Grant Agreement Number: 101015442
Project Acronym: SILVARSTAR
Project title: Soil Vibration and Auralisation Software Tools for Application in Railways

DELIVERABLE D4.1

METHODOLOGY FOR AURALISATION AND VIRTUAL REALITY OF RAILWAY NOISE

Project acronym:	SILVARSTAR
Starting date:	01/11/2021
Duration (in months):	24
Call (part) identifier:	H2020-S2RJU-2020 (Shift2Rail JU call proposals 2020) Topic: S2R-OC-CCA-01-2020
Grant agreement no:	101015442
Grant Amendments:	NA
Due date of deliverable:	30/10/2021
Actual submission date:	04/ 11/2021
Coordinator:	Pascal Bouvet, VIBRATEC
Lead Beneficiary:	Reto Pieren, Empa
Authors:	Reto Pieren[1], Fotis Georgiou[1], David Thompson[2], Kurt Heutschi[1], Giacomo Squicciarini[2], Martin Rissmann[3], Pascal Bouvet[3]
Partners:	[1]Empa, [2]University of Southampton, [3]Vibratec
Version:	Final draft
Type:	Report
Sensitivity or Dissemination level[1]:	Public
Contribution to S2R TDs or WAs[2]	WA5
Taxonomy/keywords:	Noise, Simulation, Auralisation, VR

This project has received funding from the Shift2Rail Joint Undertaking (JU) under grant agreement 101015442. The JU receives support from the European Union's Horizon 2020 research and innovation programme and the Shift2Rail JU members other than the Union.

[1] PU: Public; CO: Confidential, only for members of the consortium (including Commission Services)

[2] https://projects.shift2rail.org/s2r_matrixtd.aspx

Document history

Date	Name	Company	Details of Contribution	Position/Project Role
1/11/2020	Reto Pieren, Fotis Georgiou, David Thompson, Giacomo Squicciarini, Daniel Gremli, Dario Sala	Empa, ISVR, Bandara	Technical work towards the calculation models, software frameworks and VR demonstrator	Technical WP4 contributors
20/10/2021	Reto Pieren, Fotis Georgiou, Kurt Heutschi	Empa	Initial full draft	WP4, T4.1 and T4.3 leader, technical WP4 contributors
22/10/2021	David Thompson	ISVR	Review and minor improvements.	T4.2 leader
26/10/2021	Reto Pieren, Kurt Heutschi	Empa	Minor additions and improvements	WP4 leader, scientist
02/11/2021	Martin Rissmann	VibraTec	Review and comments	Technical WP4 contributor
4/11/2021	Reto Pieren	Empa	Minor additions and corrections	WP4 leader

Table of contents

Executive Summary

This report represents Deliverable D4.1 "Methodology of auralisation and virtual reality of railway noise" of the Collaborative Project SILVARSTAR that is funded under the European Union’s Horizon 2020 Research and Innovation Programme under the open call S2R-OC-CCA-01-2020 as part of the Cross Cutting Activities of the Shift2Rail Joint Undertaking.

This document reports on the methodologies for auralisation and Virtual Reality of railway noise with a focus on the techniques and models used in the further developments within the SILVARSTAR (SoIL Vibration and AuRalisation Software Tools for Application in Railways) project. It gives an overview and describes the status of the work being done within the second technical workstream of SILVARSTAR. This second workstream is fully covered by the project's work package WP4. The major work has been conducted within the task T4.1 on the methodology. In addition, first results from tasks T4.2 and T4.3 on model and software developments are presented.

The methodology of creating a railway noise scenario in a virtual reality environment is developed with help of a demonstrator. In a first step, simulation requirements have been identified and based on this, a suitable software architecture has been defined. Relying on experience from other projects, a prototype synthesizer structure was developed that can artificially produce audio and video for a predefined set of scenarios. The audio synthesizer structure mimicks the physical sound generation and propagation mechanisms to allow for maximal versatility in situation parameter variation.

In agreement with the project partners and the Shift2Rail FINE-2 project, a set of scenarios and noise mitigation measures has been defined and in parts implemented in the virtual reality demonstrator. A user can now dive in a scenario, freely rotate his head and dynamically switch between different variants (e.g. with or without noise barrier). In the following project phase, this set of scenarios will be extended and further refinements will be made to the audio signal synthesis, both on the emission and propagation side.

List of abbreviations, acronyms, and definitions

Abbreviation / Acronyms	Description
dB	decibel
DESTINATE	Decision supporting tools for implementation of cost-efficient railway noise abatement measures
Empa	Swiss Federal Laboratories for Materials Science and Technology
EU	European Union
GUI	Graphical Use Interface
HATS	Head and Torso Simulator
HMD	Head Mounted Display
HRTF	Head Related Transfer Functions
HpTF	Headphone Transfer Function
HVAC	Heating, ventilation, and air conditioning
ISVR	Institute of Sound and Vibration Research
SILVARSTAR	Soil Vibration and Auralisation Software Tools for Application in Railways
S2R	Shift2Rail
TWINS	Track Wheel Interaction Noise Software
UX	User experience
VBAP	Vector Base Amplitude Panning
VR	Virtual Reality
WP	Work Package

List of figures

1 BACKGROUND AND MAIN GOALS

In the second workstream of the SILVARSTAR project, a Virtual Reality (VR) tool that offers aural and visual information will be developed. This novel audio-visual VR tool will allow for comprehensive demonstrations of railway noise mitigation.

The key contribution to auralisation of railway noise in SILVARSTAR is to further develop the existing calculation models and to extend the simulation applicability to a wide variety of noise scenarios. The SILVARSTAR auralisation models will be physics-based and will allow to demonstrate different noise mitigation technologies and their possible combinations. The auralisation models will be linked to 3D visualisation solutions to realise immersive audio-visual VR experiences of railway scenes. On the hardware side, a mobile VR system involving a VR headset and headphones will be set up, calibrated and tested.

This report describes methodologies for auralisation and VR of railway noise with a focus on the techniques and models used in the further developments within SILVARSTAR.

2 INTRODUCTION

Auralisation is the acoustical counterpart to visualisation and allows users to audibly experience situations that do not necessarily exist (yet). The term was introduced by Kleiner et al. (1993) and it involves rendering an audible sound field emanating from acoustic sources located within a virtual environment. Auralisation has a long tradition in the planning process of concert halls and opera houses (Vorländer, 2020). Room impulse responses are obtained either via scale model measurements or with computer models, e.g. ray tracing (Savioja and Svensson, 2015), and combined with anechoic source recordings to render the sound at a location within the room.

During the past decade, there has been increasing interest in environmental acoustics applications of auralisation. Auralisation provides much more information than noise levels regarding the acoustics and the quality of the urban sound environment. It can model the urban sound environments with all the sound sources present; thus, it directly supports the concept of soundscapes[3] and can be used to characterise different acoustical scenarios using psychoacoustic metrics. It can also be combined with visual information, thereby achieving a more thorough evaluation of the urban environment (Georgiou, 2018). Consequently, auralisation can be a great communication tool. For example, it can be used in noise control engineering to provide an audible example of the effect of a future noise mitigation measure instead of demonstrating the reduction graphically or in terms of decibel values. This will allow the relevant parties involved in that future development (public, decision makers, vehicle customers and designers) to have a full immerse experience of the effect of the mitigation measure and obtain much better insight than sound pressure levels or other acoustic

[3] ISO definition of soundscape: "an acoustic environment as perceived or experienced and/or understood by a person or people, in context"

measures, which are difficult to communicate to an audience without an acoustic-related background.

