Audio engineering project

Experiments on matching vision and sound in virtual reality

Djordje Perinovic

Betreuung: Dr. Matthias Frank

Graz, 18.04.2022

kunst uni graz

institut für elektronische musik und akustik

Abstract

As virtual reality becomes the new widely used technology for various needs, its plausibility is the most important thing that engineers need to have in mind. The more the vision and sound represent reality, the more users will feel inside the virtual world. The goal of this work is to carry out experiments on how good the sound matches vision and vice versa. For the experiment, headphones and VR set will be used and the evaluation will take place in the virtual world using the hand controllers.

Contents

1 Introduction

About virtual reality

Virtual reality (VR) is one of the newest technologies in regards to simulations, virtual exhibitions and video games. It gave us completely new variant of enjoying art and it became a support for various industries (like in flight or acoustics simulations).

The basic setup consists of 2 base stations used for infrared tracking of the devices inside the room. These base stations are nothing else but infrared transmitters. They are not connected to the computer used for handling virtual reality, but they need to be in proximity of each other and the devices that are being tracked. Beams that they create also need to hit tracked devices without obstacles in between.

The most important parts of the equipment for virtual reality are tracked VR devices: "VR Headset" (mandatory) and "Hand controllers" (optional).

VR Headset consists of 2 lenses for each eye. As user moves head around, differences in the position and rotation are calculated in regards to the infrared beam from base stations and movement is translated to the projection that user sees. The same tracking principle is used for 2 hand controllers (see Figure 1 and online article [1]).

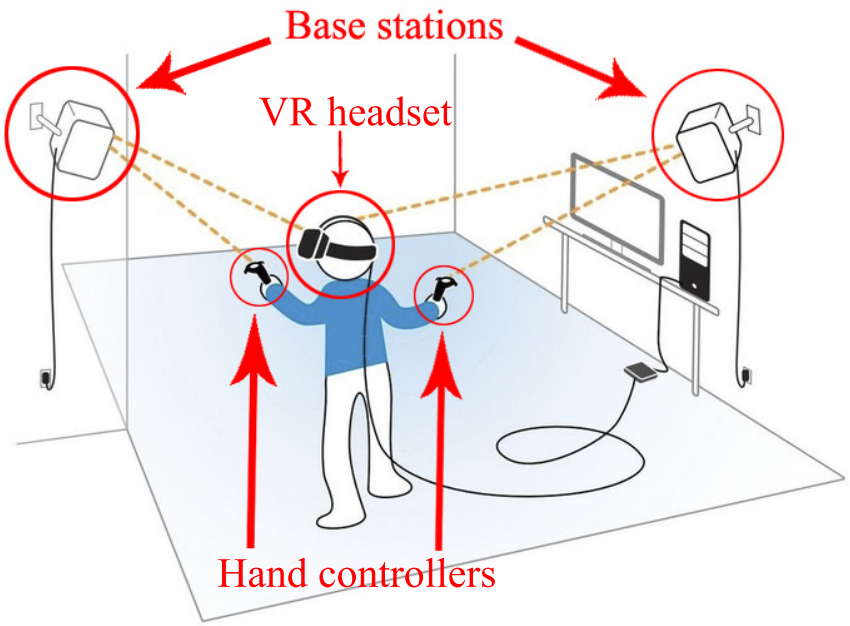

Figure 1 – Equipment and usage of virtual reality (VR)

3D Audio

3D audio (Ambisonics sound) is quite recent way of sound playback that is really important for virtual reality. It can be seen as progression of development of already familiar stereo and 5.1 sound. Stereo and 5.1 are used to play the sound in different directions, but only over horizontal axis.

Since the only requirement from the industry was to play sound to represent live music (e.g. concerts, loudspeakers or headphones for music listening) or movie sound in cinemas, stereo and 5.1 were sufficient for the needs.

As video games started becoming more and more developed, together with virtual reality, ambisonics audio was born. There were already ways of playing sound in vertical axis as well as in horizontal using only stereo, but in order to make it more realistic and plausible for virtual experience, it was needed to play sound in all directions.

One way of doing that was to speaker array that is arranged in a sphere around the listener. It is the best way to hear the most plausible audio and to really feel inside of what's happening. However, for commercial use it's not possible nor practical to have array of speakers available for every opportunity. Therefore, a method has been found to simulate the sphere of speakers and to decode it for headphones that can and have already been used for commercial purposes.

Plausibility

One of the most important conditions for the plausible virtual reality is matching between visual and auditory impression. So called room divergence effect can happen when listening to anechoic rendering while being in the more reverberant room [2]. This same effect can be noticed in VR when visual and auditory cues are induced by simulations [3]. Accordingly, experiments done in VR report that a correct auralization of a room was rated as being the most presence-inducing condition in comparison to conditions with mismatched or anechoic acoustics [4]. There were already experiments where subjects had to match pictures of listening environments with a heard speech playback [5], but VR gives us a way to completely simulate acoustic and visual environment in order to improve plausibility of the VR, which is what this project was about.

Although the plausibility of the speaker array is already pretty accurate to the real world surroundings, the only way of simulating the reality completely correct would be to have array of infinite speakers that would be powered by complex algorithm to simulate acoustics. In order to find a realistic way to improve the plausibility of ambisonics audio, algorithms for decoding of simulated and finite speaker arrays had to be improved. Many experiments have already been done on that topic, where it was confirmed on how

many speakers (which speaker array order) were needed to represent ambisonics audio without noticeable difference to the highest possible order (e.g. order 3 (2^3) for diffuse reverberation [6]).

Idea of the following experiment was to analyze virtual acoustics, like reverberation time, source distance and direction to see if the virtual acoustics represent real sound surroundings well enough.

This experiment was based on the idea of already existing experiment [7], where people used both real and virtual surroundings to evaluate the stimuli. In comparison, this project was based completely on VR in order for subjects to get used to and be longer in the virtual space. Stimuli were generated pseudo-randomly and subjects were not able to get a reference to cheat the correct result, which means that results were based purely on personal feeling.

Since the experiment was based on audio, a subject's hearing and experience regarding sound influence results. Therefore a way of differentiating between subjects was established. What was most important is to differentiate between experienced audio engineers, musicians/sound technicians and total amateurs. Every group had different experience and tendency of evaluating stimuli, but it didn't mean that amateurs did worse than experienced audio subjects. That question will be answered in later sections.

2 Experiment tools

2.1 General

Already mentioned experiment [7] was done in **Unity** real-time development platform, because of it's versatility and possibilities. This experiment was made on the same platform, as some parts could've been easily transferred to the new environment. However, the code had to be rewritten in order to fit the needs of this project, so the only parts transferred were room and speaker 3D model. In order to simulate acoustic experience in the most correct way, **IEM plug-ins** for ambisonics were needed. For that, an VST reading audio workstation had to be used and therefore, the **Reaper** DAW ("digital audio workstation") was chosen, because of it's ability to process 64 channels in one track and to read OST messages. Over **OST** ("open sound control" protocol) it was possible to establish a communication between Unity application and a DAW.

As VR set for experiment control, the **HTC Vive** was used, provided by **IEM**, the "Institute of Electronic Music and Acoustics" (part of the "University of Music and Performing Arts" in Graz). For sound playback, overear headphones **Sennheiser HD380 Pro** were connected.

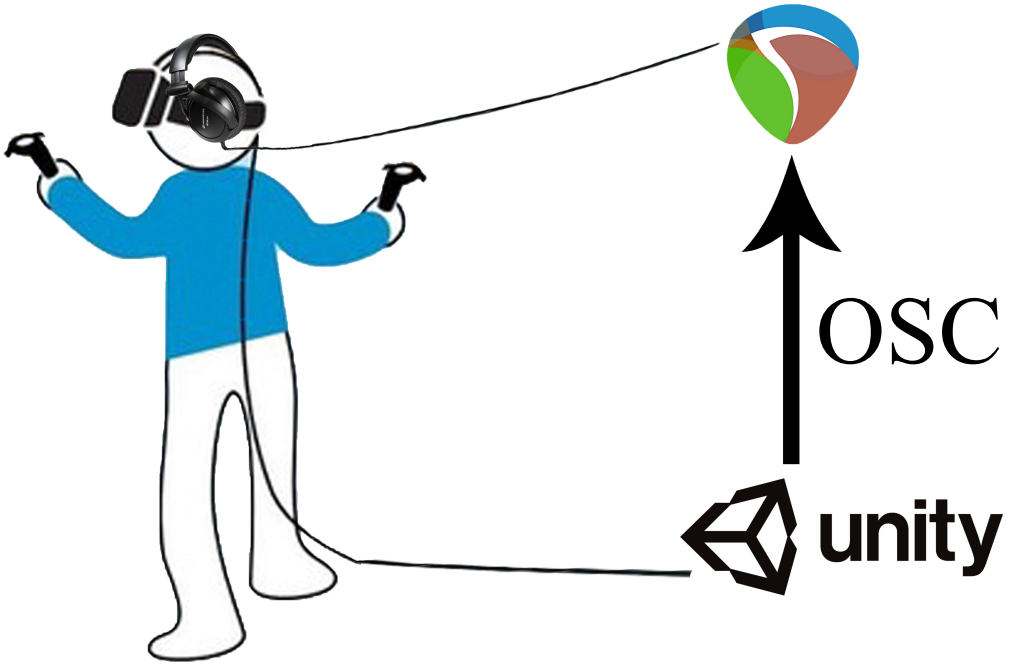

Figure 2 – Connection between Reaper and Unity

Information flow can be then described as following (refer to figure 2):

1. User moves head wearing HTC Vive headset and headphones as well as hands holding 2 motion controllers. Controllers are used for interacting with objects inside the virtual reality displayed on the headset. Positions, rotations and commands are being calculated by the HTC Vive system.

NOTE Under "HTC Vive System" it's meant HTC Vive interface. It's a small device which is the brain of the VR set and it exchanges information between computer (connected via USB) and the headset. It also processes video received over DisplayPort or HDMI to the headset, as well as the sound.

2. This information is then translated by the **SteamVR** platform and sent to Unity application, used for handling everything regarding vision and graphics of the project.
3. Position of the source (virtual speaker), receiver (subject doing the experiment) and all information regarding acoustics (like room size, reverberation etc.) is sent via OSC from Unity application to Reaper.
4. Reaper uses received information to control corresponding plug-ins and calculates ambisonics environment.
5. Resulting sound is being decoded for headphones and sent over the HTC Vive audio output (e.g. user turns head to the left while standing in front of a speaker and he will hear the sound more on the right ear).

2.2 Unity

Unity is a type of software normally used for programming video games (so called "game engine"). It makes programming of video games and expressing artistic skills much easier. What it does basically, is that it gives an user interface to a developer for easier design and programming. It has built-in physics, UI system, graphics, input/output, methods (like `Start()` being called on the first frame and `Update()` on every frame), sound handling and much more of what a developer might need. However, sound handling from Unity wasn't usable for this project, because it can't handle ambisonics (therefore Reaper was used). It's also important to mention that it uses C# as programming language and is object-oriented.

