PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Rendering Environmental Noise Planning Models in Virtual Reality

Arne NYKÄNEN[1]; Roger JOHNSSON[2]; Jonas ARÅKER[2]

[1] Luleå University of Technology, Sweden

[2] Tyréns AB, Sweden

ABSTRACT

In building and infrastructure projects sound design and requirement specification are often complicated by difficulties in understanding how the planned built environment will sound. Information about sounds is almost exclusively provided as sound pressure levels. It is difficult to understand how an environment will be perceived solely based on such data. VR models with sound give an experience much easier to comprehend. In this study VR models were developed based on first order Ambisonics recordings. Such recordings provide spatial information and can be real time rendered based on the listener's orientation. Recordings of road noise were used. The sound levels were adjusted to match calculated levels from noise planning models. VR presentations were compared with still image and binaural headphone presentations. The results showed that annoyance caused by traffic noise was significantly lower rated in the VR presentations compared with the still image and binaural headphone presentations. This shows that physically correct calibration of reproduced levels is not enough to ensure delivering a realistic experience. Therefore, VR technology may give more realistic and valid presentations of sounds in different kinds of sound quality assessments, presentations and public consultations.

Keywords: Ambisonics, Auralisation, VR, Noise Planning

1. INTRODUCTION

In the design of buildings and infrastructure sound design is often complicated by difficulties in understanding human experience. Information is complex, extensive and hard to grasp. A typical workflow for architects and designers involves rapid iterations supported by recurrent generation of numerous sketches and visualisations. Established working procedures and tools exist. Still, technology is developing, with e.g. VR tools enabling designers and stakeholders to interact with virtual models of planned buildings. However, the support for sound is still limited. Information about sound is almost exclusively provided as figures of e.g. sound pressure levels, sound reduction indices and reverberation times. This gives comprehensive descriptions of acoustic properties, but even professional acousticians have difficulties understanding how an environment will be perceived solely based on such data. Auralisation tools are evolving, and presenting VR models with properly modelled sounds could give an experience much easier to comprehend. This will most likely improve communication in development projects, accelerate design processes, and reduce the risk for mistakes.

VR tools are generally based on computer game technology. They can deliver high quality video, and also real time rendering of 3D sound. In this study realistic VR presentations were developed in order to demonstrate results from environmental noise planning models. Ambisonics recordings of traffic noise were used as input to the models. Ambisonics recordings contain spatial information that can be real time rendered based on head movements, and thereby they provide spatial information correctly with respect to the listener's orientation in the VR world (1). However, the recording must be made in a discrete point, and therefore only gives a realistic reproduction as long as the listener is not moving. If the listener moves in the VR model the Ambisonics reproduction will move with the

[1] arne.nykanen@ltu.se

listener. To avoid this, Ambisonics recordings made in different points can be mixed and cross-faded depending on the position of the listener.

1.1 Environmental Noise Planning

Environmental noise planning is commonly done by calculating sound propagation in models based on e.g. traffic flows and vehicle types (2). Equivalent sound pressure levels and maximum levels are usually reported as colour maps, see Figure 1. Such colour maps provide a good and simple overview of how high the sound levels are in different places, and can be a good help in identifying areas where special measures need to be taken. However, the colour maps are limited by their much simplified presentation of data. Even a trained acoustician may have difficulties understanding and judging how it will sound in a place based only on colour maps of maximum and equivalent sound pressure levels. It is even more difficult to explain to the public, for example during public consultation meetings, how a change in the built environment will affect sound perception and risk of annoyance.

