Virtual Acoustics: Paper ICA2016-681

Perception of the reverberation captured in a real room, depending on position and direction

Annika Neidhardt

Technische Universität Ilmenau, 98684 Ilmenau, Germany, annika.neidhardt@tu-ilmenau.de

Abstract

Virtual auditory environments are of increasing interest, in science, but also for industrial applications. With tracking devices, the auditory scene can be explored interactively. Appropriate room simulation algorithms are essential for the plausibility and naturalness of a dynamic scene, as well as for an orientation within. It is well known, that an orientation within a purely acoustical representation of a virtual room is not easy. That especially applies to people with normal vision, who usually orientate themselves mainly by visual information. In a previous study, we investigated the ability of sighted people to distinguish four different listening positions in a shoe box-model of a seminar room. The source was kept in a constant relation to the listener to focus on reverberation. The results showed, that source directivity and training effects have a significant impact. The study presented in this paper will take a closer look at the dynamic binaural auralization of a real room with more complex acoustical properties. An assignment of listening positions within the room should be easier. Furthermore, the question for an appropriate a priori training procedure is addressed. The results will help to improve the design of virtual acoustic environments for dynamic reproduction.

Keywords: dynamic binaural synthesis, virtual room, listening position, auditory perspective
Perception of the reverberation captured in a real room, depending on position and direction

1 Introduction

The demand for virtual acoustic environments is growing with its increasing quality of the auditory illusion. Additionally, it has become common that virtual auditory scenes can be explored interactively, e.g. with tracked head movements or by controlling the position of an avatar in a walk-through scenario.

Reverberation is essential for the creation of plausible acoustical impressions as well as for an externalization of the virtual sources in binaural scenes. Sophisticated room simulation algorithms have been developed over the years to supply the user with a convincing audio environment for dynamic reproduction. But do we make use of dynamic room acoustical information in a real listening environment? Can we hear if the acoustics in a dynamic virtual environment are not reproduced in a physically correct way, e.g. in terms of the correct reflection pattern?

For many applications such a dynamic audio reproduction is combined with a visual representation of the virtual environment. In that case the orientation and position is mainly estimated by visual cues. Therefore, the perception of room acoustical details have to be studied without any visual information of the virtual environment.

2 Previous studies

Perception of room acoustics and its dependency on the listening position

Shinn-Cunningham and Ram [6] studied the sensitivity to room acoustics by capturing the sound field with a stationary head and torso simulator at different positions within a class room. Participants had difficulties distinguishing listening positions with small acoustical differences, suggesting that humans do not have a high sensitivity to the exact timing and direction of arriving reflections.

Neidhardt et al. [4] conducted a similar experiment with a virtual room of approximately the same size. Furthermore, dynamic binaural reproduction was available. It was shown, that the participants were able to distinguish items due to their room acoustical properties. But no subject showed a confident assignment. The performance was better for scenes with a virtual point source than with a virtual loudspeaker.

In a study conducted by Wallmeier and Wiegrebe [8], people were able to estimate, whether they were listening to the scene with a left-handed or right-handed off-center-orientation. In that case, a $2.5m \times 27m \times 4.5m$ corridor was chosen as the room to be auralized. The participants could explore the virtual corridor by self-produced sounds and listening to the synthesized reverberation. The performance was influenced by the position in the virtual room. The distance of a lateral wall had a significant impact. Dufuo et al. [1] also found that nearby walls caused a biasing effect on sound localization. This might be caused by a masking effect of strong early reflection on the later reverberation.
The role of dynamic reproduction and self-motion  
Katz et al. [2] showed that even for blind people, who are used to attentively listening to the acoustics of their environment, interactive exploration is important for understanding the scene and creating a map in the mind. Wallmeier and Wiegrebe [8] found out, that the systematic bias of a close wall which could be observed when participants were listening without self-motion, disappeared when the rotation of body and head could be used to explore the scene.

