SIMULATION AND VISUALIZATION OF INDOOR-ACOUSTICS FOR ROBOT CONTROL

Norbert Schmitz, Jens Wettach,
and Eduard Deines
IRTG, Dep. of Computer Science
University of Kaiserslautern
Kaiserslautern, Germany
email: {nschmitz|wettach|deines}
@informatik.uni-kl.de

Peter Dannenmann and Martin Bertram
German Research Center
for Artificial Intelligence
P.O. Box 2080
Kaiserslautern, Germany
email: {Peter.Dannenmann|
Martin.Hering-Bertram}@dfki.de

Karsten Berns and Hans Hagen
Department of Computer Science
University of Kaiserslautern
P.O. Box 3049
Kaiserslautern, Germany
email: {berns|hagen}
@informatik.uni-kl.de

ABSTRACT
An autonomous robot orientates itself by using information provided by its sensor systems. Besides distance sensors, optical and acoustic sensors play a vital role in fulfilling this task. In this paper we present a novel approach to simulate and visualize the acoustic properties of an indoor scene. This simulation data is used for testing and refining the control algorithms of an autonomous robot interacting with humans in an office environment. In order to enable a robot to interact with its environment and to perform its intended tasks in a context-sensitive manner, it must be capable of interpreting the information provided by its sensor systems. For testing these interpretation capabilities, certain environmental stimuli need to be provided to the robot in a controlled and repeatable manner. However, such stimuli are not available normally. The presented work provides a virtual testing environment that permits the realistic simulation and visualization of the acoustic properties of indoor-environments. Thus, it is possible to simulate and visualize the spread of sound waves within a room and to simulate the acoustic signals a robot receives at certain positions. Using such well-defined test conditions and a visual representation of the spread of sound in the test environment, it is now possible to find positions of special acoustic properties and to use them for testing the robot's reactions to acoustic events. On this basis, the robot's control algorithms can be refined accordingly.

KEY WORDS
Virtual Environments, Acoustics, Simulation, Visualization, Robotics, Control Algorithms

1 Introduction

Mobile robots can increasingly often be found in private houses. They either mow the lawn or vacuum the floor just to mention a few examples. These robots still lack some important abilities as the ability to communicate.

The first step to man-machine communication is the recognition of humans by their voices. There are other sources of sound besides human beings which should also be recognized. A patrolling robot would also be interested in the sound of breaking glass, e.g..

Although different sound localization systems do exist, it is difficult to compare or optimize those algorithms. A simulation of sound source and sink can generate repeatable test scenarios. The overall performance of such a localization system depends on reliable tests with a realistic simulation engine. In this paper we will introduce such a simulation system for mobile indoor robots.

In the next section we give a short overview of the current state of the art in robot navigation as well as in acoustic simulation and put this into relation to previous work performed at our research groups. In section 3 and 4 we then give a more detailed description of the technologies of sound localization and simulation of acoustics, that we have developed. Finally, we describe the virtual acoustic test environment for robot control, that is currently under development.

2 Related Work

2.1 Current Research Work in Robot Navigation and Acoustic Simulation

Modeling the application environment of a mobile robot can be based on maps of different types. On the top level of abstraction a *topological* map represents interesting positions that can be recognized by the robot's sensor system as well as their spatial relationship [1], [2]. In *grid* maps the environment is divided into two-dimensional cells which are marked with attributes like *occupied*, *free* or *unknown*. Thus they are primarily used for navigation and obstacle avoidance. *Geometric* maps contain exact dimensions of objects represented by lines, polygons or rectangles. Furthermore, distance information between objects is recorded [2], [3]. Additionally, there exist several combinations of these three types of maps, for example to fuse the accuracy and consistency of grid maps with the efficiency of a topological map [4]. These maps describe the structure of the environment which has great influence on the simulation of sound.