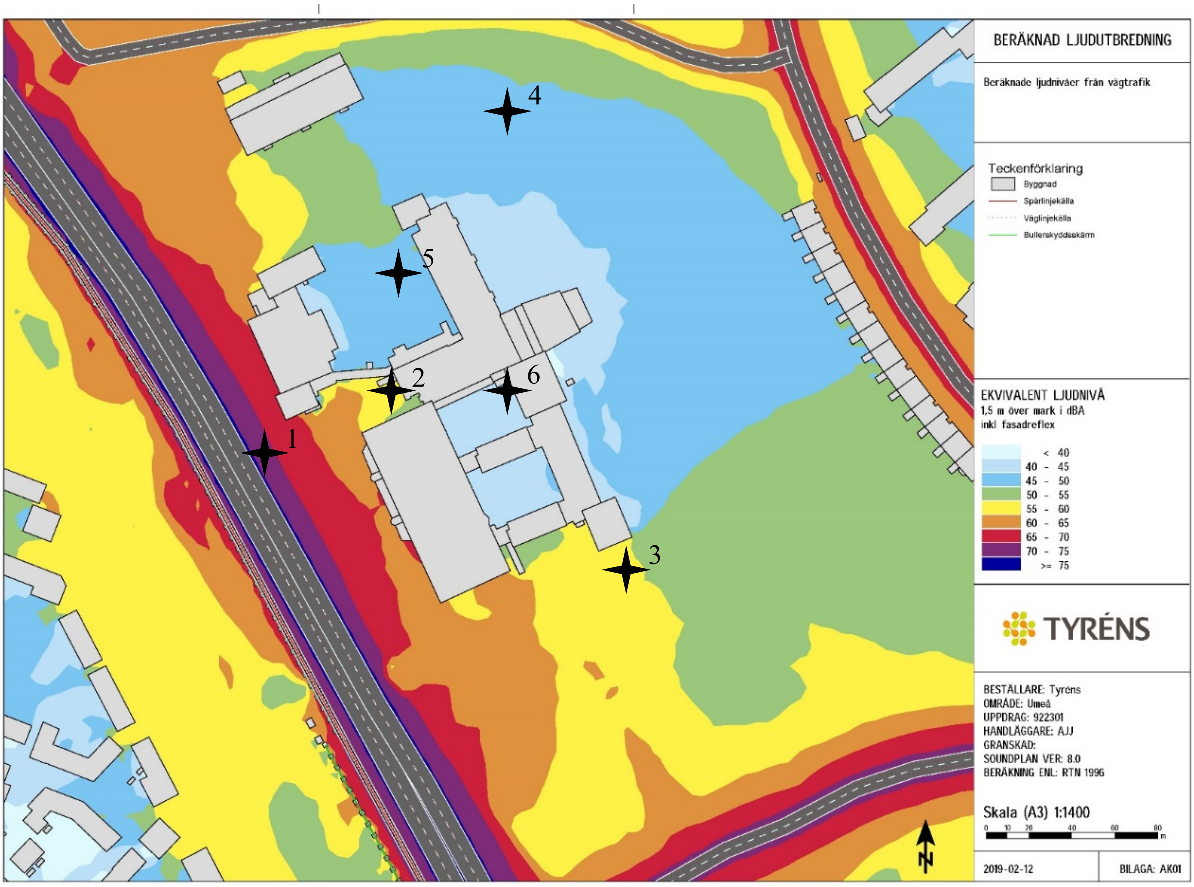


Figure 1 – Colour map showing calculated equivalent sound pressure level in an area around a school. Points of scenarios used in the listening test are marked 1 to 6.

1.2 Sound in VR and Computer Games

Creating realistic sound in VR requires binaural reproduction with high accuracy. This will allow the listener to separate sound sources, suppress background noise and hear direction as in real life. Sound in VR is generally played back through headphones. Since the 1970s, sound has been recorded with artificial heads in order to give good realism when played back over headphones. However, headphone recording and reproduction has two major limitations: it requires a sound environment to record, and it is limited to represent sound in a fixed position with the head pointing in a specific direction. The first limitation, that there must be an audio environment to record can, to some extent, be overcome by binaural synthesis. To overcome the second limitation, binaural synthesis must be rendered in real time depending on how the listener moves around in the room. This has until recently only been possible in laboratory with very powerful computers. However, today the computational power of ordinary gaming computers is sufficient for rendering binaural synthesis in real-time using conventional game engines and audio middleware for computer games. In addition, head position and orientation must be measured in real time. This can be done with gaming VR headsets. Therefore, the required hardware and software for implementing VR tools for building acoustics are currently available. However, computer games and other media do not usually require the same control as needed for design and construction of buildings and infrastructure. Still, the pursuit for realism in the computer games has led to much resources being spent on developing realistic models of 3D sound. These models perform well and are often easy to use. It is therefore worth taking a look at the models and technology used in computer games, as they can be used as tools for rapidly setting up controlled experiments on perception of real time rendered 3D sound.

