Accident Analysis and Prevention 175 (2022) 106778

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Overestimated time-to-collision for quiet vehicles: Evidence from a study using a novel audiovisual virtual-reality system for traffic scenarios

Daniel Oberfeld [1,*], Marlene Wessels [1], David Büttner

Institute of Psychology, Section Experimental Psychology, Johannes Gutenberg-Universität Mainz, Wallstrasse 3, Mainz 55122, Germany

ARTICLE INFO

Keywords:
Time-to-collision estimation
Road traffic safety
Pedestrian
Auditory perception
Electric vehicles
Virtual reality

ABSTRACT

To avoid collision, pedestrians intending to cross a road need to estimate the time-to-collision (TTC) of an approaching vehicle. Here, we present a novel interactive audiovisual virtual-reality system for investigating how the acoustic characteristics (loudness and engine type) of vehicles influence the TTC estimation. Using acoustic recordings of real vehicles as source signals, the dynamic spatial sound fields corresponding to a vehicle approaching in an urban setting are generated based on physical modeling of the sound propagation between vehicle and pedestrian and are presented via sound field synthesis. We studied TTC estimation for vehicles with internal combustion engine and for loudness-matched electric vehicles. The vehicle sound levels were varied by 10 dB, independently of the speed, presented TTC, and vehicle type. In an auditory-only condition, the cars were not visible, and lower loudness of the cars resulted in considerably longer TTC estimates. Importantly, the loudness of the cars also had a significant effect in the same direction on the TTC estimates in an audiovisual condition, where the cars were additionally visually presented via interactive virtual-reality simulations. Thus, pedestrians use auditory information when estimating TTC, even when full visual information is available. At equal loudness, the TTC judgments for electric and conventional vehicles were virtually identical, indicating that loudness has a stronger effect than spectral differences. Because TTC overestimations can result in risky road crossing decisions, the results imply that vehicle loudness should be considered as an important factor in pedestrian safety.

1. Introduction

Safe mobility requires the ability to avoid potentially dangerous collisions with objects in the environment. For instance, a pedestrian crossing a road must avoid being hit by an approaching vehicle. In such a situation, our sense of hearing provides important information. For example, we can auditorily detect a vehicle approaching us from outside our field of view. The importance of acoustic information can be expected to be even higher for persons with visual impairment. Also, the recent event of increasing electric mobility poses the question of whether and how the acoustic characteristics of quieter (electric) vehicles affect pedestrians' perception of vehicles in traffic situations.

When pedestrians want to cross a road while a vehicle is approaching, a safe crossing is only possible if the time remaining until the vehicle arrives at the pedestrian's position (time-to-collision, TTC[2]) is longer than the time needed to cross. Thus, a sufficiently accurate estimate of TTC is essential (e.g., Lee, Young, & McLaughlin, 1984; Petzoldt, 2014). Although the auditory detection of vehicles ("*Are there any vehicles near me?*") was studied extensively in recent years (e.g., Altinsoy et al., 2015; Emerson, Kim, Naghshineh, Pliskow, & Myers, 2013; Poveda-Martinez et al., 2017), resulting in recommendations for auditory vehicle alerting system (AVAS) technologies for electric vehicles (EVs) and corresponding legislative actions (NHTSA 141, 2018; UNECE R138, 2017), auditory and audiovisual TTC estimation received only limited attention. There is a large body of literature on TTC estimation and road-crossing decisions, but focused on the visual modality (e.g., Hecht & Savelsbergh, 2004). Hence, there are significant gaps in our knowledge about collision avoidance based on auditory or combined auditory and visual information.

In the visual domain, accurate TTC information is – at least in theory

* Corresponding author.
E-mail address: oberfeld@uni-mainz.de (D. Oberfeld).

[1] shared first authorship.

[2] Note that, for reasons of simplicity, we use the term "time-to-collision" even when the vehicle is not on a direct collision with the simulated observer, so that strictly speaking the term "time-to-passage" would be more appropriate.

https://doi.org/10.1016/j.aap.2022.106778
Received 3 January 2022; Received in revised form 4 July 2022; Accepted 14 July 2022

– provided by so-called optical invariants when certain assumptions are met (e.g., objects are rigid and move at a constant velocity; see Lee, 1976). For instance, the invariant τ (first described by Hoyle, 1957) valid for objects approaching an observer on a direct collision course and at a constant speed is defined as the ratio of an object's instantaneous optical size to its instantaneous rate of optical expansion (Lee, 1976), and thus relies solely on information available on the retina. However, many studies have demonstrated that even when τ-variables are available, other visual characteristics of the object can influence TTC judgments by providing heuristic cues. For instance, studies by DeLucia and colleagues (e.g., DeLucia, 1991; DeLucia, 2013) have shown that an approaching object's optical size at the moment in time when the TTC judgment is made influences these judgments, such that smaller optical sizes are associated with later arrival time estimates compared to larger optical sizes at the same actual TTC ("*size-arrival effect*"). Apart from the importance for theories of TTC estimation, the size-arrival effect was proposed to be potentially relevant for traffic safety. For instance, accidents involving motorcyclists or children might in part be related to the smaller size of a motorcycle or a child compared to a car or an adult pedestrian, respectively (Caird & Hancock, 1994; DeLucia, 2013; Horswill, Helman, Ardiles, & Wann, 2005; Petzoldt, 2016).

In the auditory domain, accurate TTC information for objects approaching on a straight direct collision path at a constant velocity is provided, e.g., by a τ-like ratio of the objects' instantaneous acoustic intensity to its instantaneous rate of change in intensity (DeLucia, Preddy, & Oberfeld, 2016; Jenison, 1997; Shaw, McGowan, & Turvey, 1991). This is sometimes referred to as *auditory τ*. Other τ-like variables can be computed on the basis of both monaural and binaural auditory cues other than intensity (Jenison, 1997), and a multitude of heuristic cues related to the motion, distance, and TTC of an object are available (e.g., Jenison, 1997; Kaczmarek, 2005; Lutfi & Wang, 1999; Rosenblum, Carello, & Pastore, 1987; Zakarauskas & Cynader, 1991), just as in the visual domain (e.g., DeLucia, Kaiser, Bush, Meyer, & Sweet, 2003; Gray & Regan, 1998; Yan, Lorv, Li, & Sun, 2011). With this in mind, it is not surprising that humans are able to estimate TTC based on auditory information alone, and that some blind individuals can use auditory TTC information with an accuracy comparable to the ability of sighted people to use visual TTC information (Rosenblum, Wuestefeld, & Saldana, 1993; Schiff & Oldak, 1990).

In two recent studies, we investigated in greater detail which cues are used in auditory and audiovisual TTC estimation (DeLucia, et al., 2016; Keshavarz, Campos, DeLucia, & Oberfeld, 2017). The relative weights of auditory and visual cues were measured, using a behavioral reverse-correlation approach. TTC estimates across a range of scenarios were collected in conditions providing only auditory, only visual, or combined audiovisual information. We measured the relative weights assigned by participants to cues that provided accurate TTC information, specifically τ, in the auditory and visual domains (Lee, 1976; Shaw, et al., 1991). We also measured the weights assigned to simpler heuristic cues, which are less reliably accurate and can be misleading under certain conditions. These included optical size (i.e., the pictorial depth cue of relative size) and sound pressure level (SPL). The results consistently showed that the accurate cues (τ) as well as heuristic cues contributed to TTC judgments in visual and auditory modalities, but the reliance on heuristic cues was greater in the auditory modality than in the visual modality. Participants relied strongly on the SPL at the moment in time when the TTC judgment was made, resulting in longer TTC estimates for approaching auditory objects with a lower acoustic intensity (i.e., lower loudness). These results represent the first evidence for an auditory analog of the visual size-arrival effect (DeLucia, 1991) discussed before, which we termed the "*intensity-arrival effect*". At identical actual TTC, participants tend to judge softer sound sources to arrive later than louder sound sources.

The intensity-arrival effect might indicate increased risks posed by quiet vehicles like electric cars. Pedestrians might overestimate the TTC of a quiet EV relative to a louder conventional vehicle with the same actual TTC, which in turn could result in risky road crossing decisions. Nonetheless, the results of the two prior experiments are limited in their generalizability to real traffic scenarios because relatively simple and somewhat artificial stimuli were used. The auditory stimuli were a pure tone (DeLucia, et al., 2016) or the sound of an idling combustion engine (Keshavarz, et al., 2017), on which the intensity profile corresponding to a direct constant-velocity approach in the auditory free field was imposed. The sounds were presented through a single loudspeaker located in front of the participants. These auditory stimuli were impoverished compared to an approaching vehicle in a real traffic scenario, which provides a dynamic spatial sound field including reflections from the ground surface and sound from different vehicle noise sources likes tires, engine, transmission, and exhaust pipe, with significant variations in the vehicle sound depending on speed or selected gear. Thus, it is an open and important question whether the effect of the acoustic intensity on auditory TTC estimates is also observed with more detailed and realistic simulations of an approaching vehicle.

In fact, the fidelity of the presented stimuli seems to play an important role in TTC estimation. Even the highly replicated size-arrival effect in the visual domain was reduced in some studies presenting real moving objects with full visual information (Cavallo & Laurent, 1988; Savelsbergh, Whiting, & Bootsma, 1991). Also, the relative weights assigned to auditory compared to visual cues were higher in our first study presenting extremely simple visual as well as auditory stimuli (expanding square on a computer screen, pure tone with increasing intensity presented via a single speaker) (DeLucia, et al., 2016) than in our second study (Keshavarz, et al., 2017), where similarly simple auditory stimuli were combined with a high-fidelity visual simulation of an urban traffic scene. A plausible explanation of this pattern is that participants placed less attention on the overly simplistic auditory stimulus when a high-fidelity visual simulation of the approaching object was available.

Thus, realistic high-fidelity auditory simulations of approaching vehicles are needed to gain reliable and ecologically valid data about the use of auditory information in TTC estimation and other traffic-relevant tasks. They should cover the following characteristics: First, realistic vehicle sounds should be presented, containing all relevant vehicle noise sources (tire, powertrain, and aerodynamic noise) as well as their dependence on speed, engine load etc. Second, when a vehicle approaches a pedestrian in the real world, there are dynamic changes in acoustic intensity (due to spherical spreading and air absorption), in frequency spectrum (due to frequency-dependent air absorption and comb-filter effects caused by interference between direct and reflected sound), as well as in interaural time and level differences (due to changes in the position of the vehicle relative to the observer). All these acoustic cues should be incorporated in the simulations. Third, participants should be able to explore the virtual scene by moving their heads as they would in the real world.

However, we are not aware of any studies on auditory TTC estimation using highly realistic and interactive spatial auralizations of moving vehicles. Previous studies presented monaural recordings of approaching vehicles or synthesized vehicle sounds on a single loudspeaker (e.g., Keshavarz, et al., 2017; Schiff & Oldak, 1990), on stereo speakers (Pugliese, Barton, Davis, & Lopez, 2020), or via headphones (e.g., Rosenblum, Gordon, & Wuestefeld, 2000; Rosenblum, et al., 1993), so that an accurate reproduction of the spatial sound scene was not available. Other experiments presented binaural recordings of real vehicles (M. S. Gordon & Rosenblum, 2005), which provides spatial information but no opportunity to interact via head movements. Wu et al. (2018) used binaural synthesis to auralize traffic at a roundabout from a pedestrian's point of view. However, no head-tracking was used and the moving sound sources were simulated based on recordings from a stationary car at a constant engine speed, and thus did neither contain tire noise, nor changes in engine noise linked to a variation in velocity or acceleration. Two recent studies used binaural recordings of real approaching vehicles in experiments on road-crossing decisions (Soares et al., 2020; Soares et al., 2021), which provide a very realistic sound quality. However, binaural recordings represent only a single, specific head orientation and receiver ear height (i.e., the head orientation and height of the dummy head during the recordings), and thus participants were not able to interact with the virtual scene via head movements. In addition, in binaural recordings parameters such as the angle of approach of the vehicle, the simulated distance of the listener from the road, or the distance and TTC of the vehicle at a given point of time in the recordings cannot be changed during the experiments, which limits experimental control and the range of experimental conditions that can be presented.