The two major challenges in environmental acoustics auralisation are that the sound radiated from the relevant sources cannot typically be recorded in an anechoic room and that the sources may move within the scene. This requires new strategies and computational models. Simple auralisation can be based on observer recordings, modified with standard signal processing operations such as mixing and filtering. Such an approach has been used to demonstrate railway noise scenarios (Geiger et al., 2018). These techniques are rather limited, however, only allowing for a static observer location. An initial tool for train pass-by noise auralisation called VAMPPASS was developed by SNCF within the SILENCE project (Bongini et al., 2009) in which the rolling noise signals were recording-based. A physics-based rolling and impact noise synthesis model was developed by Empa in the Swiss TAURA project (Pieren et al., 2017). The model separately describes the mechanical excitation due to the rail and wheel unevenness and the resulting sound radiated by the vehicle and the track. Extensions of this emission model have recently been published (Pieren et al., 2018a; Maillard et al., 2019) and partially used in the railway pass-by noise demonstrator RNX-VR developed by Empa in the DESTINATE project (Pieren et al., 2018a, Pieren et al., 2018b).

Auralisation can be combined with the visual 3D models of the simulated space within a commercial 3D game engine[4] (e.g. Unity or Unreal) providing a full immerse audiovisual simulation. In the rest of this report we will refer to VR simulation as the full audio-visual simulation of the environment.

Section **Erreur ! Source du renvoi introuvable.** gives a brief overview of the simulation requirements and the software architecture of the VR demonstrator. Section 4 outlines the information about the different simulated scenarios. In addition, it describes the auralisation methodology for train pass-bys and the auralisation of a train's interior noise. Section 5 demonstrates the user workflow. Moreover, it presents the visualisation and the graphical user interface (GUI) of the VR demonstrator and the VR station that has been developed for the audio-visual playback. Finally, Section 6 describes the methodology used for audio spatialisation, headphone equalisation and level calibration.

3 CONCEPT OF THE VR DEMONSTRATOR

3.1 SIMULATION REQUIREMENTS

The main requirement of the VR demonstrator is to create a state-of-the-art audio-visual environment that can be used for the evaluation of different scenarios of railway noise using a single train pass-by as a basis. The tool will allow the user to choose between different: environments (e.g. rural, urban), mitigation measures (e.g. barriers, rail & wheel dampers), type of trains (e.g. regional, intercity), track types (e.g. ballasted, slab) and observer locations.

[4] 'Game engine' is a software framework designed for the development of video games and many of them, such as the Unity and the Unreal game engines, are used for the design of VR applications.

Based on the input parameters, a set of audio-visual simulations will be generated which can then be compared with each other. To achieve the required high quality simulations within the real-time VR demonstration, a substantial amount of audio pre-rendering is required.

3.2 SOFTWARE ARCHITECTURE

In Figure 1 the general software architecture of the VR demonstrator is shown. Two linked software tools are being developed: the SILVARSTAR Auralisation tool and the SILVARSTAR VR tool. The input parameters related to the environment and the mitigation measures are sent to both tools.

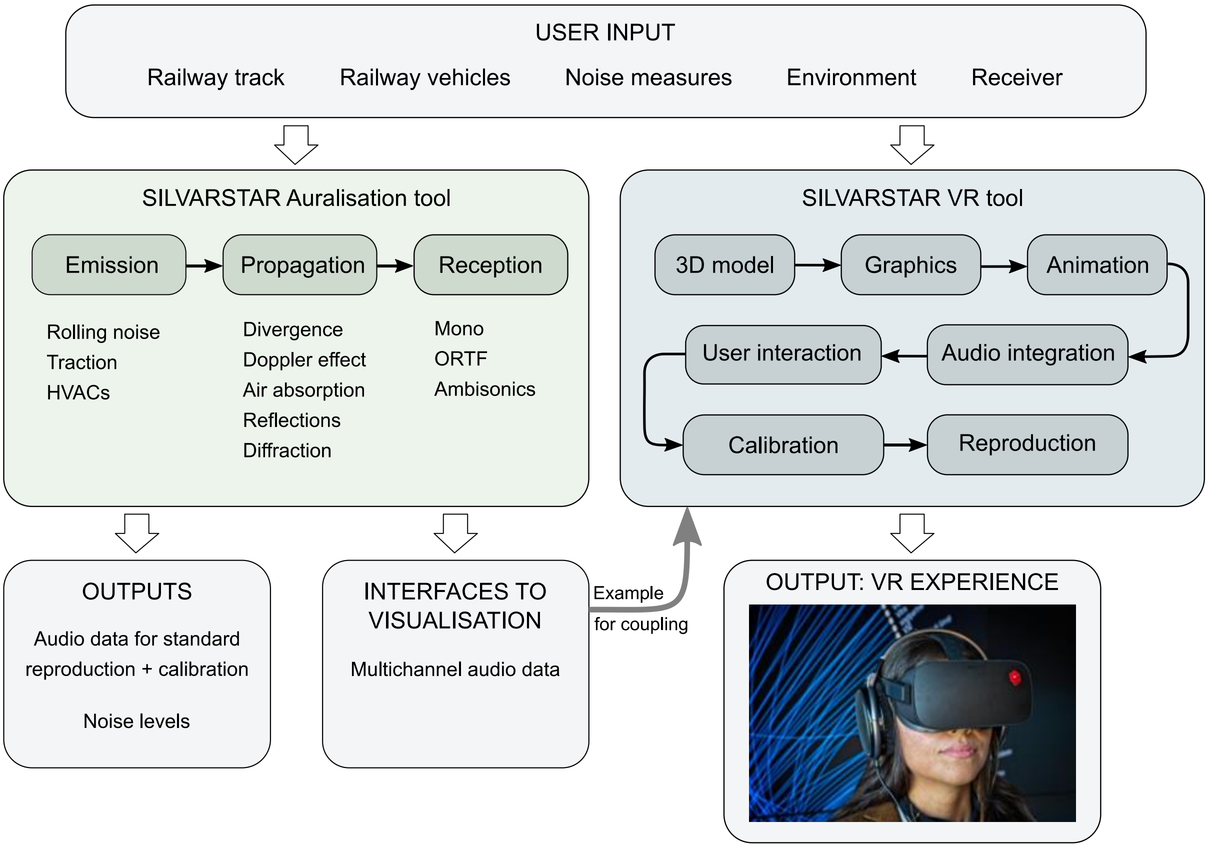

Figure 1: Software architecture of SILVARSTAR VR demonstrator

The Auralisation tool will perform the auralisation of the train pass-by based on input parameters. It will provide calibrated audio data for different reproduction systems and computed noise levels. It will also generate the multichannel audio data that will be sent as input to the VR tool. The Auralisation tool will be developed in Matlab.

The VR tool will use the user input parameters to create the 3D visual environment and the animations and then couple them with multichannel audio data. The output from the VR tool is sent to the head mounted display (HMD) and a pair of headphones for the audiovisual reproduction. The VR tool will be developed with the cross-platform gaming engine Unity. More details about the different functions of the Auralisation tool and the VR tool are given in the following sections.

The resulting VR system will facilitate communication with the public, decision makers, vehicle customers and designers through experiences in VR before delivery of projects.

4 AURALISATION OF RAILWAY NOISE

4.1 SCENARIOS RELATED TO REQUIREMENTS

The railway noise scenarios that can be considered within SILVARSTAR are based on the requirements Deliverable D9.1 of the FINE-1 project (Geiger et al., 2018), the published project call text and collaboration with the FINE-2 project partners. The work is focussed on train pass-by noise and the demonstration of mitigation measures. The scenarios will allow the comparison of different mitigation effects for different types of trains, such as varying barrier height or adding rail dampers etc. Below is a list of the input parameters for the design of the different simulated scenarios within the VR environment. A graphical representation of some of these parameters is shown in Figure 2.