2. METHOD

2.1 The Case

A school yard environment was chosen as a case with special requirements on outdoor noise levels. The Swedish Environmental Protection Agency recommends that outdoor areas intended for play, rest and teaching activities should have an equivalent sound pressure level of max 50 dBALeq and a maximum SPL of 70 dBA. For other outdoor areas of a school yard, an equivalent SPL of 55 dBALeq and maximum SPL of 70 dBA are allowed (3). Which areas that should be regarded as intended for educational activities and how large these should be are not unambiguously determined. Thus, there may be a need to interpret noise maps of the kind shown in Figure 1, and plan school yards and areas surrounding a school both from sound perspective and all other perspectives that are important for a well-functioning school environment. The possibility to walk around in a VR model and both look and listen should therefore be of help for understanding what the calculated noise levels mean.

2.2 The VR model

Unreal Engine (Version 4.21) was used to build and control the VR model. The Resonance Audio plugin was used for binaural synthesis. A semi-synthetic model was developed based on multiple Ambisonics recordings. Ambisonics recordings, as all kinds of microphone recordings, are limited to capture the sound in one point in space. Unlike artificial head head recordings, Ambisonics recordings do capture directional information and thereby contain the information required for making binaural synthesis based on the listener’s orientation in the point of recording. Unreal Engine with the Resonance Audio plugin can render Ambisonics recordings based on the listener's orientation in the virtual space, but the rendered sound field will move with the listener as the listener moves. To overcome this, either Ambisonics files recorded at different points must be mixed and cross-faded based on location, or the listener must be prevented from moving in the virtual world. One way to solve the problem is to just allow the listener to turn around in the VR world but not to move. This is common in many VR applications, not just to give realistic sound, but also because continuous movements in the VR world often give rise to motion sickness and nausea. Usually, this is solved by "teleporting" the person in the VR world, i.e. the person is moved instantly to a new position.

2.3 Recording of Traffic Noise

Traffic noise was recorded with an Soundfield SPS 200 Ambisonics microphone. The recordings were made in a pattern as shown in Figure 2. To get a sufficiently large open area the recordings were made winter time on the ice of a sea bay with a road embankment. The road was a 4 lane road with 70 km/h speed limit. The road section is close to the city centre with traffic lights in the middle of the section, so speed and traffic flow varied much, with recurrent accelerations and decelerations. Two rows of measurements were made at different distances from the road, one approximately 5 m from the road and one approximately 50 m from the road. To capture the variations at longer distance from the road, one row of measurements was made by adding four positions with 50 m spacing in between, as an extension from the two northernmost measurements closer to the road, in total 6 positions (marked 1 to 6 in Figure 2). The recordings were converted to B format using the Soundfield SurroundZone 2 VST plugin. The binaural rendering was done with the Resonance Audio plugin in Unreal Engine 4.21. Mixing was made with Unreal Engine’s built in audio mixer/controller at default settings and default settings were also used for Resonance Audio.

