Exploring the Impacts of Background Noise on Auditory Stimuli of Audio-Visual eHMIs for Hearing, Deaf, and Hard-of-Hearing People

Wenge Xu
Birmingham City University
Birmingham, United Kingdom
wenge.xu@bcu.ac.uk

Foroogh Hajiseyedjavadi
Birmingham City University
Birmingham, United Kingdom
Foroogh.Hajiseyedjavadi@bcu.ac.uk

Debargha Dey
Eindhoven University of Technology
Eindhoven, Netherlands
d.dey@tue.nl

Tram Thi Minh Tran
Design Lab
School of Architecture, Design and
Planning, The University of Sydney
Sydney, NSW, Australia
tram.tran@sydney.edu.au

Mark Colley
UCL Interaction Centre
London, United Kingdom
m.colley@ucl.ac.uk

Abstract

External Human-Machine Interfaces (eHMIs) have been proposed to enhance communication between automated vehicles (AVs) and pedestrians, with growing interest in multi-modal designs such as audio-visual eHMIs. Just as poor lighting can impair visual cues, a loud background noise may mask the auditory stimuli. However, its effects within these systems have not been examined, and little is known about how pedestrians — particularly Deaf and Hard-of-Hearing (DHH) people — perceive different types of auditory stimuli. We conducted a virtual reality study (Hearing N=25, DHH N=11) to examine the effects of background noise (quiet and loud) on auditory stimuli (baseline, bell, speech) within an audio-visual eHMI. Results revealed that: (1) Crossing experiences of DHH pedestrians significantly differ from Hearing pedestrians. (2) Loud background noise adversely affects pedestrians' crossing experiences. (3) Providing an additional auditory eHMI (bell/speech) improves crossing experiences. We outlined four practical implications for future eHMI design and research.

CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**; **Accessibility design and evaluation methods**.

Keywords

Deaf, Hard-of-Hearing, External Human-Machine Interface, Automated Vehicles, Accessibility.

1 Introduction

When crossing in front of vehicles, pedestrians rely on both implicit (e.g., speed, distance, and deceleration) [77, 81] and explicit communication (e.g., driver eye contact, hand gestures, or headlight flashes) [85]. In automated vehicles (AVs), the absence of a human driver removes these conventional channels. This creates two challenges: (1) The loss of explicit driver signals may increase uncertainty and ambiguity in interactions; (2) people are often poor at judging implicit communication cues, such as speed, stopping distance, and time-to-arrival [58, 74, 86]. To address these challenges, researchers have proposed external human-machine interfaces (eHMIs) to facilitate communication between AVs and vulnerable road users, improving safety, subjective crossing experiences, and behaviour [15, 22, 27, 59].

Yet, eHMI research has rarely involved disabled people, excluding an estimated 430 million (over 5% of the world's population) deaf and hard-of-hearing (DHH) people in the world and potentially 700 million (i.e., 10% of the whole population) DHH people in 2050[1]. A report from the UK has found that disability of pedestrians is a critical contributory factor to fatal or serious collisions with pedestrians[2], eHMI design should consider the needs of disabled people to ensure the proposed eHMI is accessible to them and create equality in transport.

One proposed approach is multi-modal eHMI [7, 18, 19, 43]. For instance, it could be an audio-visual eHMI with visual eHMI serving as the foundation [27, 35, 95] while using auditory eHMI to enhance perceived safety [29]. The choice of the auditory eHMI could also benefit low vision and blind people [18, 88] or those who experience situational impairments (e.g., being distracted by secondary activities, occluded view, etc.) [20]. However, the value of audio-visual eHMI, especially auditory stimuli among DHH people, remains unclear. Environmental factors such as background noise

---

[1]WHO: Deafness and Hearing Loss; accessed 14.08.2025
[2]Reported road casualties in Great Britain: pedestrian factsheet, 2022; accessed 14.08.2024

further complicates the issue. Ambient background noise could interfere with pedestrians' ability to detect and localise vehicles, increasing risks of injuries [34]. Loud background noise could become excessive and obscure auditory signals intended for pedestrians [61]. Despite this, only 29% of eHMI studies have included environmental noises (i.e., natural sounds and human-produced sounds), typically to enhance the simulation realism rather than to examine their interaction with auditory stimuli [89]. The investigation into how practical and useful auditory stimuli are under different noise levels for both hearing and DHH people is overlooked, which is the main gap we address in this paper. In the meantime, we aim to identify accessibility barriers that auditory eHMIs may introduce and ensure that the use of auditory cues does not inadvertently disadvantage DHH people.

Contribution Statement [92]

The main contributions of the paper include: (1) the first empirical virtual reality (VR) simulation evaluation of the effects of background noise (*Quiet*, *Loud*) on auditory stimuli (*Baseline*, *Bell*, *Speech*) on crossing experiences (trust, acceptance, perceived safety, mental load) and behaviour (eye gaze, step into the road time, early step into the road count) between hearing participants (N=25) and DHH participants (N=11); (2) four practical implications that pave the way for future eHMI design and research.