On the other hand, the assumption that the availability of dynamic binaural reproduction using head tracking would improve the capability of assigning the listening positions within a virtual room could not be confirmed [4]. Moreover, the participants seemed to prefer listening with a still head for an evaluation of the room acoustics. No significant differences in the participants performance using static and dynamic binaural synthesis could be found. But the question remained, whether people could learn to benefit from a dynamic reproduction of room acoustics.

The role of training and adaptation  
The localization of a sound source, but especially the distance perception in a room, depends on room acoustical properties like the direct-to-reverberant-energy ratio (DRR). But in different rooms, for the same distance of a source different DRRs can be measured. To accurately estimate the source position, people adapt their listening and judgment to these variations [5].

Werner et al. [9] have shown, that people can be trained to different room acoustical properties, creating a certain internal reference. The corresponding expectations have a significant impact on the judgment of externalization. Wallmeier et al. showed that sighted people are able to learn discriminating reflective surfaces echo-acoustically in VAEs with a high accuracy [7]. Additionally, in the previous experiment conducted by the author [4] training had a significant impact on the performance regarding the assignment of listening positions.

All these results point out, that in an experiment concerning the perception of room acoustical details a properly designed training procedure is essential.

3 The Experiment

The goal of this experiment is, to find out, how well people can estimate listening positions in a room using a dynamic binaural auralization of the complex acoustical information captured in a real room. The measurement and the corresponding listening experiment are described in the following paragraphs.

Measurement of the BRIRs for different positions  
Binaural room impulse responses (BRIRs) were measured using a KEMAR (45BA) head and torso simulator (HATS). One loudspeaker (Genelec 1030A) was positioned at 0° in a distance of 2.5m in front of the HATS. The HATS was positioned on an electronic turntable (Outline ET250-3D). For the measurement, logarithmic sine sweeps of 2s were used. BRIRs were captured with an angular resolution of 5°. During the recording the ears were located at 1.59m above the floor. That suits a standing person of an average height.
A conference room with a size of $10.31m \times 5.76m \times 3.07m$ was chosen for the experiment. It has a volume of $V = 182.3m^3$ and a reverberation time of $T_{60} = 0.65s$ (broadband). Thus, the critical distance is about $r_{crit} = 0.95m$. Fig. 2 provides an overview of the furniture and its arrangement within the room, as well as the different listening positions, where the sound field was captured. Table 1 contains the exact coordinates of the measured positions within the room.

Figure 1: Setup for the measurement of the BRIRs with a resolution of $5^\circ$

**Signal processing and reproduction setup**  The experiment was carried out in a listening laboratory (Rec. ITU-R BS.1116-1). For the binaural playback AKG K271 MII headphones were used. Headphone equalization was applied to the recordings. The head movements of the listener were captured using an Arttrack3 optical tracking system with an update rate of 60Hz. A graphical user interface was programmed for this application to directly switch between the different listening positions as well as for the test procedure. The dynamic reproduction was implemented using a cross fade in the time domain (duration of 44 samples) in the middle between two measured orientations.

Table 1: **Listening positions in the test**

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>$\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.10m</td>
<td>1.10m</td>
<td>1.59m</td>
<td>0°</td>
</tr>
<tr>
<td>2</td>
<td>4.50m</td>
<td>1.10m</td>
<td>1.59m</td>
<td>-90°</td>
</tr>
<tr>
<td>3</td>
<td>4.90m</td>
<td>4.70m</td>
<td>1.59m</td>
<td>90°</td>
</tr>
<tr>
<td>4</td>
<td>6.45m</td>
<td>4.60m</td>
<td>1.59m</td>
<td>0°</td>
</tr>
<tr>
<td>5</td>
<td>9.85m</td>
<td>2.80m</td>
<td>1.59m</td>
<td>140°</td>
</tr>
</tbody>
</table>
Figure 2: Conference room with a size of $10.3\,m \times 5.7\,m \times 3.1\,m$ and all its furniture: At positions 1-5 the room acoustics were captured with a KEMAR HATS and an azimuthal resolution of $5^\circ$ for an auralization in the experiment. The source was kept in the same distance and angle relation to each of the listening positions.