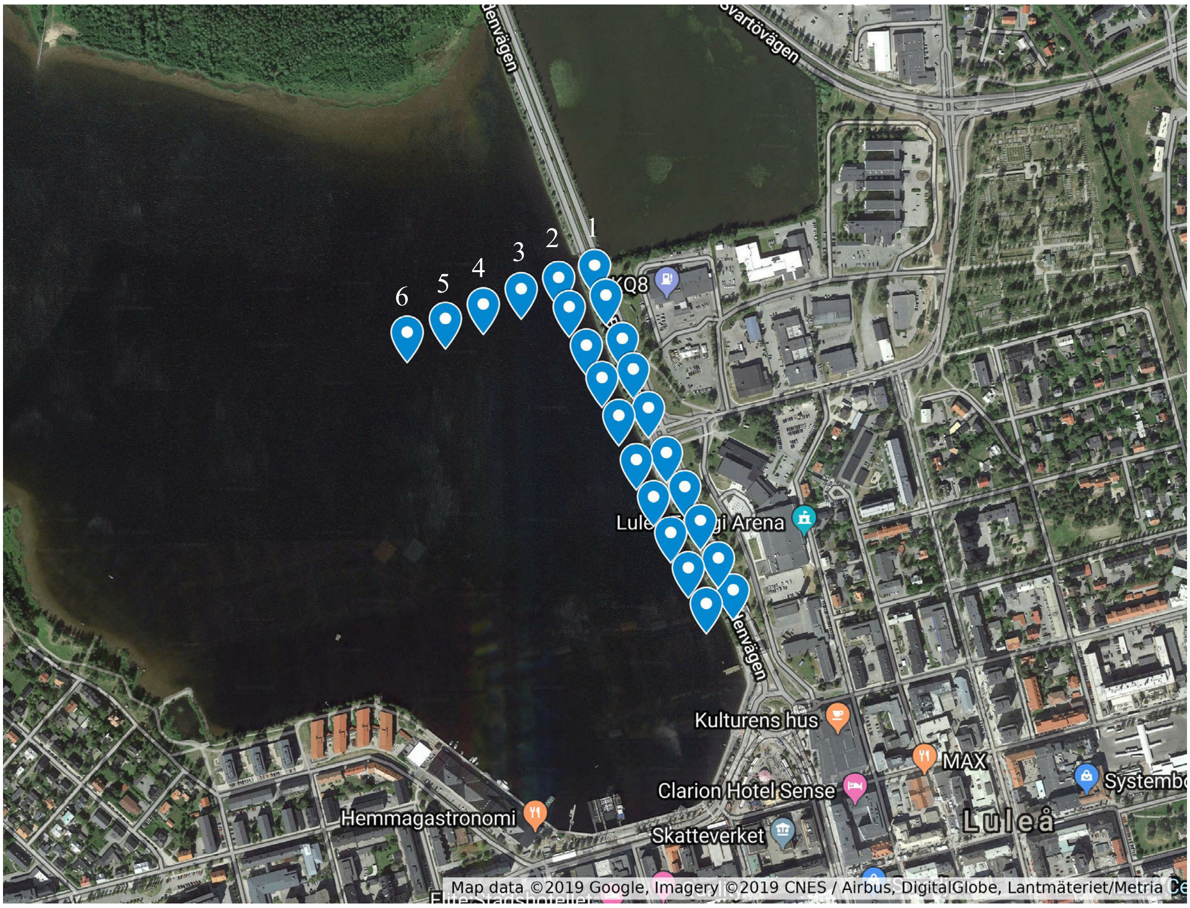


Figure 2 – Measurement positions for road noise measurements in central Luleå, Sweden. Recording points used for the listening test are marked 1 to 6.