2 Related Work

2.1 DHH People and Road Crossing

Both hearing and eyesight are important for acting and responding adequately in traffic situations [47]. DHH pedestrians would face similar visual challenges as Hearing pedestrians when crossing the road (i.e., obstructed views, low lighting, adverse weather conditions, dazzling) [4, 40]. However, they face more difficulties concerning sound direction and distance judgments due to hearing loss, which is essential to judge the location of potential threats or obstacles [54, 97]. A survey by Gür et al. [41] suggested that almost half of DHH teenagers were involved in traffic accidents as a pedestrian, 2-3 times higher than hearing teenagers. A field study in an urban environment with manually driven vehicles by Pecchini and Giuliani [71] showed that DHH pedestrians often experienced heightened apprehension when initiating a crossing and exercised greater caution toward approaching vehicles to ensure safety. Lundälv [60] suggest that pedestrians with moderate deafness are at a higher risk of being injured by a vehicle because they have difficulty in identifying the sound direction. Overall, the lack of access to auditory information has reduced feelings of safety among DHH pedestrians [46]. Over time, this could constitute a negative and fatiguing experience, discourage active travel activities like walking [70], resulting in reduced physical activity levels [14] and broadening inequities in mobility [48].

According to the World Health Organization [93], hearing aids and hearing implants are commonly used hearing technologies to help DHH people with hearing aids mainly intended to help people with mild to moderate hearing loss[3] and hearing implants to help provide a sense of sound to people with severe to profound

deafness[4]. However, these technologies come with their own issues. Hearing aids have consistently failed to improve sound localisation and could even impair it [97]. Studies suggest that only about one in five people who would benefit from a hearing aid use one, potentially due to comfort level and perceived benefits not meeting the expectation [63]. As for hearing implants, they process sounds electronically and transmit electrical stimulation to the cochlea of the individual with hearing impairment, restoring some sensation of auditory perception to help them understand sounds or speech [73]. However, data from the UK showed that only 1.3% of individuals with severe or profound deafness use hearing implants [5]. In summary, neither is the technology flawless, nor does it completely compensate for the lack of auditory cues for DHH people. Therefore, we employ the audio-visual eHMI so that visual information is still available for people when they may miss essential auditory information.

2.2 Effect of Background Noise and Road Crossing

Background (or ambient) noise has an impact on drinking behaviour [64], reading behaviour [51], office-related tasks [8], and perception of time in gyms [68]. Similar results were found in mobile interaction, where outdoor urban background noise led to more errors in the visual search task (i.e., finding an icon) and longer time for the text entry task when compared to indoor urban background noise [80]. Background noise could also affect pedestrian activities such as walking along the road or crossing the road. Korte and Grant [56] found that loud traffic noise levels and higher densities of traffic reduced pedestrians' awareness of objects placed along their route, to walk faster, and to engage more in a straight-ahead gaze fixation. Tapiro et al. [87] found urban background noise plays a negative role in pedestrian crossing behaviour, although visual distraction affected more. In this work, we were interested in how background noise in a typical urban environment (e.g., construction, social activities [23]) would impact interaction between AV and pedestrians.
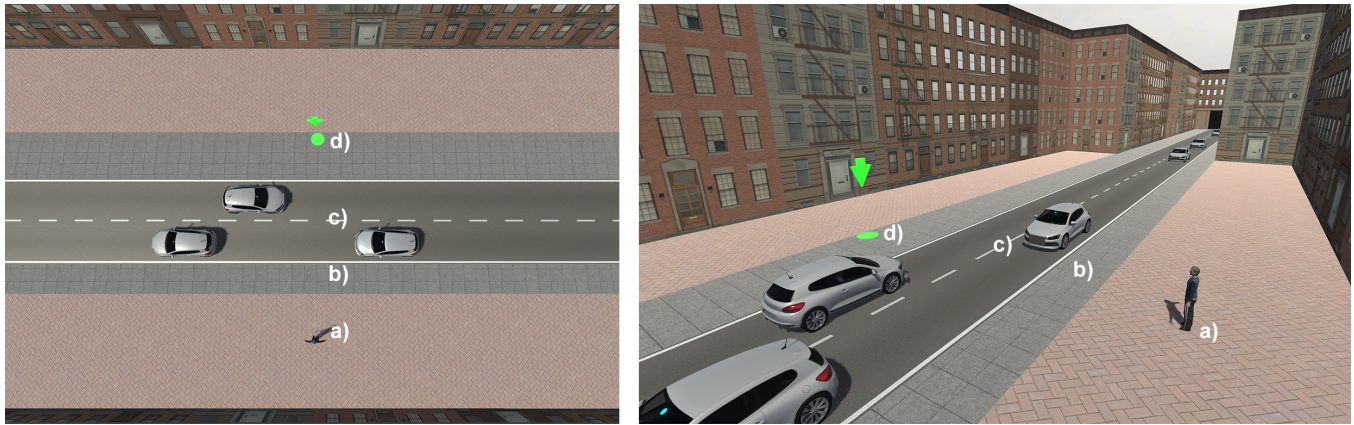
2.3 Auditory eHMIs

