Driving simulator using virtual reality tools combining sound, vision, vibration, and motion

Camilo Gil-Carvajal, Juan; Soo Jo, Eun; Chul Park, Dong; Song, Wookeun; Jeong, Cheol Ho

Published in:
Applied Acoustics

Link to article, DOI:
10.1016/j.apacoust.2024.110137

Publication date:
2024

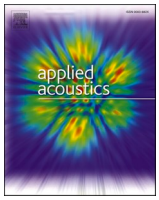
Document Version
Publisher's PDF, also known as Version of record

Applied Acoustics 224 (2024) 110137

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust



Driving simulator using virtual reality tools combining sound, vision, vibration, and motion

Juan Camilo Gil-Carvajal [a,b], Eun Soo Jo [c], Dong Chul Park [c], Wookeun Song [d], Cheol-Ho Jeong [a,*]

[a] Acoustic Technology, Department of Electrical and Photonics Engineering, Technical University of Denmark, Ørsteds Plads, Kgs. Lyngby 2800, Denmark
[b] Hearing Systems, Department of Health Technology, Technical University of Denmark, Ørsteds Plads, Kgs. Lyngby 2800, Denmark
[c] Hyundai Motor Group, 150, Hyundaiyeonguso-ro, Namyang-eup, Hwaseong-si, Gyeonggi-do, 18280, South Korea
[d] Hottinger Brüel & Kjær A/S, Teknikerbyen 26, DK-2830 Virum, Denmark

ARTICLE INFO

Keywords:
VR-based automotive simulator
Multi-modal perception
Higher order ambisonics
Binaural reproduction
Matrix inversion

ABSTRACT

Vehicle driving simulators enable the recreation of real driving experiences by simultaneously reproducing multimodal information. Here, we performed the perceptual evaluation of a virtual reality (VR)-based driving simulator in three experiments with experienced drivers with a special focus on comparing headphone-based and loudspeaker-based sound reproduction. First, we assessed how sound source localizations vary with three sound reproduction methods: Higher order ambisonics (HOA), the matrix inversion method, and binaural recordings played back over headphones without incorporating head tracking. The reproduction methods employing loudspeakers showed better performance in terms of perceived horizontal angle and distance than the reproduction over headphones without head tracking. Second, we assessed how the multimodality affected the perceived spatial immersion and powerfulness of the driving simulator. Visual stimuli were presented together with driving noise through either loudspeakers or headphones, and we controlled vibration and motion stimuli: vibration only, motion only, vibration and motion. While the inclusion of motion and vibration led to significantly higher ratings for spatial immersion and powerfulness, the choice of sound reproduction method did not have a significant impact on the ratings. Third, adding motion to the reproduced scenarios had a greater influence on immersion and powerfulness ratings than adding vibration alone. The results highlight the importance of carefully considering the multimodal information to optimize the driving simulator.

1. Introduction

Driving simulators intend to reproduce the real-world driving experience under controlled conditions convincingly. The implementation of such simulators has risen rapidly during the last decades as they provide several benefits over real vehicles. For instance, driving simulators allow the evaluation of driving skills and behavior under low-risk conditions [1,2], which is particularly relevant for the training [3] and the rehabilitation of physically and cognitively impaired drivers [4,5]. Driving simulators also enable the assessment of vehicle features during early development phases and offer great flexibility for the evaluation of the vehicle's performance under different scenarios [6]. However, despite their convenience, driving simulators have also been criticized for their lack of fidelity [2], as they fail to replicate the real driving experience accurately [7]. It is thus important to identify factors that affect perception as well as to evaluate the reproduction configurations that provide the highest perceptual validity for ensuring a realistic driving experience.

In this study, we have two main research objectives. First, we want to understand how headphone-based sound reproduction performs compared to loudspeaker-based systems in localizing and perceiving the sound event distance. Second, we want to understand how additional modalities, particularly vibration and motion, on top of audio-visual stimuli impact the immersion and powerfulness of the vehicle tested. In practice, a headphone-based sound reproduction is much more economical and easier to control compared to the loudspeaker-based systems.

Driving simulators have challenges in reproducing the multimodal stimuli representing, i.e., auditory, visual and motion. The integration of virtual reality (VR) systems with driving simulators has contributed to

* Corresponding author.
E-mail address: chje@dtu.dk (C.-H. Jeong).

https://doi.org/10.1016/j.apacoust.2024.110137
Received 30 November 2023; Received in revised form 14 June 2024; Accepted 16 June 2024
Available online 2 July 2024

the development of increasingly more realistic virtual driving environments. For example, the inclusion of head-mounted displays (HMD) for reproducing realistic visual scenes increases the realism compared to traditional displays [8], despite discomfort due to heavy HMDs [9]. The quality of the visual information is also important for the realism of the reproduced scenes. While a fast frame rate enables the correct reproduction of changes of the vehicle's speed [8], a large visual field of view induces a stronger feeling of being in the virtual environment [10]. In driving simulators, the latter could be partly due to the added peripheral vision, which is effectively exploited by experienced drivers [11]. However, high-quality visual information is not sufficient for a high-fidelity reproduction, since a lack of correspondence between the visual and vestibular cues about the perceived motion can lead to simulator sickness [12]. Therefore, VR-based driving simulators should be supplemented by coherent multimodal information, such that the driving experience is highly realistic while minimizing the potential discomfort associated with the simulator.

Sound source localization, for instance, enables drivers to maintain awareness of their surroundings. Humans rely on interaural time differences (ITDs) and intensity level differences (ILDs) between their ears to determine the azimuthal position of sound sources [13]. These binaural cues are also important for perceiving the auditory distance of nearby sources along with the low frequency content [14]. While azimuthal localization tends to be more precise for sources presented in front [13], auditory distance perception is more accurately perceived for lateral sound sources [15]. Furthermore, studies have shown that auditory distance perception is influenced by factors such as sound intensity, direct-to-reverberant ratio, and the stimulus familiarity [14]. Generally, louder and less reverberant sounds are perceived as closer, while increased familiarity with a stimulus can enhance the accuracy of distance judgements.

As a sound source reaches the eardrum, it is influenced by the listener's head, pinnae, and torso. Such position-dependent filtering is described by the head-related transfer functions (HRTF). Virtual sound environments can be recreated through headphones by accounting for HRTFs [16]. This can be achieved, for instance, through binaural recordings using Head and Torso simulators (HATS). However, any divergence between the listeners' anatomical characteristics and the physical attributes of the HATS would impact the accuracy of the sound reproduction. Moreover, the lack of high frequency individualized spectral information can result in front-back confusions [17], in which the perceived azimuthal position of a sound source typically exhibits mirror symmetry relative to the interaural axis [13].

Head turning provides an ITD change which is important for resolving front-back confusions of real sources [17,18]. However, when providing static synthesis through binaural recordings, an authentic listening experience is achieved when the listener sits still, which could result in front-back confusions. This could be avoided by using head-trackers to update the auditory image according to the listener's head movement, but this approach could introduce artifacts by continuously adjusting the signal processing [19].

Alternatively, virtual sound environments can be reproduced with loudspeaker-based methods such as higher order ambisonics (HOA; [20] or the matrix inversion method [21,22]), which have been proved successful in reproducing realistic virtual sound environments. Matrix inversion methods involve utilizing recordings from a set of microphones to calculate the inverse of a matrix representing the system's response, thereby optimizing the sound reproduction to closely match the captured recordings at specific microphone positions [22]. HOA relies on decomposing the sound field into spherical harmonics, resulting in homogeneous sound reproduction locally at the center of the loudspeaker array [20]. Other well-known sound reproduction methods are vector-based amplitude panning (VBAP; [23] and wave field synthesis (WFS; [24]). While VBAP effectively recreates specific elements of the original sound field, such as interaural differences, WFS attempts to reproduce the sound field over a large listening area but requires many loudspeakers as its accuracy heavily depends on their spacing. Given the varying characteristics of the methods, the choice of sound reproduction technique is therefore expected to influence sound source localization performance in virtual environments.

The localization of sound sources helps drivers segregate speech signals from the vehicle noise, enabling communication inside the vehicle and react more quickly to warning sound. Sound localization also helps drivers keep track of the environment and avoid potential and immediate threads of collision. Indeed, auditory alarms can be designed for vehicle warning systems to alert to danger from non-frontal positions for which visual alerts would be less effective [25]. Catchpole et al. [26], reported poorer localization performance for a simple warning signal compared to a more complex warning sound, which was attributed to the additional localization cues provided by the richer frequency content of the complex sound. Moreover, using a full chassis driving simulator, Achtemeier et al. [25] found azimuthal sound localization to be more accurate when warning signals were delivered from lateral positions compared to front and back positions. This contrasts with traditional localization experiments in which the azimuthal localization has been found to be more accurate for frontal source positions. Therefore, the localization of sound sources within the vehicle in a VR-based driving simulator could be influenced by the sound source characteristics and source.

Realistic virtual environments often induce a strong sense of being present and part of the virtually generated world [27]. This subjective experience is also known as spatial immersion [28], and here for simplicity will be referred to as immersion. Besides the strong immersive experience provided by HMDs in virtual environments [9,29], immersion is influenced by other factors, such as the coherence of the multimodal information, perceived sensory mismatch and the perception of self-motion [27]. In driving simulators, the reproduced multimodal information can also add up positively to improve the quality impression of the vehicle. For instance, powerfulness, is a desired perceptual attribute in vehicle sound design, which has been suggested to increase when the reproduced vehicle noise is accompanied by seat-floor vibration [30]. Furthermore, hearing vehicle noise while seeing photographs of the vehicle exterior can change the impression of powerfulness compared to the obtained with the noise reproduction alone [31]. Thus, in this study we varied the configuration of the reproduced multimodal information to influence the degree of immersion and powerfulness experienced in a driving simulator, and hence, it would be possible to determine the major factors that contribute the most to these perceptual attributes.

The purpose of this study was to conduct the perceptual evaluation of a VR-based driving simulator in terms of perceived sound localization, immersion, and powerfulness. Experiment 1 evaluated sound source localization in the vehicle's interior in terms of perceived horizontal angle and distance. Three sound reproduction techniques were evaluated. The assessment of sound localization was conducted for two types of sources (a speech excerpt and a warning signal), five source positions and two operating conditions (idling and constant speed at 100 km/h). Experiment 2 investigated the multimodal reproduction configurations that provide the highest immersion and powerfulness ratings of the driving simulator. For this, the recorded driving noise was reproduced through loudspeakers or headphones, whereas the scenes were reproduced with or without vibration and motion. Additionally, the effect of motion and vibration was further examined in Experiment 3 to determine which factor more substantially affected the perceived immersion and powerfulness.