To overcome these limitations, we developed a novel interactive audiovisual virtual-reality (VR) system for the presentation of moving vehicles, which provides acoustic simulations of the relevant dynamic traffic scenarios at a higher degree of realism than in many previous studies in this field. The system generates highly realistic vehicle sounds because the simulations are based on acoustic recordings of real vehicles in multiple driving conditions, collected with several microphones attached to the vehicles' chassis. The motion of the vehicle sound sources in space during the approach is accurately simulated via acoustical modeling, and sound field synthesis (e.g., Ahrens, Rabenstein, & Spors, 2014) is used to present the dynamic spatial sound fields corresponding to the real-world scenarios that are simulated. The system provides interactive simulations because listeners can interact via head movements, and the recorded vehicle sound can be presented at arbitrary approach angles and distances, making it possible to present for example exactly the same vehicle sound at different TTCs. This system enables us to present highly realistic and interactive simulations of vehicles approaching in road-crossing scenarios, containing the whole set of auditory cues (intensity and spectral changes, interaural time- and level differences, Doppler frequency shifts, etc.) as in the real world. At the same time, the system can be used to conduct highly controlled VR experiments without challenging the participants' safety.

To answer the question of whether and to what extent the vehicle loudness influences TTC estimates (i.e., a potential intensity-arrival effect), we conducted an experiment using our novel audiovisual VR system. We compared TTC estimates for vehicles with the same motion parameters but presented at two different loudness levels differing by 10 dB, to test the hypothesis that the intensity-arrival effect observed in our previous studies does also occur when realistic vehicle sounds are presented in an interactive spatial simulation. We also contrasted a condition where only auditory information about the approaching vehicle was available (auditory-only, A-only) and a condition where an interactive visual 3D simulation of the vehicle was presented in addition to the vehicle sounds (audiovisual, AV), in order to investigate whether the intensity-arrival effect occurs even when full visual information about the motion of the vehicle is available, and whether the effect is stronger when only auditory information is presented. In addition, we included sounds of a vehicle with internal combustion engine (ICEV) and an electric vehicle (EV), presented at the same loudness, to investigate if differences between the acoustic signatures of the two vehicle types, beyond a difference in loudness, affect pedestrians' TTC estimates.

2. Methods

2.1. Design and implementation of the audio-visual VR simulation system

We designed and implemented a high-performance interactive audiovisual VR system to investigate the audiovisual perception of approaching vehicles. The optical geometry of a vehicle does not change as a function of speed or acceleration. Thus, a visual VR simulation of vehicles approaching an observer can simply be obtained with readily available geometric techniques, implemented in current computer graphics.

In contrast, the tire-road noise emitted by a vehicle dynamically depends on the velocity, tire and road surface characteristics (e.g., Kropp, Sabiniarz, Brick, & Beckenbauer, 2012); the aerodynamic noise also depends on the velocity, but it is preponderant only at high speeds (more than 100 km/h). The powertrain noise depends dynamically on engine speed and engine load, which, in turn, depend on factors like the selected gear, acceleration, road inclination, etc.

Since we were not aware of completely convincing approaches for the simulation of tire, powertrain and aerodynamic noise in dynamic driving situations with changing speed, acceleration, and load condition, the auditory stimuli were based on acoustic recordings of real vehicles. The recordings were collected while driving both an internal combustion engine vehicle (ICEV) and an electric vehicle with well-defined velocity profiles (various constant speeds, various conditions with acceleration) on a test track. In the present study, only the subset of constant-speed drives is used. TTC estimation for accelerating vehicles using the recordings of acceleration drives was studied, for example, in Wessels et al. (2022). We used a source-based approach, recording the vehicle sound with four free-field microphones mounted at different positions on the chassis. During the acoustic recordings, the trajectory of the vehicle was measured with highly precise GPS position tracking. In the auditory VR simulations, the microphone signals recorded on the test track are then used as sound sources in an acoustic VR simulation software (Toolbox for acoustic scene creation and rendering – TASCAR; Grimm, Luberadzka, & Hohmann, 2019) and are animated on the basis of the GPS position tracking data. Using this approach, it is possible to present real vehicle sounds, and to simulate the exact motion of the car during the recordings. The interactive auditory VR simulations can be combined with interactive three-dimensional visual VR simulations presented stereoscopically on a head-mounted display. Further details regarding the acoustic recordings of the vehicle sounds and the audiovisual simulation system are described in the following sections.

2.1.1. Acoustic recordings of vehicle sounds

The vehicles' acoustic signals used in the simulation system are based on acoustic recordings of real cars driving on a test track of the Technical University of Darmstadt. All recordings took place on a dry asphalt road surface. The vehicles were two small passenger car models of the manufacturer Kia Motors. The ICEV was a gasoline-powered Kia Rio 1.0 T-GDI (2019, 1.0 l, 88 kW, 3 cylinders) with manual transmission. The tires on the ICEV were Continental summer tires (ContiSportContact 5, 205/45 R17). The EV was a Kia e-Niro (2019, 150 kW) with Michelin summer tires (Primacy 3, 215/55 R17). The EV was additionally equipped with an acoustic warning sound system (Acoustic Vehicle Alerting System; AVAS), which could be active at speeds between 0.5 km/h and 28 km/h, but could also be deactivated. The sound generated by the AVAS was compatible with the requirements described in UNECE R138 (2017). We made recordings of the EV with both active and inactive AVAS. However, in the present experiment, we only presented the EV without AVAS. We studied pedestrians' TTC estimation for the EV with AVAS, for example, in Wessels et al. (under revision).

Four free-field microphones (Roga MI-17) were mounted on the chassis of the vehicle at the following positions: on both sides of the vehicle above the axle of the front tires, centrally on the engine hood, and on the right side of the vehicle above the axle of the rear tire. The microphones above the tires were positioned at a height between 86 and 101 cm above ground and captured primarily the tire-road noise. In contrast, the microphone on the hood captured primarily the powertrain noise. Thus, a realistic presentation of the vehicle sound was possible in the experiment. The microphone signals were recorded on an audio recording system (Sinus GmbH Soundbook MK2) located in the vehicle (audio resolution 24 bits, sampling frequency $f_s = 51.2$ kHz). Additionally, a Garmin GPS receiver (recording rate: 1 Hz) and a Tentacle SyncE audio time code generator were connected to the Soundbook. The GPS receiver allowed for synchronization of the Soundbook system time with the GPS time, so that the temporal reference to the GPS position data in the JAVAD GPS receiver (see below) could be established.

A GPS antenna (Trimble AG25) was installed centrally on the vehicle's roof and was connected to a high-performance GPS receiver (JAVAD Triumph LS, recording rate 10 Hz) inside the vehicle. Using the Real Time Kinematic method, the GPS position of the vehicle on the test track could be recorded with a precision of a few centimeters (e.g., El-Rabbany, 2002). The high precision of the method is achieved by evaluating the carrier phase of the satellite signals, processing signals from at least 5 satellites, and matching the data from the mobile receiver with data from a geostationary reference station. Here, we used a reference station provided by the Hessian State Office for Land Management and Geoinformation within the framework of SAPOS-GPPS (https://sapos.hvbg.hessen.de/), located at a distance to the test track of about 6 km.

Post-processing of the raw GPS data from the JAVAD receiver and from the reference station was performed using the RTKLIB toolbox (https://www.rtklib.com/). In the "Kinematic" positioning mode, an extended Kalman filter was applied forward and backward to the time series of GPS data. At each time step in the GPS data, the arithmetic mean of the two filter passes was then used. A coordinate system was defined in which the vehicle's position along the road was described as the *x*-axis (pointing into the direction of travel), the lateral position on the road as the *y*-axis (pointing towards the other side of the road) and the vertical position (height above the ground) as the *z*-axis. The calculated GPS positions were converted to positions in the local coordinate system based on the World Geodetic System (WGS-84). The recorded GPS signal allowed the mapping of the vehicle position and the corresponding acoustic signal to the virtual environment used in the experiment, based on this coordinate system. In addition to the position data, the velocity vector and the acceleration vector were calculated from the GPS data at each time point, in the same local coordinate system. The position, velocity, and acceleration vectors were interpolated to a sampling rate of 1000 Hz using spline interpolation (Matlab function *interp1*) to facilitate subsequent calculations.

In addition to the acoustic recordings on the vehicle, an artificial head (Brüel & Kjaer 4100D) and a free-field microphone (Roga MI-17) were positioned stationary at a distance of 50 cm from the right edge of the road. The artificial head was located 100 cm and the free-field microphone 165 cm above the road surface. The signals from the dummy head and from the free field microphone were recorded for evaluation purposes only, they are not used in the acoustic simulations. The signals from the two microphones in the dummy head and the free-field microphone were recorded by a Sinus GmbH Soundbook MK1. In addition to the microphones, an audio time code generator (Tentacle SyncE) was connected, which was synchronized with the audio time code generator connected to the audio recording system inside the vehicle. Sensor signals were also recorded from an anemometer placed near the reference position.

The precise implementation of driving profiles with specific constant speeds or accelerations was difficult in terms of driving. In addition, there were frequent issues with undesirable noise during the recordings. In particular, significant wind noise frequently occurred in the audio recordings at higher speeds, although we used wind protection on all microphones. For these reasons, we manually checked the driving profiles (velocity and acceleration) and the audio signals from all microphones on the vehicle for each of the recordings obtained on the test track. Driving profiles with strong deviations from the intended constant speed or acceleration or significant unwanted noise were not used for the acoustic simulations. In the present study, we only presented simulations of vehicles approaching at a constant speed.

2.1.2. Auditory VR simulation and sound reproduction

A physically realistic interactive simulation of the dynamic spatial sound field corresponding to an urban traffic scene with an approaching vehicle was realized with the acoustic VR software TASCAR (Grimm, et al., 2019; https://www.tascar.org/). TASCAR offers dynamic processing of the geometry of the acoustic scene (time-variable positions of (image) sound sources, absorbers and receivers), acoustic modeling of the sound transmission from the sources to the receiver, and sound field synthesis. TASCAR models the directional characteristic of sound sources, the distance-dependent change of the sound level caused by spherical spreading and air absorption, and the distance-dependent sound travel time (which can lead to, e.g., Doppler effects). Sound reflections on the ground and other surfaces are simulated by the image sound source method (Allen & Berkley, 1979), so that, for instance, time-variant comb-filter effects due to acoustical interference between reflected and direct sound are simulated. Due to the higher-order Ambisonics rendering (see below), dynamic changes in the interaural level and time differences resulting from changes of the position of the sound sources relative to the head of the listeners are also simulated, and listeners can interact with the acoustic scene by listening around with head movements. In the present experiment, the processing of the dynamic geometry of the acoustic scene was based on the GPS position data acquired during the acoustic recordings on the test track. Thus, the motion of the cars presented in the simulations was identical to the motion of the real cars on the test track. The position of the microphones relative to the car's front along the left-right and front-back axes (see above) was taken into account in the acoustic simulations. Based on informal comparisons with the dummy head recordings obtained on the test track, the height of the virtual sound sources reproducing the signals of the microphones positioned over the axles was set to 1 cm above ground to avoid overly strong comb filter effects. This position also corresponds to the main acoustic source of the road-tire noise (contact point between tire and road). The simulated ear height of the receiver matched the actual ear height of each listener in an upright position.

A spatial sound field was generated using sound field synthesis, namely 2D 7th-order Ambisonics (Ahrens, et al., 2014; Daniel, 2000; Gerzon, 1985). The Ambisonics approach assumes sound reproduction in an acoustic free field. When preparing the laboratory space, special attention was therefore paid to reducing acoustic reflections. The rectangular laboratory area containing the speaker array (dimensions: 570 cm × 450 cm) was separated from the larger lab space (105 m^2) with sound-absorbing acoustic curtains (Gerriets Bühnenvelours Ascona 570; 570 g/m^2; absorption coefficient of 0.95 for frequencies above 400 Hz). The parts of the walls and ceiling adjacent to the speakers were lined with Basotect acoustic foam panels (BASF; 10 cm thickness, absorption coefficient of at least $\alpha = 0.9$ at frequencies above 400 Hz).