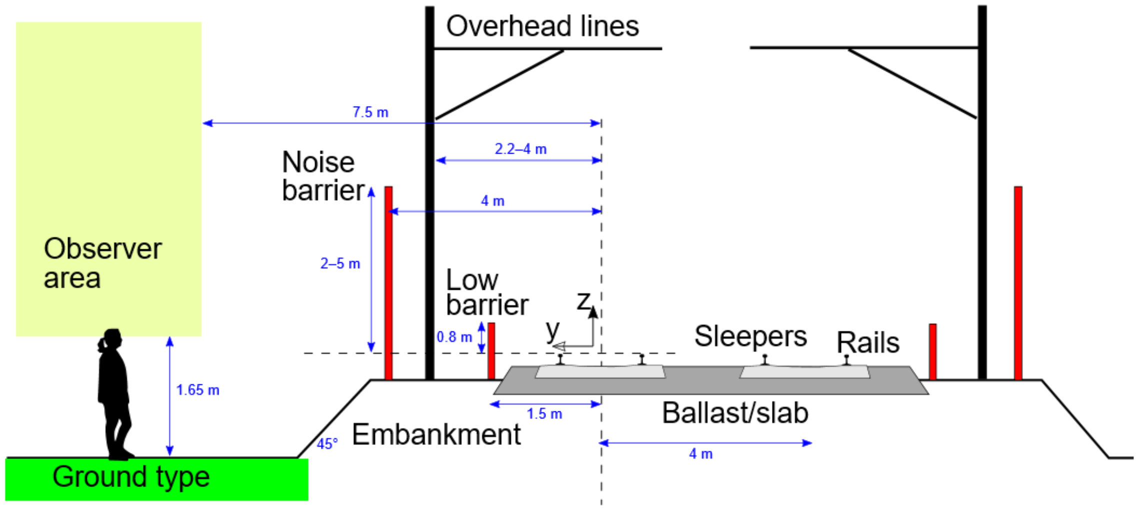

Figure 2: Sketch of the different objects/parameters that can be modified within the VR demonstrator

- **Environment:**
  - Environment type: Rural, Urban
- **Track:**
  - Embankment height: Minimum: 0 m, Maximum: 4 m
  - Track type:
    - Ballasted:
      - Sleepers type: Monoblock, Biblock
      - Rail pads type: Soft, Medium, Hard
    - Slab
- **Vehicle:**
  - Train type: Regional, Intercity, Freight
  - Train length: Short, Long
  - Travelling speed: Minimum: 50 km/h, Maximum: 160 km/h
  - Start and end position along track: Minimum: 50 m, Maximum: 500 m
  - Used track of the double-track: Close, Far
  - Driving direction: Left to Right, Right to Left
- **Observer:**
  - Distance to nearest track centre: Minimum: 7.5 m, Maximum: 50 m
  - Height above ground: Minimum: 1.65 m, Maximum: 10 m
  - Ground type: Grass, Asphalt
- **Mitigation:**
  - Freight train brake block Cast iron to K-block ratio (Ci-K-ratio): 0%, 50%, 100%
  - Wheel flats: With, Without
  - Wheel dampers: With, Without
  - Secondary sources attenuation: in dB
  - Rail roughness: Smooth, Medium, Poor
  - Rail treatment: None, Rail damper, Rail shield
  - Barrier type:
    - None
    - Low (height 0.8 m)
    - Standard
      - Barrier height: Minimum: 2 m, Maximum: 5 m

4.2 CONCEPT OF SIMULATION OF TRAIN PASS-BY NOISE

In SILVARSTAR, auralisation of train pass-bys will follow an object-based and physics-based approach. Existing models of the project partners will be improved to meet the requirements regarding simulation of combinations of different vehicles, their operation, tracks, environments and mitigation measures.

The chosen concept of the synthesis of railway pass-by noise is based on previous work of Empa within the Swiss TAURA project and the European DESTINATE project (Pieren et al. 2018a, Pieren et al. 2018b). The auralisation method follows the source-path-receiver concept illustrated in Figure 3. Virtual spatially distributed sources describe physical sound sources (see Figure 4). Acoustical sources are modelled by distinct point or line sources located in the virtual environment. Thereby, a source signal that describes the sound as radiated by the source is attributed to a certain location with a defined orientation. Based on source specifications, source signals are generated by parametric sound synthesis. Propagation effects from the source location to the observer point are simulated by processing the source signals with propagation filters. Because the source locations change over time, time-varying filters are used. In order to provide directional information to the listener, the reproduction renderer considers the directions of the incident sounds at the observer and the observer orientation to calculate the reproduction channel signals. At this processing stage recordings of ambient sounds are also integrated.

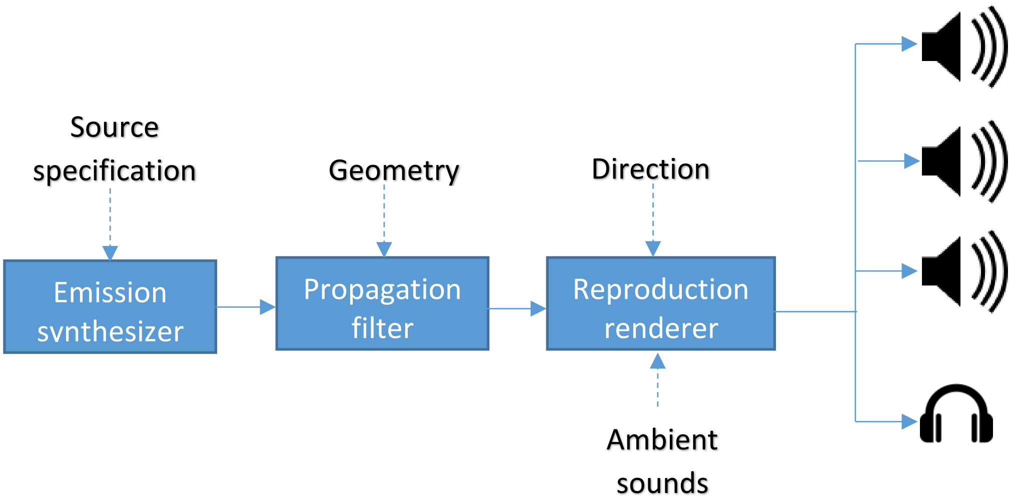

Figure 3: Block diagram of the physics-based train pass-by auralisation process with the three main model modules.

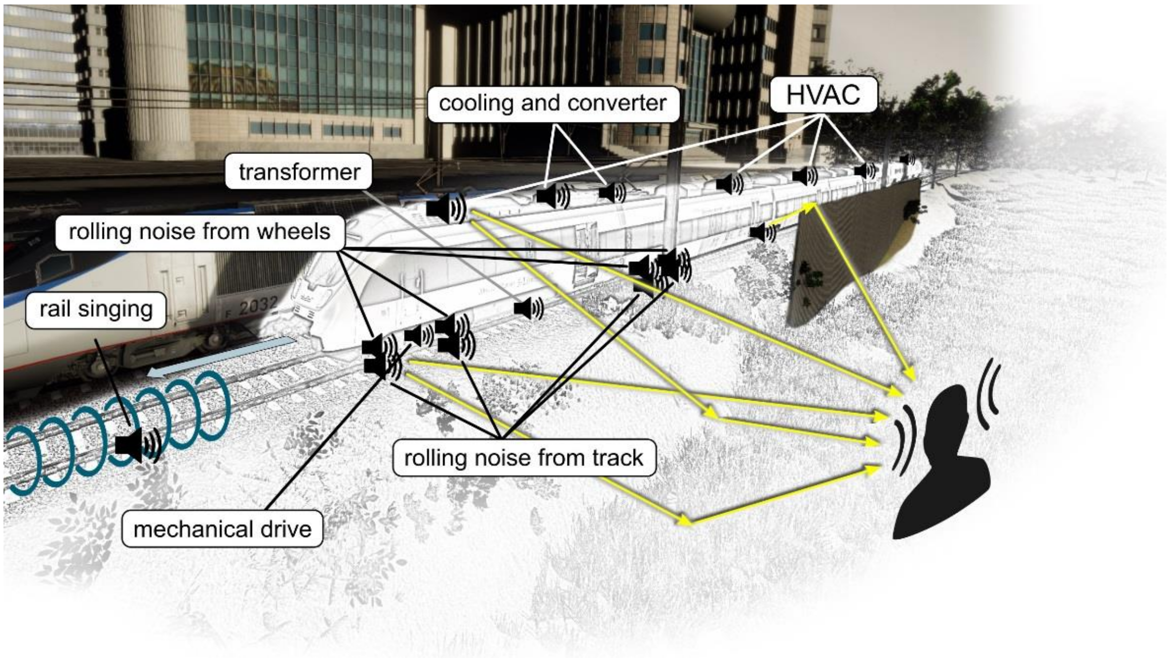

Figure 4: Rail vehicle pass-by noise auralisation concept as physics-based modelling using a collection of moving virtual sound sources representing physical acoustical sources.