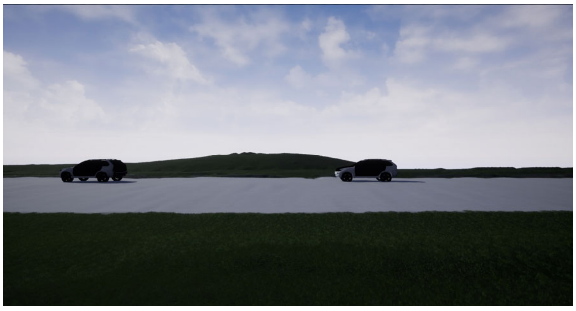
2.4 Cross-fading model

A model for cross-fading between Ambisonics recordings was developed in Unreal Engine. The model cross-faded between four adjacent recordings in a grid of arbitrary size (minimum 2 by 2). The model was assessed in informal listening tests. The cross-fading of recordings gave a realistic impression of moving around in the sound environment. It was found that with traffic noise that consists of sources moving much faster than the listener, the cross-fading was of little importance. The need for cross-fading and the size of the cells in the cross-fading matrix is much dependent on the characteristics of the sources (stationary or moving) and the distance to the sources. For outdoor environments with traffic noise the need for cross-fading arise only if the listener is moving continuously over longer distances. As this generally cause risk for motion sickness, it should be avoided. Therefore, teleporting was used in the experiment and single recordings were selected for each assessment point. No further studies of cross-fading models were made in this experiment.

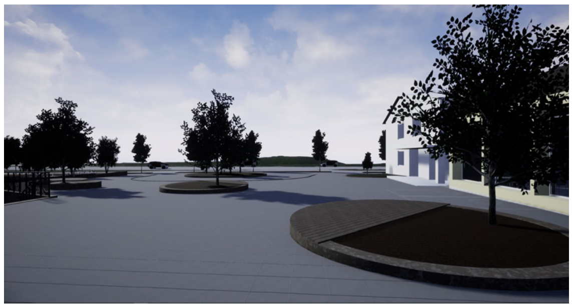
2.5 Integration of the Sound Model with a Graphical Model

A model of a secondary school developed for another project was used. The area around the school can be seen in the colour map in Figure 1. A heavily trafficked road is the main noise source in the area. Six places at different distances from the road, with different sound pressure levels and different architectural characteristics were chosen, see Table 1. The views towards the road at all places are shown in Figure 3. Ambisonics recordings from six places, 50 m apart, were chosen for the six scenarios (marked 1 to 6 in Figure 2). Scenario 1 to 4 had an open view to the road and the distances corresponded to the distance between the road and the Recording Positions 1 to 4 (see Figure 2).

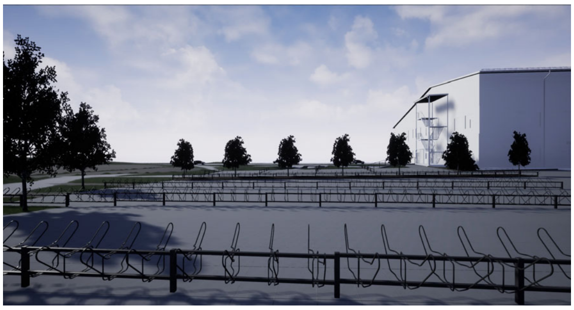
Recording Position 5, 200 m from the road, was chosen for Scenario 5. Here, the distances do not match, but in Scenario 5 the road was shadowed by a building, and hence the sound was assumed to resemble the sound at a longer distance from the road. The same argument was used for choosing Recording Position 6, 250 m from the road, for Scenario 6. Scenario 6 was inside a courtyard, and the sound was therefore assumed to resemble the sound at an even longer distance. The equivalent sound pressure level at all places were matched to the level calculated by Soundplan (Soundplan GmbH). Calibration was done by placing the headphones on an artificial head (Head Acoustics HMS IV) and measuring the equivalent sound pressure level (with the HMS IV artificial head set to LIN equalization). Measured sound pressure levels are shown in Table 1.



