DOI: 10.61782/fa.2023.0649

Auditory and audiovisual time-to-collision estimation and road-crossing decisions

Daniel Oberfeld[1*] **Marlene Wessels**[1] **Thirsa Huisman**[1]
[1] Experimental Psychology, Johannes Gutenberg-Universität Mainz, Germany

ABSTRACT

Pedestrians can only safely cross the road before an approaching vehicle if the time remaining until the vehicle arrives at their position (time-to-collision, TTC) is longer than the time needed for crossing. Using a virtual-reality (VR) system that combines physically plausible acoustic simulations of approaching vehicles with visual simulations, we investigated how the vehicle sound affects the perception and behavior of pedestrians in a road-crossing situation. Our results show that 1) for vehicles approaching at constant speed, participants estimate longer TTCs for softer compared to louder vehicles with the same actual TTC, both for auditory-only and audiovisual presentations, indicating potential risks associated with quieter vehicles. 2) When the sound of an accelerating conventional vehicle (ICEV) is presented, this largely removes the inadequate consideration of acceleration observed in visual-only TTC estimation. 3) For electric vehicles (EVs) with and without AVAS, this benefit provided by the car sound is significantly reduced compared to ICEVs. 4) In line with this, the probability of unsafe road-crossing decisions increases significantly with the acceleration level for EVs with and without AVAS, but remains low for ICEVs. Taken together, auditory information is important for pedestrians, particularly so when the approaching vehicle accelerates. Potential risks associated with EVs should be considered.

Keywords: *time-to-collision estimation, road-crossing decisions, pedestrian safety, electric vehicles, virtual acoustics.*

1. INTRODUCTION

In order for pedestrians to safely cross the road in front of an approaching vehicle, they need to ensure that the remaining time until the vehicle reaches their position (time-to-collision, TTC) is longer than the time it takes to cross the road. Hence, it is important for pedestrians to accurately estimate the TTC to adjust their crossing behavior accordingly. The sound of a vehicle plays a crucial role in providing information about its motion in real traffic. In this paper, we present the main findings from a recent series of experiments focusing on TTC estimation and street-crossing decisions with different types of sensory information presented: only auditory (A-only), only visual (V-only), or a combination of auditory and visual information (AV). The experiments examined various scenarios, including vehicles approaching at a constant speed or accelerating, and involved pedestrians' perception and behavior when interacting with internal combustion engine vehicles (ICEVs) and electric vehicles (EVs). To carry out these experiments, we utilized an innovative simulation system, which we will describe in the following.

2. INTERACTIVE AUDIO-VISUAL VIRTUAL-REALITY SIMULATION OF APPROACHING VEHICLES

When pedestrians stand at the curb and listen to an oncoming vehicle, various dynamic acoustic changes occur, such as an increase in intensity due to spherical spreading and air absorption, changes in interaural time and level differences due to changes in the azimuthal position relative

