

Perceptual comparison of different dynamic auditory Virtual Reality (VR) simulations of approaching vehicles

Jonas Krautwurm¹, Friedrich Beyer¹, Daniel Oberfeld-Twistel², Thirsa Huisman², Ercan Altinsoy¹

¹ *Institute of Acoustics and Speech Communication, Chair of Acoustics and Haptics, TU Dresden;*

² *Department of Psychology, Johannes Gutenberg Universität Mainz*

Introduction

People have to make important cognitive decisions while they're navigating through the traffic. They have to decide, whether crossing a street and have to avoid approaching vehicles. Auditive cues provide important information in this context. During the higher electrification level of currently developed cars, sounds get more unconventional and quieter. To ensure traffic safety in the future, it is necessary to understand, how different characteristics of vehicle sounds affect the pedestrian's perception of vehicles in different traffic scenarios.

To guarantee a safe experimental environment and keep the costs down with a larger sample of participants, studies in these fields are conducted in virtual reality (VR). There are different techniques to simulate realistic spatial sound events. [1] In the TU Dresden, a Wave Field Synthesis (WFS) is built in the Multimodal Laboratory, whereas in the JGU Mainz a Higher Order Ambisonic System (HOA) is used. Studies have shown, that the different simulation approaches lead to differences in the audio reproduction. [2] For getting valid results in both laboratories, it is crucial to represent the traffic scenarios as realistic as possible.

To quantify the behavior of pedestrians, there are safety-relevant values, e.g. the time-to-collision point. (TTC) A several studies of the JGU investigated the TTC estimations regarding different conditions, e.g. the vehicle type or the velocity. [3] [4] Also in Dresden such studies were conducted. [5] [6] In both experiments, a higher loudness value of the scenes leads to underestimations of the TTC. However, the TTC's were underestimated in the WFS system compared to the HOA playback environment. This leads to the question, whether it is just a calibration issue or other cues like the timbral balance is responsible for that.

Methods

For this study, vehicle pass-by-noises were played back in the two sound spatialization systems. At first the two different techniques are explained.

Laboratories

Both laboratories contain a circular loudspeaker array. In the HOA system plane sound waves come into a reference point from different directions and develop a series of spherical surface functions of the n th order. [1] The laboratory in Mainz used a 7th order Ambisonics system with 16 loudspeakers (Genelec 8020DPM-7) in past experiments. Acoustically it is mainly treated by Basotect panels, that were lined on the bottom and walls. [3] For actual studies the

system was extended to an array of 40 loudspeakers and one subwoofer, to reach a lower frequency range.



Figure 1: HOA setup at the JGU Mainz

At the TU Dresden a IOSONO WFS system is installed to reproduce spatial sound. In total, 464 loudspeakers and 4 subwoofers, that are driven individually, are built in behind perforated metal shields. The spacing between the tweeters is really small to reduce aliasing artefacts and amounts 6 cm. The laboratory is acoustically treated according recommendations of the ITU-R BS.1116, DIN 15996 and EBU 3276. Therefore, different materials (e.g. only mineral fiber or additional plastic foil) were used to damp the sound regarding specific characteristics. In the corners, Helmholtz resonators are responsible for lower frequency dampening. [7]

Calibration

The laboratory in Mainz is calibrated through the TASCAR Speaker Calibration Tool. White noise with certain characteristics is played back over every single loudspeaker and is measured with a SPL-meter. (Norsonic Nor131 with Roga MP40 free field microphone) The microphone is placed in the center of the loudspeaker array in a height of 165 cm. With the calibration tool, level differences between the speakers are compensated and the sound pressure levels are calibrated. [3] In the WFS laboratory in Dresden we used a pink noise between 125 and 8000 Hz with a sound pressure level of 80 dB as a point source. We placed this source in a height of 165 cm above the ground directly in front of the listener position. The distance amounts eight meters in this case. A freefield mic (B&K capsule type 4188; B&K preamp type 2671) was placed at the listener position in the same height as the sound source and recorded the signal. Additionally, the incoming sound level at the microphone via a certain distance can be calculated. The difference between the calculated and recorded value is then typed to the IOSONO interface, that controls the rendering computers.

Scene Generation Toolboxes

For the HOA environment in Mainz, the software TASCAR is used. With that program, it is possible to prepare scenes for virtual acoustic environments, including specific acoustic characteristics like air dampening or reflections. Also, the position change over time of certain sound sources can be simulated. [8] In Dresden, a MATLAB toolbox with all the relevant functions was created. [5] It is possible to read the .tsc files (files with the sound source characteristics for TASCAR) with this toolbox, adding the relevant acoustic phenomena as well and send the output to the WFS renderer to play back the created scenes.

Recording Driving by scenes

For the later listening experiment, driving-by scenes were recorded in both laboratories.