2. Material and methods

2.1. Subjects

Ten participants took part in Experiment 1, twenty in Experiment 2 and ten in Experiment 3 (five female participants in Experiments 1 and 2

and four in Experiment 3). All participants were experienced drivers with five or more years of car driving experience, and they reported having normal hearing and normal (or corrected-to-normal) vision. Prior to the experiments, all participants provided written consent, and all experiments in this study were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391).

2.2. Reproduction system

2.2.1. Sound

Experiment 1

A warning and a speech signal were used to evaluate sound source localization in Experiment 1. The warning signal consisted of a beeping sound emitted every 0.6 s at a centered frequency of 1 kHz. The speech signal consisted of male speech sentences, which were taken from the Danish version of the hearing in noise test [32]. The speech and warning signals were recorded while being reproduced through a Samsung Galaxy cellphone placed at five positions in a passenger car. The precise source positions in relation to the listener's location were: Front Left (FL, at 300° and 110 cm), Front Center (FC, at 357° and 60 cm), Rear Right (RR, at 150° and 55 cm), Rear Left (RL, at 210° and 55 cm), and Far Rear Left (FRL, at 215° and 128 cm). Three distinct approaches were employed to record the signals from the front passenger's seat of a sports sedan: (1) with the built-in microphones of a Head and Torso Simulator (HATS, Type 4100, HBK, DEnmark), (2) with a 32-microphone spherical array (em32 Eigenmike, mh acoustics, USA), (3) with a circular array consisting of 8 microphones (Type 4188, HBK, Denmark) placed around the HATS' head. For the idling condition, the speech and warning signals were recorded with the engine on. In contrast, for the 100 km/h condition, the signals were recorded with the engine off, but the driving noise was added during post-processing, separately from the initial recording process. Sound was recorded inside the sport sedan with binaural microphones (Type 4101B, HBK, Denmark) and with the 8-micropone array, which were placed in the driver's ears and around the driver's head, respectively.

In Experiment 1, audio recordings from the HATS were played back through headphones (Sennheiser HD-650). The Eigenmike recordings were encoded as 4th order Ambisonics using the regularized filtering approach [33]. They were then reproduced through a complete 64-loudspeaker spherical array (Type LS50, KEF, UK) using the re-encoding principle [34,35]. The 8-microphone array recordings were played through 8 horizonal loudspeakers of the spherical array using the matrix inversion method [21,22]. The loudspeaker signals are generated through convolution between the reference microphone signals and the equalization filters. The characteristics of these filters are determined by the room's frequency response between each loudspeaker and microphone. A regularization threshold of 20 dB was used, which yielded the lowest magnitude error without introducing audible artifacts. The 8 horizonal loudspeakers were located at ear level with an elevation of 0°, at the horizontal angles of 0, 315, 270, 225, 180, 135, 90 and 45° relative to the listener's position. The sound pressure level generated by each loudspeaker was adjusted to obtain the same level at the center of the array. This was done by reproducing logarithmic sweeps from each loudspeaker and measuring their impulse responses at the centre of the array.

Experiment 2

Three different vehicles were assessed: a Sports Sedan (SS), a Sports Utility Vehicle (SUV), and a Luxury Sedan (LS). The vehicle noise added in the simulations was recorded in real vehicles under three autonomous driving operating conditions: constant speed at 30 and 100 km/h and run-up (wide-open throttle, WOT). The vehicle noise was recorded with the binaural microphones and with the circular array of 8 microphones used in Experiment 1. However, the microphone array was positioned around the driver's head instead of the HATs. The noise recordings were then reproduced during the experiment over the headphones or through a circular array of 8 loudspeakers (BM6P, Dynaudio, Denmark) using the matrix inversion method with the same settings as in Experiment 1. The loudspeakers were positioned at ear level at an azimuth angle of 0, 45, 90, 135, 180, 225, 270, and 315°. The distance from the driver's seat (listening position) was 1.8 m. The individual loudspeaker levels were adjusted as in Experiment 1 to obtain the same pressure level at the center of the array.

Experiment 3

The simulated scenarios in Experiment 3 consisted of the autonomous driving conditions with constant speed at 100 km/h and WOT. The frequency and magnitude spectra of the vehicle noise was the same as in Experiment 2. However, unlike Experiment 2, in Experiment 3 the vehicle noise was only reproduced over headphones, and only two vehicles were tested: the Sports Sedan and the Luxury Sedan.

2.2.2. Visual

A head-mounted display (HMD; Vive Pro, HTC, Taiwan) was employed to display the visual scenes, which portrayed a first-person perspective from either the front passenger seat (Experiment 1) or the driver's seat (Experiment 2 and 3), as shown in Figs. 1 and 2. The scenes provided a realistic render of the vehicle's interior as well as the outdoor landscape (a straight, empty highway). The scenes were created using the Unity software platform and were reproduced with at least 60 Hz image updates.

2.2.3. Motion

The motion provided in Experiments 2 and 3 was simulated with a motion platform (iMP6-M250, InnoSimulation, Korea) that allows movement in six degrees of freedom (three translation directions and three rotational directions). The motion platform measures 80 x 80 x 60 cm, upon which a replica of a vehicle's driver's seat is mounted, complete with a steering wheel, pedals, and seat belt, see Fig. 2(A). This platform has a surge displacement of ± 0.25 m, surge acceleration of ± 6 m2, and a pitch displacement of ± 10°. All simulated scenarios involved a self-driving vehicle moving straight ahead. Consequently, the experienced motion includes both the inertia and acceleration resulting from the translation and tilting of the platform.

To control the platform, the Desktop NVH Simulator software (Type 8601, HBK, Denmark) communicates the vehicle speed profiles from Figure A1 to the Unity software platform. This information is synchronized with the vehicle noise and sent in real-time. Within the Unity platform, a motion control algorithm interprets this data, computing precise vehicle controls for the motion platform, including surge, pitch, and sway. The values, synchronized with the visual information presented through the HMD, are then transmitted to the motion platform computer via an Ethernet cable, enabling real-time monitoring of input data. Subsequently, this data is routed to a motion controller responsible for determining motion parameters, safety controls, and kinematics computation. Once computed, these parameters are sent to the AC servo motor controller, which drives the 6-axis actuators of the motion platform.

The speed profiles for the 30 and 100 km/h operating conditions were the same for all vehicles as shown in Fig. A1 For these conditions, the acceleration and deceleration periods had a similar duration of about 1.5–2 s, but the gradients were steeper for the 100 km/h condition. The constant speed interval had a duration of about 10–11 s for these operating conditions. For WOT, the speed profiles varied substantially between vehicles to reflect the differences in performance of real-world vehicles, especially in acceleration and deceleration periods, as well as in top speed.

In Experiment 2, the motion profiles for high and low intensity across operating conditions and vehicles are shown in Fig. A3. For 100 km/h, the motion profile was the same for both vehicles, whereas for WOT each vehicle had a different motion profile. For both operating conditions, the high intensity condition produced more motion than the low intensity condition.

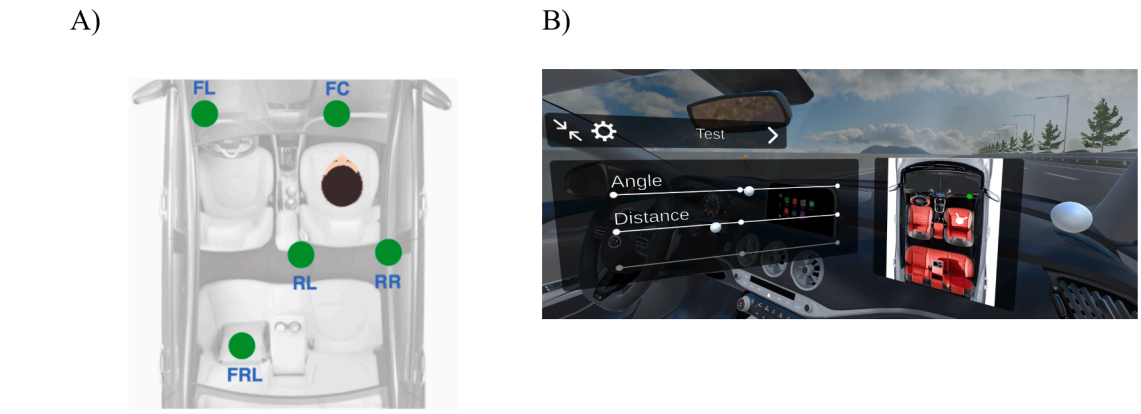


Fig. 1. A) Schematic of the tested source positions. The positions relative to the listener's position (passenger's seat) were: Front Left (FL, at 300° and 110 cm), Front Center (FC), Rear Right (RR), Rear Left (RL,), and Further Rear Left (FRL). B) Rating Interface and view from the subject's perspective. During the experiment this interface was located towards the bottom right corner.

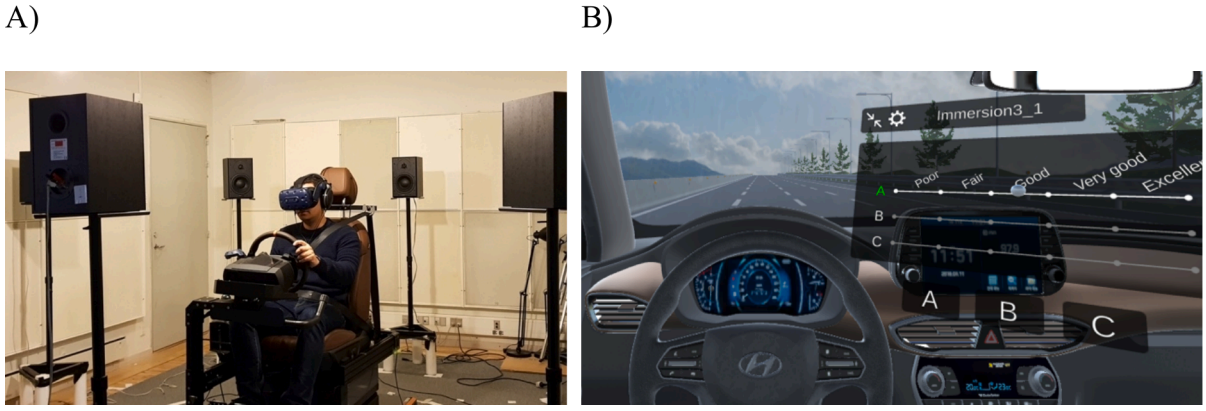


Fig. 2. A) Experimental setup in the standard listening room. B) Rating interface and view from the subject's perspective. During the experiment this interface was located towards the bottom right corner.