4.3 EMISSION SYNTHESIS

In the existing auralisation model, the major sound sources are represented by distributed point sources that move inside the virtual environment with the trajectory of the train (see Figure 4). In many railway noise situations, rolling noise is the dominant source. In addition to that, the wheel/rail contacts may also lead to transient noise, so called impact noise that occurs in cases of wheel flats or irregularities on the rail, such as switches or rail joints. Rolling and impact noise both are very characteristic in the perception of railway noise.

Following the approach from the DESTINATE project, rolling and impact noise can be artificially generated using a physics-based synthesis model (Pieren, et al., 2017). The input variables are related to the vehicle and the track. The starting point is a generated surface profile (i.e. the roughness) of the rail and the wheels. These are generated from roughness spectra described in 1/3 octave-band resolution which can be taken from the CNOSSOS-EU (European Commission, 2015) or the sonRAIL model (Thron & Hecht, 2010). Synthesised roughness signals are then combined and processed to obtain the mechanical excitation of the wheel/track structure. Using the vehicle's traveling speed, the spatial excitation signals are transformed into the time domain. Next, the modal behaviour of the structure and the radiation are considered. The energy transfer from the mechanical excitation to the radiated sound is described by transfer paths. In DESTINATE, two transfer paths have been considered, one for the contribution of the vehicle, and one for the contribution of the track, according to the CNOSSOS-EU model (European Commission, 2015). In SILVARSTAR, it is intended to extend this description to more transfer paths. A finer distinction of source contributions and thus a more refined description of the physical processes allows to appropriately include mitigation measures that affect only parts of the wheel/track system as well as an improved modelling of different source directivities. More detailed transfer functions including a separation of rail and sleeper contributions can be computed using the TWINS model (Thompson, 2009). Finally, a directivity function is applied to each source signal.

Input data for rolling noise synthesis is obtained from a combination of existing calculation models and measurements. Wheel and rail roughness, contact filters and transfer functions are required spectrally in 1/3 octave bands. These are common parameters in state-of-the-art engineering models. Information about vibrational resonances (frequency and damping), however, has to be derived from specific measurements or numerical simulations. Particularly for freight trains, variations between the axles within a train are needed to increase the realism of synthetic train pass-by sounds. A hierarchical description of the variance in the emission strengths between the axles of the train composition was proposed (Pieren et al., 2018a, Pieren et al., 2018b).

To improve the source synthesis model for rolling noise, models based on the TWINS prediction model (Thompson et al, 1996a, 1996b) will be used. The latest improvements in the calculation of track noise radiation will be included (Zhang et al., 2019). This model will allow studies of parameter variations such as rail pad stiffness or the inclusion of rail dampers or rail shields. The current synthesis model will also be improved regarding the effective damping of wheel modes, i.e. ‘rolling damping’, which is relevant for the sound radiated by the wheels.

Rail singing is a narrowband phenomenon around 1 kHz that is particularly audible when the train is approaching or receding. Rail singing can be modelled by a line source with a time-dependent synthetic source signal. The model parameters are so far obtained from specific train pass-by measurements. However, these should be linked to the rolling noise parameters.

Secondary sources include those that are related to the traction (e.g. motors or converters), to technical equipment (e.g. HVAC) and to structure-borne sound that is radiated by the car body. Source data has to be obtained from specific acoustical measurements, e.g. (Dittrich & Zhang, 2006) or with a microphone array. For each vehicle and operational condition, the source data describe the relevant source locations with corresponding narrowband sound powers and directivities. On that basis, source signals are artificially generated by spectral modelling synthesis (Pieren, et al., 2016). Secondary source data provided by rail vehicle manufacturers will be required as input.

4.4 PROPAGATION SIMULATION FOR MOVING SOURCES

Sound propagation is modelled with a time-domain 2.5D simulation. It is based on a continuous path tracing approach considering phase information and is realised by a network of time-variant digital filters. Its conception originates from the state-of-the-art engineering models that are used in environmental acoustics.

Relevant propagation effects are separately described based on a mathematical description of the physical processes. During the simulation, these effects are applied to each source signal individually. Due to source motion, all propagation effects change over time. The auralisation considers the following sound propagation effects:

- Geometrical spreading is accounted for by considering a 1/r amplitude scaling with distance r in the far-field of point sources; a $1/\sqrt{r}$ scaling is used for line sources.
- Air absorption is modelled as a frequency and distance dependent correction for a homogeneous atmosphere with a constant air temperature and constant relative humidity.
- For each propagation path, the Doppler effect is simulated by modelling the instantaneous propagation delay of the sound wave and by considering Doppler amplification.
- The ground effect is modelled by introducing additional propagation paths that are superimposed with the direct paths. The ground effect depends on the ground properties and the reflection angle.
- To reproduce shielding effects, e.g. behind a noise barrier, a diffraction model is included.

All these effects are simulated by applying digital filters that are designed based on the work in (Pieren et al., 2016).

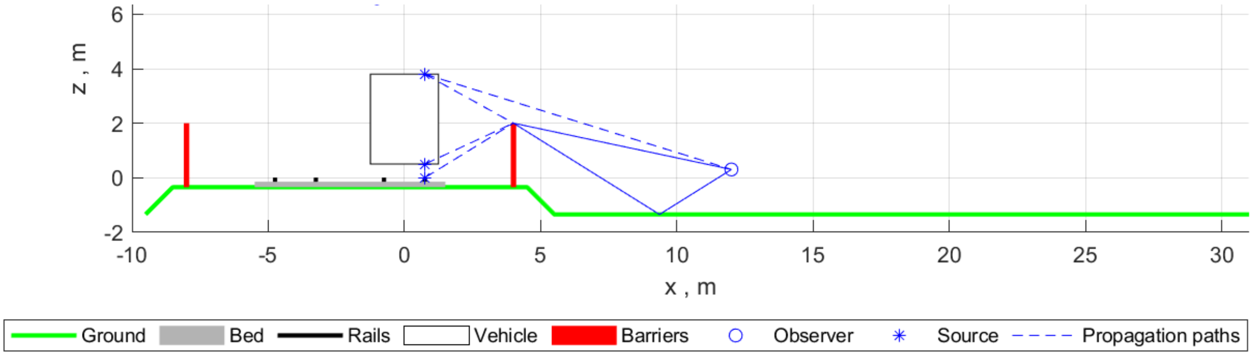

Figure 5: 2D sketch of the propagation situation in a vertical cross section with a double-track on an embankment with noise barriers and the considered propagation paths for three point sources.

Empa's current propagation simulation will be refined regarding ground effects and barrier effects, to give improvements relative to the propagation model used in the demonstrations within the DESTINATE project. Both effects are highly frequency and geometry dependent and vary during the pass-by. Ground reflections on the receiver side of the noise barrier and diffraction at the barrier edge are considered (see Figure 5). The ground effect model will account for the finite impedance of the ground. This allows consideration of the different sound reflection behaviour of acoustically soft and hard grounds (grass vs. concrete). The diffraction model will account for a single diffraction edge and will account for diffracted sound also in the 'illuminated' zone. It will be based on an analytical solution (Pierce, 1974) for point source edge diffraction (see Figure 6Figure 6).

