are acoustic modeling and auditory perception of acoustics, analysis, and synthesis of feedback systems, music information retrieval, and virtual and augmented reality. He is the recipient of multiple best paper awards including Best Paper in Journal of the Audio Engineering Society (JAES) in Jun 2020, Best Paper Award at IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in Oct 2019, 2nd Best Paper Award at International Conference on Digital Audio Effects (DAFx) in Sep 2018, and, Best Peer-Reviewed Paper at 2018 AES International Conference on Audio for Virtual and Augmented Reality (AES AVAR) in Aug 2018.

●

The biography of Ville Pulkki was published in the 2020 May issue of the Journal.