After base stations are running and seeing each other and all of the relevant equipment, SteamVR can do its job. Using a plug-in from Unity VR assets (XR-Plug-in) it's possible to easily implement head and controller movements in the application. By installing this plug-in, a dozen of models and built-in scripts become available for use, and by replacing a regular "Camera" with e.g. "CameraRig" from a plug-ins folder, virtual reality is already in place (visible inside "Hierarchy" in figure 3).

As well as for virtual reality, Unity doesn't understand OSC right away. In order to send, for sound relevant, information to Reaper, OSC script had to be installed [8]. Since Unity makes it easy for us, it's a good idea to add separate objects to hold a corresponding OSC address and port. After OSC script is added to some objects, these new options become available (see figure 3).

These objects can be later referenced by the "SendPositionOnUpdate" and "SendRotationOnUpdate" scripts (can also seen on the picture), to use correct ip addresses/ports. They can be then modified to send any information that is needed. For example, highlighted object was used to send headset rotation, but in the same way was the source rotation/position, as well as other parameters handled by Unity.

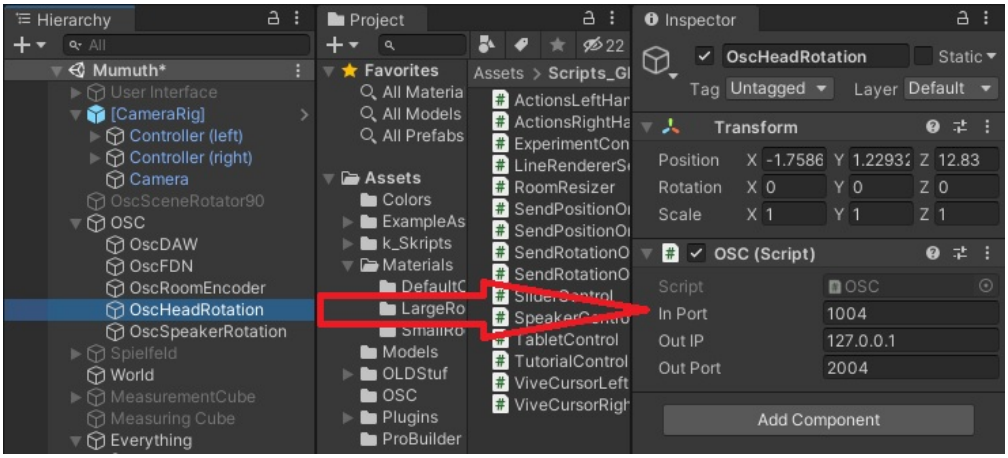

Figure 3 – OSC implementation in Unity

2.3 Reaper setup

Reaper DAW had to be set up to play the sound as realistic as it can be. As messages are being sent over OSC from Unity application to Reaper, they are being received by the corresponding plug-ins (e.g. "Scene rotator" plug-in on the master track rotates whole sound picture and is therefore responsible for the subject's head rotation). Signal flow can be then explained as following (see figure 4):

1. A male speech lossless track from [9] is imported as a source on the first track ("Room encoder").
2. In [6, 10] discovered spatial resolution was taken into account, which says that for diffuse reverberation as of 3rd order of ambisonics, the difference to the 7th order (highest, 64 speakers) can't be noticed. For the early reflections even lower order is sufficient [11]. Therefore, "kronlachner" mcfx convolver plug-in [12] for 16 channels was used. In the project prior [7] to this one, a preset for "Behritone" reference speaker was imported, so here was done the same. It basically tells for each ambisonics channel how loud it has to be. For example for the speaker directivity pattern it's clear that the channel facing the front should be the loudest. Depending on the speaker type and model, other channels are then tuned so that they represent it's directivity in real life.
3. 16 channels are coming then into the first "Scene rotator" representing speaker (source) rotation. By default it's facing front, but other directions were used during the experiment as well.
4. Direction-corrected ambisonics channels are then sent to the "Room Encoder" which is used for generating room reflections, so that subject can have impression of being in a room. By default it's 30x15x9m, which are the real dimensions

of the experiment room ("György-Ligeti-Saal" hall in Mumuth, Graz [13]). In this plug-in source and listener(subjects) position are tuned, as well as additional attenuation, so that e.g. reflections from the walls on the right or left are louder than reflections from the curtain on the front.

5. Signal is then being sent to another track and this track is responsible for the reverberation of the virtual room ("diffuse 3"). Diffuse reverberation employed 64x64 frequency-dependent feedback delay network (FDN [14, 15]). This track receives to the room size normalized audio signal and generates reverberation with repeated sequences of the signal over time. This signal is being corrected with additional EQ plug-in to reduce metallic sound.
6. Both of these tracks are then being sent to the master. On master, 3 plug-ins can be seen. First is "Scene rotator", but this one is used for head rotation, because signal from both tracks has to be rotated to replicate the room acoustics in the most realistic way. The resulting signal is, however, still in 16 channels and it needs to be played on just 2 (left and right headphones speaker). Therefore, using a state-of-the-art magnitude-least squares approach a binaural rendering from 3rd order Ambisonics was carried out [16, 17]. Additionally, output signal was being limited by the simple limiter at the end, just to reduce the possibility of distortion in the sound.

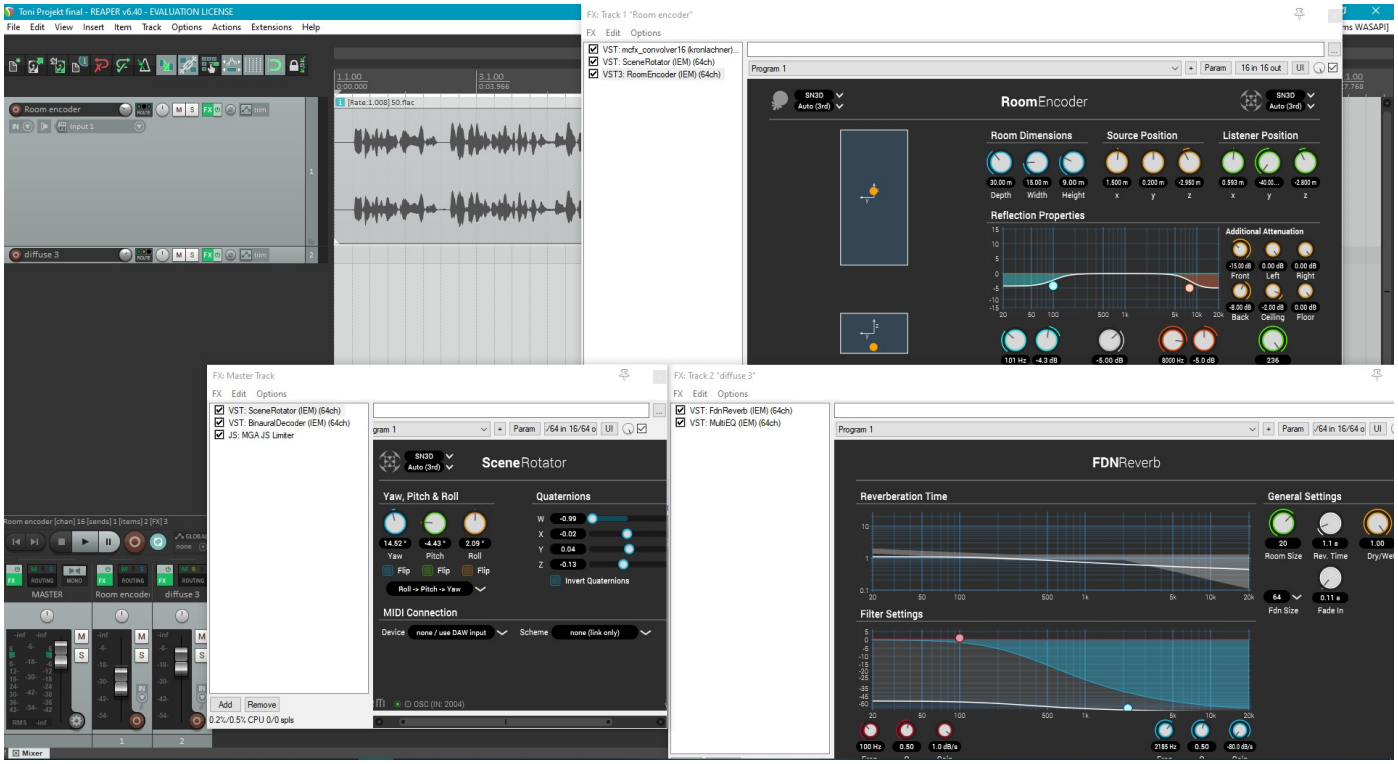

Figure 4 – Reaper setup

3 Experiment layout

In this chapter, it will be explained how the experiment was experienced by the subjects. Since many of them never had experience with a virtual reality, it was a good idea to make some kind of introduction. It was done in 2 steps, first was the personal introduction before they put VR equipment on themselves and second was after they put headset on and take controllers. They are then welcomed to the experiment and into the "Tutorial room".

3.1 Personal introduction

Personal introduction consisted of a simple explanation of which equipment will be used, what the subjects are going to see/hear and how the experiment will look like. Afterwards, help was given to adjust the headset correctly to avoid blurriness and to place fingers on the right buttons of corresponding controllers.

3.2 Tutorial room

After the headset and controllers are equipped, the experiment is started. After the splash screen, subject is in the tutorial room which consists of many slides (figure 5):

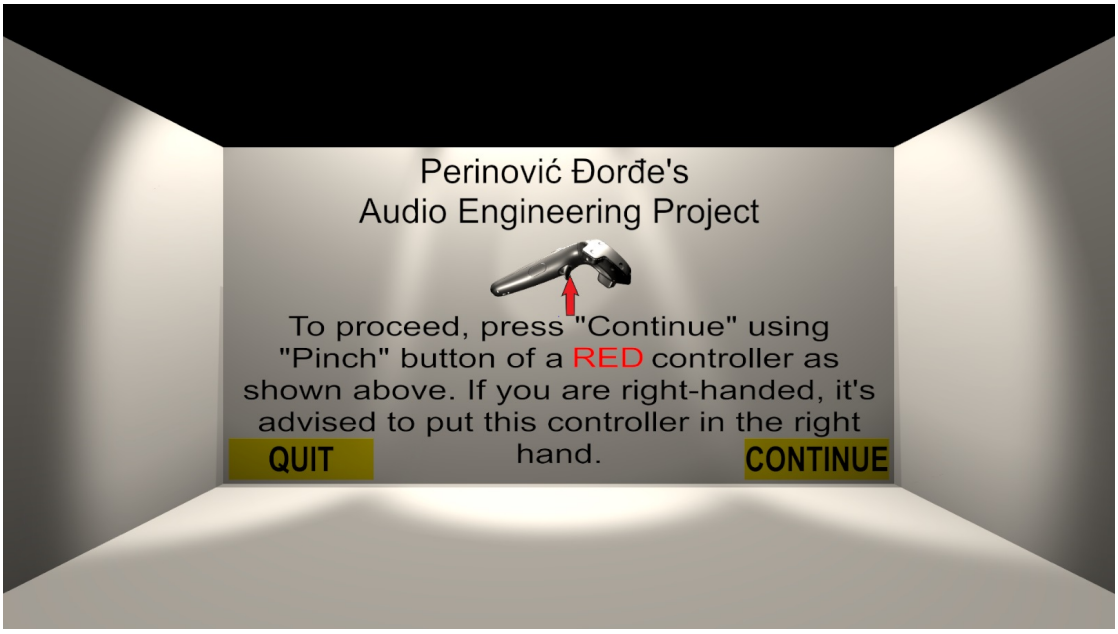

Figure 5 – Tutorial room of the experiment (first slide)

So, the controllers are named "red" and "blue" controller. Controllers are the same color in the virtual reality as in real life (black), but their respective pointers are by default in the mentioned colors. The idea was to have a **tablet** in the virtual reality, which is going to be used to control values important for the experiment analysis. In order to be able to walk around the room and adjust the values at the same time, it was necessary for the subject to be able to move the tablet.