Within the rectangular laboratory area, a circular array of 16 loudspeakers (Genelec 8020DPM-7) was installed. The radius was 2.0 m, the minimum distance of the loudspeakers to the walls was about 40 cm, and the minimum distance to the acoustic curtain was about 20 cm. The loudspeakers were positioned at an equal angular distance of 22.5°. The tweeters of the two-way speakers were located 160 cm above the floor and thus close to the average upright ear height of adults (C. C. Gordon et al., 1989). The floor within the loudspeaker array was covered with a thick carpet (IKEA Stoense), which also covered the loudspeaker bases. The monitors and computers used for the experiment were located adjacent to the loudspeaker array and were also shielded with acoustic foam panels.

The simulated auditory traffic scene was presented using the Ambisonics loudspeaker array. The 16 loudspeakers of the array were controlled by an audio converter (Ferrofish Pulse 16, 24 bit audio resolution, $f_s = 44.1$ kHz), which received the audio signals from an RME HDSPe RayDAT audio card in the Linux computer running TASCAR (Ubuntu 16.04 LTS, Intel Core i9-9900 K CPU @ 3.60 GHz, Quadro P1000). Acoustic calibration of the loudspeaker array was performed using the TASCAR Speaker Calibration Tool and a sound level meter (Norsonic Nor131 with Roga MP40 free field microphone) placed in the center of the loudspeaker array and 165 cm above the floor. During calibration, level differences between the 16 loudspeakers were compensated and the sound pressure levels of a point source and a diffuse sound field were calibrated. The sound levels from the calibrated microphones mounted on the cars during the vehicle recordings on the test track were used to set the sound levels of the simulated sound

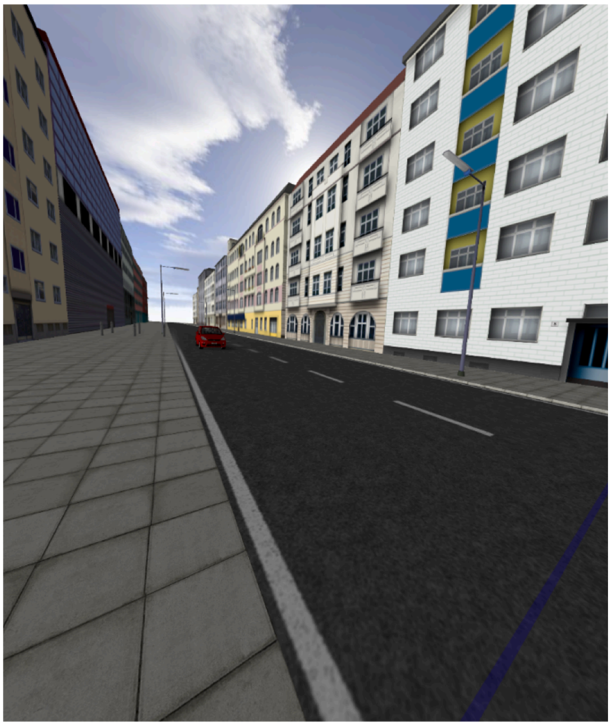

Fig. 1. Monoscopic view of the visual virtual street scene and the approaching vehicle, from the participants' perspective.

sources.

2.1.3. *Visual VR simulation*

The interactive auditory VR simulation was combined with visual VR simulations of the traffic scenes. The participants viewed the visual traffic scene stereoscopically wearing a head-mounted display (HTC Vive Pro; 1440×1600 pixels per eye, 90 Hz frame refresh rate, 110° field of view). Laser-based head and motion tracking of the HTC Vive Pro enabled the transformation of real head movements into virtual head movements, thus allowing participants to explore the visual scene. The height of the virtual camera above the simulated floor corresponded to the real eye level of the test person, as recorded by the head tracking. The simulations were created using the VR-software WorldViz Vizard 5.0 on a Windows computer (Intel Core i9-9900X CPU @ 3.50 GHz, Nvidia Quadro RTX 4000). The Vizard control script also sent commands controlling the corresponding acoustic simulations in TASCAR via the OSC network protocol (http://opensoundcontrol.org/), so that the auditory and visual VR simulations were synchronized in time.

2.2. *Stimuli and procedure*

2.2.1. *Audio-visual simulated traffic scenes*

In the experiment, audio-visual virtual traffic scenes were presented in which a vehicle approached the position of the participant at a constant speed and on a straight trajectory on the right lane of a two-lane road (see Fig. 1). Participants experienced the virtual scene from a pedestrian perspective. Their position in the virtual scene was 1 m away from the right curb, typical for a pedestrian intending to cross the road. In one of the two modality conditions, the approaching car was audible and visible (audiovisual presentation of the vehicle, AV). In the other modality condition, only the street scene was visible while the car was not visible but audible (auditory-only presentation of the vehicle, A-only).

The visual virtual road scene was modelled after the Eislebener Straße in Berlin, using 3ds Max 2020.2 and 3D data provided by the Senate Department for Urban Development and Housing of the City of Berlin (https://www.stadtentwicklung.berlin.de/planen/stadtmodelle/de/digitale_innenstadt/3d/index.shtml). The 3D model of it depicted an urban two-lane road (length approx. 300 m, width 6.5 m, lane width 3.25 m) without bends or curves as well as a uniform, gapless front of houses at both roadsides (see Fig. 1). The distance between the house fronts on the right side of the road and the right lane marker was 8.4 m. The distance between the house fronts on the left side of the road and the right lane marker was 15.6 m. Unlike in the original street, the virtual street scene did not include any bicycles, signs, parking vehicles, etc. White road markings were added as well as a blue line reaching from one side of the road to the other. The blue line was placed at a distance of 50 cm to the left of the participant's position in the virtual scene and served for orientation in the virtual environment. The approaching car was modeled after a red Mitsubishi Colt (L × W × H = 3.810 m × 1.895 m × 1.520 m). A male avatar with a neutral facial expression was presented on the driver's seat. The same visual car model was presented on all trials. It thus did not differ between the vehicle types (ICEV versus EV) or loudness levels.

In the simulated acoustic scene, the geometry of the ground surface and the house fronts were identical to the visual scene. The surfaces were simulated with plausible acoustic reflection properties. Based on ISO 9613-2:1999-10 (1999), the reflectance of the floor surface was set to $\rho = I_r/I_0 = 1.0$, where I_r is the acoustic intensity of the reflected sound wave and I_0 is the intensity of the incoming wave. Thus, we modeled both the road surface and the adjacent ground surface areas as acoustically hard. Based on the same standard, the reflectance of the house fronts was set to $\rho = 0.8$. The sound reflections were modeled with an IIR low pass filter of first order with a cut-off frequency of 5 kHz. The parameter *scattering* in TASCAR describes random deviations in the sound reflection and was set to a value of 0.5. A first-order Ambisonics recording from a quiet residential area was presented as background noise ($L_{Aeq} = 37.5$ dB).

Fig. 2 shows the sound spectra of the ICEV (left column) and the EV (right column) at the three different constant speeds (rows) in the simulated urban scene. The spectra were analyzed with the vehicles placed at a static position in the simulated urban scene, which was 10 m down the road from the listener's position against the direction of travel of the vehicles. The sounds were recorded by a virtual omnidirectional microphone placed at the same virtual position as the listener's head in the experiment. The recorded audio signals were normalized to an RMS level of 0 dB. For the ICEV, the characteristic and rotational-speed dependent harmonics are clearly visible. The EV spectra do not show any pronounced harmonics, as expected.

2.2.2. *Time-to-collision estimation: Prediction-motion task*

Time-to-collision estimates were obtained in a prediction-motion task (Schiff & Detwiler, 1979), which is one of the tasks most frequently used in the literature to study TTC estimation. Participants wore the head-mounted display, stood in the center of the loudspeaker array and experienced the simulated audiovisual traffic scene described above. Their position in the virtual scene was 50 cm away from the right curb, typical for a pedestrian intending to cross the road. When they turned their heads to the left side, they were able to see the car along the road. They were instructed to press a button on the controller to start the car's motion. The approaching car was presented for 3 s before it was "occluded", that is, it was then no longer audible and visible. On each trial, we selected a random time interval of 3 s from the available recording duration. Thus, the sound of the car differed slightly from trial to trial for each combination of vehicle type and speed, increasing the ecological validity. The temporal and spatial distance of the car at occlusion was defined by the different simulated TTCs and velocities (see below). Participants were instructed to pull the trigger of the controller when they thought that the approaching vehicle would have reached the blue line on the road, if the vehicle had continued to move towards them with the same constant velocity after occlusion. The time interval between the occlusion and the manual response was taken as the

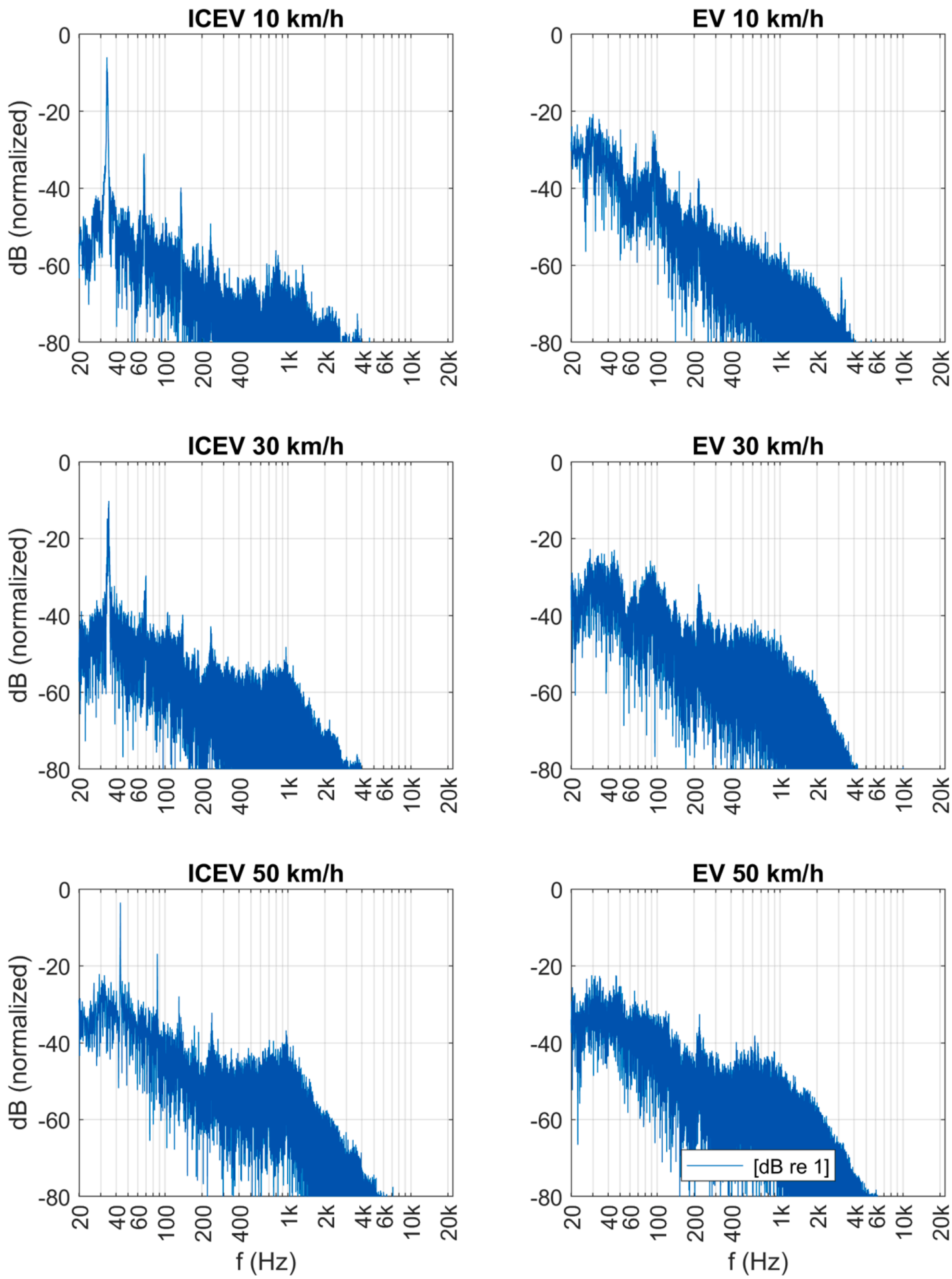

Fig. 2. Sound spectra of the ICEV (left column) and the EV (right column) at 10, 30 and 50 km/h (rows) in the simulated urban scene. The spectra were recorded with the vehicles placed at a static position in the virtual scene, at a distance of 10 m down the road from the virtual position of the listener.

participant's estimate of the TTC of the vehicle at the moment of occlusion. In one experimental condition (audiovisual, AV), the vehicles were visible and audible, while in the other experimental condition (auditory-only, A-only), the car was audible but not visible. In the A-only condition, the participants still saw the virtual street scene on the HMD. To get familiar with the experimental setting and the task, participants completed two training blocks with 15 trials each. In one training block, the car was audible and visible, in the other training block it was only audible, the order of which was also counterbalanced across participants (as described below). Subsequently, participants completed 16 experimental blocks. After each block, participants were given the opportunity to take a break. To detect potential motion-sickness symptoms throughout the experiment, the participants rated their motion-sickness on the Fast Motion-Sickness Scale ranging from 0 ("no sickness at all") to

20 (“frank sickness”) (Keshavarz & Hecht, 2011) after every second block. No issues with motion sickness occurred, except for a single participant, who did not complete the experiment due to motion-sickness symptoms.