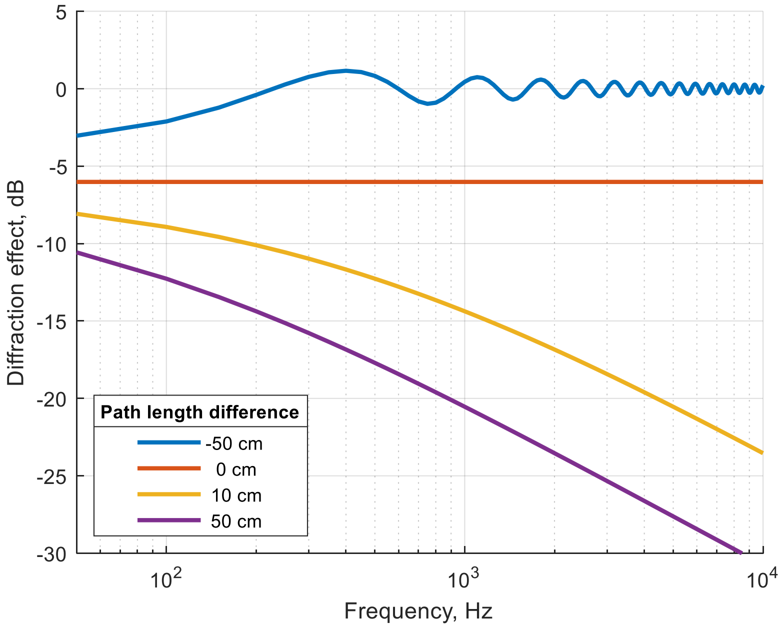

Figure 6: Frequency dependence of barrier effect calculated using an analytical solution for edge diffraction of a point source for different propagation geometries.

4.5 REPRODUCTION RENDERING

Different reproduction systems require differently rendered audio signals, so called speaker feeds. To support to the use of different reproduction systems, multiple output audio formats will be realised as illustrated in Figure 1.

Monophonic signals allow for monophonic reproduction and the computation of classical noise metrics like sound pressure levels or psychoacoustic parameters. For stereophonic loudspeaker reproduction, a 2-channel stereophonic rendering, mimicking a virtual stereophonic microphone setup, can be used. In contrast to such a channel-based format, the ambisonic B-format, a scene-based audio format, allows the output to be fed to an ambisonic decoder for playback over different loudspeaker arrays. As the interface to the SILVARSTAR VR application, where a binaural rendering over headphones is realised, a compact intermediate audio format is developed as described in Section 6.1.

The resulting audio files will contain uncompressed, pulse-code modulated (PCM) audio signals. For calibrated playback and absolute level and loudness analyses, the signal amplitudes are related to Pascals. The intended audio file format is multichannel WAV (WAVEX) with a sampling rate of 48 kHz and a 32 bit depth.

4.6 INTERIOR NOISE AURALISATION

The comfort of railway passengers is influenced by the acoustical quality while traveling (Vos, 2021). Decision making on vehicle design variants and the potential estimation of the effects of interventions in current vehicles can be supported by VR technologies. In this section, we present a framework for interior noise auralisation where 'interior' means the inside of a passenger coach.

For interior noise auralisation and VR, the same basic principles hold as for pass-by noise: sources, propagation and reproduction have to be treated separately. A major difference is that acoustic sources can be assumed to be stationary, i.e. they do not move in space in the receiver coordinate system. This substantially simplifies the source description and the propagation simulation. Typical sources are vibrating surfaces like windows, side walls, roof and floor (excited by airborne or structure-borne sound), the air conditioning system, the announcement system and speech from other passengers. Corresponding source signals and directivities can be captured by specific measurements. The sources can either be measured in an anechoic laboratory or in-situ using, for example, operational transfer path analysis (OTPA) (Garcia Ordiales et al., 2019).

Sound propagation can be simulated using existing methods from room acoustics or as used for the interior of road vehicles. Relevant propagation effects are geometrical spreading, reflections at room boundaries and objects and diffraction. Ray tracing is an established high-frequency approximation for practical applications. It can be complemented by wave-based simulations for the low frequencies. Commercial ray tracers, e.g. the software product ODEON, allow source-receiver impulse responses to be computed considering source directivity, complex propagation geometries and surfaces with different materials. Surface materials are described using a frequency-dependent sound absorption coefficient and scattering properties.

This, for instance, allows the sound absorbing effect of cushioned seats to be simulated. The computation of spatial impulse responses is required to create a spatial impression and allow the listener to localise sources in the virtual space. Use of a scene-based audio format such as an ambisonic B-format allows flexibility with respect to the sound reproduction system. B-format impulse responses for multiple source-receiver combinations and design variants can be pre-computed and stored for integration into an immersive VR environment.

Possible movements of the virtual passenger in the interior can be assumed to be extremely slow compared with the speed of sound and therefore no Doppler effects and other audible time-variant effects have to be modelled. This allows the spatial interpolation of impulse responses. This has been recently implemented in a VR environment for the architectural design of rooms such as open plan offices (Llopis, 2020). A scene-based audio format at the receiver location is then obtained by convolving each source signal with the corresponding set of interpolated B-format impulse responses and summation of the source contributions. A simpler approach that neglects reverberant effects is to consider only attenuation due to geometrical spreading in the propagation simulation. By applying time-variant gains to the source signals, an object-based audio format is obtained at the receiver.

Following the described framework, the options for the reproduction rendering and the sound reproduction system are the same as for the pass-by simulation. One can then decode the resulting receiver B-format to a real or virtual loudspeaker array. The latter option is used as an intermediate format to render for headphone reproduction using head-related transfer functions (HRTFs, see Section 6.1). In the case of an object-based format, the signals are panned to a loudspeaker array or binaurally rendered using HRTFs.

5 VIRTUAL REALITY DEMONSTRATOR

This section introduces the methodology and the current status of the VR demonstrator under development.

5.1 USER WORKFLOW AND SCENARIO INFORMATION

Figure 7 demonstrates the workflow for the user that has to be followed to design new cases and scenarios for the VR demonstrator. The demonstrator development workflow can be split into three main steps:

1. The input parameters for the simulated Case/Scenarios (see the Subsection 5.2 for details regarding the different options) need to be specified and written in an input file. As a human readable hierarchical ASCII file format the JSON file format was chosen. A template for the JSON input file has been written. The 'Case' contains the information regarding the environment (e.g. rural), track (e.g. ballasted), vehicle (e.g. long intercity) and observer (e.g. height 5 m), but not the mitigation. Each Case then contains a number of different mitigation 'Scenarios'; the number is currently limited to 3, e.g. Scenario A (without noise barrier), Scenario B (with noise barrier of 2 m height) and Scenario C (with low height noise barrier).

2. The JSON file should be sent as an input to the auralisation software, which simulates the train pass-by and produces multichannel audio signals. The auralisation software will be developed and programmed in Matlab and provided as a compiled executable or a web-based service.

3. The audio signals together with the input file must then be sent to the VR application. The VR application will be developed in Unity. Based on the input file's parameters, the VR application will search the database of predesigned 3D models, materials etc for the specified objects and generate the visualizations/animations. Next, it will couple and synchronise the audio with the visuals to create the full VR simulation of the specified Case/Scenario.