Simulation of acoustic properties of rooms can be per-

formed using a number of approaches: For a correct representation of the propagation of low frequencies, including diffraction effects, the wave equation has to be solved, using the Finite Element Method (FEM). For simulating the propagation of sound waves of medium or high frequencies, where diffraction effects are not relevant, there exist other techniques, like the Image Method [5] or Acoustic Ray Tracing [6]. Both methods aim at treating reflections of sound at a room's walls or interior in the best possible manner. The Image Method addresses this task by placing mirrored images of sound sources behind the surface where the sound has been reflected. The reflected sound is treated as if it came from this virtual sound source with the intensity of the original source, reduced by the surface's absorption coefficients. The Acoustic Ray Tracing technique, on the other hand, is based on a simulation of particles that are emitted by the sound sources, reflected at the scene's surfaces and collected around the position of the listener. First approaches to integrate such acoustic simulations into a Virtual Environment have been described by [7].

2.2 Previous Own Work

In previous research projects a SLAM[1]-based system has been built for autonomous creation of a geometric and topological map of office environments with mobile robot `Marvin` (see figure 1) [8], [9].

Furthermore a model for detecting dynamic objects as humans in the robot's vicinity has been developed. It is based on a thorough fusion of distance information from a laser scanner regarding leg positions and image information from a camera for recognizing human faces [10]. A current research topic is the design of a 3D simulation of office environments which generates realistic distance and image information based on a given scene description. This tool facilitates simulated testing of robot control strategies using 'virtual' laser scanners and cameras. As autonomous map creation requires sound exploration strategies, mechanisms for controlling the robot's focus of attention are of primary interest. In this context a system for localizing sound sources based on evaluating the time difference at the arrival of sound events using a stereo microphone system has been implemented [11] (see section 3). In order to test this system in simulation, a model of noise propagation has to be integrated to set up virtual microphones.

Our previous research on acoustics simulation includes the joint development of an acoustics simulation system integrated into a virtual environment. Within this project, an integrated simulation- and visualization concept has been realized, that computes the propagation of low frequency sonic waves using an FEM-based simulation and that covers medium and high frequencies using a particle method based on a *Phonon Map* [12]. In combining these two methods, a realistic simulation of the propagation of sound within a room is achieved. The results can be aural-

[1] simultaneous localization and mapping

Figure 1. The mobile robot `Marvin`

ized in the virtual environment, enhancing the impression of reality to the user.

3 Making Robots React to Sonic Events

In order to detect moving objects or humans in the vicinity of a mobile robot, a sound source localization system has been developed for the test platform `Marvin` [11], based on work of [13].

The robot is equipped with two microphones which are connected via a preamplifier to the sound card.

The direction of the emerging noise relative to the microphones is derived from the delay between the time of arrival of the noise at two microphones (see figure 1). This delay is calculated in the frequency domain via the Dual Delay Line (DDL) algorithm, where the history of sound samples, that have arrived at the left and right microphone, is checked for correspondence. The displacement of samples corresponding in the frequency domain can be translated into the resulting time delay via the shifting theorem of the Fourier transform.

Hence the sound samples from the left (L) and right (R) sensor are grabbed by the sound chip of the robot's control PC and transformed in the frequency domain using the Fast Fourier Transform (FFT). The corresponding delay line is given by

$$X_{Ln}^{(i)}(m) = X_{Ln}(m) \cdot e^{-j2\pi f_m \tau_i} \tag{1}$$

$$X_{Rn}^{(i)}(m) = X_{Rn}(m) \cdot e^{-j2\pi f_m \tau_i}. \tag{2}$$

Here n is the index of the actual signal time frame (execution cycle), X_{Ln} and X_{Rn} are the transformed samples, $m = 1, \ldots, \frac{M}{2}$ is the index of the M-point FFT, $i = 1, \ldots, I$ is the index of the delay in the delay-line and the frequency f_m is given as $f_m = m \cdot \frac{f_s}{M}$, based on the sampling frequency f_s. The time factor τ_i describes the delay between the left and right channel.