In the TTC-estimation task, all participants received all experimental conditions in a within-subjects design. The vehicles approached at constant *velocities* of 10, 30, and 50 km/h, all of which can be considered relevant in urban traffic scenarios. Each *vehicle type* (ICEV and EV) was presented at two different *loudness levels*, at each of the three speeds. At the loudness level L_{ICEV}, the ICEV was presented at its original sound level (as recorded on the test track), and the EV was presented at the subjectively same loudness, based on the individual loudness matches (see below). At the loudness level $L_{\mathrm{ICEV+10\ dB}}$, the level of both vehicles was increased by 10 dB relative to the level L_{ICEV}. Thus, in this condition the vehicle sounds were approximately two times louder than at the loudness level L_{ICEV} (Jesteadt & Leibold, 2011), but the two vehicles were still matched in loudness. We presented three different *TTCs at occlusion onset* (2.0 s, 3.5 s, 5.0 s), defined as the time the vehicle would have needed to arrive at the participant's position after the moment in which it disappeared from the display. The resulting 36 combinations of vehicle type, loudness level, speed, and TTC were presented once and in a randomized order within each of the 16 blocks. The *modality condition* (A versus AV) was varied blockwise and changed after each completed block. The order was counterbalanced across participants. In the first block, the car was visible and audible (AV) for participants with an odd code number, while it was only audible (A-only) for participants with an even code number. The total of 72 experimental conditions (modality condition × vehicle type × loudness level × speed × TTC) were presented 8 times, resulting in 576 experimental trials per participant.

2.2.3. Loudness-matching task

Before the TTC-estimation task, we obtained individual loudness matches between the sound of the ICEV and the EV at each of the three presented constant velocities (10, 30 and 50 km/h). These were used to present the ICEV and the EV at equal loudness in the TTC-estimation task. Because the loudness-matching task presented only auditory stimuli, participants did not wear the head-mounted display. They stood in the center of the loudspeaker array inside the lab space, just as in the main TTC-estimation task. Participants were instructed to listen to several pairs of audio signals, and to indicate for each pair by pressing one of two response buttons whether they had perceived the first or the second audio signal as louder. Each sound pair consisted of the audio signals of an EV and an ICEV, both at the same speed of either 10, 30, or 50 km/h. The acoustic recordings of the two vehicles available at the three velocities ranged in duration between 7 s and 30 s. On each trial, we extracted a random time interval of 1000 ms from each of the two vehicle recordings to be presented, in order to increase the ecological validity. The two 1000 ms vehicle sounds were presented with a silent inter-stimulus interval of 800 ms. Using TASCAR, we simulated the front of the cars to be at a static position on the simulated road at a distance of 3 m down the road from the participant. When the response had been given, the next trial started with an inter-trial interval of 500 ms.

In each experimental block, only one pair of sounds of ICEV and EV at the same velocity was presented. Using an adaptive procedure, we adjusted the sound level of one of the vehicle sounds according to the participants' responses. The sound being adapted in level is termed the *comparison* in the following. The level of the other sound, termed the *standard*, was fixed and corresponded to the originally recorded sound level of the vehicle at a distance of 3 m down the road from the listener. Using four randomly interleaved adaptive tracks for each of the three velocities, we varied whether the ICEV or the EV served as the comparison, and whether the comparison was presented first or second within the trial, aiming to reduce the effect of response biases (cf. Buus, Florentine, & Poulsen, 1997; Oberfeld et al., 2012). In each adaptive track, we applied a 1-up, 1-down adaptive rule (Levitt, 1971) tracking the 50%-point on the psychometric function. Each adaptive track started with the comparison at the original recorded sound level of the respective vehicle. If the participant responded that the comparison was louder than the standard, the level of the comparison presented on the subsequent trial was decreased; otherwise, it was increased. The initial increment or decrement in level was 5 dB. After four reversals (i.e., “peaks” and “valleys” in the adaptive track; Levitt, 1971), the track continued with a smaller step size of 2 dB until either another eight reversals had occurred or 50 trials had been presented, whichever happened first. If a track was completed according to these rules, it was still presented with a low probability to avoid presenting only a few tracks that were not yet finished at the end of the loudness matching task. For each adaptive track, the arithmetic mean of the comparison levels at the final eight reversals (with small step size) was used to determine the change in sound level required to make the comparison equally loud as the standard (i.e., the loudness match). If < 4 reversals with small step size were obtained or if the standard deviation of the comparison level at the reversals with small step size was greater or equal to 5 dB, the corresponding adaptive track was excluded from the analysis. Based on these criteria, 5 adaptive tracks were excluded.

Three blocks were presented in random order, each presenting one of the three velocities. Subsequently, these three blocks were repeated again in random order, so that each participant completed a total of six blocks. For each participant, the loudness match between the EV and the ICEV, averaged across the two blocks presented for each of the three velocities, was used for setting the sound levels of the EV in the following TTC estimation task.

2.2.4. Experimental sessions

The experiment consisted of two sessions. In the first session, participants received information about the upcoming experiment, gave written informed consent and completed the vision and hearing tests. The experimenter additionally measured the participants' ear height and inter-pupillary distance. The individual ear height was used to set the height of the simulated receiver above the ground surface accordingly, and the individual inter-pupillary distance was used to adjust the distance between the two displays of the HMD. The loudness matching task was also part of the first session, while the TTC-estimation task was presented in the second session. At the end of the second session, participants filled in a questionnaire asking for demographic data. The experiment included two obligatory longer pauses with a duration of at least 15 min each. In addition, participants had a shorter break after each experimental block (i.e., after each 15–20 min of testing). The actual time on the TTC-estimation task was about 2 h. The total duration of the experiment including both sessions and breaks was about 3–5 h.

2.3. Participants

We recruited 29 participants, but one participant aborted the experiment due to motion-sickness symptoms during testing. The data of this participant were excluded from the analyses. The sample thus comprised 28 participants (22 female, 6 male; age: $M = 24.21$ years, $SD = 5.70$ years) with normal or corrected-to-normal vision as well as normal hearing. Visual acuity, stereoscopic vision and hearing ability were assessed prior to testing. Audiometric hearing thresholds were measured bilaterally using Békésy audiometry (von Békésy, 1947) with pulsed 270 ms pure tones. The hearing thresholds on both ears of all participants were measured in a frequency spectrum between 125 Hz and 4000 Hz and did not exceed 20 dB HL. The average asymmetry in the hearing thresholds between left and right ear at octave frequencies between 125 Hz and 4 kHz was smaller than 11 dB, except for one participant with an asymmetry of 20 dB. Landolt's C test of the Freiburg Visual Acuity Test (Bach, 1996) was used to test for the required visual acuity of ≥ 1.0. The stereoscopic acuity was tested with a Titmus Test (Bennett & Rabbetts, 1998), presented on the HMD. Here, the participants had to provide the correct response on at least 6 of 9 trials, which presented binocular disparities of 800, 400, 200, 140, 100, 80, 60, 50

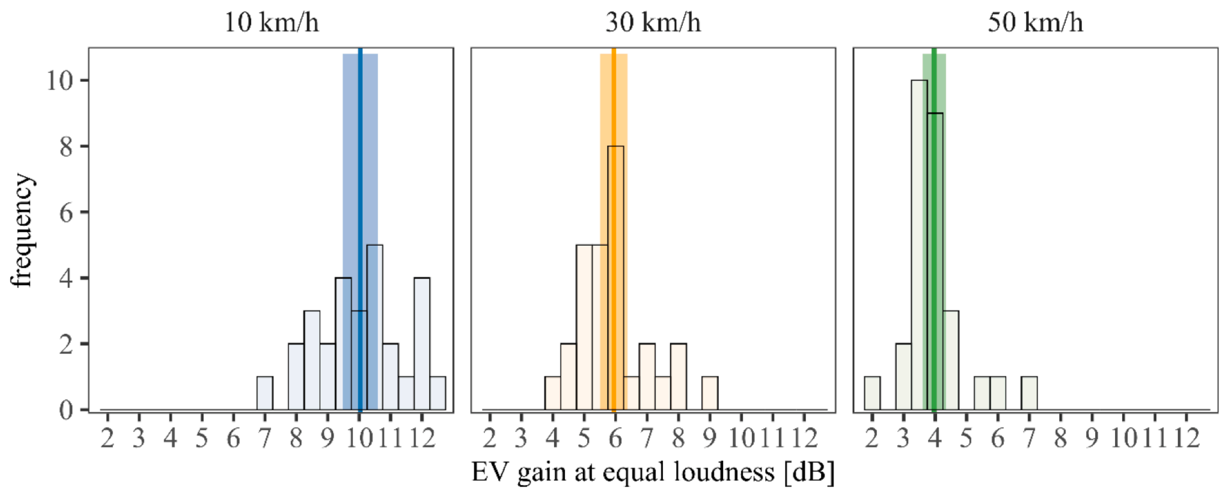

Fig. 3. Histograms of the individual change in the sound level (gain) of the EV required to match the loudness of the ICEV at the same speed. Positive values indicate that the sound level of the EV had to be increased in order to match the loudness of the ICEV at the same speed. The colored vertical lines represent the mean gain per speed. The colored transparent boxes around the mean gain indicate 95 % confidence intervals.

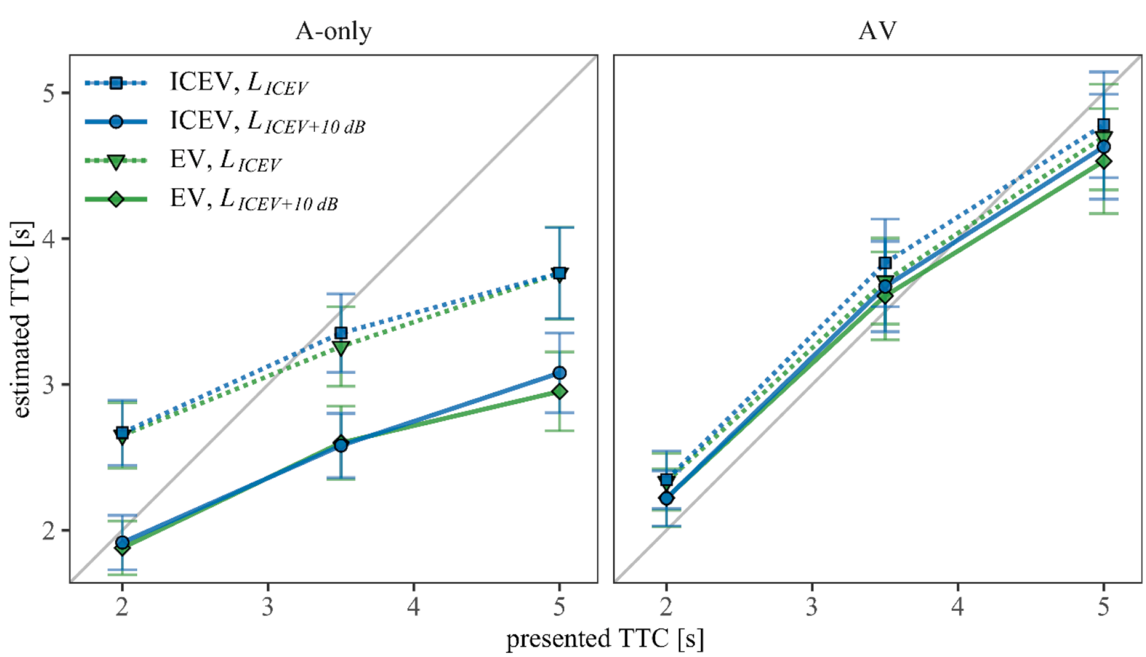

Fig. 4. Mean estimated TTC as function of the presented TTC (*x*-axis) and modality condition (left panel: A-only, car not visible; right panel: AV, car visible). Blue squares: internal combustion engine vehicle (ICEV) at its original loudness level (L_{ICEV}). Blue circles: ICEV at L_{ICEV} + 10 dB. Green triangles: electric vehicle (EV) at L_{ICEV}. Green diamonds: EV at L_{ICEV} + 10 dB. Solid lines: original loudness level of the ICEV (L_{ICEV}). Dotted lines: loudness level increased by 10 dB ($L_{ICEV+10dB}$). Error bars represent ± 1 *SE* of the mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and 40 s of arc.