**The test items** A 14s sample of a music piece played by a solo saxophone was used as the test signal in this experiment. The signal was convolved with the captured BRIRs.

**Listening experiment** 12 people took part in the experiment, 9 male and 3 female. The average age was 29.1 years. Three stated, that they had no experience with listening tests or perception of binaural audio. All reported to have a normal hearing ability as well as normal vision.

**The training procedure** Pretests have shown, that most people quickly learned to distinguish two positions. Additionally, it is well known, that the individual improvement during a training is very different and is furthermore influenced by experience, motivation, fitness and concentration. Therefore, the idea was to start with two positions and extend the choice of positions in dependency of the individual performance.

An extra appointment was set up one or two days before the actual test. First, photos of the different measurement positions and the room were shown to the participant. A schematic picture of the room and all five positions (Fig. 2) was available to the proband throughout the whole experiment.

Then the participants started listening to each of the positions in comparison to each other. They were asked to explore the different room acoustics and find differences between the positions. In the end the proband could to choose two positions he found easiest to distinguished. Then a first single stimulus test scenario was created, where only the two chosen positions
appeared as test items. Feedback was given to each answer by providing the number of the correct position. When the participant felt he was able to distinguished those two positions, he could listen to all positions in direct comparison again and select another one to be added to the choice. The single stimulus test scenario with feedback was started with the three items. The procedure was repeated for the fourth and the fifth position.

At the end of the training session, the participant decided in which order the positions would be added to the choice during the test. This was done, because it is a complex task to explore 5 positions with a dynamic reproduction. Everybody discovers different details which can be helpful in an assignment task.

The test procedure  In the second appointment, the actual test was carried out. Each participant had chosen an order of positions he thought to be the best for him in the test. At the beginning, the proband could listen to the first two of them in a direct comparison. Then, in the first test, the participant had to distinguish those two positions. The test contained 6 items in a random order. Feedback was given after all 6 items were assigned to one of the positions.

Subsequently, the third position was added to the choice. The participant could listen to the three examples in a direct comparison again. Then the next test with 9 items in a random order was carried out. The same procedure was repeated for the fourth and the fifth position with 12 and 16 items in the test.

4 Results
Pretests had shown, that the participants are not able to assign any of the positions correctly by listening to an item without training.

During the direct comparison of the acoustics at the five positions all participants confirmed to hear differences between all of them. But in the single stimulus test procedure, most participants found it quite hard to recognize a position by its acoustical properties.

Each participant had to assign 40 items to the different positions, overall 480 assignments were conducted. For 242 of them the result was correct. In the first part of the experiment, in a choice of 2 positions, 70.8% corrects answers were given. In a choice of three, 60.3% were assigned correctly, in a choice of four 45.1% and finally in the choice of all five positions 33.9%.

It has to be taken into account, that by guessing correct answer rates of 50%, 33.3%, 25% and 20% can be achieved in the different stages of the test. Three participants showed results that were below or at least close to the guessing rate. But since one of them assigned position 5 correctly almost every time, the answers of all participants were included in the analyses.

Table 2 gives an overview how the assignments were distributed over the different positions in a choice of 2 (left) or 3 (right) positions was available. During the training procedure each participant chose other positions to start with. Therefore the number of assignments per actual position varies and is given in the last column. Those numbers equal 100% in each row. It can be seen, that positions 2 and 5 were chosen most often.

The bold numbers in the diagonal show the percentage of correct answers. In a choice of 2 or
3, for each position the majority of assignments was correct. That confirms clearly audible and recognizable differences due to the varying acoustical properties at the different positions.