The correspondence check is performed on the delay-line by calculating the differences

$$H_n(i) = \frac{2\sum_{m=1}^{\frac{M}{2}} \left| X_{Ln}^{(i)}(m) - X_{Rn}^{(i)}(m) \right|}{M}. \tag{3}$$

Finally, these values are normalized to $[0..1]$ and inverted via

$$Loc_n(i) = \left(1 - \frac{H_n(i) - \min(H_n(i))}{\max(H_n(i)) - \min(H_n(i))}\right)^6 \tag{4}$$

and the peaks of these values represent the direction angle of the sound source (i. e. the index i of peak values determines the angle via $\alpha_i = 90 - \frac{i-1}{I-1} \cdot 180$, between -90° and 90°). Thus an angle of 0° describes a noise, that comes directly from the middle of both microphones. The calculated direction is then used to make the robot turn towards the detected noise.

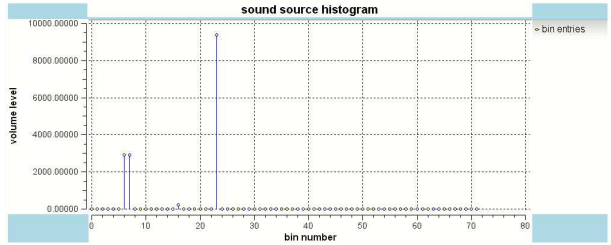


Figure 2. Histogram of localized sound sources, caused by clapping. Due to reflections, there are smaller peaks in the neighborhood of the original direction.

This approach has two main drawbacks: first, noises coming from behind the microphones cannot be distinguished from frontal ones; second, there is no appropriate relation between noise localization and robot turning motion, e. g. in case of sudden noises the robot will stop moving before it has reached the target direction. While the first problem will be solved in future revisions using a four-microphone system, the second topic has already been approached by setting up a history queue of calculated noise directions. These values are recorded as <angle, volume> pairs and used for computing an angle histogram that collects votes for the actual most interesting target direction, based on the volume levels and angles of the stored tuples. That way also short and sudden noises as hand clapping can be targeted reliably by the robot. However, as figure 2 shows, reflections of sharp noises at room walls or large furniture objects result in one main direction and different "side" directions, which will be approached successively as the queue "ages". This situation may also be rectified by applying a four sensor system and taking into account properties of sound propagation. A realistic sensor simulation based on noise reflection properties of different materials will be useful for testing and tuning this system in future stages.

4 Simulating and Visualizing Acoustic Properties of Indoor-Scenes Using Phonon Tracking

Photon mapping [14] is often used for rendering photo-realistic images, supplementing uni-directional raytracing by a variety of visual effects. We adopt a similar approach to the simulation of sound, named phonon tracing [12], which is summarized in the following.

4.1 Problem specification.

Our simulation algorithm requires the following input information:

- position of sound source s

- emission distribution E of sound source
- one or more listener positions l_i
- a triangulated scene with tagged material m_j
- an absorption function $\alpha_j : \Omega \mapsto (0,1]$ for each material
- an energy threshold ε for terminating the phonon paths

The output of our approach is a filter f_i for each listener's position l_i corresponding to the impulse response with respect to the sound source and the phonon-map containing for each phonon the energy spectrum e_p, the traversed distance d_p, the phonon's position p_p at the reflection point, its outgoing direction v_p, number of reflections n_p, and the material m_p at the current reflection.

Our simulation algorithm contains two steps, the *phonon tracing* step constructs the phonon map, and the *phonon collection and filtering* step collects the phonon's contribution to the filter for every listener position.