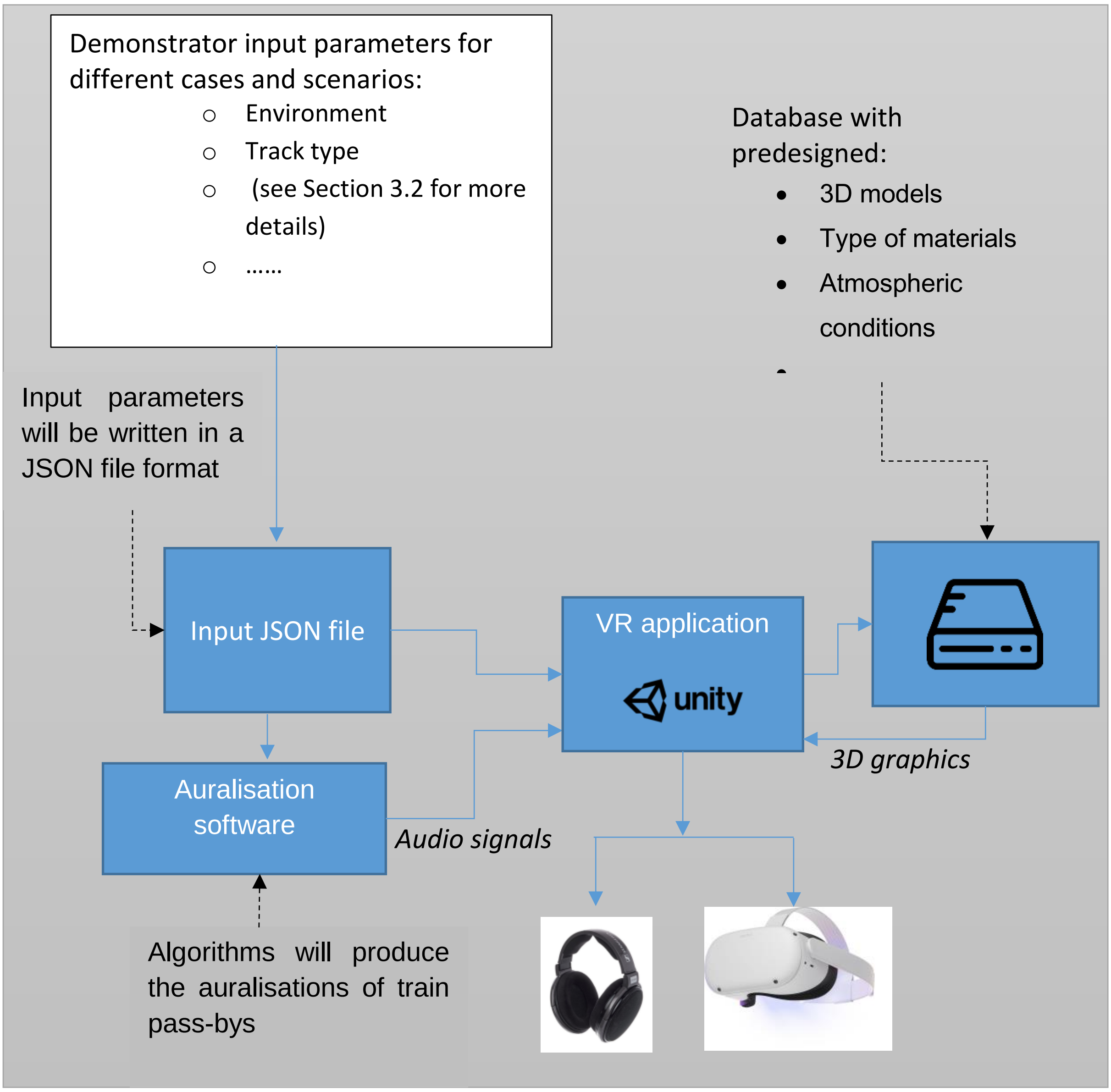

Figure 7: Block diagram of the development platform for the VR demonstrator for train pass-by events

The user will not require any programming background in order to implement the steps mentioned above. The user will only have to edit the input file based on the selected parameters for the Cases/Scenarios and import it to the auralisation software and then to the VR application together with the audio files. The template for the input file includes instructions about how the input file should be edited and the permitted values that the different parameters can take. The input file can be edited using free available software such as Notepad++ and Visual Studio Code.

5.2 VISUALISATION, GUI AND USER EXPERIENCE (UX)

The 3D animations and the GUI are designed based on previous work of Empa in the DESTINATE project (Pieren et al., 2018a; Pieren et al., 2018b) in collaboration with the Swiss company Bandara VR GmbH. The software design follows the scenarios related to the requirements document of FINE1 (Geiger et al., 2018), see Subsection 4.1. The animations are developed within the Unity game engine version 2021.1.

Visual playback is done via a HMD of the type Oculus Quest 2. This device can be considered as today's state-of-the-art commercial product. It is in the price range of 400 € and has a per-eye resolution of 1,832 × 1,920 (3.5 megapixels), which is very high for this price range. Moreover, it has a built-in head tracking device so no external sensors are required (as opposed to the Oculus Rift which was used in DESTINATE).

Once the VR application is launched, a loading window opens which asks to load the JSON input file (see Subsection 5.1 and left picture in Figure 8). Once the file is loaded, another window opens, which asks to select the case that the user would like to demonstrate (right picture in Figure 8). After selecting the case, a new window appears (see Figure 9) where the user can experience the different scenarios within that case or switch instantly to a different case. The GUI shown in Figure 9 is only visible on the operator's screen whereas the person who is wearing the HMD only sees the video shown on the bottom right of Figure 9.

In Figure 10, Figure 11, and Figure 12 examples of the 3D visuals are presented. The train type and the environment is the same in each figure, the only parameters that change being related to the track and the barriers.

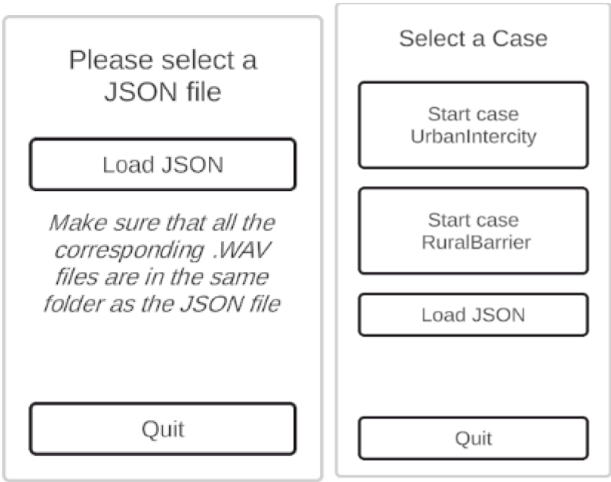

Figure 8: First two loading windows of the GUI of the VR tool prototype

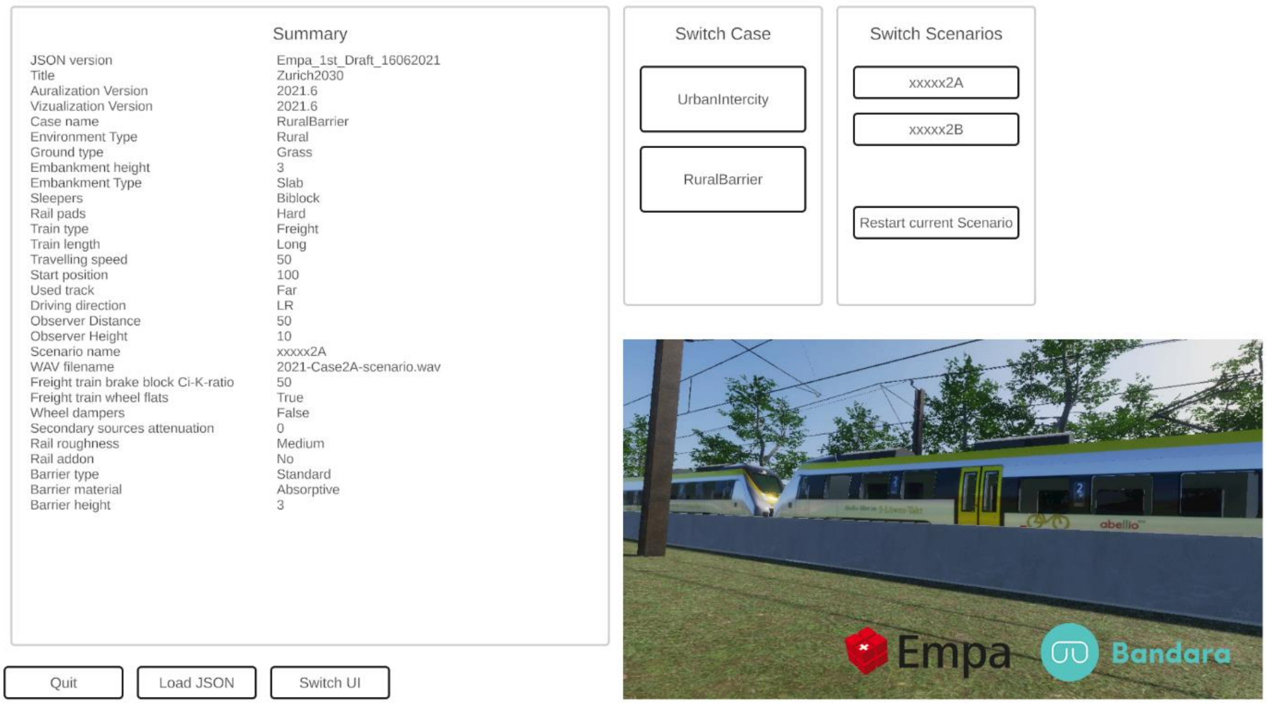

Figure 9: GUI of the VR tool prototype where the user can switch between different scenarios and cases