---

[*] ***Corresponding author***: *oberfeld@uni-mainz.de*

to the pedestrian, or dynamic comb-filter effects resulting from interference between direct and reflected sound. All these changes potentially provide cues to the arrival time of the vehicle [1]. Additionally, the sounds emitted by the vehicle vary depending on factors such as travel speed, engine type, rotational engine speed, and engine load [2]. However, most previous studies on auditory or audiovisual TTC estimation and street-crossing decisions have not adequately provided realistic acoustic simulations of approaching vehicles that incorporate the full range of these auditory cues. To address these limitations, we developed a high-fidelity interactive audio-visual simulation setup, which is described in [3]. Due to the complexity of realistically simulating dynamic vehicle sounds, we adopted a source-based approach. The acoustic source signals are recordings captured by microphones attached to the chassis of real vehicles (ICEV and EV) driving at defined constant speeds or defined positive acceleration rates on a dry asphalt surface. The vehicles recorded were one conventional and one electric passenger car models from Kia Motors (for further details refer to [3]). The EV was equipped with an Acoustic Vehicle Alerting System (AVAS), which could be active between speeds of 0.5 km/h and 28 km/h, but could also be deactivated. Recordings of the EV were made with both active and inactive AVAS. During the acoustic recordings, the vehicle's trajectory was measured using high-precision GPS tracking, so that the vehicle's position, speed, and acceleration were known at each time point in the audio signals. In our experiments, the movement of sound sources in space is simulated using the acoustic virtual reality (VR) simulation software TASCAR [4]. TASCAR creates an interactive simulation of the dynamic spatial sound field, by accounting for the changing geometry of the acoustic scene and modeling the sound transmission from sources to receivers, and renders the scene using sound field synthesis. The simulation approach presents realistic vehicle sounds and incorporates all relevant monaural and binaural distance and motion cues. Our current implementation involves rendering the simulated scenes on an array of 40 Genelec 8020DPM loudspeakers, arranged in an upper and a lower ring, along with a Genelec 7360APM subwoofer, within a large acoustically treated space. The direct sound of the vehicle is presented through 32 loudspeakers positioned at ear height plus subwoofer, using two-dimensional Higher-Order Ambisonics (15th order) [5, 6]. The reflected sound is rendered through the entire loudspeaker array, using three-dimensional VBAP [7]. These auditory VR simulations can be synchronized with three-dimensional visual VR simulations, presented stereoscopically on a head-mounted display (HTC Vive Pro Eye) with head-tracking. This system allows listeners to actively explore the simulated auditory and visual scene using head movements. The flexibility to present vehicles from various approach angles and distances enables, for instance, the presentation of identical vehicle source signals at different TTCs. As a result, the system can be used to conduct highly controlled VR experiments with a high ecological validity compared to previous studies in this field, while ensuring the safety of the participants.

3. THE EFFECT OF INTENSITY ON PEDESTRIANS' TTC ESTIMATION AND ROAD CROSSING DECISIONS

Two of our previous studies demonstrated an "*intensity-arrival effect*" [8, 9]. At the same actual TTC, participants estimated *softer* approaching sound sources to arrive *later* than louder ones. This effect might suggest a potential risk posed by quiet vehicles like EVs because pedestrians might overestimate the TTC relative to a louder ICEV, despite both have identical TTCs. As a consequence, pedestrians might cross with a shorter time gap when interacting with quieter vehicles. In **Exp. 1** (N = 28)**,** published in [3], we therefore investigated how vehicle loudness influences TTC estimation, presenting highly realistic auditory or audiovisual simulations of the approaching vehicle in the described VR system. Participants judged the TTC of an ICEV and a loudness-matched EV (AVAS inactive) driving at different constant speeds (10, 30, 50 km/h), from a position at the curb of a simulated two-lane road. We presented an auditory-only and an audio-visual condition and additionally varied the vehicle loudness levels by 10 dB. In our experiments, we use the established *prediction-motion task* [10]. The simulated car travels towards the participant for some seconds, and is then "occluded", i.e., it is no longer audible nor visible. Participants press a response button to indicate the point in time at which the approaching vehicle would arrive at their position, had it continued to approach them with the same constant speed after it disappeared. The estimated TTC is given by the time between the occlusion and the participant's button press. The TTC at occlusion was varied between 2.0 and 5.0 s. Compatible with the intensity-arrival effect, participants judged significantly *longer* TTCs for cars at the softer loudness level in the A-only condition (p < .001, d_z = 2.06). This effect was smaller but persisted in the AV condition (p < .001, d_z = 0.95), confirming that auditory information plays an important role in TTC estimation even when full visual information provided [8, 9]. No significant differences between the ICEV and the loudness-matched EV occurred, neither in the A-only condition nor in the AV condition, suggesting that the sound quality differences between the vehicle types approaching at constant speeds did not have a substantial effect on TTC estimation.