Record based approach

In a first step, real vehicles were recorded on a test track of the TU Darmstadt and the sounds were kindly provided by our project partners from Mainz. One ICEV vehicle (combustion engine and manual transmission) and an EV (with and without AVAS turned on) were recorded. A close-up microphone was placed close to the front tires, to the back right tire and to the engine of the vehicle. The GPS Data of the car was also recorded. Additionally, a dummy head and a free field microphone were placed beside the street. The data was collected and edited by the JGU Mainz. They provided us the sound samples as wav files, as well as the scene details and the GPS data per scene in excel files. [3] In the record-based simulation approach, all the recorded single source sounds of the vehicle are placed in the virtual environment according to the microphone position data at the car. The GPS data over time is added as well to simulate the driving-by-scenario. In TASCAR as well as in the MATLAB toolbox, the sound reflections of the test track (under the vehicle), air absorption effects and distance-dependent effects, like the Doppler-effect were simulated.



Figure 2: WFS system at the TU Dresden

Listener position

The car approaches the listener from the front and drives by on the right of the listener. The distance to the roadside amounts 50 cm and the shoulders are vertical to the road course. As on the original test track for later comparison possibilities, the height of the listener position is one meter.

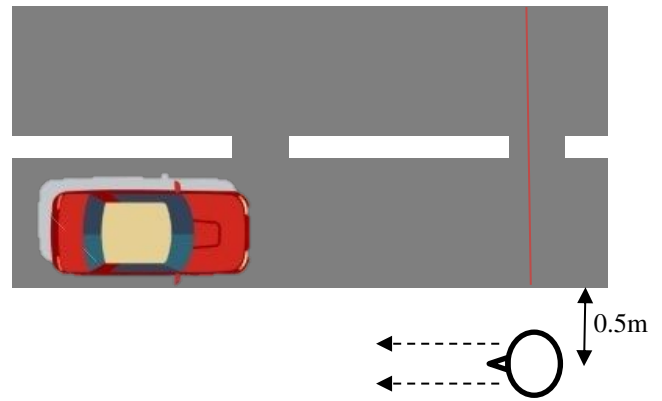


Figure 3: listener position

Recording setup laboratories

For the later experiments it is necessary to record the simulated driving-by-scenes in both laboratories binaurally. A dummy head from Head acoustics (Head HSU III.3) with its shoulder unit is used for the recordings. It is characterized by an individual ear design. The dummy head is equipped two ICP condenser microphones. During the recordings, a diffuse field (DF) equalization was used. With this individual filter, all influences on the binaural signal caused by the shoulder unit or individual ear design in a diffuse field environment (like in the laboratories) are reduced. Cause of that, the signal can be treated like a measurement microphone signal later. The Head labHSU interface was used to record the signals with the program ArtemiS Suite. A marker signal (sinus waves with additional impulses) was recorded in Mainz as well. Thus, it was possible, to identify important events of the scene, e.g. the exact point, where the vehicle arrives the listener position. The MATLAB toolbox in Dresden allows to type in the time window, that the pass-by noise should be played before the arrival at the listener position. Those two methods are crucial for the later scene-cutting. The scenes were recorded auditory only in both laboratories.

Listening experiment

For this study, a listening experiment with 24 participants was conducted. They were 19 to 61 years old (mean: 30 years) and took part voluntarily.

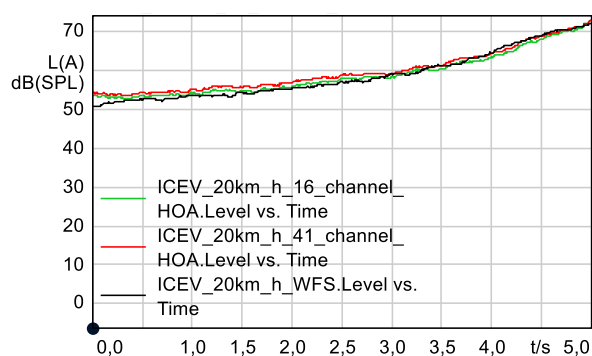
Stimuli

The cutted driving-by-scenes exhibits a time window from minus five seconds until zero seconds before the reference point of the listener. (red line in Figure 3) Because TTC experiments only includes driving-by-scenes until the reference point, the quality estimations in this study concentrates on the pass-by-noise until this point. In total, two different vehicle types, five different playback environments and five different constant velocities were used for the stimuli as shown in Table 1. The scenes with conditions, that were used for this study are marked red. In the TASCAR environment, a binaural receiver was implemented, so a virtual rendered binaural signal was added as well. It would go beyond the scope, to analyze every playback condition, so the just the spatialization systems of Mainz and Dresden were compared.