2.2.4. Vibration

Vibrations were provided in Experiments 2 and 3 by a set of three vibration-generating flat speakers embedded in the motion platform. The actuator's dimensions are Ø70 mm x 24 mm and can generate vibrations from 10 Hz to 1 kHz. One actuator was positioned under the backrest and two under the driver's seat, resulting in localized shaking or oscillatory movements.

The NVH simulator played the sound signals shown in Figure A2 via the sound card. The output of the sound card was connected to the amplifier of the actuator. The gain of the amplifier was adjusted roughly until the perceived vibration amplitude became similar to the one from the vehicles on the road. This way of adjusting the vibration is not accurate but practically suitable for the utilized actuators since more accurate frequency equalization was not possible due to the limited low-frequency performance and its limited stroke across the frequency range of interest.

The WOT condition displayed slightly more pronounced frequency components at higher frequencies (above 500 Hz), as illustrated by the slightly more noticeable red traces in Fig. A2. In contrast, for the operating condition of 30 km/h, both the magnitude and frequency spectrum were notably reduced, with frequency components only extending up to 500 Hz.

As mentioned above, in Experiment 3, the frequency and magnitude spectra of the vehicle noise, and consequently the vibration signal, was the same as in Experiment 2. The difference between high and low intensity was then an increase of the vibration level of about 5 dB for the high intensity condition.

2.3. Experimental procedure

2.3.1. Experiment 1

Experiment 1 took place at the Technical University of Denmark in the Audio-Visual Immersion Lab. This is an anechoic chamber according to ISO 26101 [36] with dimensions of 7 × 8 × 6 m. The chamber contains a spherical array of 64 loudspeakers, with a radius of 2.4 m from the center of the listener's head. The loudspeakers are mounted on seven rings at a vertical angle of 0°, ±28°, ±56° and ± 80° with respect to the listener's position, with 24, 12, 6, and 2 loudspeakers evenly distributed on each ring, respectively. The experiment comprised three testing sessions, each lasting 15–20 min. These three sessions encompassed sound reproduction using 64 loudspeakers, 8 loudspeakers, or headphones. In each session, there were trials for two operating conditions (idling and constant speed at 100 km/h), two signal types (warning or speech), and five source positions. The sequence of trials was counterbalanced within each session, and the order of the sessions was also counterbalanced across the subjects.

In every trial, subjects provided localization assessments concerning perceived horizontal angle and distance. To accomplish this, subjects were presented with a user interface via the HMD. They used a VR controller to provide the ratings while experiencing each condition for about 18 s. Their task was to position a virtual grey sphere at the perceived sound source location by adjusting the sliders corresponding to horizontal angle and distance, as shown in Fig. 1. The sphere was positioned at eye level and its starting location was randomized across trials. A second visible reference consisted of a diagram of the vehicle

which contained a green dot representing the position of the sphere. The position of the sphere and the green dot in the diagram varied as the sliders for angle and distance were adjusted. The diagram with the green dot served as a reference when the perceived sound source location was outside the visual field. Once the sphere was positioned at the perceived sound source location, the subjects could press the button next to start the following trial. This was repeated until the conclusion of the experimental session, following which subjects were instructed to have a 5-minute break.

2.3.2. Experiment 2

The experiment was conducted in a standard listening room according to [37] located at the Technical University of Denmark. This is a controlled room with a reverberation time (T30) of around 0.4 s from the 125 Hz to 4 kHz octave band, with dimensions 7.52 × 4.75 × 2.8 m and a volume of 100 m^3. Prior to the experiment, all subjects were given written instructions in English. They were asked to imagine themselves in the driver's seat of a self-driving vehicle, moving in a straight line at constant speed or at full acceleration. Their task was to rate the quality of the reproduced scenes in terms of immersion and powerfulness. Immersion was defined as the degree to which the reproduction makes the subject feel as inside a real car in the presented scenario. Powerfulness was defined as the perceived robustness of the vehicle's performance in the reproduced scene (i.e., solid engine, effortless performance, soundness, etc.).

The experiment was divided into four experimental testing sessions, each lasting 10–12 min. The subjects were asked to take breaks of 2–5 min between sessions. A different reproduction configuration was assigned to each of the four experimental sessions: with motion and vibration *on* and sound reproduced over headphones (MV_{on} + H), with motion and vibration *off* and sound reproduced over headphones (MV_{off} + H), with motion and vibration *on* and sound reproduced through loudspeakers (MV_{on} + L), and with motion and vibration *off* and sound reproduced through loudspeakers (MV_{off} + L). Immersion ratings were obtained for the three operating conditions, whereas powerfulness ratings were only tested for WOT. An experimental session then consisted of eight trials corresponding to the operating conditions (all three for immersion and only WOT for powerfulness ratings) and their repetition. The order of trials within each session was counterbalanced, and the sequence order was also counterbalanced across subjects. In a trial session, one attribute (either immersion or powerfulness) was rated for all three vehicles, and subsequently switched to the other attribute. The order of the presented vehicles was randomized.

For each of the vehicles, the ratings were provided with a user interface presented through the HMD, as shown in Fig. 2. The subjects used a VR controller to provide the ratings after experiencing each vehicle for about 20 s. The ratings were given on a continuous scale ranging from 0 to 100. Labels were provided to describe the scale categories, which were poor (0–20), fair (21–40), good (41–60), very good (61–80), and excellent (81–100). Once a rating was given, the subject could press the next button to move to the next trial.

2.3.3. Experiment 3

The test was conducted in a room also located at DTU with dimensions 6.3 × 6.8 × 3.1 m and with a reverberation time (T30) of 0.52 s averaged over the octave bands from 125 Hz to 4 kHz. The experiment consisted of three experimental testing sessions with a duration of 10–12 min each. The three experimental sessions were: with motion and vibration (MV), motion only (M) and vibration only (V). Each session contained the trials corresponding to two operating conditions (100 km/h and WOT), two vehicles (SS and LS), and two intensities (high or low). Immersion ratings were obtained for the two operating conditions and powerfulness ratings only for WOT. The trials were repeated two times and the sequence order was counterbalanced across subjects.

2.3.4. Cybersickness

In all three experiments, cybersickness was monitored between sessions, as it can negatively impact the subjects' performance in VR experiments [38]. Subjects were asked to report any degree of discomfort or sickness before the experiment and after each experimental session. Cybersickness was judged on a scale from 0 to 10, minimum to maximum values, respectively. Only one participant reported some degree of discomfort during Experiment 1 and two subjects during Experiment 2, with values of 3 or lower. In Experiment 3, one participant rated cybersickness as five, and another as four on the 10-point scale. The subjects were asked to take a longer break and reminded that they could withdraw their participation if they wished. All subjects who experienced cybersickness decided to complete the experiment after taking a break.

2.3.5. Data analysis

To assess sound source localization, errors were computed as the difference between the presented and perceived source locations for both horizontal angle and distance ratings. Front-back confusions were manually corrected before conducting statistical analyses, while distance errors were normalized to account for differences in source position. Subsequently, we conducted analyses of variance (ANOVA) to assess significant effects. These analyses followed the fitting of linear mixed-effects models [39] to the error ratings. The fixed factors were *Signal Type* (2 levels: warning, speech), *Sound Reproduction Method* (3 levels: 64 loudspeakers [64L], 8 loudspeakers [8L], Headphones [H]), *Operating Condition* (*OC*, 2 levels: Idling, 100 km/h) and *Source Position* (5 levels: Front Left [FL], Front Center [FC], Rear Right [RR], Rear Left [RL] and Further Rear Left [FRL]).

Experiments 2 and 3 evaluated immersion and powerfulness ratings using similar linear mixed-effects models. The fixed factors for Experiment 2 were *operating condition* (*OC,* 3 levels: 30kph, 100 kph, WOT), *Vehicle* (3 levels), *Motion* and *Vibration* (*MV*, 2 levels: MV_{on}, MV_{off}) and *Sound Reproduction* (2 levels: Loudspeaker [L], Headphones [H]). operating condition (*OC,* 3 levels: 30kph, 100 kph, WOT), *Vehicle* (3 levels), Motion and Vibration (*MV*, 2 levels: MV_{on}, MV_{off}) and *Sound Reproduction* (2 levels: Loudspeaker [L], Headphones [H]). For Experiment 3, the fixed factors corresponded to *Reproduction Configuration* (RC, 3 levels: Motion only, Vibration only, Motion and Vibration), *Intensity* (2 levels: High, low), *Operating Condition* (*OC,* 2 levels: 100 km/h, WOT) and *Vehicle* (2 levels: SS, LS). *Subject* was taken as a random factor in all three Experiments. Post hoc comparisons were performed with Tukey tests, and a significance level of 0.05 was used in all analyses. The overview of statistical analyses is shown in Tables A1-A6 of the supplemental material.

3. Results and discussion

3.1. Experiment 1: Sound source localization varying sound reproduction techniques

This experiment aimed to investigate in-vehicle sound source localization for a VR-based driving simulator in terms of perceived horizontal angle and distance. Three sound reproduction techniques were compared: HOA (4th order) reproduced through a spherical array of 64 loudspeakers, matrix inversion method reproduced through a circular array of 8 loudspeakers, and binaural recordings reproduced over headphones. Two types of sound sources (warning and speech) were studied, which were delivered from five positions in two driving conditions (idling and constant speed at 100 km/h).