Auditory eHMIs are typically presented through speech (i.e., verbal messages) [25, 49, 61] or non-speech (e.g., jingles, humming, bell) stimuli [11, 27, 39]. Through a video-based study, Dey et al. [27] found that different participants had completely different associations and mental models for using non-speech stimuli (i.e., bell and drone-like humming). While many people perceived the bell as a calm, inviting, and friendly signal, others experienced it as urgent and linked it with a warning. Similarly, although some people felt that a drone (i.e., humming) sound for the vehicle's engine was a natural and fitting choice, others considered it unpleasant and burdensome. Pelikan and Jung [72] investigated non-speech stimuli (e.g., humming, bell, jingle) in the wild and found that the humming sound was good for showing the presence but insufficient for other purposes. In contrast, the repeated bell was found to be powerful in indicating the vehicle's stopping.

---

[3]https://www.nidcd.nih.gov/health/hearingaids; accessed 14.08.2025

[4]https://www.nidcd.nih.gov/health/cochlear-implants; accessed 14.08.2025
[5]Hearing loss statistics in the UK: https://www.hearinglink.org/your-hearing/about-hearing/facts-about-deafness-hearing-loss/; accessed 14.08.2025

Figure 1: Setup of the virtual simulation environment: The left figure shows a top-down view, while the right figure shows a leftward view. a) The starting position of the participant, which is 4.7 m from the road. b) The 2 m wide grey pavement area, which has a trigger to inform one of the AVs to yield for the participant. c) The 6.8 m wide two-lane road that the participant needed to cross. d) The waypoint (i.e., green surface with downward arrow indicated), which is 2.5 m away from the road. Please note that the human model of the player was only a visual aid to help understand the location; the human model was disabled in the environment, and participants do not have their own avatars.

On the other hand, speech might be more reliable than non-speech auditory sounds. Deb et al. [25] investigated multiple auditory stimuli in a VR-simulated crossing study, finding that the speech was the most favoured auditory stimuli compared to horn, music, and no sound. This is supported by Hudson et al. [49], where they found speech was preferred over music. Nevertheless, none of these studies have included DHH people, nor explored how auditory stimuli would perform under different noise levels. To the best of our knowledge, this is the first study of its kind. We followed [6] to employ a speech eHMI and a bell eHMI (non-speech) in our study as the starting point of this type of research.

2.4 Multi-modal eHMIs: Accessible eHMI Solutions

Only limited eHMI research has been conducted with disabled people. Some works focus on co-designing personalised solutions based on pedestrians' own devices Asha et al. [7]. Others explored the potential of making on-vehicle eHMI accessible to wider access for disabled people. Colley et al. [18] explored the choice of auditory eHMIs with low vision and blind people through a workshop study and found that the speech auditory eHMI was best received. Their follow-up study through VR simulation suggests that having more content in the speech message could reduce the mental load. Haimerl et al. [43] compared eHMI concepts (baseline, visual-only, auditory-only, multi-modal—both audio-visual) through an online video-based survey study with participants with intellectual disabilities and participants without intellectual disabilities; they found that multi-modal eHMIs positively affect quality and inclusion Haimerl et al. [43].

Multi-modal eHMI could also benefit wider pedestrian group when compared to uni-modal eHMI. Dou et al. [29] conducted a

VR study to evaluate 12 eHMI concepts, in combinations of visual (smile/arrow), audio (human voice/warning sound), and vehicle movement style (the approaching speed decreases gradually/remains unchanged), and concluded that multi-modal eHMIs resulted in more satisfactory interaction and improved safety compared to the unimodal eHMI. He et al. [45] conducted a VR study with 12 participants and found that audio-visual modality (symbol and anthropomorphic voice) was more appealing than the eHMI with a single modality (visual or auditory). Results from the Wizard-of-Oz study by [3] suggested that a combination of audio-visual modality is most effective in understanding information.