Table 2: Results: Percentage of correct answers - a) choice of 2 - b) choice of 3

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>num</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>66.7</td>
<td>0.0</td>
<td>0.0</td>
<td>33.3</td>
<td>0.0</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>70.8</td>
<td>12.5</td>
<td>0.0</td>
<td>16.7</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>22.2</td>
<td>77.8</td>
<td>0.0</td>
<td>0.0</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>44.4</td>
<td>0.0</td>
<td>0.0</td>
<td>55.6</td>
<td>0.0</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>4.8</td>
<td>14.3</td>
<td>0.0</td>
<td>4.8</td>
<td>76.2</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 3 contains the results of the tests with a choice of 4 (left) and 5 (right) listening positions. In the final test with all positions, only few participants achieved results above the level of a choice by chance. The best participant assigned 9 out of 16 items (56.3%) correctly, the second best 8 items (50%). It has to be taken into account, that the test with 5 positions was always the last in the session and it took approximately 40min to get through other tests. Furthermore, the listening effort for solving the task is quite high. Therefore most people started to get tired. That might also play some role. But still it is obvious, that it is a very hard task to recognize each of the 5 positions within the given choice.

Since the participants could not be paid, the experiment had to be carried out within 2 appointments per person. But due to some kind of game factor, all of the subjects were motivated to achieve a high rate of correct answers.

Table 3: Results: Percentage of correct answers - a) choice of 4 - b) choice of 5

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>num</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>37.5</td>
<td>12.5</td>
<td>4.2</td>
<td>41.7</td>
<td>4.2</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>16.7</td>
<td>41.7</td>
<td>11.1</td>
<td>30.6</td>
<td>0.0</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>6.7</td>
<td>40.0</td>
<td>26.7</td>
<td>13.3</td>
<td>13.3</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>18.2</td>
<td>12.1</td>
<td>0.0</td>
<td>45.4</td>
<td>24.2</td>
<td>33</td>
</tr>
<tr>
<td>5</td>
<td>11.1</td>
<td>13.9</td>
<td>11.1</td>
<td>2.8</td>
<td>61.1</td>
<td>36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>num</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33.3</td>
<td>13.9</td>
<td>13.9</td>
<td>27.8</td>
<td>11.1</td>
<td>36</td>
</tr>
<tr>
<td>2</td>
<td>11.1</td>
<td>41.7</td>
<td>13.9</td>
<td>27.8</td>
<td>5.6</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>11.1</td>
<td>25.0</td>
<td>19.4</td>
<td>19.4</td>
<td>25.0</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>13.9</td>
<td>13.9</td>
<td>16.7</td>
<td>30.6</td>
<td>25.0</td>
<td>36</td>
</tr>
<tr>
<td>5</td>
<td>16.7</td>
<td>11.1</td>
<td>25.0</td>
<td>2.8</td>
<td>44.4</td>
<td>36</td>
</tr>
</tbody>
</table>

Observations and statements of the participants  During the exploration and training time, the participants developed different strategies of remembering and recognizing the certain acoustical characteristics of the different positions. After the test, the participants explained, which acoustical information they used to estimate the listening perspective. Table (tab. 4) provides a representative selection. Acoustical cues regarding reverberation, apparent source width, spaciousness and timbre were stated to be of help.

Especially for positions 2 and 5 some reliable cues could be found. Both were often described
Table 4: **Descriptions**

<table>
<thead>
<tr>
<th>Position</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position 1</td>
<td>nothing special, similar to Pos4</td>
</tr>
<tr>
<td>Position 2</td>
<td>high source width bright sound, most reverberant</td>
</tr>
<tr>
<td>Position 3</td>
<td>reverb from behind, surprisingly different from Pos2</td>
</tr>
<tr>
<td>Position 4</td>
<td>hard to distinguish from Pos1, no acoustical hint due to the TV screen, not as bright as 2</td>
</tr>
<tr>
<td>Position 5</td>
<td>strong low frequencies, muffled sound, dry, lack of reverb</td>
</tr>
</tbody>
</table>

to be the easiest. In the final test, those to positions were assigned correctly most often. Position 3 was usually one of the last items to be added to the choice. It is also the position with the highest error rate. And this is although several participants said that position 3 sounds surprisingly different.