In (Exp. 1), the loudness level varied from trial to trial. In **Exp. 2** (*N* = 22)**,** we varied the vehicle loudness level in a blockwise fashion to investigate if not only the loudness difference between trials, but also the "absolute" loudness of an approaching vehicle affects TTC estimation. We used a similar setup as in Exp. 1 but presented the two loudness levels separately in two blocks (blockwise condition). In a third block, we presented the two loudness levels again in a randomized order (interleaved condition). In the blockwise condition, the estimated TTCs were significantly shorter in blocks with the higher compared to blocks with the lower loudness level ($p = 0.012$, $d_z = 0.587$). Thus, the intensity-arrival effect is not restricted to conditions where the sound level varies from trial to trial, although it was much stronger in the interleaved block ($p < 0.001$, $d_z = 1.8971$).

Exp. 3 (*N* = 13), investigated the effect of vehicle loudness on *road-crossing decisions*. An ICEV travelling at different constant velocities (30, 50 and 60 km/h) was presented, each at two loudness levels as well as in an auditory-only and an audio-visual condition. The loudness levels differed again by 10 dB and were presented in an interleaved fashion. Similar to the TTC experiments, a vehicle approached for some seconds before occlusion, and participants indicated whether or not they would have crossed the road in front of the approaching vehicle at the moment of occlusion (positive or negative crossing decision, respectively). We measured the probability of a positive crossing decision ("gap acceptance") across a range of different actual TTCs. If lower loudness results in longer estimated TTCs, as demonstrated by Exp. 1 and 2, then at a given TTC at occlusion participants will think that they have more time available to cross the street in front of the vehicle when it's sound is softer and will thus make a positive crossing decision in a higher proportion of trials than for a louder vehicle. In line with this expectation, as shown in Figure 1, the probability (p_{coll}) that a positive crossing decision would have resulted in a collision with the approaching vehicle because the TTC at occlusion was shorter than the time needed to cross the road was significantly higher for softer than for louder vehicles, both in the A-only ($p = 0.002$, $d_z = 1.121$) and the AV condition ($p = 0.045$, $d_z = 0.609$).

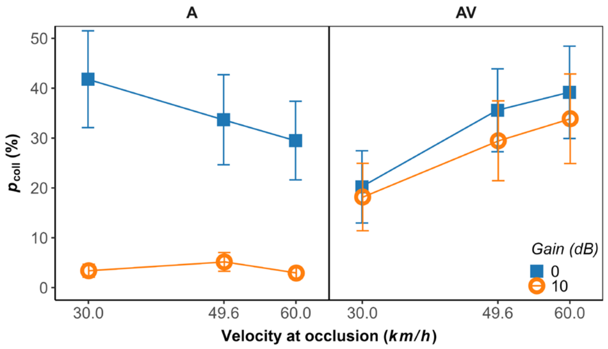

Figure 1: Exp. 3. Mean collision probability (p_{coll}) as a function of the velocity at occlusion. Blue squares: audio gain 0 dB (lower loudness level). Orange circles: gain 10 dB (higher loudness level). Left panel: auditory-only condition. Right panel: audiovisual condition. Error bars show ± 1 SEM across the 13 participants.

4. AUDITORY INFORMATION IMPROVES TTC ESTIMATION AND STREET-CROSSING DECISIONS FOR ACCELERATING VEHICLES

The literature on *visual* TTC estimation consistently shows that humans do not adequately account for the acceleration of an object [e.g., 11, 12]. Instead, they estimate the TTC of an accelerating object as if it was travelling at a constant speed (first-order estimation). For positive acceleration rates, first-order estimation leads to TTC overestimation, because the increase in velocity after occlusion is not taken into account. However, when an ICEV is accelerating, the resulting dynamic changes in the powertrain noise provide *salient acoustic cues* to acceleration. Can pedestrians use this auditory information to better account for the acceleration in TTC estimation?

In **Exp. 4** (*N* = 25), published in [13], we compared TTC estimations for an ICEV approaching at either a constant speed (a = 0 m/s^2) or accelerating during the approach (a = 2 m/s^2) between a V-only and an AV condition. In the V-only condition, the TTC estimations showed a clear first-order pattern: with increasing presented TTC, participants increasingly overestimated the TTC, compatible with the literature on visual TTC estimation. However, if the sound of the accelerating ICEV was presented in addition to the visual information (AV condition), this largely removed the first-order pattern (significant modality condition × TTC interaction, $p < .001$, $\eta_p{}^2 = 0.69$), so that the mean estimated TTC was close to the actual TTC. This result suggests that the salient acoustic signature of the accelerating ICEV helped pedestrians to account for the acceleration in their TTC estimations.