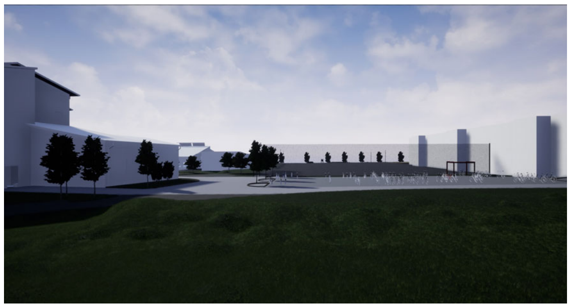
Scenario 1



Scenario 2

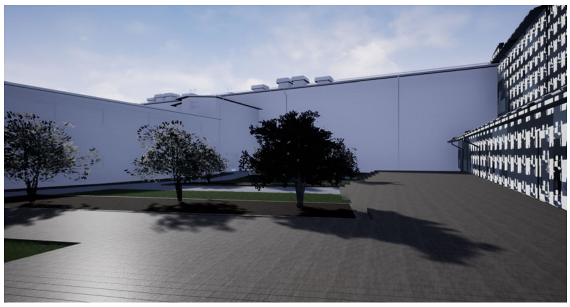


Scenario 3



Scenario 4

Scenario 5



Scenario 6

Figure 3 – The views of the six places selected for the listening tests. The pictures are from a model of Maja Beskow School in Umeå, Sweden, by courtesy of Umeå Municipality.

Table 1 – Six places chosen for the listening test.

Scenario no.	Architectural characteristics and equivalent SPL from the Soundplan model	Equivalent SPL in headphones after calibration	
1	5 m from the road, 70-75 dBA	L 73.1 dBA	R 72.4 dBA
2	50 m from the road, open view to the road, at the main entrance of the school, 55-60 dBA	L 58.8 dBA	R 56.8 dBA
3	100 m the road, open view to the road, at a bicycle parking, 55-60 dBA	L 58.7 dBA	R 56.8 dBA
4	150 m the road, open view to the road, close to a sports field, 45-50 dBA	L 47.3 dBA	R 45.5 dBA
5	80 m the road, a building shadowing the road, 45-50 dBA	L 48.2 dBA	R 47.4 dBA
6	100 m the road, in a court yard, 40-45 dBA	L 41.1 dBA	R 42.5 dBA

2.6 Listening Tests

Validating the realism of a virtual model of environmental noise is difficult. Ideally the model should be compared with reality, but such listening tests are impracticable. As artificial head recordings and renderings based on binaural synthesis are commonly used for sound quality assessments a still binaural reproduction was considered to be a suitable baseline for comparison. The still binaural reproduction was created from the same Ambisonics recording as was used for the binaural synthesis in the VR model. Resonance Audio's HRTF filters were used with the head pointing towards the road. To not create confusion between visual and auditory perception, the still binaural reproduction was presented with a still image on a computer screen (24" screen at approximately 70 cm distance from the participant). The view was pointing towards the road. All still images can be seen in Figure 3. This baseline condition was compared with a VR presentation of the same positions. In the VR presentations the participants were not allowed to move around, but they were allowed to look around as much as they liked. The sound consisted of the Ambisonics recordings rendered in real time using Resonance Audio's algorithms and HRTF filters.

The listening test was conducted in a sound recording studio. The sounds were played through a Head Acoustics PEQ V amplifier and Sennheiser HD 600 headphones. Half of the group started with the VR presentation and the other half with the still image/binaural sound condition. The participants were orally instructed to assess annoyance from traffic noise in the surroundings of a secondary school. They were told that they would do the same test, first with the environment presented through a VR headset with headphones and then with a still image of the same places on a computer screen with sound presented through headphones (or vice versa for half of the group). The six positions were presented in counterbalanced order using a Latin Square design. The participants were asked to listen to each scenario as long as they though necessary to make an accurate assessment. They were asked to assess how annoying they found the sound environment to be on an 11-point scale ranging from "not at all annoying" to "extremely annoying" (from now on called *annoyance*), and how easy or difficult they thought it was to assess the sound environment by using the computer model on an 11-point scale ranging from "very easy" to "very difficult" (from now on called *effort to assess*). The experiment was a repeated measures design where all participants assessed both conditions (VR and still image). The assessments of *annoyance* and *effort to assess* were averaged over all scenarios for each participant. The hypothesis was that the presentation technology would affect both average *annoyance* and *effort to assess* annoyance. Results were analysed with Paired-Samples t-Test.

2.7 Participants

A total of 12 subjects participated in the listening test, 5 women and 7 men. Most of them were students studying at Luleå University of Technology. All subjects had self-reported normal hearing, and all were volunteers. The mean age was 28 years (min. 22, max. 66, SD 12 years).