4.2 Phonon tracing.

Every phonon p emitted from the sound source carries the following information:

- an energy spectrum $e_p : \Omega \mapsto \mathbb{R}^+$
- the distance d_p traversed from the source
- the phonon's current position p_p
- the normalized outgoing direction v_p

Our absorption and energy functions α_j are represented by $n_e = 10$ coefficients associated with the frequencies $40, 80, 160, ..., 20480$ Hz. The basis functions for the energy spectrum are wavelets adding up to a unit impulse. Every phonon is composed of different frequencies, which is more efficient than tracing a single phonon for each individual frequency band.

Phonons are emitted from the source s according to the emission probability distribution E and have at starting point a unit energy spectrum $e_{p,i} = 1 \ (i = 1, ..., n_e)$. At the intersection of the phonon ray with the scene, the phonon direction d_p is reflected with respect to the surface normal and the absorbed energy is subtracted according to the local material m_j. The distance d_p is set to the traversed distance. The phonon is fixed at the intersection point, contributing to a global phonon map. If the maximum energy of the phonon exceeds the energy threshold, i.e. $max\{e_{p,i}\}_{i=1}^{n_e} > \varepsilon$, the next phonon re-uses the path and energy of the preceding one, saving computation time. It is started at the current position with respect to the outgoing direction d_p and contributes to the phonon map at the next surface intersection. If the threshold is not exceeded and a minimum number of reflections have been computed, then a new phonon is started from the source. After a prescribed number n_p of phonons have contributed to the global phonon map, the tracing is terminated. The phonon map is written to a file for further visualization purposes.

4.3 Phonon collection and filtering.

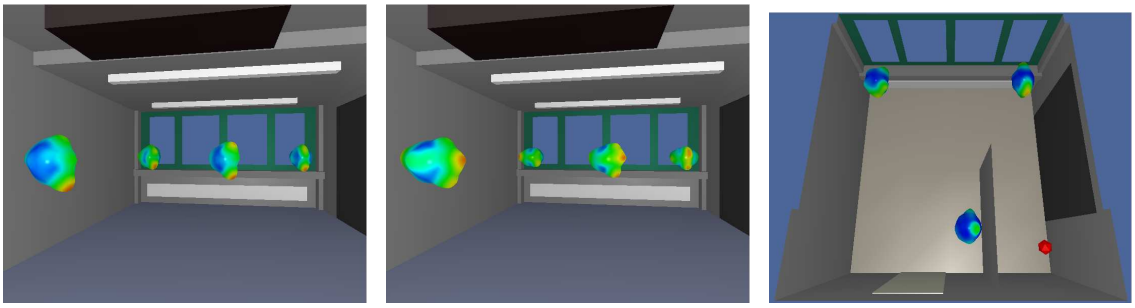
The remaining task of the phonon tracing method is collecting the phonons' contribution to a filter f for every listener's position l. This filter corresponds to the impulse response from the source, recorded at l, such that convolution with an anechoic signal, reproduces the perceived signal.
In the case of uniform absorption for all frequencies, the contribution of a phonon visible from the listener is simply a scaled, translated unit impulse (Dirac). The Dirac is shifted by the time elapsed between emission and reception of a phonon and scaled by the phonon's energy $e_{p,i}$ multiplied by a gaussian weighting of the ray's distance to the listener. In classical acoustic raytracing [15], a sphere is used to collect rays at the listener position. Using a gaussian, however, provides much smoother filters, since more phonon rays contribute to the filter, weighted by their shortest distance.
In the more general case of frequency dependent absorption, the unit impulse is subdivided into wavelets representing the individual frequency bands. The filter then becomes a sum of this wavelets scaled by $e_{p,i}$ and shifted by the elapsed time. A description of our phonon tracing algorithm can be read in more detail in [12].

4.4 Listener-based Visualization

The phonon map characterizes the acoustic behaviour of a scene considering the location of a specific sound source. However, in our application we are interested in visualizing the characteristics of an acoustic signal received at a certain position within the room. In this section we present such a visualization method capable of depicting the received energy at a listener position. With this approach we can visualize the energy spectrum received at this position and detect from which direction most of the energy reaches the listener.