**Use of direction dependent information** In a previous experiment [4], people seemed to prefer solving this kind of task by listening with a still head. In the present study, again, this phenomenon could be observed. During the training session all participants explored the virtual perspective by turning into different directions, but about the half of the participants went back to listening with a still head in the test. When they knew which position they were listening to, some participants were sure to hear certain walls or e.g. the plant. One participant described that for position 1, in an 180° orientation there was no reverberation on the left ear. But during the test, such cues did not turn out to be of any help. The participants could not confidently assign the corresponding positions.

Another participant tried to use the change of loudness, spaciousness and timbre when turning the head. But he pointed out the difficulty to exactly reproduce a certain movement in order to find differences regarding the room acoustics. That is probably an important reason, why subjects might like to listen with a still head.

About the half of the participants avoided head movements during the test. It remains an open question whether they could gain some additional room acoustical information from very small changes in azimuth.

**5 Conclusions**

In order to study the perception of simulated room acoustics in a dynamic reproduction like for the walk-through scenario, we need to know more about the perception of room acoustics in a real environment and of room acoustics captured in a real environment. This study provides a further step in that direction.

The experiment was designed to investigate the perception of room acoustics depending on the position within a real room. Another goal of the study was to get an idea, whether direction dependent information is of any help in solving this task.
All participants confirmed to hear clear differences between the different positions in a direct comparison. In particular, differences in reverberation, spaciousness, timbre and apparent source width were named. During the actual test, the single stimulus design reduced direct comparison to a minimum. Each subject relied on remembering recognizable acoustical properties. Although this was a hard task, most participants were able to distinguish two or three positions, suggesting audible and recognizable acoustical differences.

But in a choice of 5 positions, even the best participants assigned only about the half of the items correctly. It was hard to find reliable cues to recognize a certain position. This suggests that small random changes in room acoustics might be perceived as plausible as well.

In a walk-through scenario the avatar does usually not jump from one position to another. There will be a continuous change of acoustics. When the participant already have difficulties to estimate the listening position while immediate changes of acoustics are presented, it seems to be unlikely, that they can use room acoustical information for an orientation within the scene. But this should be tested in another study.

The results give an impression of the potential use of room acoustical information in the chosen type of room. A general statement is not possible, further studies are necessary.

Furthermore, this experiment focused on room acoustics. Having exactly the same relative distance and orientation of listener and sound source in different positions within a room is not the usual case. In typical scenes the position and direction of the sound source probably play an important role regarding the estimations of the own listening perspective within the room.

The experiment is designed in a similar way to that presented in [4]. Instead of simulating room acoustics with a simplified room model, this time room acoustics were captured in a real room of a similar size. Furthermore, a step-by-step approach was chosen for the training procedure. Compared to the results in [4] the performance of the participants slightly improved. Probably both, the detailed room acoustics as well as the modified training procedure contribute to that. For a more precise statement, a better understanding of the adaptation and training processes is required.

Without or with only little training the people will not be able to discover small differences like applying the acoustical properties of a nearby listening position. That confirms the finding of Shinn-Cunningham et al. [6]. So far, for this task no clear benefit of head movements can be found.

However, after dealing with the same virtual room for some time and interactively exploring the acoustical properties at different positions in the room, experienced listeners might be able to discover missing sufficiency in the simulation of room acoustical details. Further investigations are necessary to learn more about the listening capabilities after a long-term training.

5.1 Future work

To study the perception of details in the room acoustics in-depth, qualified participants are essential. During the relatively short training long term experience in analyzing and judging room acoustics is of advantage. Therefore, on the one hand a method to check for the qualification and ability of a potential participant would be helpful.
On the other hand, more effort needs to be put into the development of an efficient training method. There might be ways to support the listener in finding essential cues. A similar experiment should be conducted with other rooms. With a different size or shape, the differences between single positions might be more obvious. Wallmeier et al. [8] worked with a long virtual corridor and the participants performed well in the echo-localization task. Furthermore, in order to study the perceived differences in more detail, it will be helpful to record the perceptions using a repertory grid method or working with a certain vocabulary, e.g. SAQI [3].

Acknowledgements

The author would like to thank the Fraunhofer IDMT for providing some of the equipment. Furthermore, thanks to all participants for taking part and putting so much effort in this experiment.

References