Table 1: Stimuli conditions for the listening test

Playback condition	Vehicle type	Velocity
16 channel HOA	ICEV	10 km/h
41channel HOA	EV with AVAS	20 km/h
Virtual HOA		30 km/h
WFS		40 km/h
Real test track		50 km/h

The scenes were level matched according their maximum A-weighted level. That is shortly explained by the ICEV stimuli, where the vehicle drives a constant velocity of 20 km/h. The mean value of the original A-weighted level of the three playback conditions for each vehicle type and velocity was calculated. A new even value close to the mean value was chosen and the scenes were calibrated. Loudness differences through calibration issues in the laboratories were reduced through this procedure. In total the combinations of the scene conditions led to 30 different driving-by-stimuli. The combinations used for this study led to 18 different scenes.

**Figure 4:** level (A) over time of the ICEV; 20km/h stimuli

Attributes

To compare the audio reproduction methods according their reproduction characteristics, the perception has to be quantified. For that, a number of attributes were selected. To cover a wide range of the perceptual space, the attributes were chosen out of different groups. In this paper we only concentrate on two out of the six attributes. The first word pair is called soft-loud and gives information about the perceived loudness of the scenes. To analyze the timbral balance, the pair dark-bright was chosen. In the experiment, the two attributes could be rated according intensity rating scales from 0 to 100 percent. [9]

Procedure

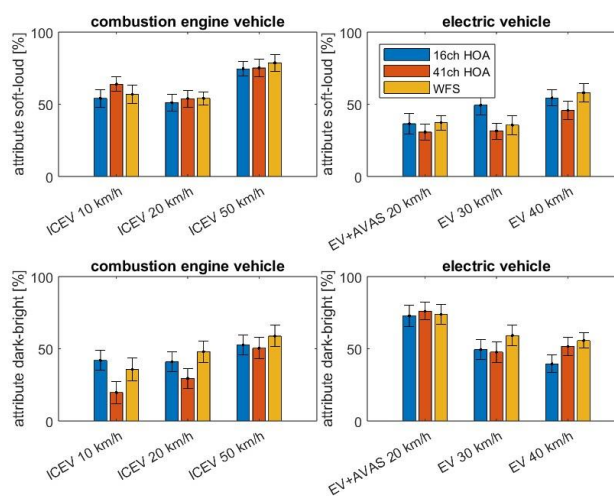
The participants sat down on a chair first and chose a comfortable position first. The stimuli were played back binaurally over a Head labO2-V1 combined with Sennheiser HD headphones. To restore the original binaural sound out of the *measurement signal* (chapter Recording setup), the playback system was equalized by a playback DF filter. The listening experiment was divided into two parts with three rating attributes. The participants had to listen to the 30

stimuli described above in each part and could repeat them as often as they wanted. Before they started the actual ratings, they had to rate 8 training-stimuli, which not were respected in the final results. The two attributes, that are relevant for this study, were split up in the two different experimental parts. The results concentrate on the 18 chosen stimuli and the perceptual results of the two attributes described above.

Results&Discussion

At first, the mean perceptual ratings of the participants are described below. For the analysis, a repeated measurements ANOVA with the two factors *velocity* and *playback condition* was performed for each vehicle type.

Results of attributes soft-loud and dark bright

**Figure 5:** rating results of the attributes soft-loud and dark-bright

The rmANOVA reveals, that for the ICEV condition the loudness perception doesn't differ significantly. $F(46,2)=1.875, p=.165$, but the loudness perception of the EV scenes show a significant difference, $F(40.087, 1.743)=8.045, p=.002$. The reason for that could be, that the participants are more familiar with the conventional engine sound and can estimate the loudness of the ICEV better, than from the EV. The perception of the attribute dark-bright differ significantly for both the EV, $F(37.617, 1.636)=12.364, p<.001$ and ICEV stimuli, $F(43.017, 1.870)=10.876, p<.001$. For the timbral perception, at least for the ICEV condition, a tendency is recognizable. The scenes were perceived darker in the 41 channel HOA environment than in the other two environments. This effect got stronger with declining velocity. Because in a combustion engine vehicle, motor orders lead to lower frequency content and could cause this shift in perception. Especially with a lower engine RPM (ICEV 10 km/h) the effect is noticeable. However, this effect is not visible, when the conventional combustion engine is missing (EV stimuli). There, the stimuli with the electric vehicle and the AVAS sound turned on was perceived brighter, than when it was turned off.

Psychoacoustic values

Also, some important psychoacoustic values were calculated, to compare the ratings of the perceptual attributes. At first the psychoacoustic value *loudness* was

calculated according to the ISO 532-1. The psychoacoustic value *sharpness* was chosen for analyzing the timbral balance. The calculation follows the *Aures* method. In the diagrams, the maximum values were calculated.