For this purpose we render a triangulated sphere deformed according to the weighted phonons received at the listener position. The phonons are collected using the collection step described in section 4.3. For each phonon which contributes to the total energy at the listener position, we first calculate the intersection point $p_{intersec}$ between the ray from the sphere's center c_s to the phonon's postion p_p and the sphere itself. We then increase the radial displacement $disp_{sp}$ of that point sp of the intersected triangle whose distance to $p_{intersec}$ is minimal, according to the energy contributed by the collected phonon.

The color of the sphere points is calculated according to the spectral energy distribution of the phonons received at this position. For the color coding, we provide the option



(a) (b) (c)

Figure 3. Energy distribution at listener positions (a) at 80 Hz, (b) at 1280 Hz, and (c) at 5120 Hz at three position in a room with a separating wall with total apsorption.

to consider all frequency bands in total or each of them separately. In the first case the points are color coded by using the RGB components, such that red corresponds to the average of $e_{p,8},...,e_{p,10}$ (5120, 10240, 20480 Hz), green corresponds to the average of $e_{p,5},...e_{p,7}$ (640, 1280, 2560 Hz), and blue to the average of $e_{p,1},...,e_{p,4}$ (40, 80, 160, 320 Hz). In the second case, considering only one frequency band, we encode the energy of this band using the HSV model. We interpolate the color of the spheres between red (full energy) and blue (energy equals zero) corresponding to the energy $e_{p,i}$ of the i-th frequency band.

The following figures present the results of the introduced visualization approach. Figure 3 depicts the sonic energy received at certain positions in a reference room. In figure 3 (a), we can observe that most of the engergy at a frequency of 80 Hz is reflected from floor and ceiling and at 1280 Hz (figure 3 (b)) most of the energy is reflected from the walls.

Figure 3 (c) shows a room with a separating wall with total absorption. This wall prevents most reflections from the surfaces on the left hand side of the room to the right one and vice versa. We placed a listener position behind the wall and two in the corners of the room, where we visualized the energy distribution at 5120 Hz. As we can see, no reflections are received from the separating wall and only a few reflections reach the listener from the other surfaces.

5 Virtual Acoustic Test Environments for Robot Control

Several simulations for indoor robots can be found in literature. Most of them are 2D representations and only few include the simulation of sensors. We propose a 3D robot simulation which is able to simulate sensor measurements. Figure 4 shows a scene of the current simulation. It shows our laboratory with several tables and cupboards. Several objects like computers, displays and chairs will enhance the model in the future.

The software is able to simulate the measurements of laser scanners and cameras. Laser scanners can be simulated in any plane parallel to the ground. `Marvin` uses

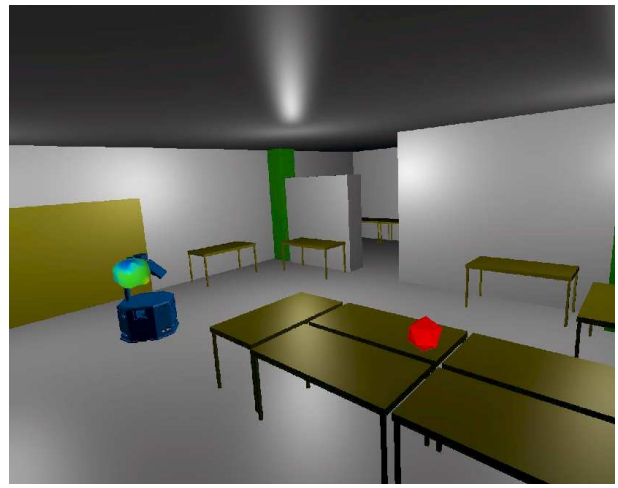


Figure 4. Mobile robot simulation in our laboratory with sound source and microphone visualization

two scanners, one pointing forward, the other one backward. They are mounted in a height of about 15 cm above the floor. The web camera is mounted on a pan-tilt unit. The simulation is able to provide camera images in any orientation of the camera. Two microphones are mounted on Marvin in a height of about $1m$ for sound localization. This localization system can be tested based on the sound localization (see section 3) and the sound simulation (see section 4) presented before.