3. RESULTS

Means and standard errors of the assessments of *annoyance* for each of the six scenarios under both conditions (VR or still image/binaural sound) can be found in Figure 4. Means and standard errors of the assessments of *effort to assess* annoyance for each scenario and condition are found in Figure 5.

For the statistical analysis an average of each participant's assessments of *annoyance* and *effort to assess* was used. On average, the participants rated annoyance significantly higher in the still image/binaural sound condition (M = 4.4, SE = 0.51) than in the VR condition (M = 3.8, SE = 0.56, t(11) = 2.22, $p < .05$, r = 0.56). On average, the participants did not rate the effort to assess annoyance significantly different in the still image/binaural sound condition (M = 2.3, SE = 0.31) compared with the VR condition (M = 1.6, SE = 0.26, $t(11) = 2.04$, $p > .05$).

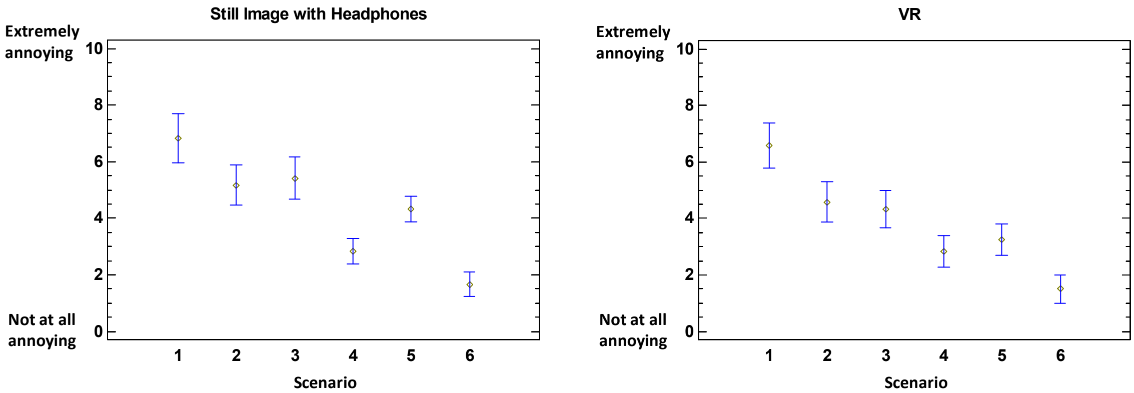


Figure 4 – Mean and Standard Error (internal s) of *annoyance*.