So, in order to avoid confusion with many buttons that controllers can offer, red controller was used to interact with progression of the experiment (adjust stimuli values, click on yellow buttons like "Continue" or "Next" and play/pause of the playback) and blue controller was used more as a help (for moving a tablet).

One more feature for controller pointers was added and it was for them to change colors when corresponding objects were hit. So when red controllers **hits** any yellow button (or slider on a tablet) it will turn yellow as well, so that user knows that the interactable object is being pointed at. The same logic was with blue controller and cyan tablet. After subjects click the "**pinch**" button (figure 5) on the corresponding object, it will fire

its function and turn green just like the controllers pointer (see example on figure 6).

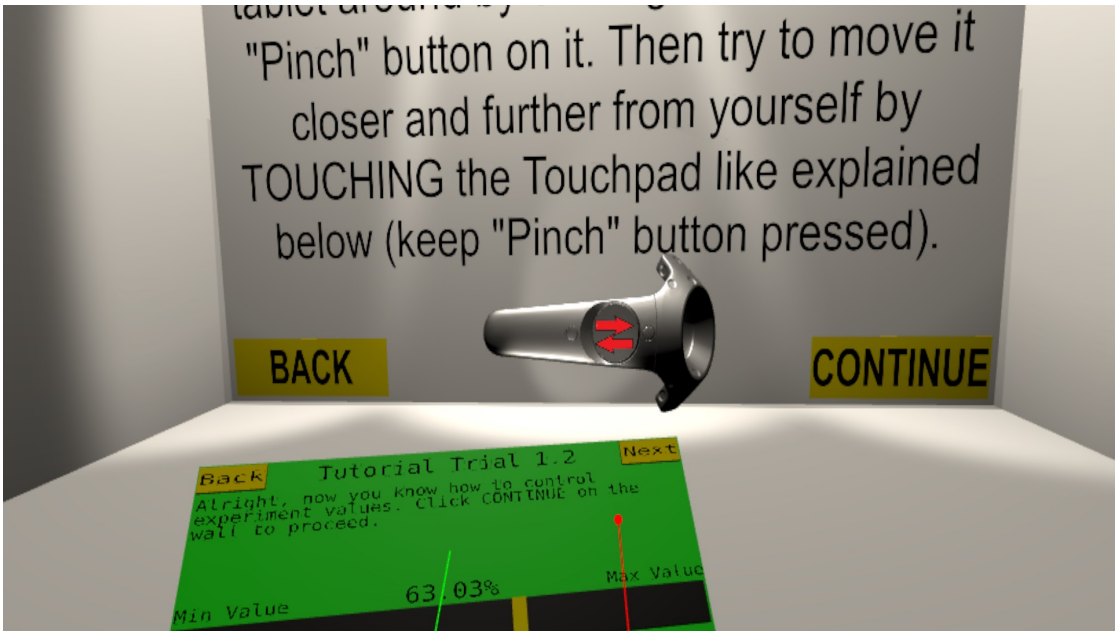

Figure 6 – Tablet being selected and moved around during tutorial

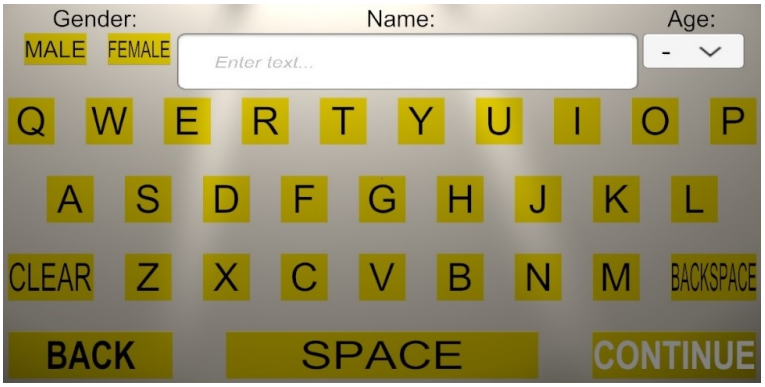

Figure 7 – 3rd tutorial slide (info input)

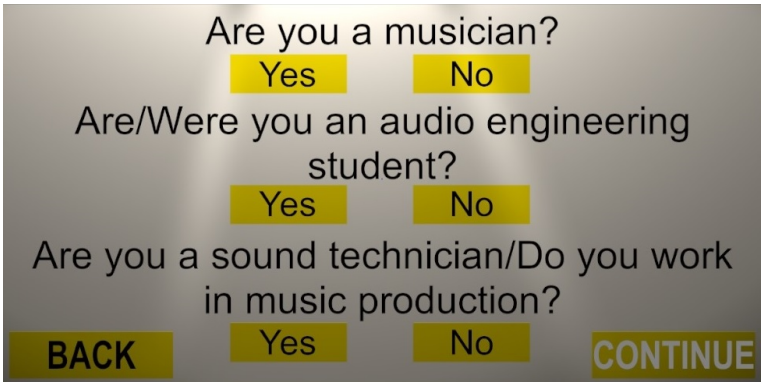

Figure 8 – 4th tutorial slide (questions)

Besides numerous explanations of the usage of the equipment, few tutorial slides were used for input of basic subjects information like name, gender, age and experience with the audio, like seen in the figures 7 and 8. All of this information is later added to the exported results of the subject, so that comparison between results of subjects with different experience can be analyzed. As seen in the figures, "Continue" button has grey letters and can't be clicked unless all of the information is properly entered, since it would corrupt generation of the necessary exported results.

On the last slide of the tutorial part, subject is being informed that the tutorial is complete and that by clicking on the "Continue" button, he/she will be teleported to the experiment room. In the experiment room, already in tutorial introduced tablet is being shown and it can be interacted with. All further information for the experiment is being explained on the tablet and the subject is ready to proceed without interruptions. However, if any further instructions were needed, experiment designer was present throughout the whole experiment.

3.3 Experiment room (György-Ligeti-Saal)

In the experiment room (virtual "György-Ligeti-Saal" hall in Mumuth of KUG [13]) a subject is presented with a virtual speaker and a tablet where the first experiment trial is explained(see figure 9).

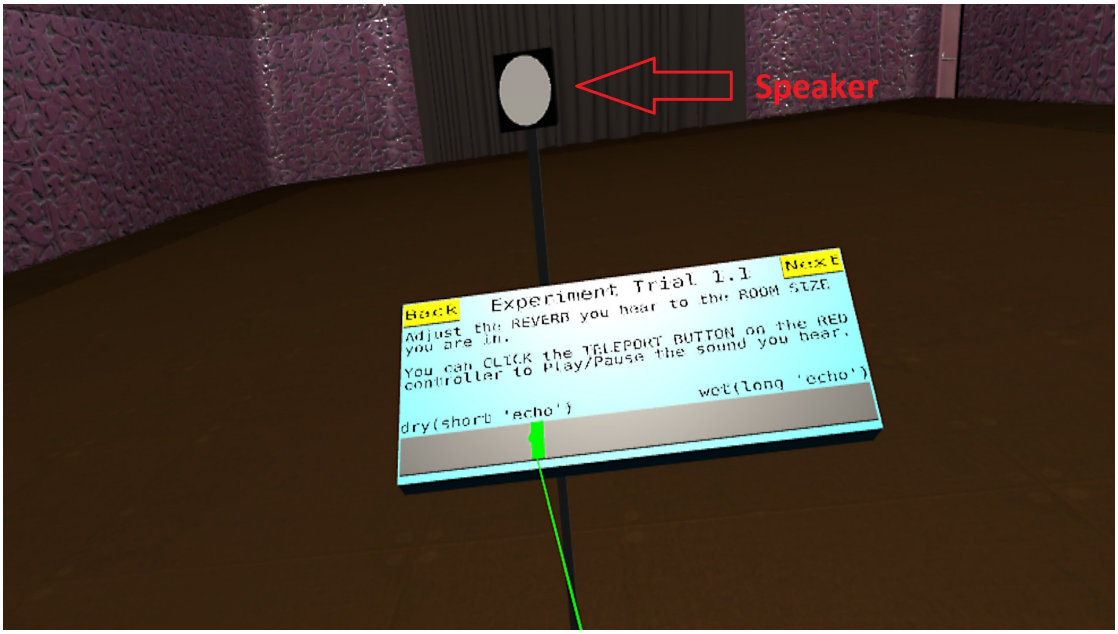

Figure 9 – Subject adjusting slider (FDN-Reverb) according to the room size

Experiment was divided in 6 parts which were further divided in trials. Every trial was clearly explained as well as which stimuli value is being controlled. However, subject could only tell in which direction the value is being changed (e.g. longer or shorter reverberation), but not the exact value that is being set, which meant that the subject could only rely on his/her hearing. This part was the most important because here were generated the values for analysis. Experiment consisted altogether from 40 trials and the complete duration for every subject was around 20 minutes, which depended the most on the experience and willingness to do experiment in detail.

For all parts except part 1 and 2, room size was at a default value, which are the already mentioned real life dimensions of Mumuth. For all parts except part 3 and 4, a virtual speaker was 1.5 meters away from the subject (default value). Also, for all parts except part 5 and 6, the virtual speaker was facing the subject (rotation of 0 degrees).

First 2 parts of experiment consisted of 10 trials and other 4 parts of 5 trials each, which can be seen detailed in the following table (table 1):

Experiment trial	Stimuli value	Detailed explanation
From 1.1 to 1.10	FDN-Reverb	Subject sees a different room size for every trial. Adjusting a slider, sound reverberation is being changed to match the room size.
From 2.1 to 2.10	Room size	Subject hears a different sound reverberation for every trial. Adjusting a slider, room size is being changed to match the sound reverberation.
From 3.1 to 3.5	Speaker distance	Subject hears a sound being played at a different distance every trial. Adjusting a slider, the virtual speaker is being moved to match the position the sound is being heard from.
From 4.1 to 4.5	Source distance	Subject sees a virtual speaker at a different distance every trial. Adjusting a slider, the position of the sound being heard is moved to match the position of the virtual speaker.
From 5.1 to 5.5	Speaker rotation	Subject hears a sound being played at the specific position, but facing different direction every trial. Adjusting a slider, the virtual speaker at that position is being rotated to match the corresponding direction.
From 6.1 to 6.5	Source rotation	Subject sees a speaker at the specific position, but rotated differently every trial. Adjusting a slider, the direction of the sound being heard is being rotated to match the direction of the virtual speaker.