Fig. 3 and Fig. 4 show the absolute localization error for perceived horizontal angle across the five source positions for the warning and speech signals, respectively. The values are shown for idling (top panel) and 100 km/h (bottom panel) across the three sound reproduction methods. Front-back confusion errors were manually corrected before computing the absolute angle errors. These confusions were frequent for

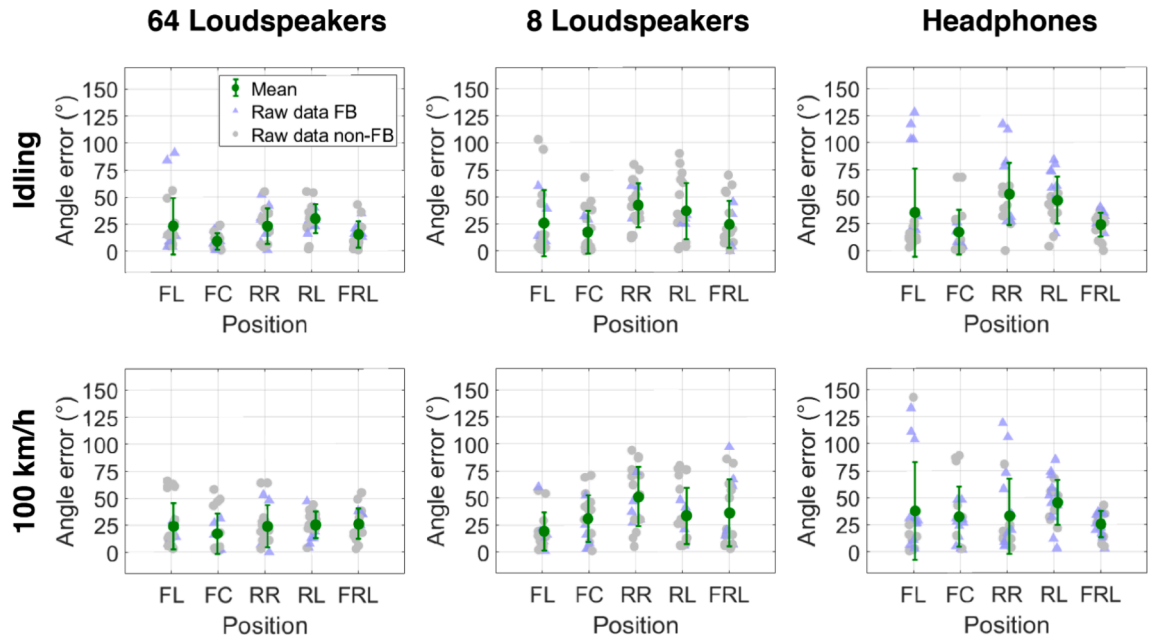


Fig. 3. Mean absolute horizontal angle error for the warning signal. The results are presented for idling (top panel) and 100 km/h (bottom panel) across the three tested sound reproduction methods. The purple triangles and grey dots denote individual errors where a front-back confusion occurred (corrected) or did not occur, respectively. The error bars show plus and minus one standard deviation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

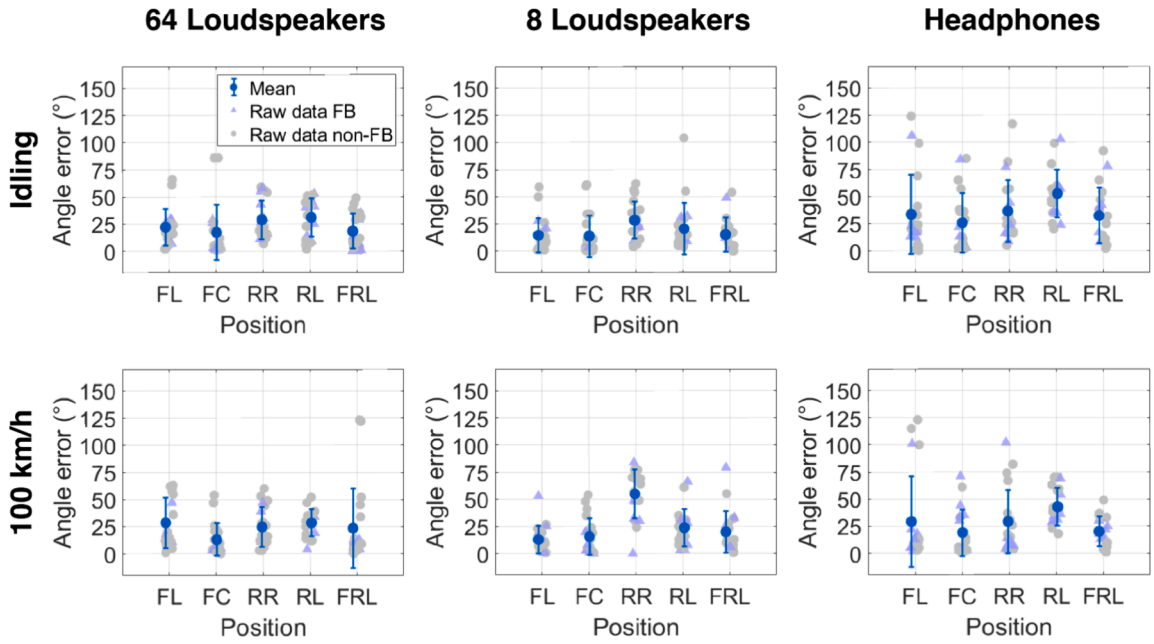


Fig. 4. Mean absolute horizontal angle error for the speech signal. The results are presented for idling (top panel) and 100 km/h (bottom panel) across the three tested sound reproduction methods. The purple triangles and grey dots denote individual errors where a front-back confusion occurred (corrected) or did not occur, respectively. The error bars show plus and minus one standard deviation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

all three sound reproduction methods, appearing on average in one-third of the trials for the three methods (64L, Mean = 29.5 %; 8L, Mean = 29 %; H, Mean = 32.7 %).

Absolute angle errors were influenced by the sound reproduction method, signal type, position, and operating conditions. This was confirmed by a 4-way ANOVA which showed significant main effects of *Sound* [$F(2, 1131) = 23.5$, $p < 0.0001$], *Signal* [$F(1, 1131) = 7.3$, $p < 0.0001$], and *Position* [$F(4, 1131) = 24.5$, $p < 0.0001$], as well as significant two-way interactions between *Sound* and all other predictors. The three-way interactions between *Sound* × *OC* × *Position* and between *Signal* × *OC* × *Position* were only marginally significant. An overview of the results of the ANOVA is presented in Table A1 in the Appendix.

Overall, absolute angle errors were lowest for the 64L method (Mean = 22.6°, SD = 19.5), followed by 8L (Mean = 26.8°, SD = 24.3) and were largest for the H method (Mean = 33.5°, SD = 29.1). This was supported by post hoc analysis, revealing significant pairwise differences between

all three methods ($p < 0.05$). For each sound reproduction method, the absolute angle errors did not vary with signal type, except for the 8L method, which exhibited localization errors approximately10° larger for the warning signal than for the speech signal ($p < 0.0001$). Moreover, while absolute angle errors were similar for the 64L and H methods across reproduction conditions, they were on average 6° larger for the 8L method under the 100 km/h operating condition compared to the idling condition ($p = 0.0102$). The results suggest that sound source localization in terms of perceived angle was more accurate for the 64L sound reproduction method and less accurate for the H method. Moreover, the 8L method appeared to be the most susceptible to changes in signal type and acoustic environment.

Mean angle errors also varied across source positions. The frontal source positions FC and FRL yielded the lowest errors with 19° and 23°, followed by FL with 25°, and the rear positions RR and RL with 36° and 35°, respectively. These findings were supported by post hoc analysis, which revealed significant differences for the pairwise comparisons involving frontal source positions with rear source positions ($p \leq 0.023$). This indicates that sound localization in terms of perceived angle was more accurate when the sound was delivered from frontal source positions compared to rear ones. Furthermore, these azimuthal localization errors appear to be larger than those reported in traditional localization experiments (below 10°, [40,41]), which could be attributed to reflections and diffractions inside the vehicle potentially affecting the recorded localization cues.

The angle errors across source positions depended on the sound reproduction method. Positions FL, FC and RL exhibited larger angle errors (13°, 7° and 18°, respectively) with the H method compared to the 64L and 8L methods ($p \leq 0.0001$, except between 8L and H for RL, which was not significant). In contrast, RR yielded, on average, 13° larger errors with the 8L method than with the other methods, and this difference was statistically significant when compared to the 64L method ($p \leq 0.0001$). Notably, for FRL all three sound reproduction methods performed similarly ($p \geq 0.395$). The results suggest that the 64L method outperformed the other two methods in perceived horizontal angle at all source positions.

Fig. 5 and Fig. 6 illustrate the relative localization error for perceived horizontal distance across the five source positions for the warning and speech signals, respectively. The values are shown for idling (top panel) and 100 km/h (bottom panel) across the three sound reproduction methods. Distance errors were normalized to account for the substantial differences in source position that could influence the error. The obtained relative distance errors varied with sound reproduction method, operating condition, and source positions. This was evidenced in the results of the ANOVA, which showed significant main effects of *Sound* [$F(2, 1131) = 31$, $p < 0.0001$], *Signal* [$F(1, 1131) = 9.2$, $p = 0.0025$] and *Position* [$F(4, 1131) = 7.9$, $p < 0.0001$]. Two-way interactions involving *Position* were also significant, as well as the two-way interaction of *Sound* × *OC* and the three-way interaction *Sound* × *OC* × *Position*. An overview of the results of the ANOVA is presented in Table A2 in the supplemental material.

Relative distance errors were most prominent with the H sound reproduction method (Mean = 0.5, SD = 0.27), followed by the 8L method (Mean = 0.4, SD = 0.33), and the 64L with the smallest errors (Mean = 0.3, SD = 0.24). The post hoc analysis showed significant differences when comparing H with the 8L and 64L methods ($p < 0.0001$), but not when comparing the methods with loudspeakers ($p = 0.0397$). The sound reproduction methods had similar performance across operating conditions, except for the 8L method which exhibited slightly larger distance errors for the 100 k/m compared to the idling condition ($p = 0.0403$). Moreover, the distance errors were slightly larger for the warning (Mean = 0.41, SD = 0.30) than for the speech signal (Mean = 0.37, SD = 0.27), suggesting that perceived sound source distance was slightly more accurate for speech signals.

Relative distance errors also depended on the source position. Sounds delivered from position FC had the lowest mean errors with 0.3, followed by positions RR, RL and FRL with 0.4 and position FL with the largest mean error of 0.5. The post hoc analysis showed significant differences for the pairwise comparisons between FL and all other positions ($p \leq 0.0354$), except RL ($p = 0.2315$), and between FC and RL ($p = 0.0051$). Furthermore, for the frontal positions, the H sound reproduction method had significantly larger mean distance errors than the methods including loudspeakers ($p \leq 0.0005$). In contrast, for the rear position RR the method 8L had the largest errors compared to the other two methods ($p \leq 0.001$), whereas for RL all methods performed similarly ($p \geq 0.1557$).