But does this benefit provided by the vehicle sound also apply to electric vehicles without and with AVAS? In **Exp. 5** (*N* = 30), published in [14], we obtained TTC estimations for an ICEV and an EV with or without activated AVAS, at acceleration rates between 0.4 and 2.6 m/s^2. At a given actual TTC, the mean estimated TTCs increased substantially as a function of acceleration rate for the EV without AVAS (p = .031), thus exhibiting a first-order pattern and suggesting inadequate consideration of acceleration. The effect of acceleration was reduced when the AVAS was activated on the EV although it did not reach the level of the ICEV, where the estimated TTC showed no significant effect of the acceleration rate. This suggests that, participants were better able to use the information about acceleration communicated by the sound of an ICEV, compared to an EV.

In Exp. 5, the presented acceleration rates and speeds at occlusion of the accelerating EV and ICEV differed to a certain extent because they exactly corresponded to the vehicles' motion during the recordings made for our simulation system on the test track. These manual drives showed deviations from the intended velocity profiles in conditions with acceleration, particularly so for the ICEV with manual transmission, because the driver needed to perform gear shifts during acceleration. In **Exp. 6** (*N* = 15), we therefore presented the recorded source signals of the different vehicles, but the motion of the sound source simulated in the virtual scene corresponded exactly to an initial phase of 2.0 s with a constant speed of 10 km/h ($a = 0$ m/s^2), followed by an acceleration phase of 3.0 s with exactly $a = 2.0$ m/s^2. Thus, the motion was *identical* for all vehicle types. We additionally presented vehicles approaching at a constant speed that matched the speed at occlusion of the accelerating vehicle (v_{occ} = 31.6 km/h; note that at this speed the AVAS could not be not activated), and added a V-only condition. The data for the V-only condition and for the EV without activated AVAS both showed a first-order estimation pattern. In contrast, the TTCs estimated for the ICEV indicated that participants differentiated between constant-speed and accelerating ICEV approaches. For the ICEV, the mean estimated TTCs for the accelerating approaches were closer to the actual TTCs than for the EVs. When the AVAS of the accelerating EV was activated, the TTC estimations shifted towards those for the ICEV, but without reaching the same accuracy level, as in Exp. 5. The results thus confirmed that the sound of an ICEV promoted better consideration of acceleration than the sound of an EV.

Even though the simulated motion was the same for all three vehicle types in Exp. 6, there was a distinct difference in the sound produced by the ICEV. Due to the selected presentation duration, the sound of the ICEV stopped shortly after a gear shift, which meant that the final 500 ms of the sound represented a phase where the ICEV increased its acceleration rate from less than 2 m/s² to over 2 m/s². On the other hand, EVs maintained a nearly constant acceleration rate throughout the entire acceleration phase. Did the higher acceleration rate audible in the final part of the ICEV sound contribute to a more adequate consideration of acceleration during TTC estimation for this vehicle type?