Table 1 – Experiment layout

After every trial the subject clicks button "Next" on the tablet and the adjusted value is being stored for the export. Subject was able to click "Back" and go to any trial from before or even to tutorial room to change data. After the subject has adjusted the last trial, he/she is informed that the end of experiment has been reached and that it's still possible to change any values from before if needed. If the subject is satisfied, a button "Finish" can be clicked and the results are then exported as a .txt file to the application's folder.

Every subjects result file is named after the format: 'Subjects Name'+'Gender'+'A (if audio engineer)'+'S (if sound technician)+'M (if musician)'+'Subjects age'+'.txt'. As an example, 26 years old male subject with all questions on figure 8 answered with "Yes" generated a results file with a name: "DJORDJE PERINOVIC M ASM 26.txt"

4 Result evaluation

As already mentioned in the previous section, for every subject one .txt file was generated for later analysis (see one of the results in Appendix A). Only the numbers were exported so that it is easier to read the files in the matlab, but for easier understanding, referring to the mentioned appendix, it should be added:

1. for first 10 values ("FDNReverb to RoomSize:"), a unit "second".
2. for second 10 values ("RoomSize to FDNReverb:"), a unit "cubic meter".
3. for third 10 values, a unit "meters to listener".
4. for last 10 values, a unit "degrees to listener".

In the same way reference values are generated which represent mathematically correct values (see Appendix B). These files were then used in the matlab to generate median values with 95% confidence intervals. To additionally do a statistical analysis of the results, p-values were generated, which use Bonferroni-Holm-corrected Wilcoxon signed-rank tests. Let's start with all results, regardless of their experience with audio/music:

4.1 All results

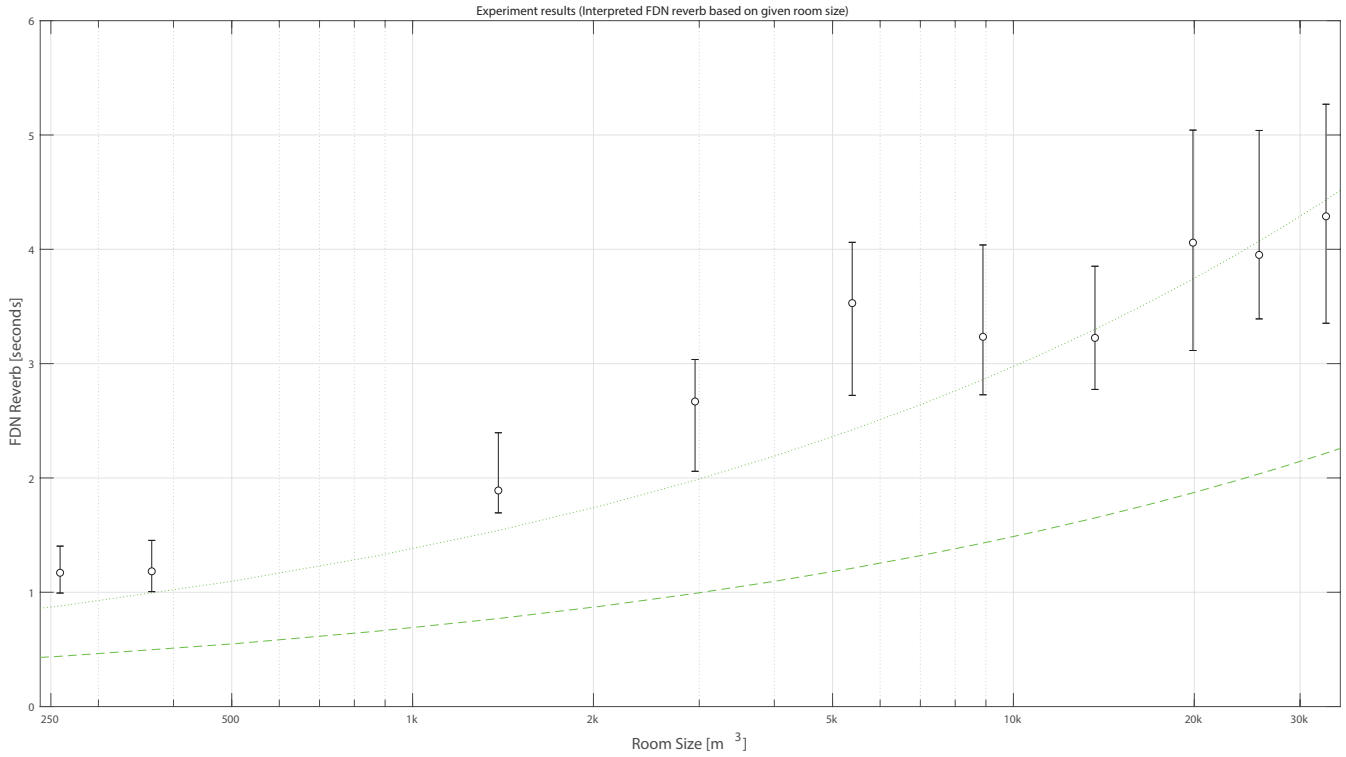

Figure 10 – FDN reverb to room size (all results)

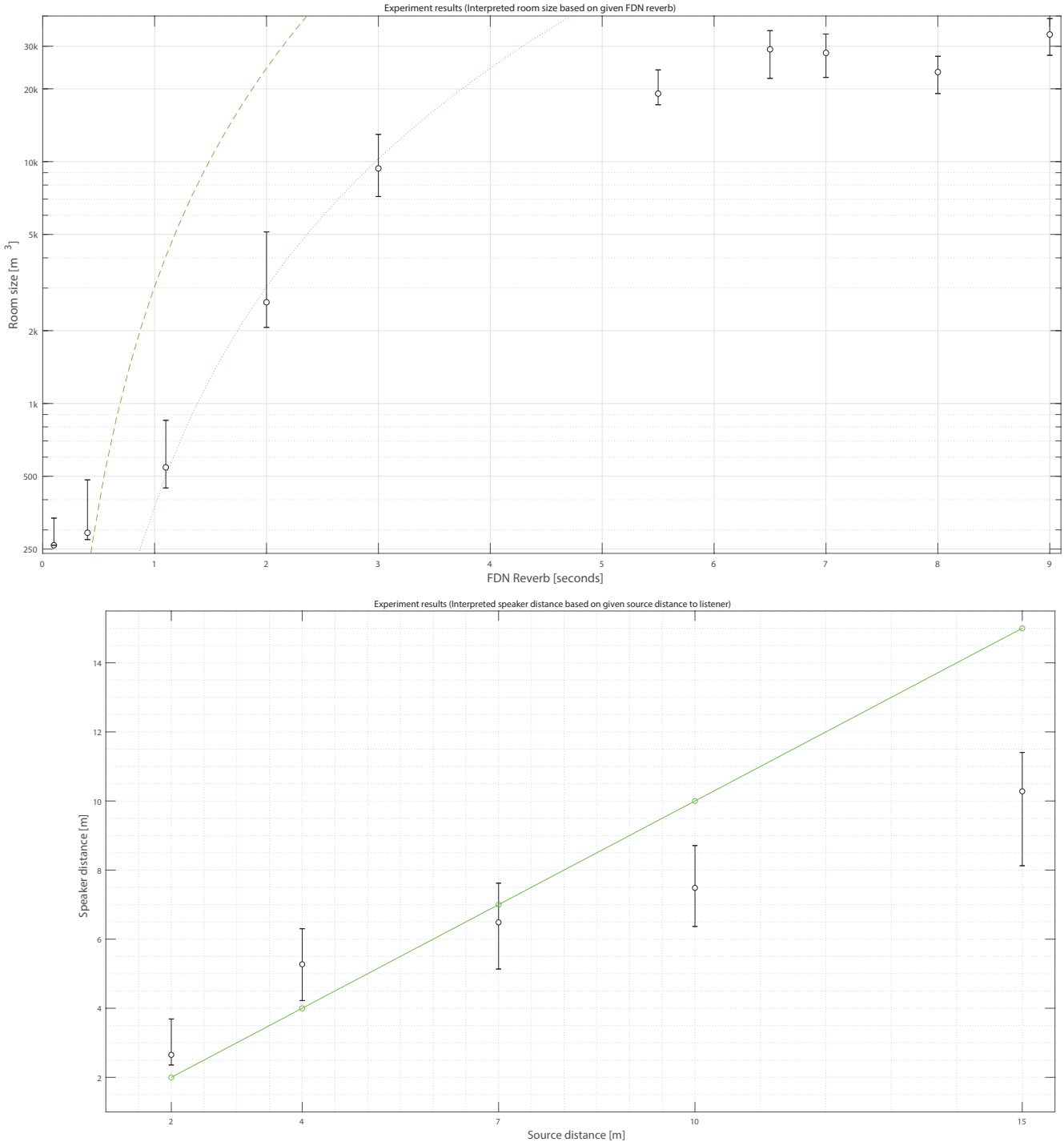

Figure 11 – Top: room size to FDN reverb (all results)
Bottom: speaker distance to source distance (all results)