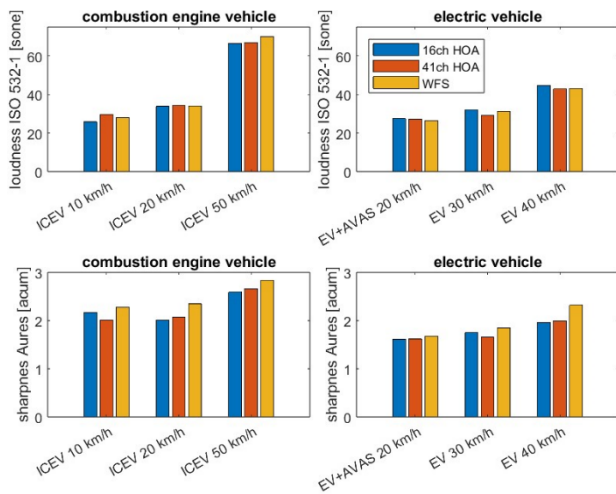


Figure 6: psychoacoustic values

Through the level match of the scenes, also the loudness of the stimuli nearly shows the same values. By building a linear regression between the perceived loudness with the calculated loudness, a correlation between those values is visible. ($R^2=0.632$, $F(1,16)=27.507$, $p<.001$) The loudness perception is even stronger correlated with the calculated sharpness. ($R^2=0.843$, $F(1,16)=85.619$, $p<.001$) It was not possible, to detect a correlation between the perceived timbral balance and the sharpness. The perceptions show some similar tendencies (e.g. ICEV 10 km/h stimuli) to the psychoacoustic values, but not a significant correlation.

Conclusion

Before analyzing the results, no perceptual differences between the certain playback environments were expected, because the stimuli were level matched, were simulated with the same source signals and position data of the vehicles and the same recording and playback setup was used regarding the vehicle stimuli. In contrast to that, the results in this study show significant perceptual differences regarding the two chosen attributes.

Outlook

This study uses stimuli of driving-by-scenes, that are characterized by a time window until zero seconds before the arrival point at the listener position. Many studies until now (as mentioned above) investigated the estimations of TTC's in different lengths, because this value is important for traffic safety inquiries. In the conducted listening test, the aim of this study was a first overview about the sound characteristics of the actual spatialization systems in Dresden and Mainz. Future experiments should include the TTC estimations as well. One open question is, whether not only the loudness, but also the timbral balance of a signal affects the TTC estimations. If that was the case, beside the calibration of the systems, also the equalization of the systems plays a central role, when simulating traffic scenarios in the virtual environments.

Acknowledgments

This project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 444809588. The project is part of the priority program AUDICTIVE - SPP2236.

The authors want to thank all subjects that took part in the experiment.

Literatur

- [1] Blauert, J.: 3-D-Lautsprecherwiedergabemethoden. DAGA 2008, Dresden, 2008.
- [2] Spors, S.; Ahrens, J.: A Comparison of Wave Field Synthesis and Higher-Order Ambisonics with Respect to Physical Properties and Spatial Sampling. In: *Audio Engineering Society Convention 125*. Audio Engineering Society, 2008.
- [3] Oberfeld, Daniel; Wessels, Marlene; Büttner, David. Overestimated time-to-collision for quiet vehicles: Evidence from a study using a novel audiovisual virtual-reality system for traffic scenarios. *Accident Analysis & Prevention*, 2022, 175. Jg., S. 106778.
- [4] Wessels, M.; Hecht, H; Huisman, T; Oberfeld, D.: Trial-by-trial feedback fails to improve the consideration of acceleration in visual time-to-collision estimation. *PLoS one*, 2023, 18. Jg., Nr. 8, S. e0288206.
- [5] Beyer, F; Fischer, S.; Steinbach, L.; Altinsoy, M. E.: Comparison of Recorded and Synthesized Stimuli of Traffic Scenarios in an Auditory Virtual Reality Environment Using Wave Field Synthesis. DAGA 2023, Hamburg, 2023.
- [6] Steinbach, L.; Beyer, F.; Altinsoy, M. E.; Oberfeld-Twistel, D.; Wessels, M.: Safety Investigation on Traffic Scenarios using Virtual Environments in a Wave Field Synthesis Laboratory. DAGA 2022, Stuttgart, 2022.
- [7] Altinsoy, M. E.; Jekosch, U.; Landgraf, J.; Merchel, S.: Progress in auditory perception research laboratories—Multimodal measurement laboratory of Dresden University of Technology. In: *Audio Engineering Society Convention 129*. Audio Engineering Society, 2010.
- [8] Free Software Foundation, 1991. *TASCAR-Toolbox for Acoustic Scene Generation And Rendering* [online]. Boston. [Zugriff am: 19.03.2024]. Verfügbar unter: <http://www.tascar.org/manual.pdf>
- [9] International Telecommunication Union, 2017. ITU-R BS.2399-0: Methods for selecting and describing attributes and terms, in the preparation of subjective tests. Geneva, 2017