3.2. *Experiment 2: Combined effect of motion, vibration and sound reproduction method on perceived immersion and powerfulness*

The purpose of this experiment was to evaluate if adding motion and vibration has a significant effect on the perceived immersion and

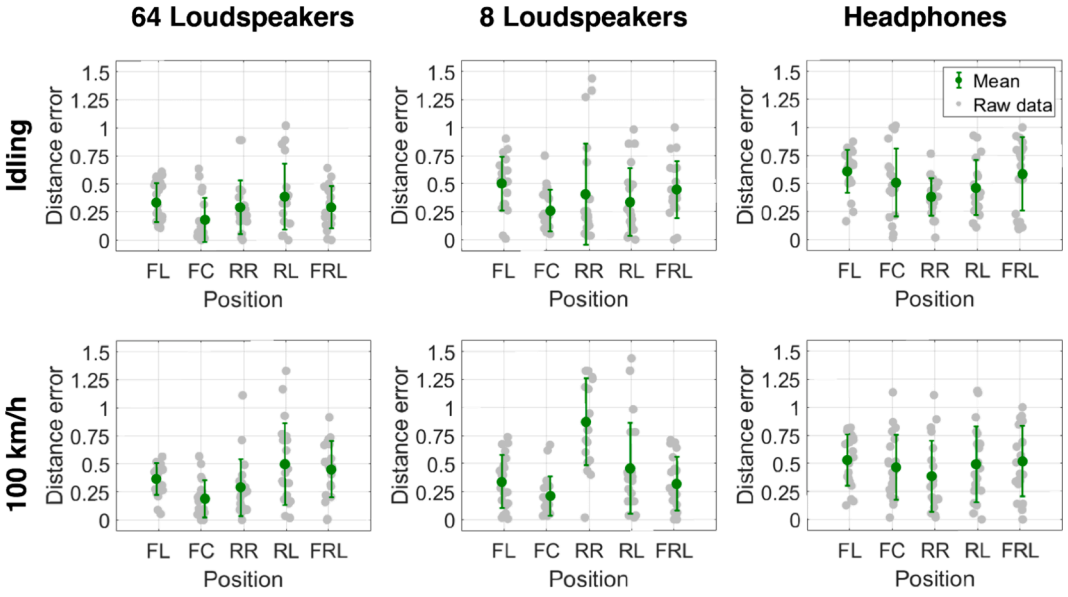


Fig. 5. Mean absolute values of the relative distance error for the warning signal. The results are presented for Idling (top panel) and 100 km/h (bottom panel) across the three tested sound reproduction methods. The gray dots show the individual errors. The error bars show plus and minus one standard deviation.

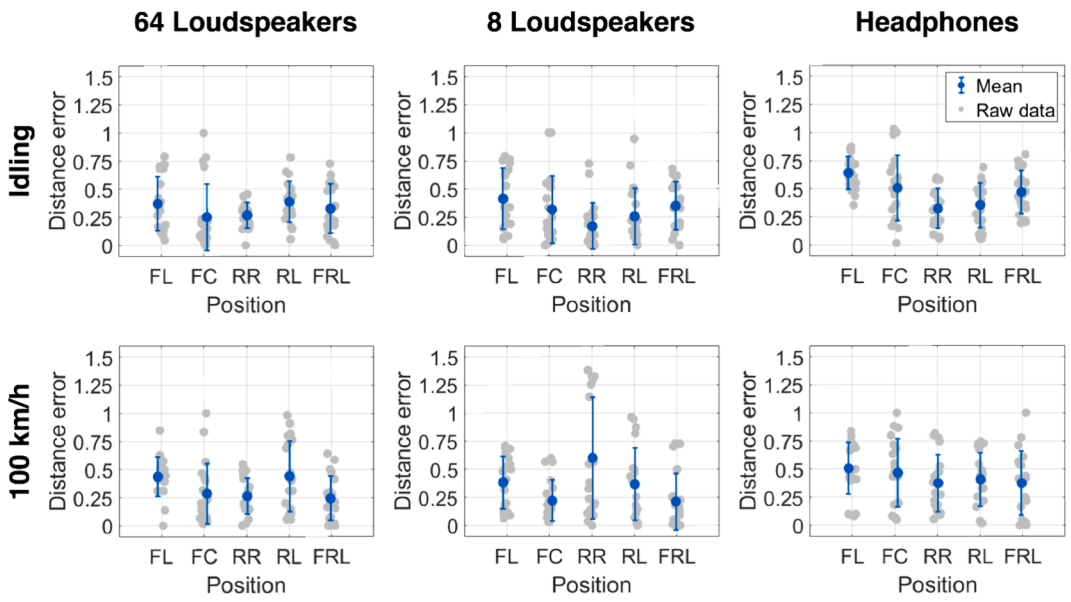


Fig. 6. Mean absolute values of the relative distance error for the speech signal. The results are presented for Idling (top panel) and 100 km/h (bottom panel) across the three tested sound reproduction methods. The gray dots show the individual errors. The error bars show plus and minus one standard deviation.

powerfulness of a VR-based driving simulator. Furthermore, it was assessed whether the sound reproduction method substantially influences these two perceptual attributes. The effect of motion and vibration was tested in two conditions (on/off), whereas the sound reproduction was evaluated for the circular array of 8 loudspeakers using the matrix inversion method and the headphone reproduction using binaural recordings. Three driving conditions were assessed (constant speed at 30 and 100 km/h and run-up) for three vehicles and two repetitions of each scenario.

Fig. 7 shows the immersion ratings across operating conditions (30 km/h [blue], 100 km/h [green] and WOT [red]) and tested vehicles (SS, SUV and LS) for each reproduction configuration. Immersion ratings were generally higher for reproduction configurations containing vibration and motion (MV_{on} + L and MV_{on} + H) than for those without (MV_{off} + L and MV_{off} + H). This indicates that the addition of motion and vibration enhanced the immersion of the driving experience. In contrast, immersion ratings did not appear to be substantially affected by the sound reproduction method, as reflected by the overall similar immersion ratings between the conditions MV_{on} + L and MV_{on} + H, and between MV_{off} + L and MV_{off} + H.

The 4-way ANOVA performed on the immersion ratings revealed significant main effects for *MV* [$F(1, 1385) = 109.2$, $p < 0.0001$], *OC* [$F(2, 1385) = 97.6$, $p < 0.0001$] and *Vehicle* [$F(2, 1385) = 12.1$, $p < 0.0001$], as well as for the two-way interactions *MV* × *OC* [$F(2, 1385) = 7$, $p = 0.0009$] and *OC* × *Vehicle* [$F(5, 1385) = 6.1$, $p < 0.0001$], whereas the main effect of *Sound Reproduction* was not significant [$F(1, 1385) = 2.2$, $p = 0.1409$].

Moreover, across operating conditions, immersion ratings were generally higher for WOT and lower for 30 km/h. This pattern is consistent with the fact that the speed profiles exhibited more prominent motion and acceleration in the WOT condition, as illustrated in Figure A1. Post hoc comparisons confirmed the significant differences across all operating conditions for each reproduction configuration ($p < 0.0001$), except between WOT and 100 km/h presented with vibration and motion ($p = 0.2245$). Immersion ratings also varied across operating conditions and vehicles, as reflected by the significant interaction of *OC* × *Vehicle*. These differences are likely due to the motion profiles and vibration which were different across vehicles and operating conditions (see Figs. A1 and A2), particularly for WOT, for which the motion profiles and vibration were different across vehicles.

Fig. 8 shows the powerfulness ratings for WOT across tested vehicles (SS, SUV and LS) for each reproduction configuration. Perceived powerfulness ratings were higher when the reproduction configuration contained vibration and motion. The 3-way ANOVA showed significant main effects for *MV* [$F(1, 449) = 40.3$, $p < 0.0001$] and *Vehicle* [$F(2, 449) = 83.1$, $p < 0.0001$], as well as for their two-way interaction [$F(2, 449) = 4.3$, $p = 0.0138$]. In contrast, the main effect of *Sound Reproduction* was not significant [$F(1, 449) = 0.0005$, $p = 0.0981$], indicating that the perceived powerfulness remained consistent regardless of the sound reproduction method.

Powerfulness ratings also varied across reproduction configurations and vehicles, which is consistent with the significant interaction of *MV* × *Vehicle*. As in the case of immersion, the differences in powerfulness ratings across vehicles are likely due to the differences in motion profiles and vibration, although seeing the different vehicles might have also influenced the ratings by varying the expected powerfulness.

3.3. *Experiment 3: Combined and separate effects of motion and vibration on perceived immersion and powerfulness*

The purpose of this experiment was to determine which factor, either motion or vibration, influenced more immersion and powerfulness ratings of a VR-based driving simulator. This was unclear from Experiment 2 since motion and vibration were turned on and off simultaneously. Only the operating conditions of 100 km/h and WOT, which had higher immersion ratings in Experiment 2 were tested here. The reproduction configurations of motion only, vibration only, and motion and vibration were assessed. Two intensities of motion and vibration (high and low) and two repetitions of each condition were evaluated. Since the sound reproduction method did not influence the perceptual attributes in Experiment 2, the vehicle noise was only reproduced over headphones.

Fig. 9 shows the immersion ratings for each tested vehicle (Sports Sedan and Luxury Sedan) for the operating conditions of 100 km/h (green) and WOT (red) and reproduction configurations (with motion only, with vibration only, and with motion and vibration). Filled and empty boxplots represent the conditions reproduced with high or low intensity (of motion and/or vibration), respectively. For 100 km/h, immersion ratings were generally similar between high and low intensity, whereas for WOT immersion ratings were higher when the intensity of vibration and motion was high. The 4-way ANOVA revealed