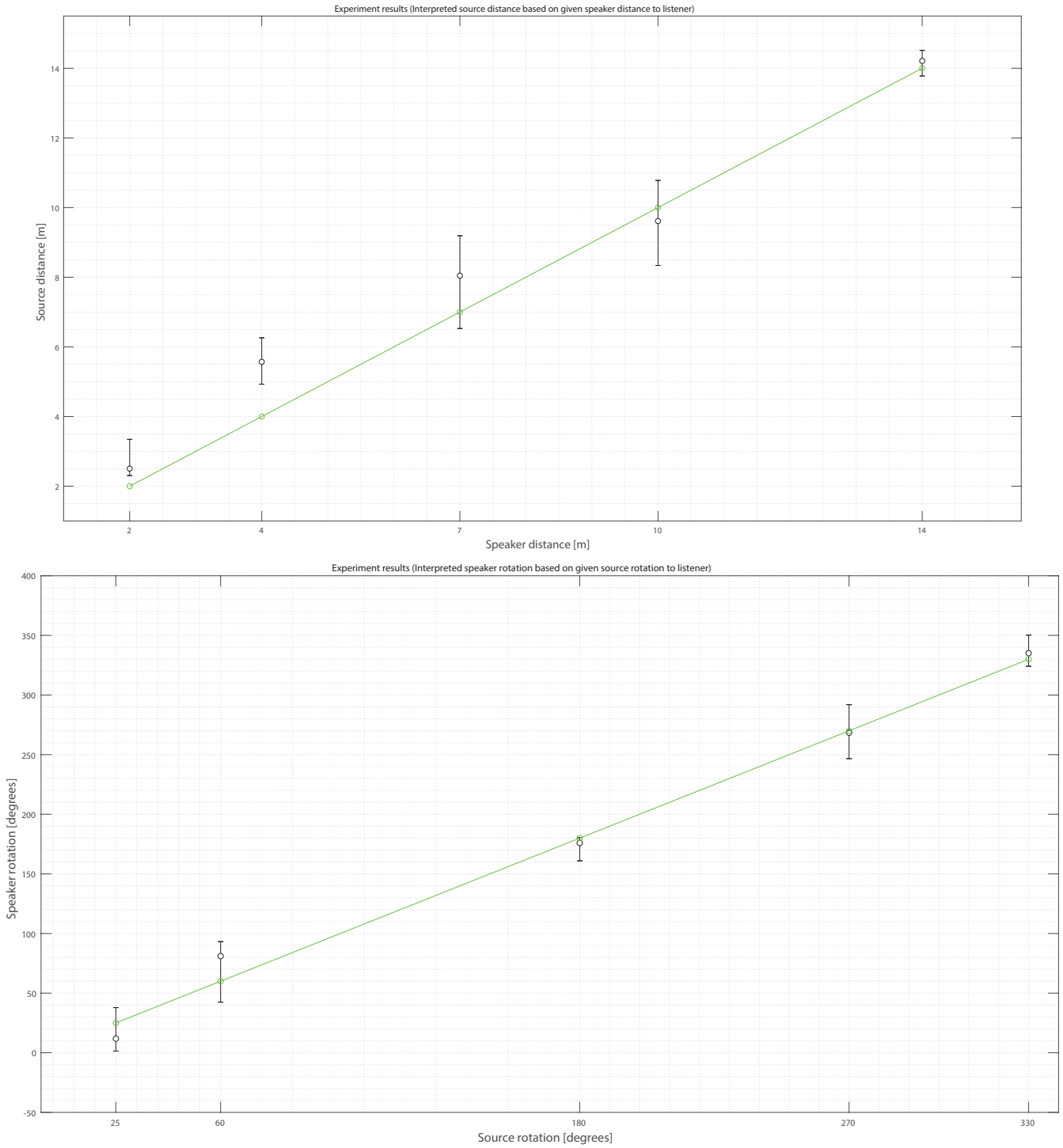

Figure 12 – Top: source distance to speaker distance (all results)
Bottom: speaker rotation to source rotation (all results)

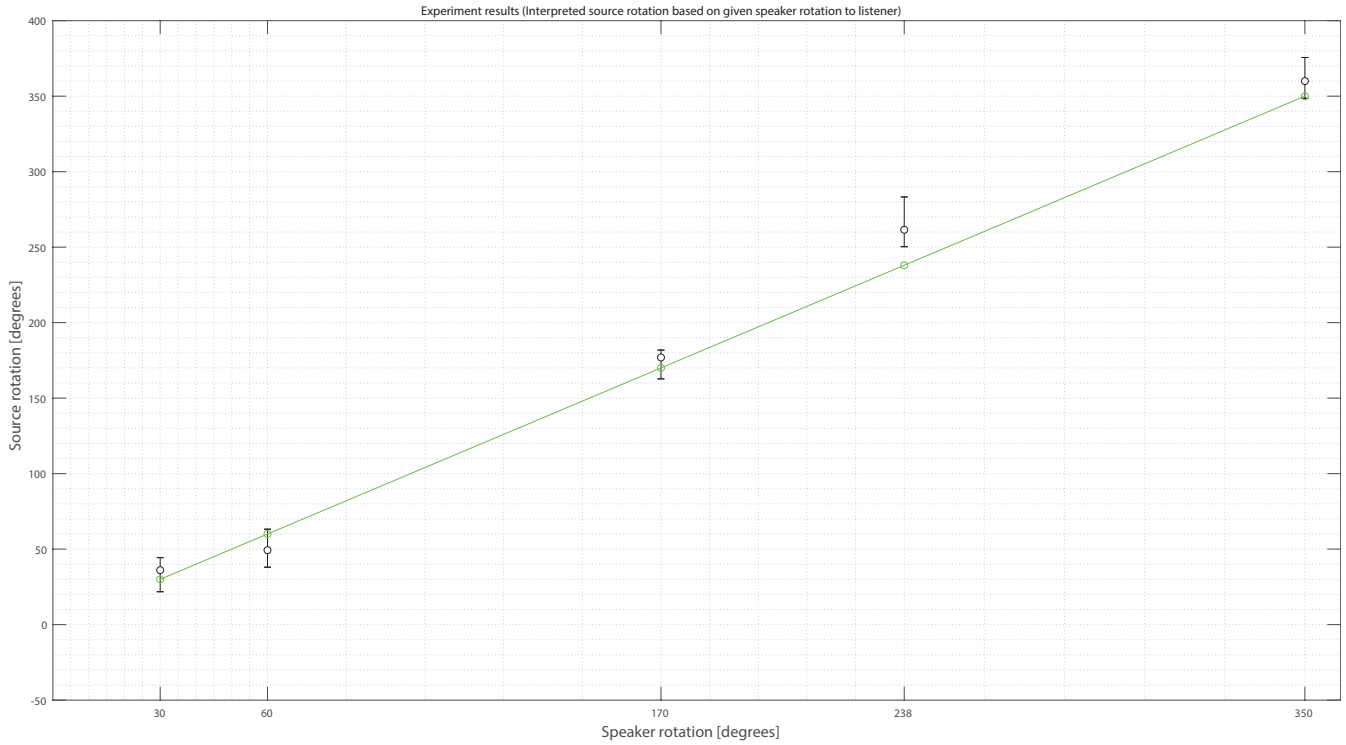

Figure 13 – Source rotation to speaker rotation (all results)

From the figure 10 to figure 13, results from different trials are shown and sorted ascending. It was altogether 34 results, from which 11 results of audio engineers and 14 results of musicians/sound technicians were gathered. Green line connecting circles of the same color can also be seen in the plots and they represent reference values for every trial. Additionally, a second dotted green line can be seen on reverberation evaluation and it represents a reverberation 2 times longer than the correct one. As seen from figure 10 it can be noticed that all of the subjects tended to evaluate reverberation for all room sizes between 1s and 5s, probably because it was the first part of the experiment and they didn't have an audio reference of how the room should approximately sound like in real life, which would ruin the point of the experiment.

When comparing how subjects evaluated the part where they need to adjust reverberation to the part where they adjust room size (top picture on figure 11), it can be concluded that subject found it much easier to adjust their visual surroundings to the audio than vice versa. As seen from figure 11(bottom) to figure 13 it can be clearly seen that subjects found it much easier to match the sound with regarding distance and direction in comparison with the reverberation and room size.

These were the results of all 34 subjects. In the following plots, results will be separated in few groups so that it can be seen if the experience of the subjects made any difference.

4.2 Musicians/Sound technicians

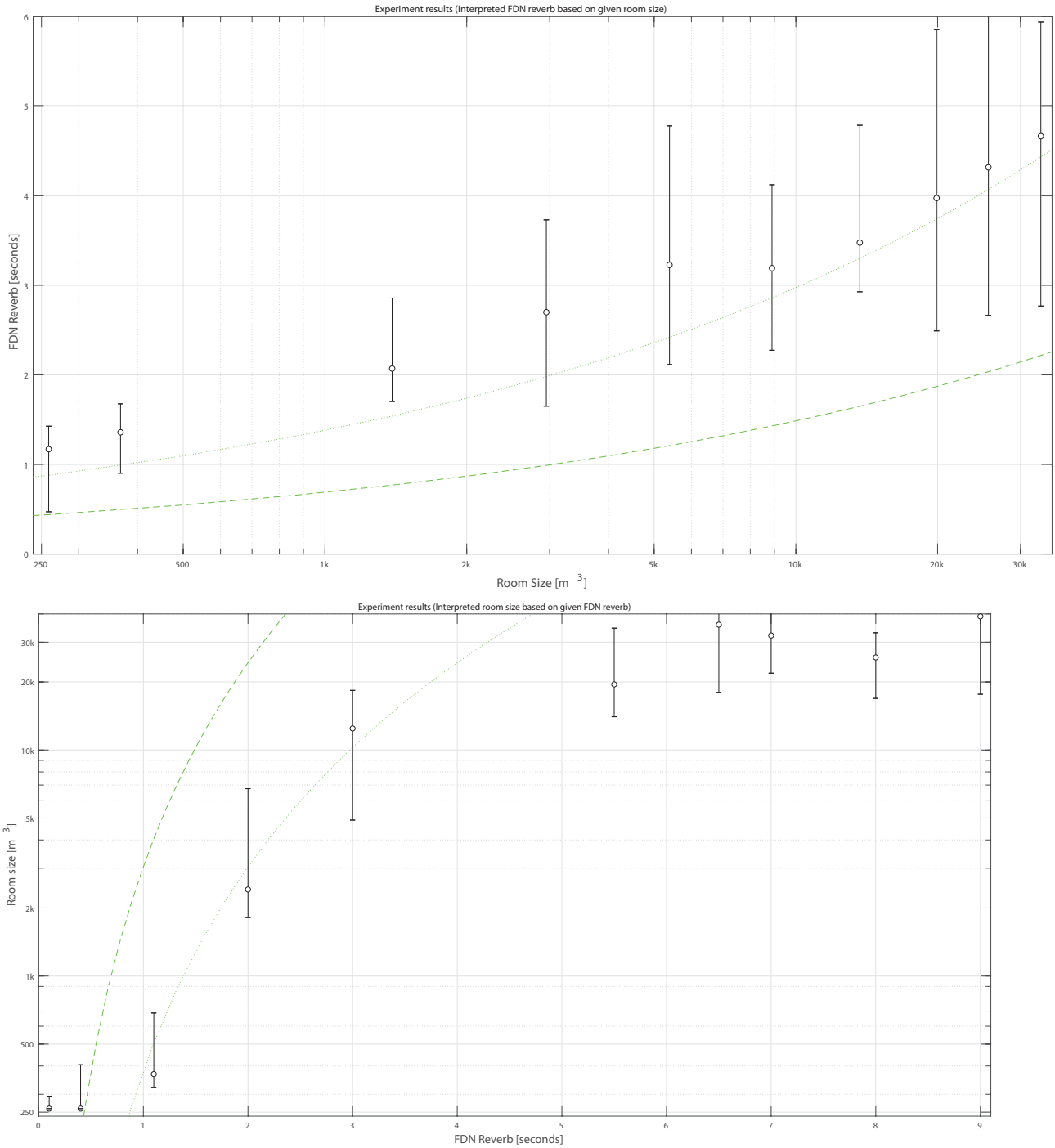

Figure 14 – Top: FDN reverb to room size (musicians and sound technicians)
Bottom: room size to FDN reverb (musicians and sound technicians)