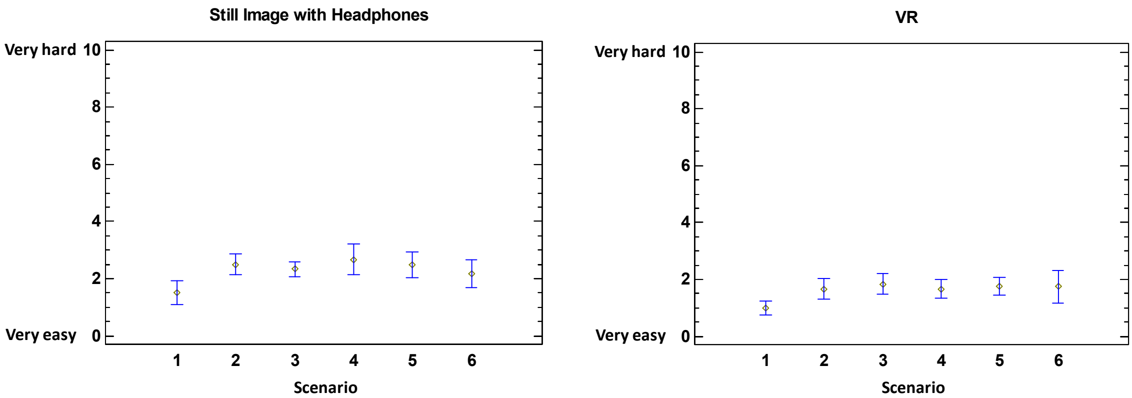


Figure 5 – Mean and Standard Error (internal s) of *effort to assess* annoyance

4. DISCUSSION

The results are in line with previous research showing that vision can affect the experience of sound. For example, Patsouras et al. (4) showed that red trains were judged as louder than green trains in a laboratory listening test accompanied by photographs of trains. Later Menzel and Fastl (5) showed a similar effect of colour on the perceived loudness of cars. The results are also in line with a previous study by Iachini et al. (6) of how annoyance assessments in VR presentations differ from annoyance assessments based on sound only. They showed that participants assessed the same sound stimuli as less annoying when presented in VR compared to presented without any visual stimuli. Our study shows the same effect. Subjects judge traffic noise as less annoying when presented in VR compared to presented with a still image. The difference between our study and Iachini et al. is that we compared the VR condition with a condition containing still image visual information. A higher degree of realism achieved by using VR may be an explanation for the lower annoyance ratings. Another difference between our study and Iachini et al. is that we used the head tracking information from the VR headset to render the binaural signals from the Ambisonics recordings in real time based on the listener's orientation in the VR condition, while Iachini et al. presented a still binaural reproduction also in their VR presentation. In the still image condition we used a still binaural signal that resembles the binaurally recorded signal used by Iachini et al. in both their conditions (VR and no visual stimuli). Our objective was to study whether real time rendering of binaural sound based on the listener's orientation has an effect on sound quality assessments. The results suggest that such an effect may exist, but from the present experiment effects caused by the binaural sound rendering cannot be distinguished from effects caused by the more realistic visual VR presentation. Iachini et al. suggest that the visual VR presentation may itself cause lower annoyance ratings. However, a conclusion from both studies is that presenting stimuli in VR does effect annoyance ratings. None of the studies make any comparison with reality, so no conclusions on which cases that are more representative can be drawn. A hypothesis is that the more realistic VR presentation would also give a result more representative for reality.

5. CONCLUSIONS

Annoyance caused by traffic noise was significantly lower rated in VR presentations compared with still binaural headphone presentations with still images. This study cannot determine whether the effect is caused by real time binaural synthesis based on head-tracking data from the VR headset, by the higher level of realism in the visual VR presentation, or a combination of both. New experiments are required in order to be able separate the effects of the possible causes.

The study shows that physically correct calibration of reproduction levels is not enough to ensure listening test results to correspond to reality. However, it suggests that VR technology may give more realistic and valid presentations of sounds in different kinds of sound quality assessments, presentations and public consultations. Graphical representations in VR as well as real time rendered 3D sound may both contribute to better tools for working with sounds in building and civil engineering applications.

ACKNOWLEDGEMENTS

The research was conducted in collaboration between Luleå University of Technology and Tyréns AB, and was funded by the Swedish Foundation for Strategic Research and Sven Tyréns Foundation.

REFERENCES

1. Frank M, Zotter F, Sontacchi A. Producing 3D Audio in Ambisonics. AES 57th International Conference: The Future of Audio Entertainment Technology - Cinema, Television and the Internet: Graz, Austria 2015.
2. Karantonis P, Gowen T, Simon M. Further Comparison of Traffic Noise Predictions Using the CadnaA and SoundPLAN Noise Prediction Models. International Congres on Acoustics, ICA 2010: Sydney, Australia 2010.
3. Riktvärden för buller på skolgård från väg- och spårtrafik. Stockholm, Sweden: Natruvårdsverket, NV-01534-17; 2017.
4. Patsouras C, Filippou TG, Fastl H. Influence of color on the loudness judgement. Forum Acusticum; Sevilla, Spain 2002.
5. Menzel D, Fastl H. Influence of color on loudness judgements. J Acoust Soc Am. 2008;123(5):2477-2479.
6. Iachini T, Maffei L, Ruotolo F, Sense VP, Ruggiero G, Masullo M, Alekseeva N. Multisensory Assessment of Acoustic Comfort Aboard Metros: a Virtual Reality Study. Appl Cognit Psychol. 2012;26:757-767.