The need to implement multi-modal eHMIs has been (1) a common agreement among studies involved with disabled people [7, 18, 19, 43], (2) suggested by a recent review paper on accessibility of eHMI concepts [35], and (3) identified as key opportunity to address gaps for disabled people [10]. A core reason is that each modality has specific trade-offs [62]. Multi-modal interfaces or feedback designs have been commonly used beyond the field of eHMI research to make the interaction accessible for disabled people [5, 21, 91]. Therefore, in this work, we decided to make the evaluation more practical by employing audio-visual eHMI, i.e., having a constant visual eHMI in addition to the auditory eHMIs (i.e., Bell and Speech; throughout the paper, we considered the baseline condition as a natural auditory stimuli from the running vehicle) that we wanted to explore. Details of the visual eHMI discussion can be found in Section 3.3.

3 Virtual Simulation Environment

We used Unity v2022.3.44 to develop the virtual simulation environment, featuring a straight, two-way, two-lane urban road (see Figure 1). In the UK, where the study was conducted, traffic law dictates drivers to yield at zebra crossings and at red traffic signals, which would predetermine pedestrian expectations of the vehicle's behaviour based on right of way. To focus on the impact of the

eHMI, we therefore used a mid-block location without markings, where pedestrians may legally cross but are advised to do so with caution. This type of setting is widely used in VR studies of pedestrian–vehicle interaction [2, 89], as it requires participants to attend to approaching vehicles and their communication rather than rely on traffic control infrastructure.

The maximum speed was set at 50 km/h, while the simulated AV drove in a range of 40 km/h to 50 km/h. We set 13 AVs to drive in the first lane (i.e., the one closer to the participant). When one AV exit through the tunnel on the right-hand side, a new AV would be initialised to enter the first lane through the tunnel on the left. The time gap between each AV was around 2.4 seconds to 2.8 seconds. In the second lane, 2 AVs drive towards the right-hand side tunnel when launching the environment; these 2 vehicles would disappear after entering the tunnel and leave the second lane empty. This setting for the second lane was to raise awareness among the participants of vehicles potentially approaching from the left-hand side. Simultaneously, we did not want any vehicles in the second lane to affect the crossing decision and behaviour of the participants.

Participants started at the position shown in (Figure 1a), walked and waited at the grey pavement (see Figure 1b) until they felt it was safe for them to cross the road (see Figure 1c) and reached the green waypoint (see Figure 1d). Each participant crossed twice for each condition.

3.1 AV

3.1.1 Appearance. Figure 2 shows the AVs used in the user study. In line with prior work [15, 38, 61], (1) a round cyan light positioned at the top centre of the windshield indicates that the AV is driving autonomously, (2) a light strip located at the bottom of the bumper region displays the light design, and (3) a display located at the grill region of the bumper to display the text/symbol. Unless the AV yielded for participants, neither the light strip nor the display would show any additional information. The visual eHMI was positioned on the front grille of the vehicle, which (1) aligns with current pedestrian expectations and experience, as they typically look towards the location of the driver's head or vehicle movement [28, 32], and (2) follows standard practice of the eHMI research [15, 27].
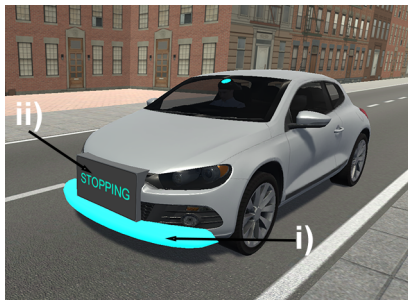
Figure 2: The appearance of the AV used in our study, where a human avatar sits in the driver's seat but does not engage with the vehicle. Active mode of i) the Light Strip and ii) the Display when the AV was fully stopped.

3.2 Parameters Testing

We conducted an iterative pilot testing with eight testers (4M; 4F; including 1 DHH tester with severe hearing loss of both ears) to ensure that (1) the visual eHMI design is reasonable and (2) both the choice and volume set for the two urban background noise conditions are reasonable. In addition, they were also satisfied with the other parameters set in the environment, e.g., choice of words and audio, message repetition, interval of the message, tire-road surface sound, and Acoustic Vehicle Alerting System (AVAS). We did not recruit these testers for the formal user study. In our formal user study, none of the participants argued that the background sound and vehicle sound were not realistic.

3.2.1 Vehicle Sound Implementation. **AVAS**: We embedded the AVAS sound from BMW into all vehicles for their slowdown process, activated when their speed is from 20 km/h to 0 km/h. The original sound showed a spectrum and frequency shift. When fully stopped, we loop the last second of the sound where they were in a similar frequency range. The average volume of the AVAS was about 64 dB (larger than the minimum requirement of 56 dB). Our implementations followed the regulation [31] and a similar implementation by Dey et al. [27]; refer to drone auditory eHMI.

Tire-Surface: To improve the realism of the study [89], we also implemented a tire-surface sound system attached to each AV. We ensured the sound[6] had a volume of 62 dB when driving at 50 km/h. To simulate a realistic slowdown tire-surface effect, we took approximate values found in [50, 69], applied a feed-out effect when the speed reached around 20 km/h (50 dB), and another feed-out effect so the sound reached 0 dB when full stopped.