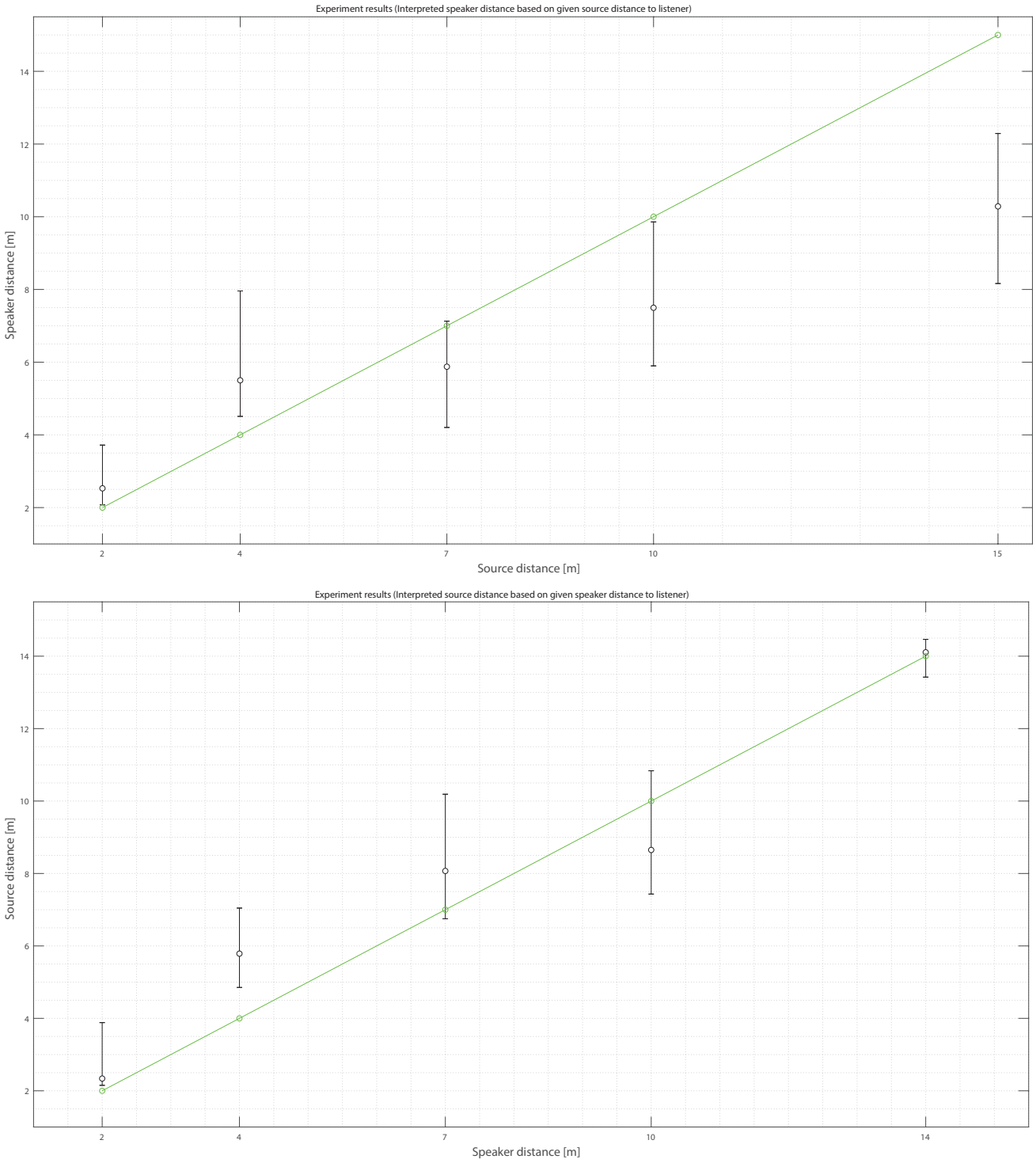

Figure 15 – Top: speaker distance to source distance (musicians and sound technicians) Bottom: source distance to speaker distance (musicians and sound technicians)

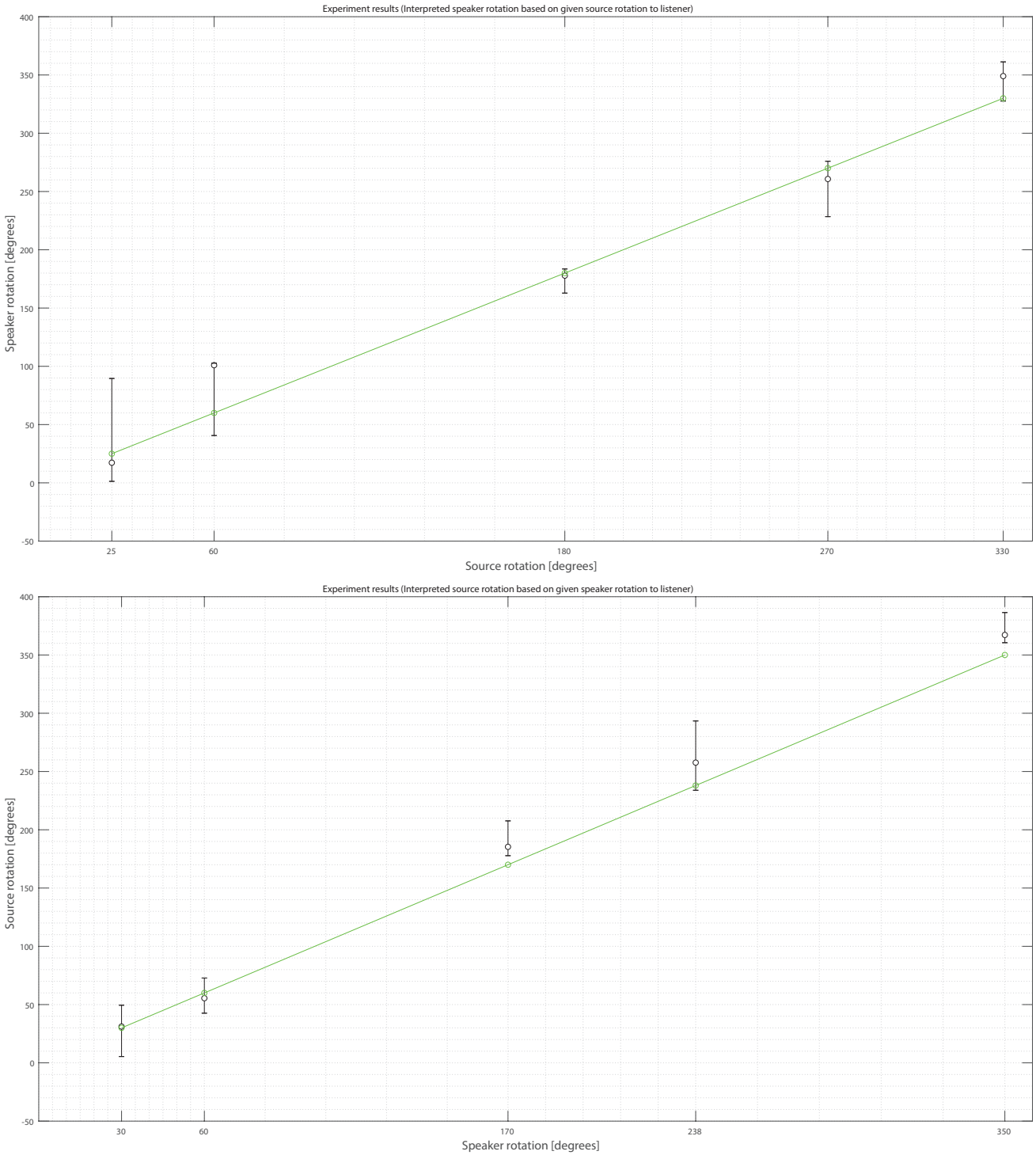

Figure 16 – Top: speaker rotation to source rotation (musicians and sound technicians) Bottom: source rotation to speaker rotation (musicians and sound technicians)

4.3 Audio engineers

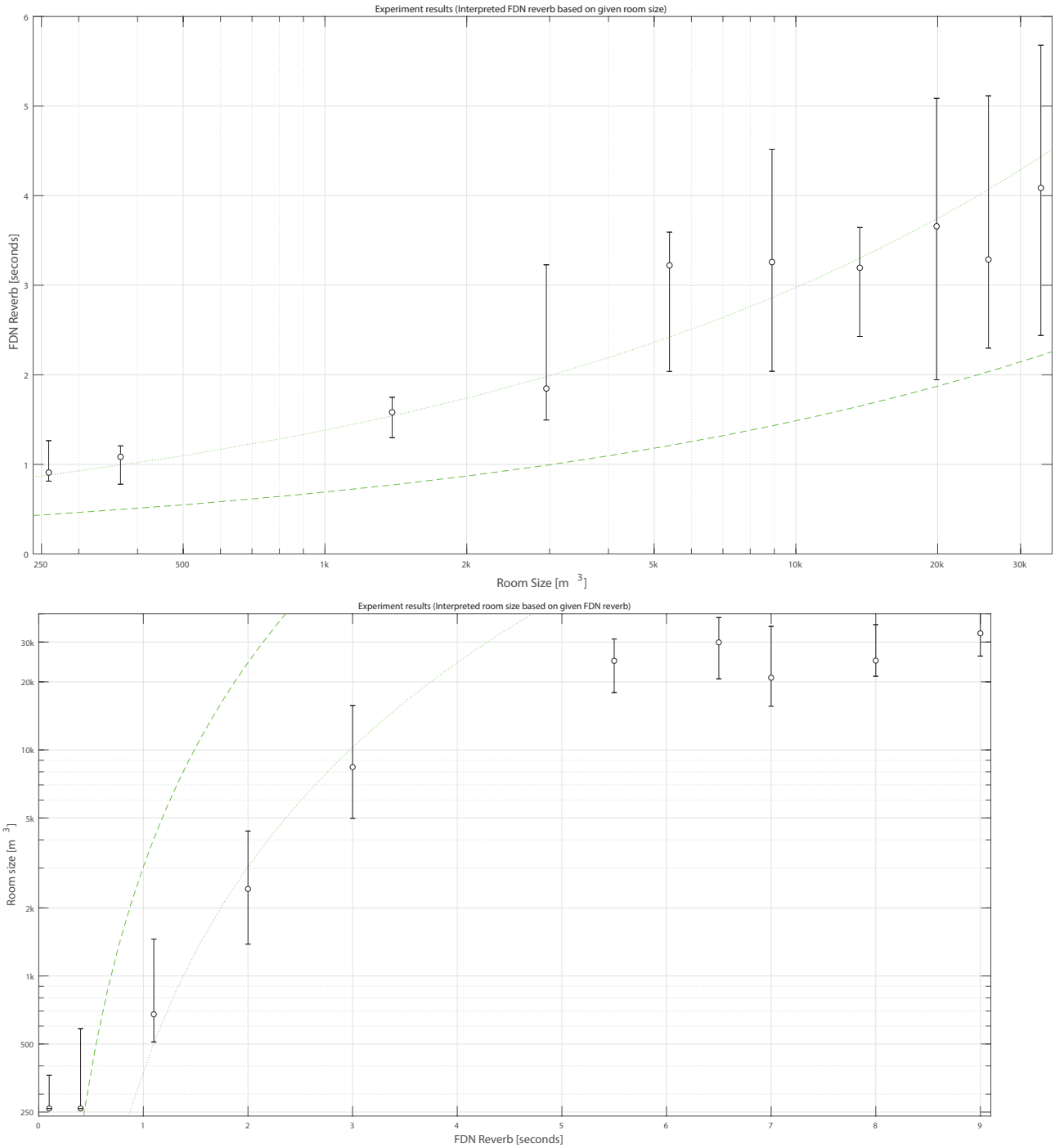

Figure 17 – Top: FDN reverb to room size (audio engineers)
Bottom: room size to FDN reverb (audio engineers)