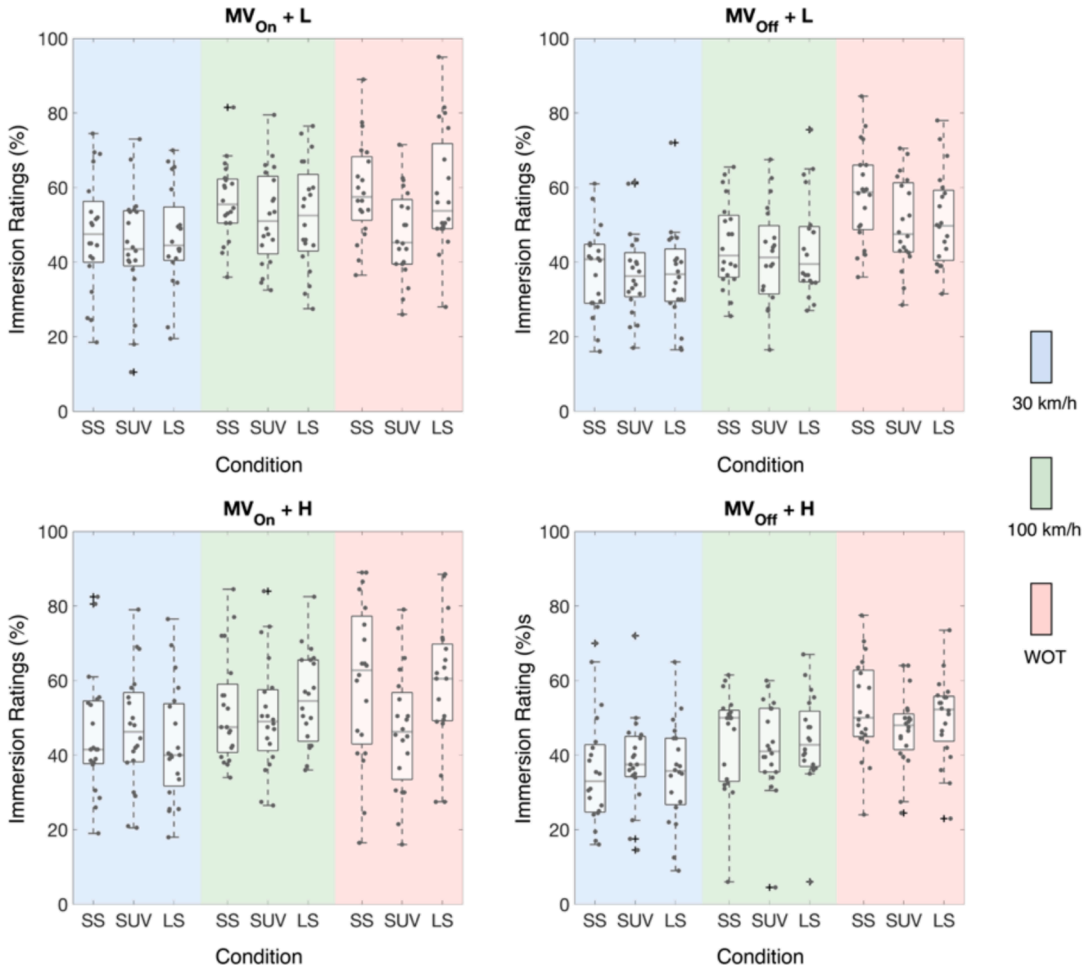


Fig. 7. Immersion ratings across tested vehicles (Sports Sedan [SS], Sports Utility Vehicle [SUV] and Luxury Sedan [LS]) and operating conditions (30 km/h [shaded blue], 100 km/h [green] and WOT [red]). The ratings are shown for each reproduction configuration: with motion and vibration and sound through loudspeakers (MVon + L), with motion and vibration and sound over headphones (MVon + H), without motion and vibration and sound through loudspeakers (MVoff + L) and without motion and vibration and sound over headphones (MVoff + H). Boxes contain the median (horizontal gray line) and delimit the first and third quartiles. The whiskers show the upper and lower 1.5 interquartile range, and crosses depict outliers. The gray dots show the individual ratings. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

significant main effects for *Reproduction Configuration* [$F(2, 447) = 17.9$, $p < 0.0001$], *OC* [$F(1, 447) = 40.9$, $p < 0.0001$], *Intensity* [$F(1, 447) = 21.4$, $p < 0.0001$], and *Vehicle* [$F(1, 447) = 16.2$, $p < 0.0001$], as well as for the two-way interactions for *OC* × *Intensity* [$F(1, 447) = 16.4$, $p < 0.0001$] and *OC* × *Vehicle* [$F(1, 447) = 8.2$, $p = 0.0033$]. Post hoc multiple comparisons showed significant differences between high and low intensity for WOT ($p < 0.0001$) but not for 100 km/h ($p = 0.6817$).

The significant interaction of *OC* × *Vehicle* is reflected in the slight differences in the immersion ratings across operating conditions for the two vehicles. These differences might be related to the different motion and vibration patterns of the vehicles. Moreover, the Post hoc comparisons for the *Reproduction Configurations* showed significant differences between Vibration only and Motion only ($p < 0.0001$) and between Vibration only and Motion and Vibration ($p < 0.0001$), but not between Motion only and Motion and Vibration ($p = 0.9771$). This indicates that the immersion ratings were more influenced by the addition of motion than vibration.

Fig. 10 shows the powerfulness ratings for the operating condition of WOT across reproduction configurations. Filled and empty boxplots represent the conditions reproduced with high or low intensity. In general, powerfulness ratings were slightly lower when only vibration was present and tended to be higher for high than low intensity for all reproduction configurations. Moreover, the vehicle Luxury Sedan had generally higher powerfulness ratings across all operating conditions and for both intensities than the Sports Sedan. The 4-way ANOVA confirmed the statistical significance of these observations, showing significant main effects for *Reproduction Configuration* [$F(2, 219) = 3.1$, $p = 0.0451$], *Intensity* [$F(1, 219) = 41.3$, $p < 0.0001$] and *Vehicle* [$F(1, 219) = 80.7$, $p < 0.0001$]. No significant differences were found for the post hoc comparisons between *Reproduction Configurations*, although the comparison between motion only and vibration only was close to significance ($p = 0.0574$). Thus, as in the case of immersion ratings, adding motion had more influence on the powerfulness ratings than adding vibration.

4. General discussions

4.1. Effect of sound reproduction method on perceived horizontal angle and distance

4.1.1. Perceived horizontal angle

The perceived horizontal angle was most accurate when using the 64L sound reproduction method, followed by the 8L method and the binaural headphone reproduction, which was the least accurate. The reduced accuracy of the headphone reproduction method could be due to the lack of individualized HRTFs recordings, resulting in suboptimal localization cues [42]. Moreover, the lack of head tracking in the headphone reproduction and the positioning of the rating interface in

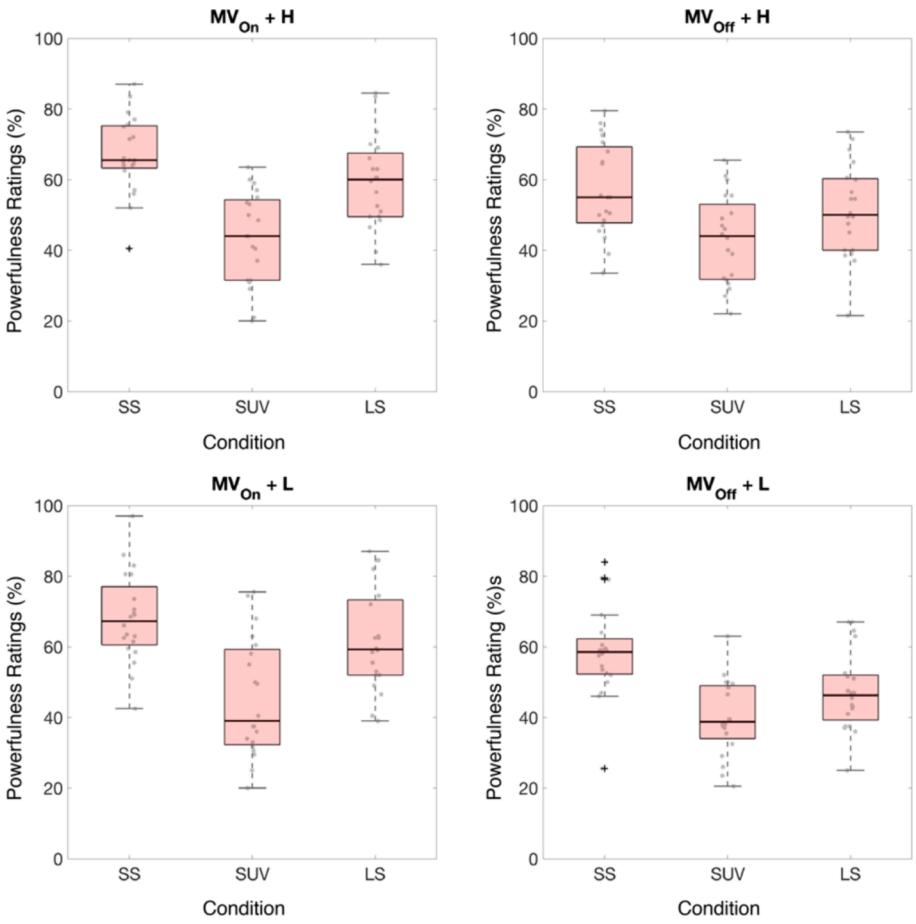


Fig. 8. Powerfulness ratings across tested vehicles: Sports Sedan (SS), Sports Utility Vehicle (SUV) and Luxury Vehicle (LS). The ratings are shown for each reproduction configuration: with motion and vibration and sound through loudspeakers (MVon + L), with motion and vibration and sound over headphones (MVon + H), without motion and vibration and sound through loudspeakers (MVoff + L) and without motion and vibration and sound over headphones (MVoff + H). Boxes contain the median (horizontal gray line) and delimit the first and third quartiles. The whiskers show the upper and lower 1.5 interquartile range, and crosses depict outliers. The gray dots show the individual ratings.

front, slightly tilted towards the left (Fig. 1), may have contributed to additional localization errors. Involuntary movements or tilting of the listener's head could have occurred, potentially causing minor shifts in the virtual sound environment. The difference in performance between the 8L and 64L methods was also expected, given the higher spatial resolution available with the increased number of loudspeakers in the latter method.

All three evaluated sound reproduction methods, however, produced frequent front-back confusions in the present study. Factors contributing to these confusions may include the absence of high-frequency individualized spectral cues and the presence of background noise. The latter could have interfered with the perception of low-frequency dynamic cues for localization through energetic masking, as evidenced in the spectrograms shown in Fig A4. Additionally, in this study, the recorded target signals were reproduced from a smartphone, and thus, the interior surfaces of the car (i.e., windshield, seats, etc.) may have acted as substantial acoustic "obstacles" and reflectors, affecting the recorded ILDs and ITDs for localization. Individual differences resulting from mismatches in localization cues as well as different strategies for solving the localization task may have also influenced the perception of front-back confusions. The latter could be partly attributed to the fact that the provided visual reference for the localization task was presented in front, potentially leading some participants to keep their heads still instead of turning to resolve perceptual ambiguities when sound was reproduced using loudspeaker-based methods.

The results for perceived horizontal angle also suggest that the speech signal was localized more accurately than the warning signal. This could be attributed to the engine noise being more effective in masking the warning signal perceptually, possibly due to its narrower spectral content compared to the speech signal, as shown in Figure A4. The spectral masking was particularly pronounced for the 100 km/h condition, where the noise exhibited an extended frequency range up to about 5 kHz. This could explain why this condition resulted in less accurate angle perception, significantly influencing perceived localization with the 8L sound reproduction method.

Furthermore, perception of horizontal angles was more accurate for frontal source positions compared to rear positions, which is consistent with previous studies indicating better localization performance for frontal sources [14,43]. Mean angle errors of 28° in the present study, however, were substantially larger than those reported in previous studies, where localization errors of less than 10° were observed for most source positions [40,41]. Still, perceived angle in this study was better compared to findings for in-vehicle sound localization with mean values larger than 40° [25]. The difference in angle errors between traditional sound localization studies and those conducted in-vehicle could be attributed to the challenging nature of the sound localization task in the latter case.