The experiment was conducted in accordance with the ethical principles of the Declaration of Helsinki and the Ethics Committee of the Institute of Psychology of the Johannes Gutenberg University Mainz (approval number: 2019-JGU-psychEK-S011). All participants volunteered for course credit or monetary reward (7 € per hour). Prior to testing, but after study information was provided and possible risks were explained, they gave their written consent.

3. Results

3.1. Loudness matches

For each sound pair of the EV and ICEV at the same speed of either 10, 30 or 50 km/h, we calculated the change in sound level (gain) required to make the EV equally loud as the ICEV, individually for each participant. Positive values indicate that the sound level of the EV had to be increased in order to match the loudness of the ICEV at the same speed. As Fig. 3 shows, the level of the EV had to be increased at all speeds to make it equally loud as the ICEV. On average, the required gain was highest at the slowest speed and lowest at the fastest speed. This is the expected pattern, because the tire noise, which can be assumed to be rather similar for the two car types, dominates at speeds above 20 km/h, so that differences in the engine sound levels should have the strongest effect at the slowest speed. At the slowest speed, the sound level of the EV had to be increased in level by 10.04 dB on average to make it equally loud as the ICEV. Thus, at its original sound level, the EV was about half as loud as the ICEV at a speed of 10 km/h. A repeated-measures ANOVA (rmANOVA), using a multivariate approach, showed a significant effect of speed on the loudness matches, $F(2, 26) = 255.97$, $p < .001$, $\eta_p^2 = 0.95$.

Table 1
Results of the rmANOVA for the A-only condition (cars not visible). Displayed are *F*-values, numerator degrees of freedom, denominator degrees of freedom, *p*-values and partial η^2 (η_p^2).

Factor	F	df_{Num}	df_{Den}	p	η_p^2
Loudness	**119.41**	**1**	**27**	**<0.001**	**0.82**
Vehicle	0.87	1	27	0.359	0.03
TTC	**56.26**	**2**	**26**	**<0.001**	**0.81**
Velocity	**34.72**	**2**	**26**	**<0.001**	**0.73**
Loudness × vehicle	0.05	1	27	0.823	0.00
Loudness × TTC	0.51	2	26	0.610	0.04
Vehicle × TTC	0.28	2	26	0.762	0.02
Loudness × velocity	**12.47**	**2**	**26**	**<0.001**	**0.49**
Vehicle × velocity	**3.72**	**2**	**26**	**0.038**	**0.22**
TTC × velocity	**14.72**	**4**	**24**	**<0.001**	**0.71**
Loudness × vehicle × TTC	2.12	2	26	0.140	0.14
Loudness × vehicle × velocity	0.43	2	26	0.656	0.03
Loudness × TTC × velocity	2.69	4	24	0.055	0.31
Vehicle × TTC × velocity	1.34	4	24	0.282	0.18
Loudness × vehicle × TTC × velocity	1.94	4	24	0.136	0.24

Three two-sided paired-samples *t*-tests with a correction for multiple comparisons (Hochberg, 1988) showed that the mean change in gain of the EV differed significantly between each pair of speeds at an α-level of 0.001 (10 vs. 30 km/h: $t(27) = 16.10$, Cohen's (1984) $d_z = 3.04$; 10 vs. 50 km/h: $t(27) = 23.01$, $d_z = 4.35$; 30 vs. 50 km/h: $t(27) = 10.64$, $d_z = 2.01$).

3.2. TTC estimates

Prior to the analyses, we applied a Tukey (1977) criterion that excluded outlying TTC estimates in each combination of participant and experimental condition (modality condition × vehicle type × speed × loudness level × TTC). From a total number of 16,128 TTC estimates, 65 were located more than three interquartile ranges above the third quartile or below the first quartile and were therefore excluded (0.40%). We aggregated the remaining TTC estimates for each combination of

Table 2
Results of the rmANOVA for the AV condition with visible cars. Displayed are *F*-values, numerator degrees of freedom, denominator degrees of freedom, *p*-values and partial η^2 (η_p^2).

Factor	F	df_{Num}	df_{Den}	p	η_p^2
Loudness	**26.19**	**1**	**27**	**<0.001**	**0.49**
Vehicle	4.21	1	27	0.050	0.13
TTC	**83.56**	**2**	**26**	**<0.001**	**0.87**
Velocity	**36.83**	**2**	**26**	**<0.001**	**0.74**
Loudness × vehicle	0.16	1	27	0.694	0.01
Loudness × TTC	0.43	2	26	0.654	0.03
Vehicle × TTC	2.38	2	26	0.112	0.15
Loudness × velocity	**19.77**	**2**	**26**	**<0.001**	**0.60**
Vehicle × velocity	2.52	2	26	0.100	0.16
TTC × velocity	**6.73**	**4**	**24**	**0.001**	**0.53**
Loudness × vehicle × TTC	0.14	2	26	0.873	0.01
Loudness × vehicle × velocity	0.29	2	26	0.750	0.021
Loudness × TTC × velocity	0.97	4	24	0.442	0.14
Vehicle × TTC × velocity	0.72	4	24	0.588	0.11
Loudness × vehicle × TTC × velocity	1.47	4	24	0.241	0.20

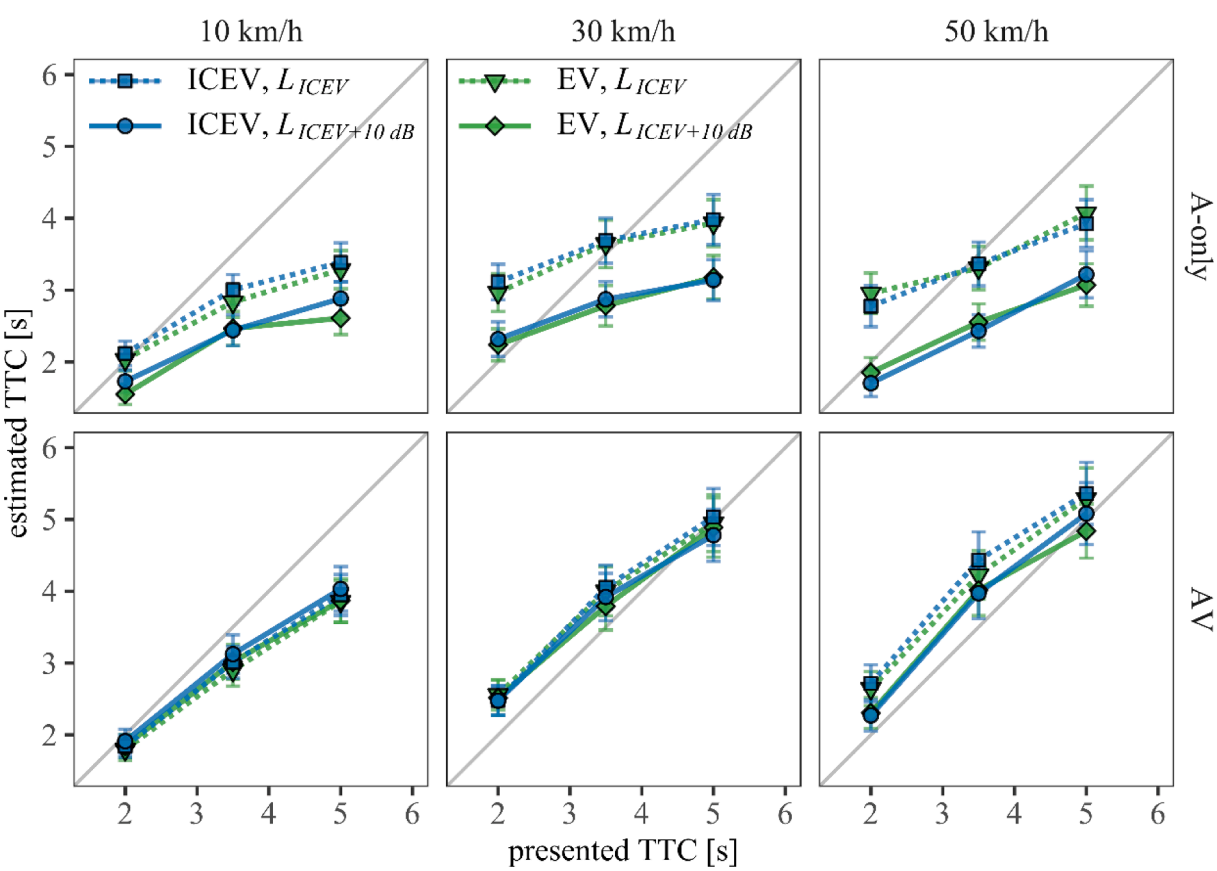

Fig. 5. Mean estimated TTC as function of the presented TTC (*x*-axis), velocity (columns), and modality condition (upper row: A-only, car not visible; lower row: AV, car visible). Blue squares: internal combustion engine vehicle (ICEV) at original loudness (L_{ICEV}). Blue circles: ICEV at L_{ICEV} + 10 dB. Green triangles: electric vehicle (EV) at L_{ICEV}. Green diamonds: EV at L_{ICEV} + 10 dB. Dotted lines: L_{ICEV} + 10 dB. Solid lines: L_{ICEV}. Error bars represent ± 1 *SE* of the mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table A1
Results of the five-factorial rmANOVA using a multivariate approach. Displayed are *F*-values, numerator degrees of freedom, denominator degrees of freedom, *p*-values and partial η² (η_p^2).

Factor	F	df_{Num}	df_{Den}	p	η_p^2
Modality	**47.88**	**1**	**27**	**<0.001**	**0.64**
Loudness	**117.77**	**1**	**27**	**<0.001**	**0.81**
Vehicle	**4.44**	**1**	**27**	**0.045**	**0.14**
TTC	**75.83**	**2**	**26**	**<0.001**	**0.85**
Velocity	**40.06**	**2**	**26**	**<0.001**	**0.75**
Modality × loudness	**89.71**	**1**	**27**	**<0.001**	**0.77**
Modality × vehicle	0.13	1	27	0.724	0.00
Loudness × vehicle	0.01	1	27	0.911	0.00
Modality × TTC	**73.75**	**2**	**26**	**<0.001**	**0.85**
Loudness × TTC	0.58	2	26	0.567	0.04
Vehicle × TTC	1.69	2	26	0.205	0.11
Modality × velocity	**32.90**	**2**	**26**	**<0.001**	**0.71**
Loudness × velocity	**18.94**	**2**	**26**	**<0.001**	**0.59**
Vehicle × velocity	3.67	2	26	0.039	0.22
TTC × velocity	**9.32**	**4**	**24**	**<0.001**	**0.61**
Modality × loudness × vehicle	0.20	1	27	0.656	0.01
Modality × loudness × TTC	0.37	2	26	0.697	0.03
Modality × vehicle × TTC	0.44	2	26	0.650	0.03
Loudness × vehicle × TTC	1.20	2	26	0.318	0.08
Modality × loudness × velocity	2.42	2	26	0.109	0.16
Modality × vehicle × velocity	2.66	2	26	0.090	0.17
Loudness × vehicle × velocity	0.40	2	26	0.675	0.03
Modality × TTC × velocity	**12.08**	**4**	**24**	**<0.001**	**0.67**
Loudness × TTC × velocity	1.38	4	24	0.271	0.17
Vehicle × TTC × velocity	1.41	4	24	0.261	0.19
Modality × loudness × vehicle × TTC	0.41	2	26	0.673	0.03
Modality × loudness × vehicle × Velocity	0.40	2	26	0.673	0.03
Modality × loudness × TTC × velocity	0.92	4	24	0.460	0.13
Modality × vehicle × TTC × velocity	0.67	4	24	0.620	0.10
Loudness × vehicle × TTC × velocity	2.19	4	24	0.101	0.27
Modality × loudness × vehicle × TTC × velocity	0.36	4	24	0.836	0.06

participant and experimental condition. Fig. 4 shows the mean estimated TTC as a function of the presented TTC, for each combination of modality condition, car, and loudness level. For the A-only condition with invisible vehicles (left panel), the TTC was estimated considerably shorter at the higher loudness level ($L_{\mathrm{ICEV+10\ dB}}$; solid lines and symbols) than at the lower loudness level (L_{ICEV}; dashed lines and open symbols). For the AV condition with visible vehicles (right panel), we observed a similar pattern, although the difference between the estimated TTCs at the two different loudness levels was less pronounced. At each loudness level, the TTC estimates for the loudness-matched ICEV and EV were virtually identical. Thus, the descriptive data show longer TTC estimates for quieter vehicles, even when full visual information was available, while the vehicle type (ICEV versus EV) had virtually no effect at a given loudness level.