The environments for the simulation are based on real rooms with furniture. Additionally, absorption coefficients are experimentally evaluated for each material in the modeled rooms. Typical sounds and noises in these rooms have been recorded. These include the closing of a door, clapping with hands, scratching with a cup on a table and several other sounds.

Phonon tracing computes the pulse response for a given source and listener in a virtual room. The fast convolution (based on the FFT) of this pulse response with an anechoic signal (associated with the source) provides the signal received at the listener's position.

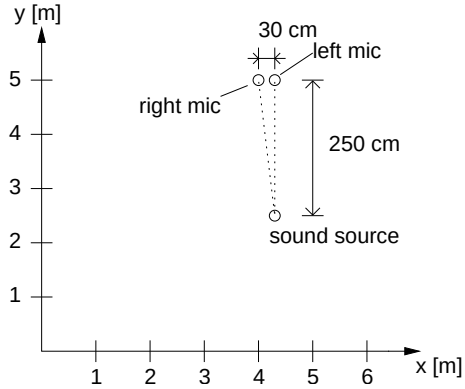


Figure 5. Test setup with microphones arranged frontal to the sound source. Simulated height was about 1 m. The microphones are labeled with respect to the robot.

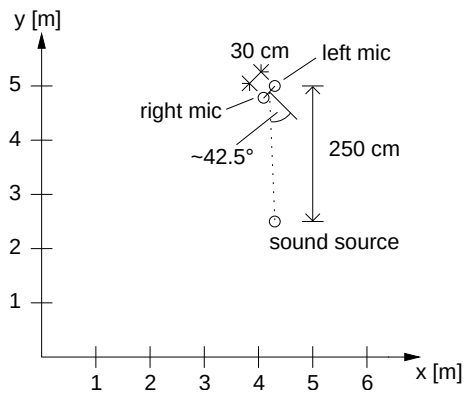


Figure 6. Test setup with sound emerging from about 42.5° to the right of the two microphones. Simulated height was about 1 m. The microphones are labeled with respect to the robot.

With the aid of this simulation, repeatable experiments can be performed. The behavior of the robot can be optimized and different localization algorithms can be compared. The following section will present the results of a localization algorithm based on the simulated microphone sounds.

6 Experiments

In order to validate the accuracy of the sound propagation simulation experiments in three static configurations have been performed (see figures 5, 6 and 7).

In each case the sound source and the two microphones were at fixed positions and a spoken text of about 78 seconds was simulated. The resulting sound at the microphone positions was recorded into a corresponding mono file and each set of two files was used as input for the sound source localization algorithm (see section 3) in turn.

As this algorithm starts with estimating the background noise level to activate the localizer only when "real" noise is present, a section of about 5 seconds of small noise have been prepended to each recorded mono file. Thus the implemented localizer could be used as in reality with the only difference that the sound input came from the two files instead of the two microphones. To simulate a real-time localization the number of samples n that are processed in each execution cycle depends on the execution period T and the sample rate f_s according to equation 5:

$$n = \frac{f_s}{T}. \tag{5}$$

Figure 8 shows the calculated direction of the sound source with respect to the microphones for the frontal test setup (test case a). As can be seen from the figure the mean value is about 0.4962° and the standard deviation is about 5.2706° .