Achtemeier et al. [25] also noted poorer sound localization performance and lower confidence in judgements for sounds delivered from rear positions in-vehicle. However, in contrast to the present findings, the authors found that frontal positions also yielded poorer localization accuracy compared to lateral ones. While in the current study, visual

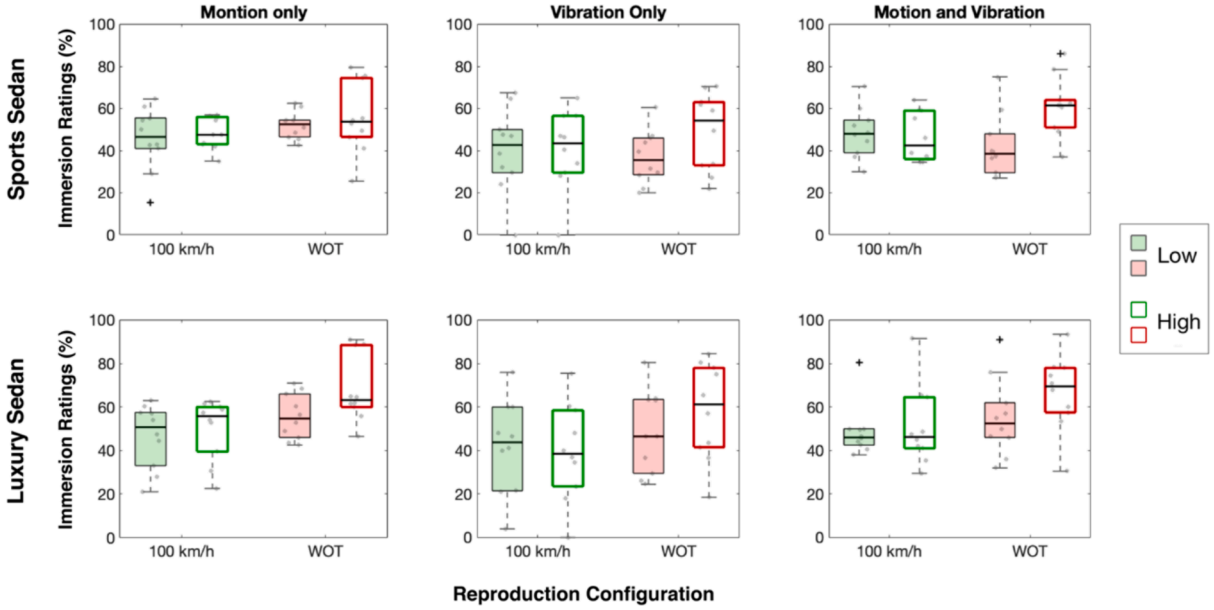


Fig. 9. Immersion ratings for each tested vehicle: Sports Sedan and Luxury Sedan. The ratings are shown across operating conditions, 100 km/h (green) and WOT (red), and reproduction configurations: with motion only, with vibration only, and with motion and vibration. Filled and empty boxplots represent the conditions reproduced with high or low intensity (of motion and/or vibration), respectively. Boxes contain the median (horizontal gray line) and delimit the first and third quartiles. The whiskers show the upper and lower 1.5 interquartile range, and crosses depict outliers. The gray dots show the individual ratings. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

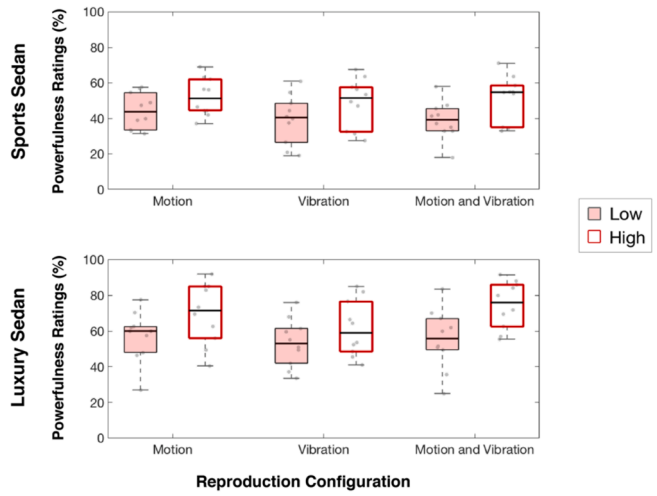


Fig. 10. Powerfulness ratings for each tested vehicle: Sports Sedan and Luxury Sedan. The ratings are shown across reproduction configurations: with motion only, with vibration only, and with motion and vibration. Filled and empty boxplots represent the conditions reproduced with high or low intensity (of motion and/or vibration), respectively. Boxes contain the median (horizontal grey line) and delimit the first and third quartiles. The whiskers show the upper and lower 1.5 interquartile range, and crosses depict outliers. The grey dots show the individual ratings.

information was provided through an HMD, Achtemeier et al. [25] conducted their tests inside a full-chassis driving simulator with a static image of a sound booth projected on a screen. In both studies, however, the sound was likely influenced by the acoustic properties of the chassis and the interior acoustics of the vehicle, either during the recordings (present study) or during the playback (reference study).

4.1.2. Distance perception

The outcomes of Experiment 1 also show that distance perception was less accurate when the sound was delivered over headphones than through loudspeakers. As in the case of angle perception, the lack of individualized HRTFs may have led to less precise localization cues for distance perception compared to those provided by the loudspeaker-based methods. While research suggests that the use of binaural synthesis with non-individualized HRTFs is adequate to achieve good auditory source localization in VR [44,45], the loudspeaker-based methods appear to have better preserved in-vehicle localization cues in the present study.

The results also showed that distance perception was more accurate for the speech signal than for the warning sound. One reason for this could be that humans are more familiar with speech signals than with warning-like signals, which may have facilitated their perception. Moreover, the speech signal has a richer spectral content compared to the warning sound, which may have provided stronger cues for distance perception that were not masked by the vehicle noise. This is evident in the frequency spectra depicted in Fig. 4A, illustrating that the most salient frequency components for the warning signal are concentrated around its fundamental frequency and the first two harmonics, with limited spectral content above 4 kHz. In contrast, the speech signal exhibits a broader frequency distribution and prominent frequency components above 4 kHz, making it less susceptible to energetic masking than the warning signal.

Notably, the warning signal lacks frequency components below its 1 kHz fundamental frequency, whereas male speech signals typically have fundamental frequencies around 100 Hz. This difference in spectral characteristics might also explain the observed differences in distance perception performance. In fact, Kopčo and Shinn-Cunningham [15] suggested that the lowest frequency present in stimuli strongly influenced distance perception of nearby sources. In their study, they examined perceived distance in a simulated classroom using noise bursts with varying center frequencies (300 – 5700 Hz) and bandwidths (200–5400 Hz). They found that stimuli with energy at low frequencies were judged more accurately than those with energy only at higher frequencies. Interestingly, whereas the low-frequency energy presence influenced distance perception in the reference study, stimulus bandwidth had no significant effect. This suggests that the poorer perceived distance for the warning signal may also be due to its higher

fundamental frequency compared to the speech signal.

The results also indicated that the idling condition was more accurate in terms of distance perception compared to the 100 km/h condition in the 8L method. This difference was likely due to the higher power and broader frequency spectrum of the noise at 100 km/h compared to the idling condition (as shown in Figure A4). This may have increased the energetic masking for the 100 km/h condition, extending up to about 5 kHz compared to the idling condition, which only extended up to about 500 Hz.

Interestingly, distance perception was less accurate for the front left (FL) source position, despite being one of the most lateralized positions relative to the listener. This contrasts with previous reports suggesting that distance perception is generally more accurate for lateral sounds source positions [14,15]). One reason for this is that the FL position was not perfectly lateralized with respect to the listening position, as it was positioned towards the front. Another reason could be the influence of vehicle interior surfaces on distance perception cues, particularly the windshield located just above the source position. These nearby surfaces do not exist in traditional sound localization studies conducted in typical rooms.

4.1.3. Effect of vibration and motion on immersion and powerfulness

The results of Experiment 2 showed that adding vibration and motion to a VR-based car driving simulator increases the perception of immersion and powerfulness. The results agree with the study of Hashimoto and Hatano [30], who showed that adding seat-floor vibration to vehicle noise strengthens the subjective impression of powerfulness. Unlike their study, in the current study vibration was added to the backrest and seat, which further supports the influence of vibration on powerfulness perception in driving simulators. The results are also consistent with Vailland et al. [46], who found that adding acceleration and centrifugal effects to a VR-based wheelchair simulator using a motion feedback platform contributed to enhanced immersion, resulting in an increased sense of presence. The reason for the increased immersion ratings when adding vibration and motion in the present study could be that the addition of coherent multisensory information may have reduced the sensory conflict between visual, vestibular, and somatosensory cues, which have been suggested to contribute to simulator and motion sickness [47,48]. Differences in perceived sensory conflicts could have influenced the results even if few test subjects explicitly reported experiencing substantial discomfort or sickness in the tested conditions of the present study.

Furthermore, the results of Experiment 3 indicated that the addition of motion had a greater contribution to the perception of powerfulness and immersion than the addition of vibration. This seems consistent with the fact that increased perception of self-motion (i.e., the perception of one's own body movement through the virtual environment) has been identified as one of the factors that improve perceived immersion in virtual environments [27]. Although self-motion in vehicle driving might rely primarily on visual information, it may also be affected by other multisensory information [49]. Since the same visual information was provided here for each vehicle, the perceived self-motion could have been stronger when the simulations included motion than when only vibration was added.

4.1.4. Effect of sound reproduction method on immersion and powerfulness

Unlike Experiment 1, in which sound reproduction influenced perceived localization, in Experiment 2, the sound reproduction method had no substantial effect on immersion and powerfulness ratings. This indicates that both reproduction methods, binaural recordings reproduced over headphones, and matrix inversion method reproduced through a circular array of eight loudspeakers, elicit a similar perception of immersion when applied to VR-based driving simulators. The reason for this could be that both methods have similar performance in capturing and reproducing vehicle noise, resulting in no noticeable differences in terms of perceived immersion and powerfulness. Alternatively, the sound reproduction methods could induce differences in these perceptual attributes but were not captured in the present study. This could occur if the additional multisensory information "overshadowed" the effect of the sound reproduction method. For example, wearing headphones may have led to decreased immersion ratings due to discomfort, but the discomfort associated with wearing the HMD in all conditions could have induced greater discomfort that outweighed the effect of the headphones. Similarly, the effect of adding/removing motion and vibration might have outweighed the effect of varying the sound reproduction method on the immersion and powerfulness ratings. Therefore, further research would be needed to clarify whether the sound reproduction method significantly influences perceived immersion and powerfulness in driving simulators. This involves exploring reproduction configurations that can potentially induced stronger immersion effects, such as incorporating individualized HRTFs into headphone reproduction, or addressing potential discomfort effects impacting immersion and powerfulness ratings.

4.2. Limitations

It is likely that the translation and tilting movements of the motion platform may have caused small oscillations, potentially affecting the perceived immersion and powerfulness in the form of perceived vibrations, even in the motion only condition in Experiment 3. However, the motion and vibration conditions were clearly distinguishable. In the vibration condition, the oscillations were localized in the driver's seat and backrest and were not accompanied by tilting or translation movements.