3.3 Visual eHMI

As a visual eHMI represents the standard of the field, we included a fixed visual eHMI candidate across all experimental conditions. We believed multi-modal eHMI should be the priority eHMI representation as (1) its advantage over unimodal eHMI [3, 29, 45] and (2) the potential to be more accessible to disabled pedestrians [18, 43]. Since there is still no explicit agreement on which visual design is the most beneficial and how many visual signals might be suitable or sufficient for the other road users [27], we employed a combination of Abstract Light [26] and Text [33] as our visual eHMI concept, this design was widely accepted by participants in prior work [95].

The visual eHMI remained inactive during standard driving. When a yielding command is triggered, the light strip would pulse in Cyan with its pulse frequency going from fast (pulsates between on and off at a rate of 1 Hz) to slow (pulsates between on and off at a rate of 0.5 Hz) to indicate a speed change. Meanwhile, during this slowdown, the display would show "STOPPING" to convey that the vehicle is slowing down. When the AV fully stopped, the light strip would stop pulsing and remain static; meanwhile, the text displayed on the display would change to "STOPPED". Both words were presented in bold cyan letters.

---

[6]Tire rolling sound: https://drive.google.com/file/d/1SsnY6dP10cM6LmpIZsTqE5y-3L_Q_taM/view?usp=sharing; accessed: 09.04.2025

3.4 Background Noise

We implemented two typical urban background noise environments: quiet and loud.

Quiet: This condition represents a quiet part of the city with relatively light social activities around the environment [23]. We used two audio clips [7] from BBC Sound Effects, with each background noise audio source located near each side of the building. These files form a mixture of overlapping human conversation (e.g., conversational speech, laughter, calls) and incidental background clutter. Standing at the starting point, the perceived sound volume was approximately an average of 58 dB with maximum volume reached around 64 dB, which is roughly the standard city noise sound [13, 17] and is the volume implemented by prior work in eHMI literature [24].

Loud: This condition represents a noisy part of the city with dense social activities and a construction site [23]. We used three audio clips [8] from BBC Sound Effects and Free Sound organisation. In line with the *Quiet* condition, each background noise audio source is located at one side of the building. The construction file was located 1 meter next to the background noise audio source at the same end of the starting point. These files combine construction-site activity (i.e., machinery, hammering, and high-frequency industrial sounds) and a busy market environment (i.e., dense human chatter, crowd movement, and general ambient bustle). Standing at the starting point of the environment, the participant would perceive an average sound volume of 70 dB, with the maximum volume reach 79 dB [17]. A real world city environment may consist of even louder sound or sudden unexpected sound [78], however, we did not include this and control this to be the volume as we described as (1) exposing loudness of over 85 dB could cause hearing loss [82] which is not ethical, and (2) testing sudden unexpected sound is beyond the focus of this study.

To mitigate participant familiarity with the background noise and introduce natural variation across trials, we randomised the playback onset of each audio source. Specifically, the system initiated playback at a random point within the first 0 – 30 seconds of the recording.

3.5 Auditory Stimuli

As discussed in the Section 2.3, we selected *Speech* and *Bell* (no-speech) as our Auditory eHMIs.

Speech: Faas et al. [38], Zhang et al. [96], ISO technical report [1], and Volvo[9] suggested that AVs should not give cross advice as it creates liability and legibility issues. Therefore, we used intention-based messages such as "I'm stopping" and "I'm stopped" for the slowdown and fully stopped state, respectively. These messages were generated by an online text-to-audio website[10]. When a yielding command was activated in the AV, it would first wait for 1.5 seconds and then play "I'm stopping" (0.75 seconds long) with a

0.5-second interval. The choice of a 1.5-second delay is to allow the AV to get closer to the pedestrian. This message would play three times during the slowdown process. When the AV fully stopped, the AV would immediately play the message "I'm stopped" (0.7 seconds long) with a 1.1-second interval. The AV would keep repeating this verbal message with a proposed interval until the participant passed the first lane of the road.

Bell: We employed a Bell sound downloaded from BBC Sound Effects[11], which is a similar Bell sound as in [27]. Following the implementation from [27], the Bell sound (1.1 seconds long) used for stopping and full stop was the same. When a yielding command was activated, the AV would wait for 1.5 seconds and then play the Bell sound with a 0.3-second interval. The AV would play the Bell sound three times during the slowdown process. When the AV fully stopped, the AV would immediately play the Bell with a 1-second interval to indicate the difference between stopping and being stopped. Similar to the speech auditory design, the AV would repeat this bell sound with a proposed interval until the participant passed the first lane of the road.

All files have an average volume of around 65 dB, with the maximum volume of around 74 dB. We tested it by placing the audio source 2 meters away from the main camera in Unity, and placed a volume measurement device in the middle of the headset.