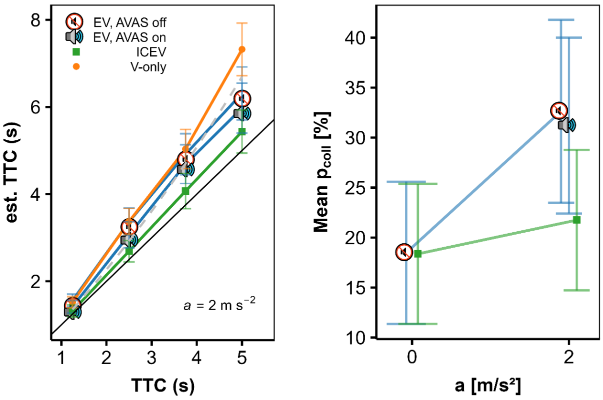

Figure 3: Left panel: Exp. 7. Mean estimated TTC as a function of presented TTC and vehicle type. The dashed gray line corresponds to first-order estimation. The solid black line represents the actual TTC. Orange symbols: visual-only condition. Green symbols: audiovisually presented ICEV. Blue symbols: audiovisually presented EV. Loudspeaker symbols indicate whether the AVAS was activated or not. Right panel: Exp. 9. Mean collision probability p_{coll} as a function of acceleration and vehicle type. Error bars show ± 1 SEM across the 24 (Exp. 7) and 15 (Exp. 9) participants, respectively.

To answer this question, **Exp. 7** (*N* = 24) replicated the methodology of Exp. 6, but with a modification: the presentation duration was extended for all vehicle types, allowing the sound of the ICEV to be presented for an additional second after the gear shift. During this extra second, the acceleration rate of the ICEV remained close to 2.0 m/s², as was the case for both EVs. The results of Exp.

7, shown in the left panel of Figure 3, were consistent with the findings of the previous experiments, exhibiting first-order patterns in the visual only-condition and for the EV with and without AVAS, but not for the ICEV (vehicle type × TTC interaction $p < .001$, $\eta_p{}^2 = 0.35$). Therefore, they demonstrated that the different patterns observed in the estimated TTCs for an accelerating ICEV compared to an EV were unlikely to be caused by the fact that the ICEV sound indicated a higher acceleration rate towards the end of the presentation compared to the EV sounds.

Finally, we examined street-crossing decisions of pedestrians in interaction with accelerating ICEVs and EVs. Based on the findings from the TTC estimation-Exp. 4-7, we hypothesized *riskier* street-crossing decisions for accelerating EVs than for accelerating ICEVs.

In **Exp. 8** ($N = 25$), published in [15], we presented simulations of an ICEV and an EV with and without activated AVAS. The presented scenarios included approaches with acceleration rates ranging from 0 to 2.5 m/s. Consistent with the findings from Exp. 4 and 5, for the ICEV, there was no significant effect of the acceleration rate on p_{coll} ($p = .085$), which remained relatively low across all acceleration rates. However, for the EVs with and without AVAS, p_{coll} showed a significant increase with the acceleration rate ($p < .001$).

In Exp. 8, the simulations presented the actual vehicle trajectories driven on the test track, again resulting in slightly different acceleration rates between the ICEV and the EVs. In **Exp. 9** ($N = 15$), we used the same experimental design as Exp. 6 and contrasted audiovisually presented approaches with an acceleration rate of 2.0 m/s² to constant-speed approaches with a matched v_{occ} of 31.6 km/h, but simulated identical motion for all three vehicle types. As expected and consistent with the TTC estimation results, the right panel in Figure 3 shows that p_{coll} increased only slightly for the constant-speed and accelerated approaches of the ICEV, but increased significantly more strongly when the EVs with and without AVAS accelerated than when they travelled at a constant-speed (vehicle type × a interaction $p = .006$, $\eta_p{}^2 = 0.42$).

Overall, this series of experiments clearly demonstrates the importance of vehicle sound for pedestrians to consider the acceleration of an approaching vehicle, even when full visual information is available. This is evident in TTC estimations and street-crossing decisions. However, the benefit provided by the vehicle sound is diminished for EVs compared to ICEVs, even when an AVAS is activated.