The frequency spectra of the vibration signals were not adjusted, only the overall level relative to the real vehicles on road. Future studies employing precise calibrations of vibration signals could thus yield results that deviate from those presented here. The lack of individualized HRTFs in the binaural synthesis, and possibly the absence of head-tracking, might have led to poorer performance in the headphone sound reproduction method compared to the loudspeaker-based methods. Hence, future studies could assess the advantages of incorporating individualized HRTFs and head-tracking in binaural synthesis compared to loudspeaker-based methods for in-vehicle sound localization. Moreover, due to logistical constraints and the resource-intensive nature of the VR-based driving simulation experiments, this study had a limited sample size, and hearing function was assessed through self-reporting rather than formal audiometric testing. While statistical methods were employed to mitigate individual variability, a larger sample size and more controlled hearing screenings would enhance the generalizability of findings in future studies.

5. Conclusion

In conclusion, this study aimed at investigating the multimodal configurations of a VR-based driving simulator which provide the best performance in terms of perceived sound localization, immersion and powerfulness. The results indicate that in-vehicle sound source localization can be challenging, which could limit the implementation of vehicle features relying only on the acoustic perception. For sound source localization, HOA with 64L provided the best performance compared to the matrix inversion method with 8L and the binaural headphone reproduction without head tracking support. Additionally, both loudspeaker-based methods outperformed the binaural headphone reproduction method in terms of angle perception and distance perception. Adding vibration and motion increased the perception of immersion and powerfulness. For both perceptual attributes, the inclusion of motion had a greater contribution than the inclusion of vibration. However, no difference was found in the perception of immersion and powerfulness when comparing the sound reproduction methods using the circular array of 8 loudspeakers and the binaural reproduction over headphones. Overall, the results presented here favored the

implementation of more comprehensive VR-based driving simulators in terms of the presented multimodal information to improve the realism of the driving experience.

CRediT authorship contribution statement

Juan Camilo Gil-Carvajal: Writing – original draft, Validation, Resources, Methodology, Investigation, Formal analysis. **Eun Soo Jo:** Writing – review & editing, Validation, Resources, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Dong Chul Park:** Writing – review & editing, Resources, Funding acquisition. **Wookeun Song:** Writing – review & editing, Supervision, Software. **Cheol-Ho Jeong:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

We would also like to thank the visual developer from Perfect Storm for providing the visuals for the VR simulator.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apacoust.2024.110137.

References

[1] Chan E, Pradhan AK, Pollatsek A, Knodler MA, Fisher DL. Are driving simulators effective tools for evaluating novice drivers' hazard anticipation, speed management, and attention maintenance skills? Transport Res F: Traffic Psychol Behav 2010;13(5):343–53.
[2] De Winter, J., van Leeuwen, P. M., & Happee, R. (2012, August). Advantages and disadvantages of driving simulators: A discussion. In *Proceedings of measuring behavior* (Vol. 2012, p. 8th).
[3] Roenker DL, Cissell GM, Ball KK, Wadley VG, Edwards JD. Speed-of-processing and driving simulator training result in improved driving performance. Hum Factors 2003;45(2):218–33.
[4] Schultheis MT, Mourant RR. Virtual reality and driving: The road to better assessment for cognitively impaired populations. Presence Teleop Virt 2001;10(4):431–9.
[5] Cox, D. J., Davis, M., Singh, H., Barbour, B., Nidiffer, F. D., Trudel, T., … & Moncrief, R. (2010). Driving rehabilitation for military personnel recovering from traumatic brain injury using virtual reality driving simulation: A feasibility study. *Military medicine*, *175*(6), 411-416.
[6] Kang, H. S., Jalil, M. A., & Mailah, M. (2004, June). A PC-based driving simulator using virtual reality technology. In *Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry* (pp. 273-277).
[7] Godley ST, Triggs TJ, Fildes BN. Driving simulator validation for speed research. Accid Anal Prev 2002;34(5):589–600.
[8] Mourant RR, Schultheis MT. A HMD-based virtual reality driving simulator. Paper presented at the International driving symposium on human factors in driver assessment, training and vehicle design, Smowmass Village at Aspen, Colorado. 2001.
[9] Michael D, Kleanthous M, Savva M, Christodoulou S, Pampaka M, Gregoriades A. Impact of immersion and realism in driving simulator studies. Internat J Interdiscipl Telecommun Network (IJITN) 2014;6(1):10–25.
[10] Lin JW, Duh HBL, Parker DE, Abi-Rached H, Furness TA. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In: Proceedings IEEE virtual reality 2002. IEEE; 2002. p. 164–71.
[11] Mourant RR, Rockwell TH. Strategies of visual search by novice and experienced drivers. Hum Factors 1972;14(4):325–35.
[12] Bles W, Bos JE, De Graaf B, Groen E, Wertheim AH. Motion sickness: only one provocative conflict? Brain Res Bull 1998;47(5):481–7.
[13] Middlebrooks JC, Green DM. Sound localization by human listeners. Annu Rev Psychol 1991;42(1):135–59.
[14] Kolarik AJ, Moore BC, Zahorik P, Cirstea S, Pardhan S. Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. Atten Percept Psychophys 2016;78:373–95.
[15] Kopčo N, Shinn-Cunningham BG. Effect of stimulus spectrum on distance perception for nearby sources. J Acoust Soc Am 2011;130(3):1530–41.
[16] Kim SM, Choi W. On the externalization of virtual sound images in headphone reproduction: A Wiener filter approach. J Acoust Soc Am 2005;117(6):3657–65.
[17] Jiang J, Xie B, Mai H, Liu L, Yi K, Zhang C. The role of dynamic cue in auditory vertical localisation. Appl Acoust 2019;146:398–408.
[18] Wallach H. The role of head movements and vestibular and visual cues in sound localization. J Experiment Psychol 1940;27(4):339.
[19] Hammershøi D, Møller H. Binaural technique—Basic methods for recording, synthesis, and reproduction. Commun Acoust 2005:223–54.
[20] Zotter F, Frank M. Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality. Springer Nature; 2019. p. 210.
[21] Song W, Marschall M, Corrales JDG. Simulation of realistic background noise using multiple loudspeakers. Proc. Int. Conf. on Spatial Audio (ICSA) Graz, Austria. 2015.
[22] Kirkeby O, Nelson PA, Orduna-Bustamante F, Hamada H. Local sound field reproduction using digital signal processing. J Acoust Soc Am 1996;100(3):1584–93.
[23] Pulkki V. Virtual sound source positioning using vector base amplitude panning. J Audio Eng Soc 1997;45(6):456–66.
[24] Berkhout AJ, de Vries D, Vogel P. Acoustic control by wave field synthesis. J Acoust Soc Am 1993;93(5):2764–78.
[25] Achtemeier JD, Craig CM, Morris NL, Davis B. Superior side sound localisation performance in a full-chassis driving simulator. Ergonomics 2020;63(5):538–47.
[26] Catchpole KR, McKeown JD, Withington DJ. Localizable auditory warning pulses. Ergonomics 2004;47(7):748–71.
[27] Witmer BG, Singer MJ. Measuring presence in virtual environments: A presence questionnaire. Presence 1998;7(3):225–40.
[28] Weibel D, Wissmath B. Immersion in computer games: The role of spatial presence and flow. Internat J Comput Games Technol 2011. 2011.
[29] Alshaer A, Regenbrecht H, O'Hare D. Immersion factors affecting perception and behaviour in a virtual reality power wheelchair simulator. Appl Ergon 2017;58:1–12.
[30] Hashimoto, T., & Hatano, S. (2019, September). Effect of Multimodal Input to the Perception of Car Interior Noise. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* (Vol. 259, No. 8, pp. 1338-1347). Institute of Noise Control Engineering.
[31] Ellermeier, W., & Legarth, S. V. (2006, May). Visual bias in subjective assessments of automotive sounds. In *Proceedings of Euronoise* (Vol. 30).
[32] Nielsen JB, Dau T. The Danish hearing in noise test. Int J Audiol 2011;50(3):202–8.
[33] Moreau S, Daniel J, Bertet S. 3d sound field recording with higher order ambisonics–objective measurements and validation of a 4th order spherical microphone. In: 120th Convention of the AES; 2006. p. 20–3.
[34] Daniel, J., Moreau, S., & Nicol, R. (2003, March). Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging. In *Audio Engineering Society Convention 114*. Audio Engineering Society.
[35] Favrot S, Buchholz JM. LoRA: A loudspeaker-based room auralization system. Acta Acust Acust 2010;96(2):364–75.
[36] ISO 26101 (2012). Acoustics – Test methods for the qualification of free-field environments.
[37] IEC268-13. (1985). Sound system equipment. *Part 13: Listening tests on loudspeakers*.
[38] Weech S, Kenny S, Barnett-Cowan M. Presence and cybersickness in virtual reality are negatively related: a review. Front Psychol 2019;10:158.
[39] Bates, D., Mächler, M., Bolker, B., Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.
[40] Zahorik P, Brungart DS, Bronkhorst AW. Auditory distance perception in humans: A summary of past and present research. Acta Acust Acust 2005;91(3):409–20.
[41] Bronkhorst AW. Localization of real and virtual sound sources. J Acoust Soc Am 1995;98(5):2542–53.
[42] Wightman FL, Kistler DJ. Factors affecting the relative salience of sound localization cues. Binaural Spatial Heari Real Virtual Environ 1997;1:1–23.
[43] Makous, J. C. and J. C. Middlebrooks (1990). "Two-dimensional sound localization by human listeners". In: The journal of the Acoustical Society of America 87.5, pp. 2188–2200.
[44] Begault DR, Wenzel EM, Anderson MR. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. J Audio Eng Soc 2001;49(10):904–16.
[45] Berger CC, Gonzalez-Franco M, Tajadura-Jiménez A, Florencio D, Zhang Z. Generic HRTFs may be good enough in virtual reality. Improving source localization through cross-modal plasticity. Front Neurosci 2018;12:21.
[46] Vailland G, Gaffary Y, Devigne L, Gouranton V, Arnaldi B, Babel M. March). Vestibular feedback on a virtual reality wheelchair driving simulator: A pilot study. In: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction; 2020. p. 171–9.

[47] Lucas G, Kemeny A, Paillot D, Colombet F. A simulation sickness study on a driving simulator equipped with a vibration platform. Transport Res F: Traffic Psychol Behav 2020;68:15–22.
[48] Reason JT. Motion sickness adaptation: a neural mismatch model. J R Soc Med 1978;71(11):819–29.
[49] Kemeny A, Panerai F. Evaluating perception in driving simulation experiments. Trends Cogn Sci 2003;7(1):31–7.