Figure 10: Freight train in rural environment, with slab track, rail shields and 2 m standard barrier

Figure 11: Same as Figure 10 but with ballasted track, no rail shields, a low-height barrier and an embankment

Figure 12: Same as Figure 10 but with ballasted track, no rail shields and 2 m transparent barrier (only a basic representation of the ballast is used which will not be visible at realistic observer locations with larger distance to the track)

To make the experience interactive, virtual buttons are added inside the virtual environment so even the user wearing the HMD is given the opportunity to switch scenarios by clicking them using the Oculus controllers. The switch between the scenarios (by clicking the virtual buttons) is done dynamically, e.g. assuming that we have a case with Scenario A (without noise barrier) and B (with noise barrier) and the user is playing Scenario B (see Figure 13) and when the train is just in front of him/her, by pressing the left virtual button (see Figure 14) the barriers disappear (switch to reference scenario A) without the simulation starting again from the starting position. It should be noted that this is not the final GUI since it is still under development.

Figure 13: Rural VR environment with virtual buttons (scenario with barriers) of the VR tool prototype

Figure 14: Rural VR environment with virtual buttons (scenario without barriers) of the VR tool prototype

5.3 VR STATION

A mobile VR station (see Figure 15 and Figure 16) has been designed and put in operation which is in line with requirements specified in FINE1 report (Geiger et al., 2018) as listed below:

- Spatial requirements:
    - Small construction for exhibitions, trade fairs (maximal 2 m x 2 m)
    - Extension for larger rooms (presentation situation)
- Size / mobility of technical equipment (transport / setup-time):
    - Fits in a van
- Setup-time:
    - Under 4 hours
- Aesthetics:
    - Solid, professional, technical appearance
    - Possibility to apply company logos
- Time frame:
    - One sequence should not be longer than 3 minutes
- Presentation frame:
    - Moderation & discussion: possibility for pause and interaction necessary
- Internet capabilities

Figure 15: VR station of Empa

The station consists of two main units, the laptop stand and the speaker-and-monitor stand. The laptop stand was designed to insulate the noise produced from the laptop fans, which operate most of the time at 100% due to computational demands from the real-time VR rendering. Therefore, if they are not insulated, they would produce so much noise that it would be audible to the user while wearing the headphones. The laptop stand has also built-in fans to ensure sufficient cooling for the laptop, but they produce much lower noise levels. The speaker-and-monitor stand has dimensions of 1.90 m (H) × 1.29 m (W). To inform the operator and possible spectators about the current demonstration condition, a 32-inch screen is attached to it and two Neumann KH 80 4-inch speakers. A multichannel USB audio interface is installed, which is responsible for the audio playback via the headphones and the speakers. Two motion sensors have also been attached (to the left and right of the speakers) to allow head/motion tracking using the Oculus Rift but since the system will be upgraded with the Oculus Quest 2, which has built-in motion tracker, these will no longer be necessary. In order for the VR demonstrator to operate with the current visuals/animations and the Oculus Quest 2 it must be ensured that the graphics card (GPU) is at least an NVIDIA GEFORCE RTX 20 series or equivalent.

Figure 16: Operation of Empa's current VR station with a user wearing a HMD and headphones.

6 AUDIO REDENDERING AND HEADPHONE PLAYBACK CORRECTION

6.1 SPATIAL AUDIO RENDERING, HRTFS AND HEAD TRACKING

To achieve high immersion and plausibility of the VR experience, the auralisations are spatially rendered allowing the user to localise sound sources in the virtual space. The simulation shall also allow for head rotations of the user.

In the SILVARSTAR VR demonstration, the audio reproduction is implemented via headphones. As a compact intermediate channel-based audio format, an array of virtual speakers is used. Firstly, the concept of head related transfer functions (HRTFs) and virtual speakers is explained and then the method used to render the auralisation is demonstrated.

HRTFs contain the information about how sound propagates to the eardrum in free space, which includes interaural time delays and interaural level differences as well as spectral modifications caused by the head, outer ear and body. For a point source in the far field, the two HRTFs for left and right ear are a function only of the incidence angles, e.g. azimuth and elevation. Several HRTF data sets are published and available. They vary in size and quality with respect to angular resolution and equalisation. By convolving a monophonic signal with the corresponding HRTFs of left and right ear, a binaural (two-channel) signal is rendered. A virtual loudspeaker can be binaurally rendered using HRTFs. For example, to playback binaurally via headphones a signal coming from a loudspeaker located at 30° azimuth and 0° elevation, the mono signal that is being played by the speaker must be convolved with the HRTF measured at that angle. Using this method, different speaker layouts can be rendered binaurally.

The virtual loudspeaker set up that has been used for this VR demonstrator is 2D and is shown in **Erreur ! Source du renvoi introuvable.**. It covers only the region in front of the receiver, which is sufficient for simulating a train passing-by in front of the listener. The process followed for spatial rendering is as follows: The signal for each of the 7 virtual loudspeakers is rendered using a modified version of Vector Base Amplitude Panning (VBAP) (Pulkki, 1997), which is a generalisation of the (stereo) tangent panning law (pair-wise amplitude panning). Pair-wise amplitude panning requires an object-based sound scene description and can handle almost any irregular loudspeaker layout. As only a minimum number of feeds are simultaneously active, it produces good source localization, average sized sweet spot and only little coloration. The VBAP is modified to introduce frequency dependence to account for the sound field distortion introduced by the virtual head. In addition, these virtual speakers are binaurally rendered using HRTFs. The HRTFs that are used are generic, meaning that they are not individualised.

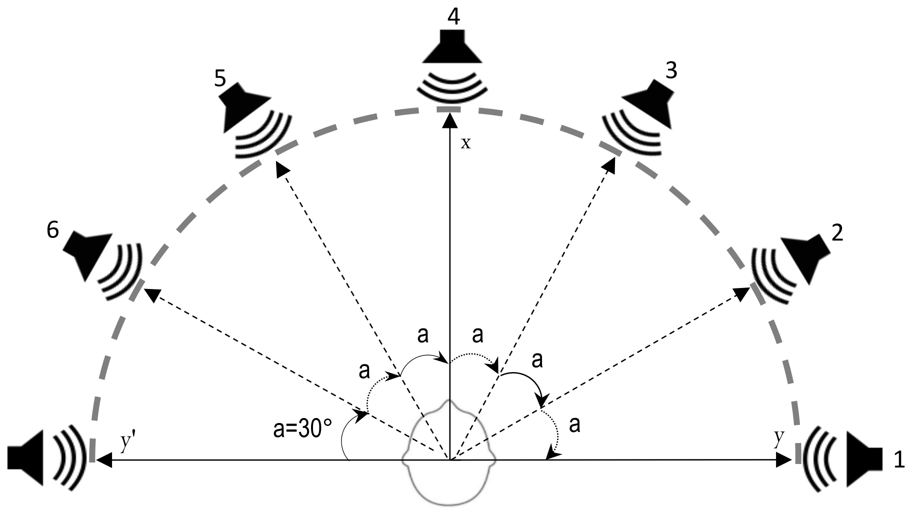

Figure 17: Regular semi-circular virtual loudspeaker set-up consisting of 7 channels as intermediate audio format

When listening via standard headphones, the reproduced sound field moves with the head. This is highly unnatural and thus weakens the credibility of the scene. To overcome this, a head tracker is used to capture the head movements and this information is used to update the listener's HRTFs according to their head orientation, thereby improving realism and sound externalization (Böhm et al., 2018). The head tracker is built in within the Oculus Quest 2. The 7 HRTFs that correspond to the orientation of the head and the 7 virtual speakers shown in Figure 17**Erreur ! Source du renvoi introuvable.** are updated in real-time within the game engine.