To examine the effect of loudness on the TTC estimations when the car was not visible and when it was visible, we conducted separate rmANOVAs, one for each modality condition, with a multivariate approach, and followed up with pairwise paired-samples *t*-tests with Hochberg-correction (Hochberg, 1988), where necessary. We first focus on the effects of loudness level and vehicle type, which are central to the present study, and subsequently discuss effects that are related to TTC estimation in more general terms.

Table 1 shows the results of the rmANOVA for the A-only condition. In this condition, we observed a strong significant effect of the loudness level, as expected. The TTC was estimated significantly shorter for louder vehicles ($M = 2.50$ s, $SD = 1.19$ s) compared to quieter vehicles ($M = 3.24$ s, $SD = 1.41$ s), which is compatible with an intensity-arrival effect (DeLucia, et al., 2016). The effect of loudness level on the TTC estimates was large in terms of the difference in milliseconds (742 ms, $SD = 359$ ms) between the mean estimates. It was also observed very consistently across participants, indicated by a large statistical effect size of $d_z = 2.07$. The loudness level × velocity interaction was significant. As shown in Fig. 5, the effect of loudness on the TTC estimates was smaller at the lowest compared to the two higher velocities in the A-only condition.

The effect of vehicle type was not significant. Thus, at equal loudness, the different acoustic signatures of the ICEV and EV did not play a significant role. As displayed in the upper left panel of Fig. 5, the largest but still only slight descriptive difference between the TTC estimates for ICEVs and EVs in the A-only condition was observed at a velocity of 10 km/h. This pattern is compatible with the more prominent difference between the acoustic signature of the ICEV and EV, because tire noise dominates at higher speeds. A dependency of the effect of vehicle type on the velocity was, indeed, indicated by a significant vehicle type × velocity interaction in the rmANOVA, but follow-up paired samples *t*-tests between the TTC estimates for the ICEV and the EV at each velocity were all non-significant with Hochberg correction at an α-level of 0.050.

Table 2 shows the results of the rmANOVA for conditions with visible cars (AV) and loudness-matched ICEV and EV. Even when full visual information about the approaching vehicle was available, we still observed a significant effect of the loudness level, compatible with an intensity-arrival effect, although reduced in comparison to the A-only condition (see Fig. 4). As in the A-only condition, the TTC of louder vehicles ($M = 3.48$ s, $SD = 1.50$ s) was estimated shorter than for quieter vehicles ($M = 3.62$ s, $SD = 1.49$ s). The statistical effect size was also reduced in comparison to the A-only condition, $d_z = 0.97$, but can still be considered large according to Cohen (1988). There was a non-significant trend towards shorter TTC estimates for the EV compared to the ICEV. As in the A-only condition, the loudness × velocity interaction was significant, showing a stronger effect of loudness level at the higher velocities compared to the lowest. Post-hoc paired samples *t*-tests with Hochberg correction showed that at an α-level of 0.05 the effect of loudness was significant and most prominent at 50 km/h, $t(27) = 7.53$, $d_z = 1.42$, and also significant at 30 km/h, $t(27) = 3.10$, $d_z = 0.59$. At 10 km/h, the loudness effect did not reach significance, $t(27) = 1.92$, $d_z = 0.36$. These effect sizes highlight the relevance of vehicle loudness at higher velocities in the AV condition.

3.3. Effects involving presented TTC and velocity

Turning to the effects not involving loudness level or vehicle type, in the A-only condition, the effect of presented TTC on the estimated TTCs was significant. Thus, the participants were able to perceive the differences in presented TTC, even when only auditory information was available. However, the TTC estimates showed a pronounced regression-to-the-mean pattern. An increase of the presented TTC by 3.0 s (2.0 s versus 5.0 s) resulted in a change in the mean estimated TTC of <1.5 s. There was a significant effect of velocity on the TTC estimates. On average, the TTC estimates were longest at 30 km/h, shortest at 10 km/h, and intermediate at 50 km/h. The TTC × velocity interaction was also significant. At the longest presented TTC, the mean TTC estimates at the two highest speeds were virtually identical, while at the shorter TTCs, the mean TTC estimate was longer at 30 km/h compared to 50 km/h (Fig. 5).

In the AV condition, the effect of presented TTC on the TTC estimates was also significant, and the mean estimates were closer to the presented TTC, showing only a weak regression-to-the mean pattern. Participants slightly overestimated presented TTCs of 2.0 s and 3.5 s by approximately 200 ms, while they underestimated the longest TTC of 5.0 s by approximately 400 ms. The effect of speed was significant. In contrast to the A-only condition, the estimated TTC increased monotonically with increasing velocity in the AV condition. The increase in estimated TTC with increasing velocity is compatible with a size-arrival effect (DeLucia, 1991). At a given presented TTC and given the constant size of the simulated visual object, the optical size of faster-travelling vehicles was smaller than for slower-travelling vehicles. Because in the AV condition the cars were also audible, the intensity-arrival effect could have played an additional role. The TTC × velocity interaction was also significant. The effect of velocity on the TTC estimates increased with the presented TTC.

3.4. Effects involving the modality condition (A-only versus audio-visual)

To confirm the differences in the TTC estimates between the A-only and the AV condition, we additionally computed a five-factorial rmANOVA with a multivariate approach, including the additional factor modality condition. The results are shown in the Appendix (Table A.1). This analysis confirmed that the loudness level had a significant effect on the TTC estimates across the two modality conditions. A significant loudness × modality condition interaction confirmed that this effect was more pronounced in the A-only condition where the cars were not visible (mean difference between the TTC estimates at the two loudness levels $M_{Dif} = 0.74$ s) than in the AV condition where the cars were both audible and visible ($M_{Dif} = 0.14$ s). That is, participants relied more strongly on the vehicle loudness when no visual information were present. Across modality conditions, the loudness × velocity interaction was significant, showing a stronger effect of loudness at the two higher velocities compared to the lowest velocity. Also, the effect of vehicle type was significant across the two modality conditions, although statistically and numerically small, with slightly longer mean TTC estimates for the ICEV than for the EV.

The modality condition × TTC interaction was significant, confirming the stronger regression-to-the-mean pattern in the A-only compared to the AV condition. The modality condition × velocity interaction was also significant, confirming the different dependence of the mean TTC estimates on velocity in the two modality conditions, which we described above. Finally, there was a significant modality condition × TTC × velocity interaction. As Fig. 5 shows, the mean TTC estimates increased between 10 km/h and 30 km/h for all presented TTCs and both modalities. For visible cars (AV), the TTC estimates increased slightly between 30 km/h and 50 km/h at the two longer TTCs but were at a similar level at the shortest TTC. In contrast, in the A-only condition, the TTC estimates at 30 km/h and 50 km/h were at a similar level at the longest TTC, but even decreased between 30 km/h and 50 km/h at the two shorter TTCs indicating a strong underestimation of TTC. As different combinations of velocity and TTC result in different distances, this three-way interaction might hint at a different reliance on distance information in the AV compared to the A-only condition. DeLucia, et al. (2016) and Keshavarz, et al. (2017) reported that auditory TTC estimation is more strongly based on distance-related heuristics than visual TTC estimation.

4. Discussion

We introduce a novel interactive audiovisual virtual-reality system for traffic scenarios that allows us to study TTC estimations for approaching vehicles with a higher degree of realism than in previous studies on TTC estimation and road crossing decisions. The system presents highly realistic and physically plausible auditory simulations of approaching vehicles, based on recordings of real vehicles, acoustic modeling of the sound propagation from the sound sources to the receiver, and rendering of the dynamic spatial sound field via sound field synthesis (higher-order Ambisonics). The simulated auditory scenes provide a full range of dynamic monaural and binaural cues available in a real traffic scenario. The auditory simulations are complemented by interactive high-fidelity visual simulations presented on a VR headset with head tracking.

Using this system, we examined TTC estimations from a pedestrian perspective for ICEVs and EVs with the same loudness, approaching at a constant speed (10 to 50 km/h) and presented only auditorily (A-only; car not visible) or audiovisually (AV; car audible and visible). We varied the loudness level of the approaching vehicles to investigate whether the TTC overestimation for softer compared to louder approaching sound sources observed in previous studies (DeLucia, et al., 2016; Keshavarz, et al., 2017), which used somewhat impoverished and artificial auditory stimuli, is also found when realistic acoustic vehicle simulations are presented. As expected, the results showed TTC overestimation for softer compared to louder approaching sound sources (intensity-arrival effect; DeLucia, et al., 2016), which not only occurred within the A-only but also within the AVcondition. Thus, this study contributes to the understanding of auditory heuristics processing with and without additional visual information, which has important scientific and practical implications.

With highly realistic and thus ecologically valid vehicle simulations, the intensity-arrival effect was massive in the A-only condition of the present study (mean difference of 740 ms between the TTC estimates at the lower compared to the higher loudness level, $d_z = 2.06$). It was still significant and relatively large in terms of statistical effect size ($d_z = 0.97$) even when full visual information about the approaching vehicle was available. From the viewpoint of fundamental research, the results of the present study thus confirm that audiovisual TTC estimates are to a significant extent based on the loudness of the approaching object (DeLucia, et al., 2016; Keshavarz, et al., 2017), which is a heuristic cue because the loudness of a sound source does not directly specify its TTC. From a practical perspective, our results indicate that in real traffic scenarios, quieter vehicles might lead to longer TTC estimates, and thus potentially to riskier road crossing decisions compared to louder vehicles.

The role of loudness as a heuristic cue used for TTC estimation is particularly relevant in the context of increasing electric mobility. EVs are more silent than ICEVs, particularly at lower speeds (see Fig. 3). This results in impaired auditory detectability of EVs compared to ICEVs (e.g., Altinsoy, 2013; Grosse et al., 2013). To ensure road safety, legislative actions were therefore necessary to improve the auditory detectability of EVs by implementing acoustic vehicle alerting systems (AVAS), which emit artificial sounds at lower speeds. Nonetheless, the strong effect of vehicle loudness on the TTC estimates found in the present study shows that the effect of vehicle loudness is not restricted to detectability but also affects the perception of a vehicle's motion in a suprathreshold situation where the vehicle is clearly audible and has already been detected. This finding emphasizes the need to investigate and potentially adapt the design of AVAS technology not only with regard to detectability, but also regarding vehicle motion perception (see Wessels at al., under revision). In this context, it is noteworthy that there was no substantial difference between the TTC estimates for EVs and ICEVs when presented at the same loudness, which underlines the central role of loudness for TTC estimation. In other words, the effect of loudness on the TTC estimates did not interact with the different spectral characteristics of the two vehicle types. Thus, if an AVAS puts the loudness of an EV in the same range as the loudness of an ICEV, and not only at lower speeds, we would no longer expect a systematic difference between TTC estimates for the two vehicle types.