4 Virtual Reality User Study

This study was guided by the following research questions (RQs):

> **RQ1**: How do the ratings for experience (i.e., trust, acceptance, perceived safety, mental load) and behaviour (i.e., gaze behaviour, step-in road time, early step into the road count) differ between Hearing and DHH participants?
> **RQ2**: What impact does the Background Noise have on pedestrians regarding experience and behaviour?
> **RQ3**: What impact do the Auditory Stimuli of the Audio-Visual eHMI have on pedestrians regarding experience and behaviour?

4.1 Study Design and Outcome Measures

We employed a 2×3 within-subjects design with two within-subjects factors (1) Background Noise (Quiet and Loud) and (2) Auditory Stimuli (Baseline, Bell, Speech). The order of Background Noise × Auditory Stimuli was counterbalanced in the study.

Subjective: Crossing experience was measured via questionnaires after each condition.

- We used a 21-point single question regarding the mental workload from the NASA-TLX [44] questionnaire to measure mental workload (the lower the number, the smaller the workload).
- We employed the 5-point Likert scale Trust in Automation questionnaire [55] to measure the trust in AV (Understandability and Trust) through the 4-item Understandability and 2-item Trust subscales (the higher the number, the better).

- Perceived Safety was measured by a 7-point Likert scale that ranged from -3 (anxious/agitated/unsafe/timid) to +3 (relaxed/calm/safe/confident) [37].
- We employed the van der Laan acceptance scale with the subscales 'usefulness' and 'satisfying' to measure Acceptance [90].

Objective: Crossing behaviour was recorded by the developed program. For eye gaze, we counted the fixations when the AV was 20 meters away from the participant's crossing point till the participant stepped into the road. We choose 20 meters as the starting point for calculating the results in this paper because early research suggests that pedestrians start to look more at the vehicle than at the road ahead from this range [28] and it provides higher eye gaze accuracy. Light Strip, Display, and the Whole Vehicle are the areas of interest we are keen to explore as they allow us to understand if gaze behaviour of participants would change under different sound and noise conditions when interacting with the eHMI (e.g., would they pay more attention to the vehicle in general or the active eHMI component due to its perceptual salience) [87].

- Step Into the Road Time: The time taken by the participant to start crossing the road from the moment an AV starts to slow down on the nearest lane.
- Early Step Into the Road Count: The number of times the participant stepped onto the road before the AV fully stopped.
- Eye Gaze
  - **Light Strip Duration**: Fixation on the Light Strip measured in second.
  - **Display Duration**: Fixation on the Display measured in second.
  - **Whole Vehicle Duration**: Fixation on the combination of Light Strip, Display, and other parts of Vehicle, measured in second.
  - **Active Visual eHMI Duration**: duration on active visual eHMI for conditions that employ Visual eHMI, i.e., Light Strip for *Abstract Light*, combination of Light Strip and Display for both *Abstract Light + Text* and *Abstract Light + Symbol* conditions.
  - **Active Visual eHMI Percentage**: (Active Visual eHMI Duration / the Whole Vehicle Duration) × 100%.

At the end of the study, participants rated their perception of the Necessity and Reasonability of (1) Bell and (2) Speech concepts in a 7-point (1=Totally Disagree to 7=Totally Agree) [15] and then ranked the Audio eHMIs (i.e., Baseline, Bell, Speech), the lower the number, the better. This was followed by a semi-structured interview. We first asked the participants questions about the Baseline—"Were you able to hear and understand?" and "Do you think the background sound impacts how you perceive it?" Then the interview moved to the proposed auditory eHMI design—"Could you hear it when the vehicle (1) slowed down and (2) fully stopped?", "Overall, what do you think about it?", "Anything you liked/disliked about it?", and "Do you think the background sound impacts how you perceive it?". The final stage of the interview followed two questions "Can you share some examples of sound that affect you in a road crossing or street walking activity?" and "Would you think you can cross without the auditory eHMI?". In the end, we offered an open question to ask if participants had anything to add.

4.2 Apparatus and Setup

A Varjo XR-4 focal edition was used as the VR headset, which offers a 90 Hz refresh rate with 3840 × 3744 resolution and a 120° × 105° field of view. We enabled the built-in 200 Hz eye tracker during the study, which provides and records eye gaze visualisation and eye measurements such as pupil iris diameter, openness, and interpupillary distance [12]. Cyberith Virtualizer Elite 2 was used as our walking solution for locomotion. Varjo XR-4 and Cyberith Virtualizer Elite 2 were connected to a high-end PC with an i9 CPU, 64 GB RAM, and a GeForce RTX 4090 Graphics card to provide the best immersive experience. The sound experience was provided directly through the VR device's built-in speakers, as wearing additional headphones or earphones would clash with hearing aids or implants and may result in an uncomfortable experience. The study was conducted in an indoor, well-illuminated, quiet laboratory room that could not be seen from outside.