Figure 9 shows the corresponding results for the setup with sound emerging from about 42.5° to the right of the two microphones (test case b). The mean direction value is about 39.1353° and the standard deviation is about 6.2402°

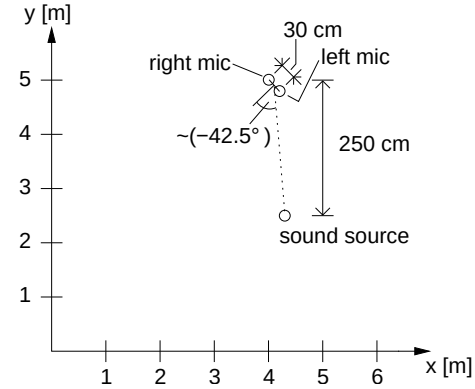


Figure 7. Test setup with sound emerging from about 42.5° to the left of the two microphones (symmetric test case with respect to figure 6). Note: Due to the room geometry the sound propagation conditions are different from that in figure 6

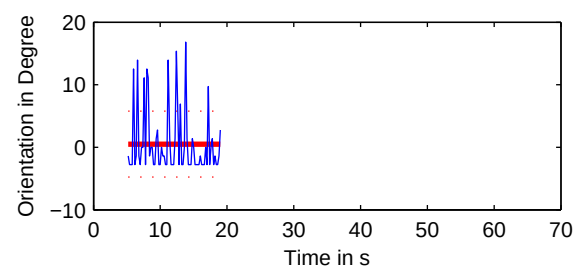


Figure 8. Calculated sound source direction relative to the microphones for test setup of figure 5. The calculation starts after 5 seconds due to background noise prepended to simulated mono files.

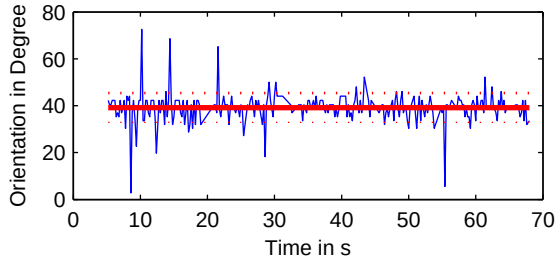


Figure 9. Calculated sound source direction relative to the microphones for test setup of figure 6. The calculation starts after 5 seconds due to background noise prepended to simulated mono files.

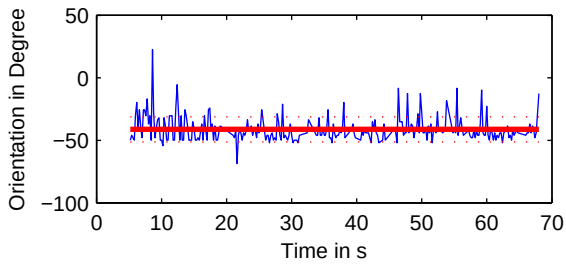


Figure 10. Calculated sound source direction relative to the microphones for test setup of figure 7. The calculation starts after 5 seconds due to background noise prepended to simulated mono files.

.

In figure 10 eventually the localization results for the setup with sound emerging from about 42.5° to the left of the microphones are reported, with a mean direction value of about -41.2373° and a standard deviation of 9.9850° (test case c). Although this test is symmetric to the setup of figure 7, the sound propagation is different to due to asymmetric room geometry. This fact also influences the localization algorithm which produces less severe outliers than in test case b (comp. fig. 9 and fig. 10).

From these experiments it gets obvious that the sound propagation simulation produces qualitatively realistic results, as the simulated sound direction can be estimated by the localization algorithm with an average accuracy of less than 5° . The visible outlier direction values can be traced back to the inaccurate assignment of material constants in the simulation environment which produces a higher reverberation level than in reality.

Thus the experiments prove the validity of the simulation and of the localization strategy with outlier filtering based on source orientation histogramming. Also, two problems have been revealed by these tests. First, the realistic assignment of material constants needed for the sound reflection calculation is not straightforward. Second, the simulation by now is restricted to static source and sink configurations. That means in future developments the simulation should take a microphone movement on a given trajectory into account. This way the simulated audio files can trigger a simulated robot motion, e. g. from a relative start orientation of 42.5° to the right as in test case b to a frontal end orientation, based on a correct source focusing via the presented localization algorithm.