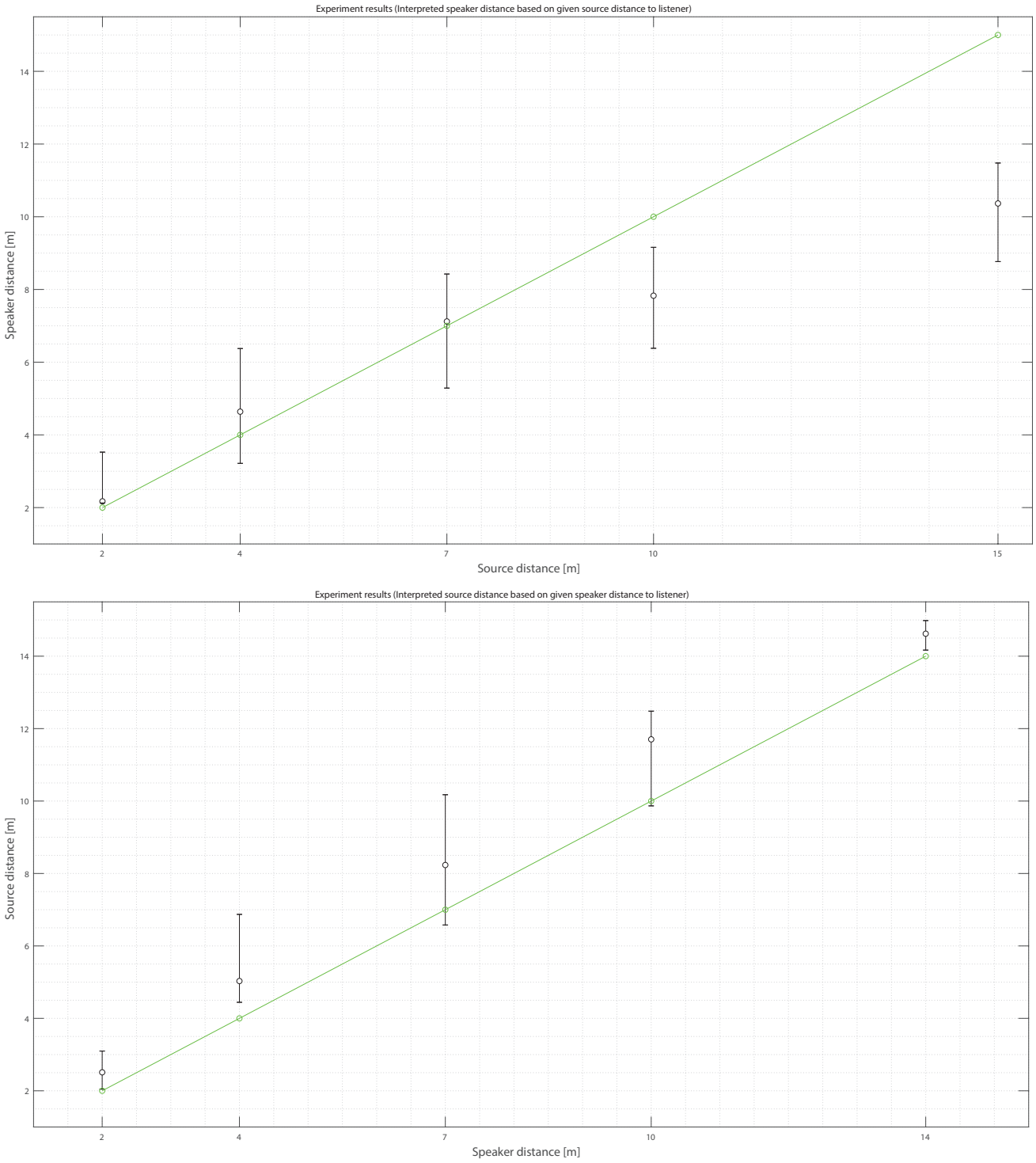

Figure 18 – Top: speaker distance to source distance (audio engineers)
Bottom: source distance to speaker distance (audio engineers)

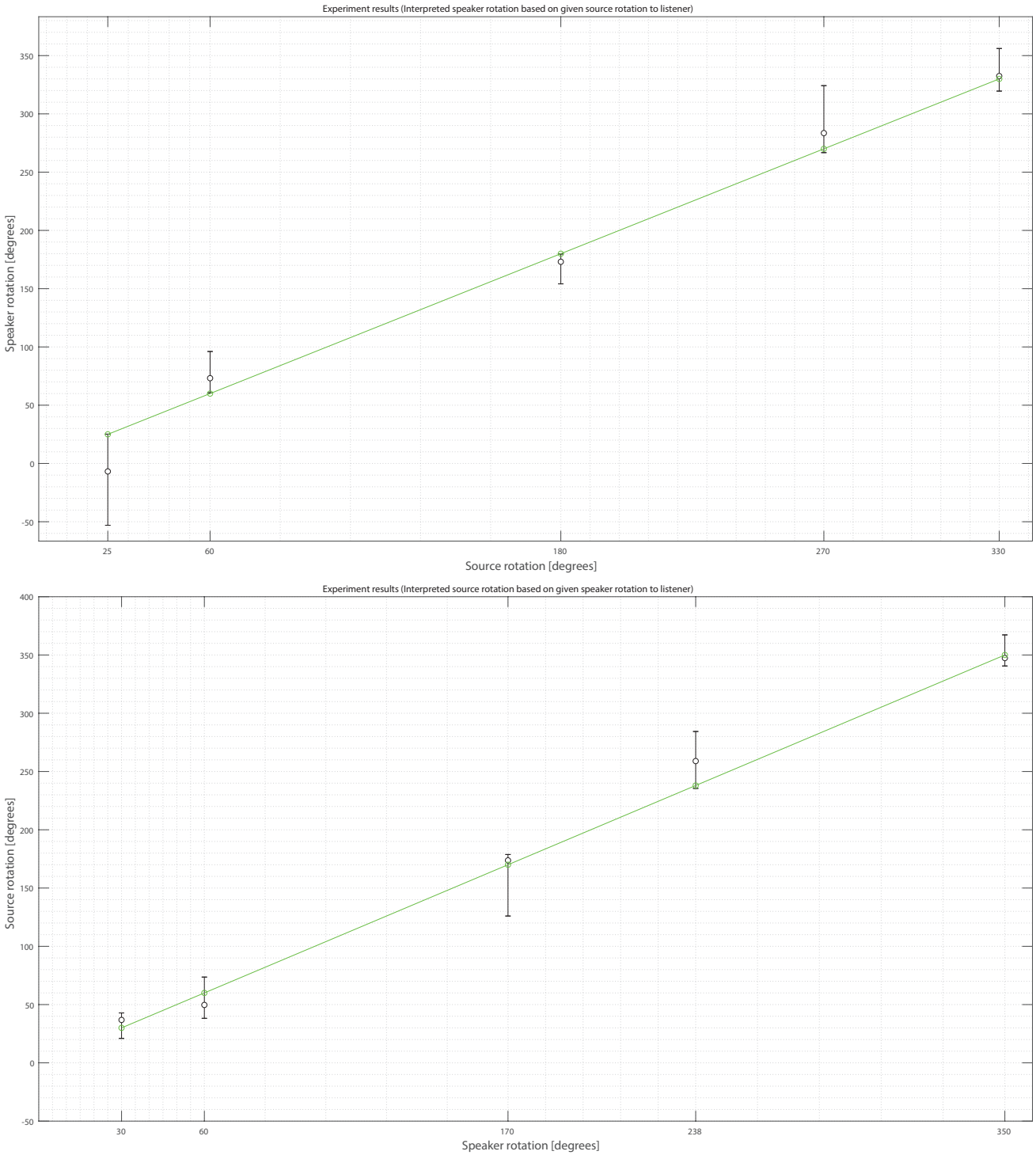

Figure 19 – Top: speaker rotation to source rotation (audio engineers)
Bottom: source rotation to speaker rotation (audio engineers)

4.4 Comparison and ranking

While comparing even the first parts of experiment (figure 10 and figure 14, top) it can be concluded that musicians and subjects with more audio experience were able to evaluate stimuli closer to the correct values. However, whiskers of few stimuli are still not covering the dotted reference.

While comparing mentioned results further with audio engineers (figure 17, top), it can be seen that their median values are even closer to the reference and this time, whiskers of every stimulus are covering the dotted reference line. Even for them, rooms represented bigger than $15000m^3$ were not evaluated accurate enough. These results are probably due to simply not being able to tell the difference between the rooms that big. It's much easier for the subject to tell if the room wall or ceiling is 2 or 3 meters from him/her than if it's for example 22 or 23 meters. One more obstacle is that reverberation is not that easily understood and accuracy for that stimuli is not that high, regardless of subjects experience prior to the experiment, because this information doesn't represent something evolutionary important.

As already mentioned, dotted reference (2 times longer reverberation than real reference) was observed instead of the real one (dashed). In this case, the room itself has a reverberation shorter than expected, simply because it's built that way. Even if the subjects could find themselves in the real room, they wouldn't be able to evaluate the room correctly. Acoustic absorbers of the evaluated room are covered with artistic patterns and it's easy to underestimate their effect, which can also be seen in the analysis of the first 2 parts of the experiment. Therefore, the second reference line was added to evaluate the virtual reality more correctly and this way, the tendency of the evaluated results corresponds much more correctly to the reference.

As seen from figure 10, results of the first 2 room sizes are not significantly different, which is shown by p-values (p = 0.688). Next 3 enlargements of the room until around 5k cubic meter, show that these stimuli were in fact more significantly different ($p \leq 0.022$). Until around 14k cubic meter stimuli are not evaluated differently ($p \geq 0.81$). Afterwards, reverberation stimulus that represents room of 20k cubic meter, gives significantly different results ($p \leq 0.022$), after which stimuli are evaluated more or less the same ($p \approx 1$). It can be then concluded that there were 5 significant steps, which means that this part could've consisted of 5 trials (and not 10). However, additional arguments regarding this part will follow in next section.

Relative increase was between 25% and 60%, while insignificant steps were between 1% and 10%. The results agree therefore with just noticeable difference of around 27% [18]. Further, the relation between evaluation of room size and reverberation could

be also confirmed in the literature about venues for amplified music [19].