4.3 Procedure

The study started with a brief introduction from the experimenter. Subsequently, the participants needed to sign the consent form and complete a demographic questionnaire. Participants were given two trials to get familiar with the Virtualizer Elite 2 device and the Varjo XR-4. Once the participants were ready, they needed to complete the formal study conditions (we counterbalanced the order across participants), with their task being to cross the street twice. Eye calibration was made/checked at the beginning of each condition to ensure accuracy. After each condition, participants had to answer the required questionnaires (See Section 4.1). At the end of the study, they participated in a semi-structured interview, which lasted about 10-15 minutes. The study lasted about 60 minutes for English users and 90 – 120 minutes for sign language users (longer time for questionnaire and interview). Participants were under the observation and supervision of an experimenter. A BSL interpreter was always present to assist with the study for BSL users.

4.4 Participants

Participants were recruited via physical posters, social media platforms, charities, and informal referrals through word-of-mouth. In total, we recruited 36 participants to the user study, with 25 participants (12 male, 13 female; Mean age = 26.4, SD = 3.96, range 21 to 35) who self-identified as no hearing loss people, including 1 participant who had one ear for Mild hearing loss level and 1 participant who had Moderate hearing loss for one ear; and 11 participants (5 male, 6 female; Mean age = 53.73, SD = 18.84, range 21 to 78) who identified themselves as DHH people (see Table 1). We categorised the hearing loss levels, normal (<20 dB), mild (21 - 40 dB), moderate (41 - 70 dB), severe (71 - 95 dB), and profound (>95 dB), based on the recommendations from the National Health Services[13]. All DHH participants participated in our study used a pair of hearing aids or implant during the study. In a 5-point Likert scale (1 = Not at all, 5 = Definitely), Hearing participants showed medium interest in AVs (M = 3.2, SD = 1) and had more knowledge about AVs (M = 4.28, SD = 0.84). In contrast, Participants in

---

[12]Varjo Developer Eye Tracking
[13]National Health Service: https://www.esht.nhs.uk/service/audiology/diagnosis-and-testing/; accessed: 08.03.2025

Table 1: Demographic Information of the DHH Participants.

ID	Age	Gender	Left & Right Ear Hearing Loss Level	Identity	Preferred Communication	Using Hearing Technology?
P25	33	Female	Profound; Profound	Deaf	BSL	Yes
P26	71	Male	Severe; Severe	HoH	English	Yes
P27	41	Male	Profound; Profound	Deaf	BSL	Yes
P28	60	Female	Profound; Severe	HoH	English	Yes
P29	78	Male	Severe; Severe	HoH	English	Yes
P30	21	Male	Profound; Profound	deaf	English	Yes
P31	69	Female	Moderate; Severe	HoH	English	Yes
P32	71	Female	Moderate; Severe	deaf	English	Yes
P33	60	Male	Profound; Profound	deaf	BSL	Yes
P34	34	Female	Severe; Severe	deaf	English	Yes
P36	53	Female	Severe; Profound	deaf	English	Yes

DHH reported less interest ($M = 2.82, SD = 1.40$) and knowledge ($M = 3.64, SD = 0.92$) in AVs.

We adhere to all institutional safety measures and data protection guidelines throughout the experiment. We also consulted experienced accessibility researchers who specialise in working with DHH participants to review our study set-up and design to ensure it is accessible and friendly to all participants. Informed consent was obtained from all participants and the study was approved by our university's Research Ethics Committee.

5 Results

We focused on the main and interaction effects of the three independent variables: *Background Noise* (*Quiet* and *Loud*; within-subjects) and *Auditory Stimuli* (*Baseline*, *Bell*, and *Speech*; within-subjects), and *Hearing Group* (*Hearing* and *DHH*; between-subjects). Shapiro-Wilk test suggested that our data were not normally distributed. We therefore used *nparLD*, which can handle small samples and unequal group sizes [67], and has been widely used by similar studies [18, 52, 79]. The Modified ANOVA-type statistic are reported for the whole-plot factor (i.e., main effect of *Hearing Group*) as suggested by [12]. Otherwise, ANOVA-type statistics are reported by default. We employed Bonferroni correction for all post-hoc tests.