5. SUMMARY

Using a system that offers more realistic audiovisual simulations of approaching vehicles compared to previous studies, we conducted research on time-to-collision (TTC) estimation and road-crossing decisions. Our findings unequivocally demonstrate the importance of auditory information in traffic scenarios for accurate TTC estimation and safe road-crossing decisions. The first series of experiments revealed that participants perceived quieter vehicles as arriving later than louder ones when the actual TTC was identical, even when provided with complete visual information about the approaching vehicle's motion. Moreover, riskier road-crossing decisions were made for quieter vehicles compared to louder ones. In the second series of experiments, consistent results highlighted the crucial role of vehicle sound in conveying acceleration information that is not readily available in the visual domain. Only when the sound of an ICEV saliently signaled that the vehicle was positively accelerating as it approached were participants able to make relatively accurate TTC estimations and safer road-crossing decisions. However, for EVs, this benefit provided by the vehicle sound was significantly diminished compared to ICEVs, even when an AVAS compliant with UNECE R138 was activated. In short, our findings demonstrate that pedestrians rely on auditory cues to judge the motion of an approaching vehicle in street-crossing situations, particularly when the vehicle accelerates, and even when the vehicle is fully visible to them.

6. ACKNOWLEDGEMENTS

This work was supported grants from Unfallforschung der Versicherer (www.udv.de/udv-en) and Deutsche Forschungsgemeinschaft (www.dfg.de; OB 346/8-1; priority program AUDICTIVE - SPP2236: Auditory Cognition in Interactive Virtual Environments) awarded to Daniel Oberfeld.

7. REFERENCES

[1] Jenison, R.L., "On acoustic information for motion". *Ecological Psychology*, 1997. **9**(2): p. 131-151.

[2] Zeller, P., ed. *Handbuch Fahrzeugakustik*. 2018, Springer.

[3] Oberfeld, D., M. Wessels, and D. Büttner, "Overestimated time-to-collision for quiet vehicles: Evidence from a study using a novel

audiovisual virtual-reality system for traffic scenarios". *Accident Analysis & Prevention*, 2022. **175**: p. 106778.

[4] Grimm, G., J. Luberadzka, and V. Hohmann, "A toolbox for rendering virtual acoustic environments in the context of audiology". *Acta Acustica United with Acustica*, 2019. **105**(3): p. 566-578.

[5] Gerzon, M.A., "Ambisonics in multichannel broadcasting and video". *Journal of the Audio Engineering Society*, 1985. **33**(11): p. 859-871.

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

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

[8] DeLucia, P.R., D. Preddy, and D. Oberfeld, "Audiovisual integration of time-to-contact information for approaching objects". *Multisensory Research*, 2016. **29**(4-5): p. 365–395.

[9] Keshavarz, B., J.L. Campos, P.R. DeLucia, and D. Oberfeld, "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*, 2017. **79**(3): p. 929-944.

[10] Schiff, W., and M.L. Detwiler, "Information used in judging impending collision". *Perception*, 1979. **8**(6): p. 647-58.

[11] Lee, D.N., D.S. Young, P.E. Reddish, S. Lough, and T.M.H. Clayton, "Visual timing in hitting an accelerating ball". *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology*, 1983. **35**(MAY): p. 333-346.

[12] Benguigui, N., H. Ripoll, and M.P. Broderick, "Time-to-contact estimation of accelerated stimuli is based on first-order information". *Journal of Experimental Psychology: Human Perception and Performance*, 2003. **29**(6): p. 1083-1101.

[13] Wessels, M., C. Zähme, and D. Oberfeld, "Auditory information improves time-to-collision estimation for accelerating vehicles". *Current Psychology*, 2022.

[14] Wessels, M., S. Kröling, and D. Oberfeld, "Audiovisual time-to-collision estimation for accelerating vehicles: The acoustic signature of electric vehicles impairs pedestrians' judgments". *Transportation Research Part F: Traffic Psychology and Behaviour*, 2022. **91**: p. 191-212.

[15] Oberfeld, D., M. Wessels, and S. Kröling, *Risiko hohe Beschleunigung? Straßenquerungsverhalten von Fußgänger:innen in Interaktion mit E-Fahrzeugen (mit und ohne AVAS) im Vergleich zu Verbrennern [Risks imposed by acceleration? Road crossing behavior of pedestrians in interaction with electric vehicles with and without AVAS in comparison to conventional vehicles]*. Unfallforschung der Versicherer: Forschungsbericht. Vol. 76. 2021, Berlin: Gesamtverband der Deutschen Versicherungswirtschaft e.V.