The role of loudness for TTC estimation is further emphasized by a consistent effect of loudness on the TTC estimates at all constant velocities. The effect of loudness on TTC estimates was stronger at the faster speeds compared to the lowest speed of 10 km/h. This finding is somewhat surprising because the difference in loudness caused by the 10 dB sound level difference between the two loudness levels can be assumed to be similar for the – on average – quieter vehicle sounds at 10 km/h compared to the – on average – louder vehicle sounds at the higher speeds. At all speeds, the sound levels were well above the detection threshold, and thus in a region where log loudness is a linear function of the sound level (e.g., Jesteadt & Leibold, 2011), so that an increase in sound level by 10 dB should have resulted in an increase in loudness by approximately a factor of 2 at all speeds. Therefore, additional research is required to identify the origin of this effect. For instance, pedestrians might rely on different acoustic cues when the vehicle travels at a low speed compared to when it travels at higher speeds.

In addition to the insights into the intensity-arrival effect and potential effects of vehicle type, the present experiment provides important fundamental information about TTC estimation based on auditory cues, because the degree of realism of the presented auditory stimuli was higher than in previous experiments on auditory TTC estimation. Although participants' TTC estimates increased when the presented TTC increased in the A-only condition (left panel in Fig. 4), an increase of the presented TTC by 3.0 s resulted in a change in the mean estimated TTC of $<$1.5 s. Thus, the auditory TTC estimates showed a pronounced "regression-to-the-mean" pattern, which is compatible with previous results (DeLucia, et al., 2016; Keshavarz, et al., 2017; Schiff & Oldak, 1990). Thus, the availability of a full set of physically plausible auditory TTC cues in the present experiment did not remove this effect. In fact, a regression-to-the-mean pattern has also been observed in many studies on visual TTC estimation (e.g., Hecht, Brendel, Wessels, & Bernhard, 2021; Heuer, 1993; McLeod & Ross, 1983; Oberfeld & Hecht, 2008), where it was particularly prominent when the fidelity of the presented stimuli was limited, as for example when viewing the approaching objects through an aperture (DeLucia & Liddell, 1998). In contrast, in the AV condition of the present experiment, where both auditory and visual information was presented with high fidelity, the TTC estimates showed only a weak regression-to-the-mean pattern (right panel in Fig. 4). Thus, the additional visual information promoted the accuracy of the TTC estimates in comparison to the A-only condition.

Another difference between auditory-only and audiovisual TTC estimates concerned the effect of velocity. In the AV condition, the estimated TTC increased monotonically with increasing velocity, which is compatible with a size-arrival effect (DeLucia, 1991) or a distance bias (Law et al., 1993). In contrast, in the A-only condition, the mean TTC estimates depended on the velocity in a non-monotonic way. It remains to be investigated whether this difference in the effect of velocity is related to the reduced distance information available in the A-only compared to the AV condition. In the A-only condition, the distance in the simulated traffic scenario can be assumed to have been signaled mainly by acoustic intensity (Kolarik, Moore, Zahorik, Cirstea, & Pardhan, 2016) and probably to a lesser extent by motion parallax (Genzel, Schutte, Brimijoin, MacNeilage, & Wiegrebe, 2018), while in the AV condition a full set of monocular and binocular visual distance cues was additionally available.

Several questions remain open due to limitations of our study and should be addressed in future research. First, it is unclear whether the reduced accuracy of TTC estimates in the A-only condition was mainly due to the fact that all participants possessed normal vision and can thus be assumed to use predominantly visual information in TTC judgments in everyday life, like when crossing a road. In a study by Schiff and Oldak (1990), congenitally blind or early-onset blind adults showed a much reduced regression-to-the-mean pattern in auditory-only TTC judgments for approaching vehicles compared to participants with normal vision. This finding is particularly striking since the auditory recordings presented in the study by Schiff and Oldak were of rather reduced quality (presentation of monophonic vehicle recordings on a single loudspeaker). We would expect early-onset blind persons to be able to estimate TTC even more precisely with more high-fidelity auditory simulations as used in our study. Additionally, it would be interesting to investigate whether extensive training of normally sighted people in auditory-only TTC estimation could improve the use of auditory cues and, thus, render the pattern of estimates similar to those of the visual-only or audiovisual TTC estimates.

Second, our results demonstrated that visual in addition to auditory information reduced the regression-to-the-mean pattern in participants with normal vision and hearing. Put differently, with additional visual information about the approaching vehicle, participants were able to judge its TTC considerably more accurately. However, we did not include a visual-only condition in the present experiment, in order to maintain a reasonable experimentation time. For this reason, the data do not provide detailed information about a potential benefit when auditory cues are available in addition to visual cues. Put differently, it remains to be investigated how auditory and visual cues are weighted relatively to each other when highly realistic audiovisual simulations as in the study at hand are presented rather than rather simplistic simulations as in our two previous studies (DeLucia, et al., 2016; Keshavarz, et al., 2017). In any case, the significant intensity-arrival effect observed in the AV condition clearly shows that our participants factored auditory cues into their TTC estimates even when full visual information was available.

Third, in terms of the acoustical simulations, it would be interesting to compare the sound quality of our simulation approach as well as the TTC estimation results to other simulation approaches, for example to pass-by auralizations that use source signals based on an inversion process of pass-by recordings that undo the effects of Doppler frequency shifts, spherical spreading, air-attenuation, etc. (Forssén, Hoffmann, & Kropp, 2018; Hoffmann & Kropp, 2019). Improvements of our simulations could potentially be achieved by using close-proximity recordings of the tire-road noise (Pereira, Soares, Silva, Sousa, & Freitas, 2021), and by modeling the directivity of the vehicle noise sources, which are at present implemented as point sources, similar to other auralization approaches (Pereira, et al., 2021).

Fourth, the road-tire noise can vary considerably across different tire types and road surfaces. The two recorded vehicles (ICEV and EV) were equipped with different tires (see Methods), which might have resulted in differences between the level of tire-road noise for the two vehicles. However, the important feature of the present experiment is that the two vehicles were presented at equal loudness, which was achieved by the individual loudness matches that were obtained before the TTC-estimation task started. For this reason, differences in the road-tire noise level in the original recordings of the ICEV and the EV played no role in the TTC-estimation task. Still, potential differences in tire-road noise due to the different types of tires may have contributed to the loudness difference between the two vehicles in the original recordings (see Fig. 2). Potentially, the different tires might have caused not only a different tire-sound level but also a different tire-road-noise frequency spectrum, which would then have increased the spectral difference between the two vehicle types. Before this background, it is even more interesting that the vehicle type had no substantial effect on the TTC estimations, indicating that when the two vehicles were presented at equal loudness, spectral differences did not play an important role, as we discussed earlier. Other road surfaces like cobblestones would also create a different situation, as the tire-road noise would be more dominant than the powertrain noise (Soares et al., 2017), and could also provide stronger speed cues than on a smooth asphalt surface. A higher level of tire-road noise would reduce the sound level difference between ICEVs and EVs (Fig. 3). Again, this does not affect the results of the present TTC experiment, because the two vehicle types were loudness-matched.

5. Conclusions

The results of the present study confirm the central role of vehicle loudness for TTC judgments in traffic scenarios. Using a novel simulation system that provides auditory and audiovisual simulations of approaching electric and conventional vehicles with a higher degree of realism than in previous studies, the data confirmed a) that humans are capable of estimating TTC based only on auditory cues (albeit with somewhat lower accuracy compared to audiovisual estimation), b) that the loudness of the approaching sound source has a strong effect on TTC estimates based on only auditory information but also when full visual information is available, and c) that vehicle loudness should therefore be considered as an important factor in traffic safety.

CRediT authorship contribution statement

Daniel Oberfeld: Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Writing – original draft, Writing – review & editing. **Marlene Wessels:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – original draft, Writing – review & editing. **David Büttner:** Conceptualization, Investigation, Methodology, Writing – review & editing.

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.

Acknowledgements

This work was supported by funding from Unfallforschung der Versicherer (udv.de) and Deutsche Forschungsgemeinschaft (www.dfg.de; OB 346/8-1; priority programme AUDICTIVE – SPP2236: *Auditory Cognition in Interactive Virtual Environments*) granted to Daniel Oberfeld. The funders had no role in study design, in the collection, analysis and interpretation of data, in the writing of the paper, and in the decision to submit the article for publication.

We are very grateful to Joachim Bös, Sören Wenzel, Micha Süs, Felix Reissner, Nicole Langer, Elena Slomski, and Christian Adams (Research group System Reliability, Adaptive Structures, and Machine Acoustics, Technical University of Darmstadt) for their help in organizing and conducting the vehicle recordings on the test track, to Stefan Leinen and Björn Reuper (Institute of Geodesy, Technical University of Darmstadt) for their support in the GPS position tracking, to Wolfgang Ellermeier (Institute of Psychology, Technical University of Darmstadt) for providing the dummy head, to Giso Grimm (Medical Physics, Uni Oldenburg) for his help in setting up the acoustic simulations in TASCAR, to Agnes Münch (Experimental Psychology, Uni Mainz) for her support in implementing the visual VR simulations, to Florian Völk (WindAcoustics, Windach) for his services in the post-processing of the vehicle recordings, to Isabela Araujo for her contributions to the quality control of the vehicle recordings, to Anna-Lena Spies, Christoph Keßler, Jonas Krämer and Sina Neuwardt for their help during the recordings, and to Vanessa Dimmler, Julia Guldan, Sophia Glaszner and Christoph Keßler for their help in data collection.

Appendix A

Table A.1

References

Ahrens, J., Rabenstein, R., Spors, S., 2014. Sound field synthesis for audio presentation. Acoustics Today 10 (2), 15–25.

Allen, J.B., Berkley, D.A., 1979. Image method for efficiently simulating small-room acoustics. Journal of the Acoustical Society of America 65 (4), 943–950. https://doi.org/10.1121/1.382599.

Altinsoy, E., 2013. *The detectability of conventional, hybrid and electric vehicle sounds by sighted, visually impaired and blind pedestrians.* Paper presented at the INTER-NOISE and NOISE-CON Congress and Conference Proceedings.

Altinsoy, E., Landgraf, J., Rosenkranz, R., Lachmann, M., Hagen, K., Schulze, C., Schlag, B., 2015. Wahrnehmung und Bewertung von Fahrzeugaußengeräuschen durch Fußgänger in verschiedenen Verkehrssituationen und unterschiedlichen Betriebszuständen. *FAT Schriftenreihe: Berichte der Bundesanstalt für Straßenwesen* (278).

Bach, M., 1996. The Freiburg Visual Acuity test: Automatic measurement of visual acuity. Optometry and Vision Science 73 (1), 49–53. https://doi.org/10.1097/00006324-199601000-00008.

Bennett, A.G., Rabbetts, R.B., 1998. Bennett and Rabbetts' clinical visual optics, 3rd ed. Butterworth-Heinemann, Oxford, Boston.

Buus, S., Florentine, M., Poulsen, T., 1997. Temporal integration of loudness, loudness discrimination and the form of the loudness function. Journal of the Acoustical Society of America 101 (2), 669–680.

Caird, J.K., Hancock, P.A., 1994. The perception of arrival time for different oncoming vehicles at an intersection. Ecological Psychology 6 (2), 83–109. https://doi.org/10.1207/s15326969eco0602_1.

Cavallo, V., Laurent, M., 1988. Visual information and skill level in time-to-collision estimation. Perception 17 (5), 623–632.

Cohen, J., 1988. Statistical power analysis for the behavioral sciences, 2nd ed. L. Erlbaum Associates, Hillsdale, N.J.

Daniel, J., 2000. Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia. Ph.D., Université Pierre et Marie Curie (Paris VI).

DeLucia, P.R., 1991. Pictorial and motion-based information for depth perception. Journal of Experimental Psychology: Human Perception and Performance 17 (3), 738–748.

DeLucia, P.R., 2013. Effects of size on collision perception and implications for perceptual theory and transportation safety. Current Directions in Psychological Science 22 (3), 199–204. https://doi.org/10.1177/0963721412471679.

DeLucia, P.R., Kaiser, M.K., Bush, J.M., Meyer, L.E., Sweet, B.T., 2003. Information integration in judgements of time to contact. Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology 56 (7), 1165–1189. https://doi.org/10.1080/02724980343000053.

DeLucia, P.R., Liddell, G.W., 1998. Cognitive motion extrapolation and cognitive clocking in prediction motion tasks. Journal of Experimental Psychology: Human Perception and Performance 24 (3), 901–914.