7 Conclusions and Future Work

We presented a sound localization system for mobile robots based on the Fast Fourier Transform. It is able to detect the direction of non-moving sound sources like voices with a resolution of less than 5° . We verified these results with the sound from a transistor radio.

The presented sound simulation is able to calculate how sound spreads in well known environments. This simulation allows to calculate the sound a listener would hear at any position in a given room. It also implements the visualization of spreading sounds within a room.

Combining these two parts results in a simulation of noisy environments for indoor robots. Localization algorithms can be optimized and compared because all experiments a repeatable. The presented experiments show that it it possible to localize a sound source with the help of the simulated sound files.

Future work will include the simulation of moving sound sources and moving listeners. These extensions are necessary for the recognition of humans and the sound localization while the robot is moving.

8 Acknowledgments

This work is funded by the Ministry of Science, Continuing Education, Research, and Culture of the German Federal State of Rhineland-Palatinate within the Rhineland-Palatine Cluster of Excellence ”Dependable Adaptive Systems and Mathematical Modeling”.

References

[1] E. Fabrizi and A. Saffiotti. Augmenting Topology-Based Maps with Geometric Information. *Robotics and Autonomous Systems*, pages 91–97, 2002.

[2] P. Althaus and H. I. Christensen. Automatic Map Acquisition for Navigation in Domestic Environments. In *Proceedings of the IEEE/International Conference on Robotics and Automation (ICRA)*, pages 1551–1556, 2003.

[3] L. J. Latecki, R. Lakaemper, X. Sun, and D. Wolter. Building Polygonal Maps from Laser Range Data. In *Proceedings of the ECAI Int. Cognitive Robotics Workshop*, 2004.

[4] S. Thrun. Learning Metric-Topological Maps for Indoor Mobile Robot Navigation. *Artificial Intelligence*, 99(1):21–71, 1998.

[5] J. B. Allen and A. Berkeley. Image method for efficiently simulating small-room acoustics. *Journal of the Acoustic Society of America*, 65(4):943–950, 1979.

[6] U. Kulowski. Algorithmic Representation fo the Ray Tracing Technique. *Journal of Applied Acoustics*, 18:449–469, 1984.

[7] T. Funkhouser, P. Min, and I. Carlbom. Real-Time Acoustic Modeling for Distributed Virtual Environments. In *Proceedings of ACM SIGGRAPH*, pages 365–374, 1999.

[8] H. Schäfer. Mapping of Indoor Environments with Mobile Robots. Project thesis, University of Kaiserslautern, Germany, 2005.

[9] D. Schmidt, T. Luksch, J. Wettach, and K. Berns. Autonomous Behavior-Based Exploration of Office Environments. In *submitted to 3rd International Conference on Informatics in Control, Automation and Robotics*, 2006.

[10] T. Braun, K. Szentpetery, and K. Berns. Detecting and Following Humans with a Mobile Robot. In *Proceedings of the EOS Conference On Industrial Imaging and Machine Vision*, 2005.

[11] J. Lantto. Sound Source Detection System for Control of an Autonomous Mobile Robot, a Behaviour-Based Approach. In *16-th CISM-IFToMM Symposium on Robot Design, Dynamics, and Control (ROMANSY)*, 2006.

[12] M. Bertram, E. Deines, J. Mohring, J. Jegorovs, and H. Hagen. Phonon Tracing for Auralization and Visualization of Sound. In *Proceedings of IEEE Visualization 05*, 2005.

[13] L. Calmes. A binauraul sound source localization system for a mobile robot. Master's thesis, RWTH Aachen, 2002.

[14] H. W. Jensen. Global Illumination using Photon Maps. In *Rendering Techniques '96 (Proceedings of the 7th Eurographics Workshop on Rendering)*, pages 21–30, 1996.

[15] U. Krockstadt. Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibrations*, 8(18), 1968.