6.2 HEADPHONE EQUALISATION

For a realistic and adequate auralisation, the simulated sound should be reproduced with the correct spectral content. While wearing headphones, the sound pressure at the eardrums should be as close as possible to the situation where the listener is exposed to a spatial sound field without wearing headphones. For that, the headphones must be of sufficient quality, reliable in their acoustical response, and their characteristics have to be known and compensated for.

Since headphones are not designed to have a flat frequency response, they have to be equalised for an auralisation or psychoacoustical application. Headphone equalisation is accomplished by first measuring the headphone transfer function (HpTF) using a head-and-torso simulator (HATS). For robust measurement results it is recommended to measure the HpTF multiple times by removing and placing the headphone back on the HATS in order to account for the changes in the coupling between the ears and the headphones each time the headphones are placed back on the head, which causes a change in the HpTF. The equalisation filter is computed by inverting the measured HpTF (Lindau & Brinkmann 2012). This will be integrated as a final processing step in the auralisation software before audio export.

The open-back circumaural headphones Sennheiser HD650 (see Figure 18) are chosen for the reproduction, as in the RNX-VR application from DESTINATE. The HpTFs of the Sennheiser HD650 will be measured using a HATS. It will be made clear in the VR demonstrator user guide that all users must use the Sennheiser HD650 headphones because the applied equalisation filter works only on this headphone model.

Figure 18: Headphone HpTF measurements at Empa with a Sennheiser HD650 using a HATS wearing an HMD of type Oculus Rift.

6.3 LEVEL CALIBRATION

For a realistic and adequate auralisation it is important to achieve the correct loudness perception. One of the main challenges in binaural reproduction is the headphone calibration which is required to be able to reproduce the correct sound pressure level at the eardrums. While wearing the headphones, the sound pressure at the eardrums should be as close as possible to the situation where the listener is exposed to a spatial sound field without wearing headphones.

Headphones are calibrated by adjusting the playback volume based on measurements with a HATS, a free-field measurement microphone and a sound level meter (as reference) and a broadband loudspeaker under anechoic conditions (see Figure 19). Once the headphones are calibrated, the audio output voltage can be measured with a voltmeter while a reference signal is played back. This reference voltage value can be used for future re-calibrations in order to avoid the use of the HATS, which is an expensive measurement device.

Figure 19: Level calibration measurement inside Empa's semi-anechoic laboratory with a HATS, the considered headphones and a reference loudspeaker at sufficiently large distance

7 CONCLUDING REMARKS

The implemented VR demonstrator already shows a high level of maturity in artificial image and sound generation for train pass-by events from a pedestrian perspective. Due to the physics-based modelling of the sound generation and propagation phenomena, a high degree of flexibility is achieved, which is able to simulate a large number of possible scenarios.

In the subsequent project phase, refinements will be made to the signal synthesis and propagation filtering, and the catalogue of scenarios and noise protection measures that can be modelled will be expanded.

8 REFERENCES

Böhm J., Heutschi K., Pieren R., Martner O., Isert N., Lutzenberger S., … Ulianov C. (2018). Review on the results of the requirement analysis and assessment. Technical report, DESTINATE project H2020-S2RJU-OC-2015-01-2, Deliverable D2.1, Report to the EC.

Bongini, E. & Bonnet, E. (2009). Railway noise sources definition within the scope of passby sound synthesis. In: *Proceedings of Euronoise 2009, 8th European Conference on Noise Control*, Edinburgh, United Kingdom.

Dittrich M. & Zhang X. (2006). The Harmonoise/IMAGINE model for traction noise of powered railway vehicles. *Journal of Sound and Vibration, 293*, p. 986–994.

European Commission, (2015). Commission Directive (EU) 2015/996 of 19 May 2015 establishing common noise assessment methods according to Directive 2002/49/EC of the European Parliament and of the Council. *Official Journal of the European Union, 58*.

Garcia Ordiales J.I., de Vos P. H. D., Franklin F., Isert N., Martner O., Komorski P. & Marsilla M., Albero R. (2018). Validation and integration of proposed solutions. Technical report, DESTINATE project H2020-S2RJU-OC-2015-01-2, Deliverable D4.1, Report to the EC.

Geiger P., Kirchner K.-R., Garburg R. & Weissenbach O. (2018). Requirements for Auralisation and Visualisation. Technical report, FINE-1 project, Deliverable D9.1, Report to the EC.

Georgiou F. (2018). Modeling for auralization of urban environments, Ph.D. Thesis, Eindhoven University of Technology, Eindhoven, The Netherlands.

Kleiner M., Dalenbäck B.-I. & Svensson P. (1993). Auralization-an overview. *Journal of the Audio Engineering Society*, 41(11):861–75.

Lindau A. & Brinkmann F. (2012). Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings. *Journal of Audio Engineering Society*, 60(1/2):54-62.

Llopis H. S., Pind F., & Jeong C. H. (2020). Development of an auditory virtual reality system based on pre-computed B-format impulse responses for building design evaluation. *Building and Environment*, *169*, 106553.

Pierce A.D. (1974). Diffraction of sound around corners and over wide barriers. *Journal of the Acoustical Society of America*, 55, 941.

Pieren R., Bütler T. & Heutschi K., (2016). Auralization of Accelerating Passenger Cars Using Spectral Modeling Synthesis. *Applied Sciences, 6, 5*, pp. 1-27.

Pieren R., Heutschi K., Aalmoes R. & Simons D. (2017). *Evaluation of Auralization and Visualization Systems for Railway Noise Scenes.* Hongkong, InterNoise 2017, International Congress and Exposition on Noise Control Engineering.

Pieren R., Lauper D. & Heutschi, K. (2018a). *Demonstrator for rail vehicle pass-by events.* Crete, Greece, EuroNoise 2018.

Pieren R., Heutschi K., Lauper D. Zon, R. Aalmoes, R. & Lourens, J. (2018b). Demonstration of railway noise auralisation and visualization. Technical report, DESTINATE project H2020-S2RJU-OC-2015-01-2, Deliverable D3.2, Report to the EC.

Pulkki, V. (1997). Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45 (6), pp. 456-466.

Savioja, L. & Svensson, U. (2005). Overview of geometrical room acoustic modeling techniques. *Journal of the Acoustical Society of America*, 138(2):708–30.

Thompson, D.J., Hemsworth, B. & Vincent, N. (1996a). Experimental validation of the TWINS prediction program for rolling noise, part 1: description of the model and method. *Journal of Sound and Vibration*, 193(1), 123-135.

Thompson, D.J., Fodiman, P. & Mahé, H. (1996b). Experimental validation of the TWINS prediction program for rolling noise, part 2: results. *Journal of Sound and Vibration*, 193(1), 137-147.

Thompson D. (2009). Railway noise and vibration: mechanisms, modelling and means of control, Oxford, UK, Elsevier Science.

Thron T. & Hecht M. (2010). The sonRAIL emission model for railway noise in Switzerland. *Acta Acustica united with Acustica, 96*, p. 873 – 883.

Vorländer M. (2020). Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality, Second Edition (RWTHedition), Springer, Berlin.

Vos P. H. D., Andringa T. & Hagen M. V. (2021). A Soundscape Approach to Assess and Predict Passenger Satisfaction. *Notes on Numerical Fluid Mechanics and Multidisciplinary Design* 150 (pp. 150-157).

Zhang, X., Thompson, D., Quaranta, E. & Squicciarini, G. (2019). An engineering model for the prediction of the sound radiation from a railway track. *Journal of Sound and Vibration* 461, 114921