In the case of adjustment of the room size to the given reverberation (for example figure 11, top) it can be also concluded that there were only 5 significant steps with values of $p \leq 0.009$ (between 1.1s and 6.5s). Other stimuli had values of $p \geq 0.236$ and therefore had insignificant differences. As already mentioned in the analysis of the results from section 4.1, when the room size is increased, it becomes more difficult to evaluate the stimulus correctly and this effect is already familiar from literature where is concluded that typical reverberation times for concert halls (which this room actually is) are in range between 1.5s and 3s and everything above sounds for subjects rather unnatural [20, 21].

In the next figures representing evaluation of source and speaker distances (see bottom half of figure 11 and top part of figure 12, as well as figure 15 and figure 18), it can be seen that subjects regardless of their audio experience did much better in evaluating sound distance. As well as for bigger room sizes, for longer distances (10 meters to source/speaker and above), it can be seen that whiskers of the plot do not go over reference values. This effect was observed already in real life experiments, where subjects would overestimate shorter source distances and underestimate longer distances [22, 23], which also corresponds to virtual rooms [24]. Additionally, at least for the sources further away, underestimation can be explained through one more fact and that's a increase of the just noticeable difference of the direct-to-reverberant (DDR) energy ratio for low DDR values [25]. However, this experiment showed interesting difference to mentioned facts and that's when adjusting auditory source (heard sound distance) to visual sound distance (seen speaker distance), like in top half of figure 12, subjects didn't underestimate far distance values (stimuli further than 8 meters).

One more interesting discovery is that a standard deviation of auditory distance perception is about 1.6 times the distance [26], but in this experiment the standard deviation was much smaller (0.37 to 0.69). P values in these parts showed that all stimuli were significantly different, which means that 5 stimulus evaluations in this case were optimal.

Last figures of the project were evaluation of source and speaker rotation (see bottom half of figure 12 and 13, as well as figure 16 and figure 19). Here, as well as for previously mentioned figures, no significant difference between results of different groups can be noticed. For certain stimuli can be said that audio engineers and sound technicians evaluated results with slightly more deviation from the reference than when compared to all results. It still can't be concluded that people with audio experience had trouble with evaluating stimuli, but it can be mentioned that audio experience is not

always a plus. In this case it could make audio engineers and sound technicians question their evaluation for too long, resulting in the more deviation from the reference. It has to be added that subjects could walk around the virtual loudspeaker to adjust orientation more easily. However, in the similar experiments without visual cues, subjects have been able to adjust the orientation approximately the same [27, 28].

5 Conclusion and further development

As seen from previous section, results could be analyzed well enough and the tendency of the evaluation made sense. Still, for the first few stimuli of the whole experiment, more deviation from the reference could be noticed.

Simple reason behind it has been already mentioned by some subjects themselves. They didn't know how to adjust evaluation for the first few trials simply because they didn't know what is the maximal and minimal room size, so that they have some kind of reference. By some subjects with more audio experience was even mentioned that it would be good to have some prior experience in the room itself, so that they know what is the reverberation of the real room. Although it would be easier to do the first few trials, this kind of experience would be considered as cheating in this case. Whole point of this experiment was to have no reference in the beginning, so that every subject simply must rely completely on their personal feeling and how the room surrounding them should sound like in their opinion.

Evaluation of the FDN Reverb based on the given size and vice versa (first 2 parts of the experiment) consisted of 10 trials and all other parts had only 5. In addition, values that show how much every stimuli gives new information to the previous one regarding evaluation have been analyzed (in last section described p-values). These values showed that 10 trials were not really needed for the analysis, since subjects didn't really notice any difference between neighboring stimuli. However, for the first parts of the experiment, it was good to have more trials, since majority of the subjects used the VR equipment for the first time and they needed time to get used to it. Despite having a tutorial, even for the audio experienced subjects, they got the point of the experiment as of around 4th trial (1.4 in this case), so that increased number of trials did actually have a purpose.

One of the subjects even mentioned that it would be a good idea to have some kind of ranking system, so that every subject knows at the end of the experiment how many "points" they got from the evaluation or how accurate they were during the experiment. Some other subject mentioned that it would be good to have a limited time for every

trial, so that too much thinking can't ruin the evaluation. That would also mean that every subject would have the same amount of time to finish the experiment, which would made it even more fair.

As conclusion it can be said that this type of experiment not only generated interesting results and showed that virtual acoustics is getting closer to real acoustics, but it was also fun for every single subject. It made sense, since this experiment was almost like a video game and the audio experienced people enjoyed it even more. Many of them showed interests to be part of other similar experiments and made suggestions on how to develop this experiment even more. For the next similar experiments, the same coding can be developed to have more sound sources, different rooms or even more subjects doing the same experiment at the same time with 2 VR sets, when the technological requirements are fulfilled. As seen from this project, virtual reality as well as game engines like Unity allow us infinite possibilities, so why not experiment with virtual acoustics as well.

6 Appendix A

Experiment results:

FDNReverb to RoomSize:
1.59342
1.72603
1.1769
2.02329
2.19239
1.4706
2.15056
1.51955
1.88801
1.5507

RoomSize to FDNReverb:
19024.205804
259.200012
45519.666195
27552.188516
27777.542996
2169.600785
45519.666195
37861.804533
259.200012
21048.400497

SpeakerDistance to SourceDistance:
7.2294
2.7717
9.865951
11.5143
6.2817

SourceDistance to SpeakerDistance:
7.3887
2.30595
14.79885
6.0981
10.360049

SpeakerRotation to SourceRotation:
169.919998
205.09201
297.252014
123.659996
326.664001

SourceRotation to SpeakerRotation:
238.17601
3.312
268.523987
237.888
282.347992

7 Appendix B

Reference values:

FDNReverb to RoomSize:
3.7125
1.4641
0.0704
2.4167
5.4043
0.8019
8.999489
0.09997
0.3773
6.964787

RoomSize to FDNReverb:
7363.636863
1472.727395
11045.455295
29454.547453
20250.001374
4050.000363
33136.365885
23931.819806
368.181849
25772.729021

SpeakerDistance to SourceDistance:
7
2
15
10
4

SourceDistance to SpeakerDistance:
10
2
14
4
7

SpeakerRotation to SourceRotation:
180
60
270
25
330

SourceRotation to SpeakerRotation:
170
30
350
60
238

References

[1] R. Holly, "The least painful way to set up htc vive lighthouses!" 2017. [Online]. Available: https://www.vrheads.com/least-painful-way-set-htc-vive-lighthouses

[2] S. Werner, F. Klein, T. Mayenfels, and K. Brandenburg, "A summary on acoustic room divergence and its effect on externalization of auditory events," in *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2016, pp. 1–6.

[3] K. Enge, M. Frank, and R. Höldrich, "Listening experiment on the plausibility of acoustic modeling in virtual reality," in *Fortschritte der Akustik, DAGA*, 2020.

[4] P. Larsson, D. Västfjäll, P. Olsson, M. Kleiner *et al.*, "When what you hear is what you see: Presence and auditory-visual integration in virtual environments," in *Proceedings of the 10th annual international workshop on presence*, 2007, pp. 11–18.

[5] C. Chen, R. Gao, P. Calamia, and K. Grauman, "Visual acoustic matching," *arXiv preprint arXiv:2202.06875*, 2022.

[6] D. Perinovic and M. Frank, *Spatial Resolution of Diffuse Reverberation in Binaural Ambisonic Playback*. DEGA, 2021.

[7] K. Enge, *Hoerversuch zur Plausibilitaet akustischer Modellierungen in virtuellen Umgebungen*. IEM, 2019.

[8] T. Fredericks, "Unity osc scripts." [Online]. Available: https://thomasfredericks.github.io/UnityOSC

[9] EBU, "Ebu sqam test cd for audio evaluation," 2008. [Online]. Available: https://tech.ebu.ch/publications/sqamcd

[10] M. Frank and M. Brandner, "Perceptual Evaluation of Spatial Resolution in Directivity Patterns," in *Fortschritte der Akustik, DAGA*, Rostock, Mar. 2019.

[11] M. Frank, M. Brandner, and F. Zotter, "Perceptual evaluation of spatial resolution in early reflections," in *Fortschritte der Akustik, DAGA*, 2022.

[12] M. Kronlachner, "Mcfx convolver plug-in," 2014. [Online]. Available: http://www.matthiaskronlachner.com/?p=1910

[13] KUG, "György-ligeti-saal hall in mumuth," 2009. [Online]. Available: https://www.kug.ac.at/universitaet/campus/mumuth/

[14] J.-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in *90th AES Conv., prepr. 3030*, Paris, February 1991.

[15] J. Stautner and M. Puckette, "Designing Multi-Channel Reverberators," *Computer Music Journal*, vol. 6, no. 1, pp. 52–65, 1982.

[16] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural rendering of ambisonic signals via magnitude least squares," in *Fortschritte der Akustik, DAGA*, Munich, March 2018.

[17] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3616–3627, 2018.

[18] Z. Meng, F. Zhao, and M. He, "The just noticeable difference of noise length and reverberation perception," in *2006 International Symposium on Communications and Information Technologies*. IEEE, 2006, pp. 418–421.

[19] N. W. Adelman-Larsen and J. J. Dammerud, "A survey of reverberation times in 50 european venues presenting pop & rock concerts," in *Proceedings of Forum Acusticum*, 2011.

[20] L. L. Beranek, "Concert hall acoustics - 1992," *The Journal of the Acoustical Society of America*, vol. 92, no. 1, pp. 1–39, 1992.

[21] M. Skålevik, "Reverberation time–the mother of all room acoustic parameters," in *Proceedings of 20th International Congress on Acoustic, ICA*, vol. 10, 2010.

[22] A. J. Kolarik, B. C. Moore, P. Zahorik, S. Cirstea, and S. Pardhan, "Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss," *Attention, Perception, & Psychophysics*, vol. 78, no. 2, pp. 373–395, 2016.

[23] P. Zahorik and F. L. Wightman, "Loudness constancy with varying sound source distance," *Nature neuroscience*, vol. 4, no. 1, pp. 78–83, 2001.

[24] A. Kuusinen and T. Lokki, "Investigation of auditory distance perception and preferences in concert halls by using virtual acoustics," *The Journal of the Acoustical Society of America*, vol. 138, no. 5, pp. 3148–3159, 2015.

[25] E. Larsen, N. Iyer, C. R. Lansing, and A. S. Feng, "On the minimum audible difference in direct-to-reverberant energy ratio," *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 450–461, 2008.

[26] P. W. Anderson and P. Zahorik, "Auditory/visual distance estimation: accuracy and variability," *Frontiers in psychology*, vol. 5, p. 1097, 2014.

[27] F. Zotter, M. Frank, A. Fuchs, and D. Rudrich, "Preliminary study on the perception of orientation-changing directional sound sources in rooms," in *Proc. of forum acusticum, Kraków*, 2014.

[28] F. Zotter and M. Frank, "Investigation of auditory objects caused by directional sound sources in rooms," *Acta Physica Polonica A*, vol. 128, no. 1, pp. A5–A10, 2015.