DeLucia, P.R., Preddy, D., Oberfeld, D., 2016. Audiovisual integration of time-to-contact information for approaching objects. Multisensory Research 29 (4–5), 365–395. https://doi.org/10.1163/22134808-00002520.

El-Rabbany, A., 2002. Introduction to GPS: the global positioning system. Artech house.

Emerson, R.W., Kim, D.S., Naghshineh, K., Pliskow, J., Myers, K., 2013. Detection of quiet vehicles by blind pedestrians. Journal of Transportation Engineering 139 (1), 50–56. https://doi.org/10.1061/(asce)te.1943-5436.0000478.

Forssén, J., Hoffmann, A., Kropp, W., 2018. Auralization model for the perceptual evaluation of tyre-road noise. Applied Acoustics 132, 232–240. https://doi.org/10.1016/j.apacoust.2017.11.023.

Genzel, D., Schutte, M., Brimijoin, W.O., MacNeilage, P.R., Wiegrebe, L., 2018. Psychophysical evidence for auditory motion parallax. Proceedings of the National Academy of Sciences of the United States of America 115 (16), 4264–4269. https://doi.org/10.1073/pnas.1712058115.

Gerzon, M.A., 1985. Ambisonics in multichannel broadcasting and video. Journal of the Audio Engineering Society 33 (11), 859–871.

Gordon, C.C., Churchill, T., Clauser, C.E., Bradtmiller, B., McConville, J.T., Tebbetts, I., Walker, R.A., 1989. Anthropometric survey of US Army personnel: Summary statistics, Interim report for 1988: U.S. Army Natick Research, Development & Engineering Center.

Gordon, M.S., Rosenblum, L.D., 2005. Effects of intrastimulus modality change on audiovisual time-to-arrival judgments. Perception & Psychophysics 67 (4), 580–594. https://doi.org/10.3758/bf03193516.

Gray, R., Regan, D., 1998. Accuracy of estimating time to collision using binocular and monocular information. Vision Research 38 (4), 499–512.

Grimm, G., Luberadzka, J., Hohmann, V., 2019. A toolbox for rendering virtual acoustic environments in the context of audiology. Acta Acustica United with Acustica 105 (3), 566–578. https://doi.org/10.3813/aaa.919337.

Grosse, J., Weber, R., Van de Par, S., 2013. Measurements and model predictions of detection thresholds for a conventional and electric vehicle in traffic and pink noise. In D. G. f. Akustik (Ed.), *AIA-DAGA 2013: Proceedings of the International Conference on Acoustics, Merano. March 18-21. 2013.* (pp. 538-541). Merano.

Hecht, H., Brendel, E., Wessels, M., Bernhard, C., 2021. Estimating time-to-contact when vision is impaired. Scientific Reports 11 (1). https://doi.org/10.1038/s41598-021-00331-5.

Hecht, H., Savelsbergh, G.J.P. (Eds.), 2004. Time-to-contact. Elsevier Science Publishers B V, Amsterdam, Netherlands.

Heuer, H., 1993. Estimates of time to contact based on changing size and changing target vergence. Perception 22 (5), 549–563.

Hochberg, Y., 1988. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 75 (4), 800–802.

Hoffmann, A., Kropp, W., 2019. Auralization of simulated tyre noise: Psychoacoustic validation of a combined model. Applied Acoustics 145, 220–227. https://doi.org/10.1016/j.apacoust.2018.08.026.

Horswill, M.S., Helman, S., Ardiles, P., Wann, J.P., 2005. Motorcycle accident risk could be inflated by a time to arrival illusion. Optometry and Vision Science 82 (8), 740–746. https://doi.org/10.1097/01.opx.0000175563.21423.50.

Hoyle, F., 1957. The Black Cloud. Heinemann, London.

ISO 9613-2:1999-10, 1999. Acoustics — Attenuation of sound during propagation outdoors — Part 2: General method of calculation.

Jenison, R.L., 1997. On acoustic information for motion. Ecological Psychology 9 (2), 131–151.

Jesteadt, W., Leibold, L., 2011. Loudness in the Laboratory, Part I: Steady-State Sounds. In: Florentine, M., Popper, A.N., Fay, R.R. (Eds.), Loudness. Springer, pp. 109–144.

Kaczmarek, T., 2005. Auditory perception of sound source velocity. Journal of the Acoustical Society of America 117 (5), 3149–3156. https://doi.org/10.1121/1.1880832.

Keshavarz, B., Campos, J.L., DeLucia, P.R., Oberfeld, D., 2017. Estimating the relative weights of visual and auditory tau versus heuristic-based cues for time-to-contact judgments in realistic, familiar scenes by older and younger adults. Attention Perception & Psychophysics 79 (3), 929–944. https://doi.org/10.3758/s13414-016-1270-9.

Keshavarz, B., Hecht, H., 2011. Validating an Efficient Method to Quantify Motion Sickness. Human Factors 53 (4), 415–426. https://doi.org/10.1177/0018720811403736.

Kolarik, A.J., Moore, B.C.J., Zahorik, P., Cirstea, S., Pardhan, S., 2016. Auditory distance perception in humans: A review of cues, development, neuronal bases, and effects of sensory loss. Attention Perception & Psychophysics 78 (2), 373–395. https://doi.org/10.3758/s13414-015-1015-1.

Kropp, W., Sabiniarz, P., Brick, H., Beckenbauer, T., 2012. On the sound radiation of a rolling tyre. Journal of Sound and Vibration 331 (8), 1789–1805. https://doi.org/10.1016/j.jsv.2011.11.031.

Law, D.J., Pellegrino, J.W., Mitchell, S.R., Fischer, S.C., McDonald, T.P., Hunt, E.B., 1993. Perceptual and cognitive factors governing performance in comparative arrival-time judgments. Journal of Experimental Psychology: Human Perception and Performance 19 (6), 1183–1199.

Lee, D.N., 1976. Theory of visual control of braking based on information about time-to-collision. Perception 5 (4), 437–459.

Lee, D.N., Young, D.S., McLaughlin, C.M., 1984. A roadside simulation of road crossing for children. Ergonomics 27 (12), 1271–1281. https://doi.org/10.1080/00140138408963608.

Levitt, H., 1971. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*(2). Suppl 2, 467–477.

Lutfi, R.A., Wang, W., 1999. Correlational analysis of acoustic cues for the discrimination of auditory motion. Journal of the Acoustical Society of America 106 (2), 919–928. https://doi.org/10.1121/1.428033.

McLeod, R.W., Ross, H.E., 1983. Optic flow and cognitive factors in time-to-collision estimates. Perception 12 (4), 417–423.

NHTSA 141, 2018. Federal Motor Vehicle Safety Standard No. 141: Minimum Sound Requirements for Hybrid and Electric Vehicles.

Oberfeld, D., Hecht, H., 2008. Effects of a moving distractor object on time-to-contact judgments. Journal of Experimental Psychology: Human Perception and Performance 34 (3), 605–623. https://doi.org/10.1037/0096-1523.34.3.605.

Oberfeld, D., Heeren, W., Rennies, J., Verhey, J., Snyder, J., 2012. Spectro-Temporal Weighting of Loudness. Spectro-temporal weighting of loudness. *PLOS One* 7 (11), e50184.

Pereira, F., Soares, F., Silva, C., Sousa, E., Freitas, E., 2021. CPX based synthesis for binaural auralization of vehicle rolling noise to an arbitrary positioned stander-by receiver. Applied Acoustics 182, 108211.

Petzoldt, T., 2014. On the relationship between pedestrian gap acceptance and time to arrival estimates. Accident Analysis and Prevention 72, 127–133. https://doi.org/10.1016/j.aap.2014.06.019.

Petzoldt, T., 2016. Size speed bias or size arrival effect-How judgments of vehicles' approach speed and time to arrival are influenced by the vehicles' size. Accident Analysis and Prevention 95, 132–137. https://doi.org/10.1016/j.aap.2016.07.010.

Poveda-Martinez, P., Peral-Orts, R., Campillo-Davo, N., Nescolarde-Selva, J., Lloret-Climent, M., Ramis-Soriano, J., 2017. Study of the effectiveness of electric vehicle warning sounds depending on the urban environment. Applied Acoustics 116, 317–328. https://doi.org/10.1016/j.apacoust.2016.10.003.

Pugliese, B.J., Barton, B.K., Davis, S.J., Lopez, G., 2020. Assessing pedestrian safety across modalities via a simulated vehicle time-to-arrival task. Accident Analysis and Prevention 134, 105344.

Rosenblum, L.D., Carello, C., Pastore, R.E., 1987. Relative effectiveness of three stimulus variables for locating a moving sound source. Perception 16 (2), 175–186. https://doi.org/10.1068/p160175.

Rosenblum, L.D., Gordon, M.S., Wuestefeld, A.P., 2000. Effects of performance feedback and feedback withdrawal on auditory looming perception. Ecological Psychology 12 (4), 273–291.

Rosenblum, L.D., Wuestefeld, A.P., Saldana, H.M., 1993. Auditory looming perception: Influences on anticipatory judgments. Perception 22 (12), 1467–1482. https://doi.org/10.1068/p221467.

Savelsbergh, G.J.P., Whiting, H.T.A., Bootsma, R.J., 1991. Grasping Tau. Journal of Experimental Psychology: Human Perception and Performance 17 (2), 315–322.

Schiff, W., Detwiler, M.L., 1979. Information used in judging impending collision. Perception 8 (6), 647–658.

Schiff, W., Oldak, R., 1990. Accuracy of judging time to arrival: Effects of modality, trajectory, and gender. Journal of Experimental Psychology: Human Perception and Performance 16 (2), 303–316.

Shaw, B.K., McGowan, R.S., Turvey, M.T., 1991. An acoustic variable specifying time-to-contact. Ecological Psychology 3 (3), 253–261.

Soares, F., Freitas, E., Cunha, C., Silva, C., Lamas, J., Mouta, S., Santos, J.A., 2017. Traffic noise: Annoyance assessment of real and virtual sounds based on close proximity measurements. Transportation Research Part D-Transport and Environment 52, 399–407. https://doi.org/10.1016/j.trd.2017.03.019.

Soares, F., Silva, E., Pereira, F., Silva, C., Sousa, E., Freitas, E., 2020. The Influence of Noise Emitted by Vehicles on Pedestrian Crossing Decision-Making: A Study in a Virtual Environment. Applied Sciences-Basel 10 (8). https://doi.org/10.3390/app10082913.

Soares, F., Silva, E., Pereira, F., Silva, C., Sousa, E., Freitas, E., 2021. To cross or not to cross: Impact of visual and auditory cues on pedestrians' crossing decision-making. Transportation Research Part F-Traffic Psychology and Behaviour 82, 202–220. https://doi.org/10.1016/j.trf.2021.08.014.

Tukey, J.W., 1977. Exploratory Data Analysis. Addison-Wesley Pub. Co, Reading, Mass.

UNECE R138. (2017). Regulation No 138 of the Economic Commission for Europe of the United Nations (UNECE) — Uniform provisions concerning the approval of Quiet Road Transport Vehicles with regard to their reduced audibility.

von Békésy, G., 1947. A new audiometer Acta Oto-Laryngologica. 35 (5–6), 411-422.

Wessels, M., Zähme, C., Oberfeld, D., 2022. Auditory information improves time-to-collision estimation for accelerating vehicles. Current Psychology. https://doi.org/10.1007/s12144-022-03375-6.

Wessels, M., Kröling, S., Oberfeld, D., under revision. Audiovisual time-to-collision estimation for accelerating vehicles: The acoustic signature of electric vehicles impairs pedestrians' judgments.

Wu, H.J., Ashmead, D.H., Adams, H., Bodenheimer, B., 2018. 3D sound rendering in a virtual environment to evaluate pedestrian street crossing decisions at a roundabout. Paper presented at the 2018 IEEE 4th VR Workshop on Sonic Interactions for Virtual Environments.

Yan, J.-J., Lorv, B., Li, H., Sun, H.-J., 2011. Visual processing of the impending collision of a looming object: Time to collision revisited. J. Vision 11 (12). https://doi.org/10.1167/11.12.7.

Zakarauskas, P., Cynader, M.S., 1991. Aural intensity for a moving source. Hearing Research 52 (1), 233–244. https://doi.org/10.1016/0378-5955(91)90203-l.