5.1 Subjective Crossing Experience

5.1.1 Trust in Automation. We could not find any significant difference among the **Understandability** ratings.

Regarding **Trust**, Figure 3 illustrates trust ratings for all conditions across *Hearing* and *DHH* participants. The non-parametric variance analysis (NPVA) revealed a significant main effect of *Background Noise* ($F = 4.000, df = 1, p = .045$) and *Auditory Stimuli* ($F = 9.101, df = 1.879, p < .001$). Post-hoc analysis for *Background Noise* confirmed that participants gave higher trust ratings in *Quiet* conditions ($M = 3.73, SD = 0.95$) than *Loud* conditions ($M = 3.55, SD = 1.00, p = .016$). Post-hoc pairwise comparisons for the main effect of *Auditory Stimuli* showed that participants gave lower rating for *Baseline* ($M = 3.39, SD = 0.97$) than *Bell* ($M = 3.67, SD = 0.95, p = .032$) and *Speech* ($M = 3.86, SD = 0.96, p = .002$).
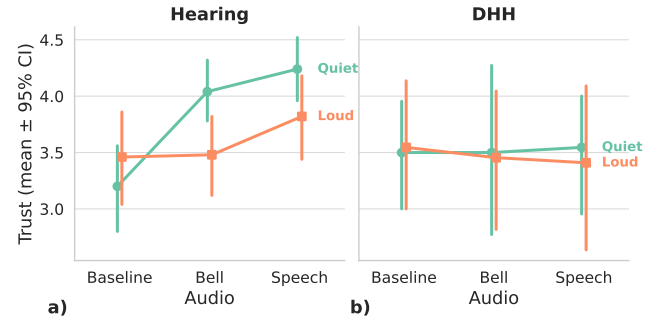
Figure 3: Mean Trust ratings of audio stimuli types under Quiet vs. Loud background noise across a) Hearing participants and b) DHH participants. Data are shown with 95% confidence intervals.

The NPVA showed a significant interaction effect of *Hearing Group × Background Noise* ($F = 4.047, df = 1, p = .044$). Post-hoc results showed that among the *Hearing* group, higher trust ratings were given in *Quiet* conditions ($M = 3.83, SD = 0.91$) than *Loud* conditions ($M = 3.59, SD = 0.95, p = .021$). The NPVA also yielded a significant interaction effect of *Auditory Stimuli × Background Noise* ($F = 6.309, df = 1.837, p = .002$), post-hoc pairwise comparisons showed that (1) among *Bell*, trust was rated higher in *Quiet* conditions ($M = 3.86, SD = 0.94$) than *Loud* conditions ($M = 3.59, SD = 0.94, p = .0133$), (2) within *Quiet* conditions, *Baseline* ($M = 3.29, SD = 0.94$) was rated significantly lower than *Bell* ($M = 3.88, SD = 0.94, p = .007$) and *Speech* ($M = 3.69, SD = 0.95, p = .001$). In addition, the NPVA yielded a significant interaction effect of *Hearing Group × Auditory Stimuli* ($F = 8.476, df = 1.879, p < .001$). Post-hoc results showed that among *Hearing* people, *Baseline* ($M = 3.36, SD = 0.96$) was rated significantly lower than *Bell* ($M = 3.76, SD = 0.83, p = .024$) and Speech ($M = 3.94, SD = 0.91, p = .003$).

5.1.2 Acceptance. As for **Usefulness**, the NPVA revealed a significant main effect of *Background Noise* ($F = 10.869, df = 1, p < .001$), *Auditory Stimuli* ($F = 8.177, df = 1.907, p < .001$), and
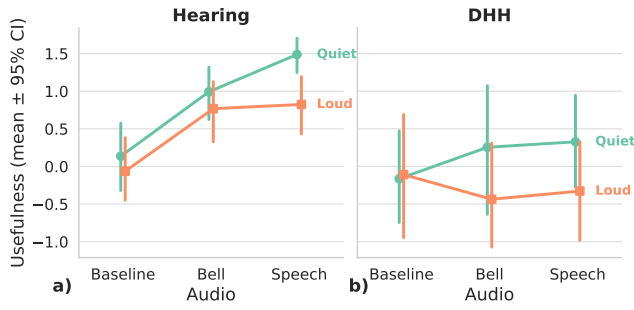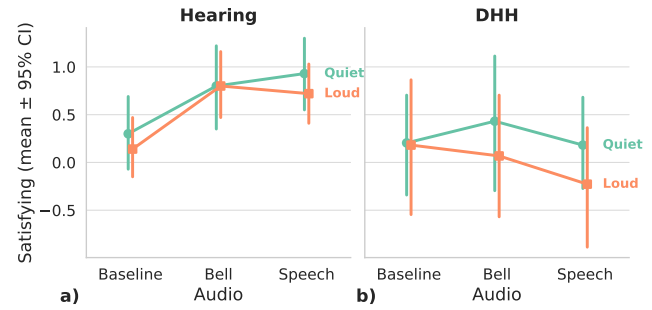
Figure 4: Mean Usefulness ratings of audio stimuli types under *Quiet* vs. *Loud* background noise across a) *Hearing* participants and b) *DHH* participants. Data are shown with 95% confidence intervals.

Figure 5: Mean Satisfying ratings of audio stimuli types under Quiet vs. Loud background noise across a) Hearing participants and b) DHH participants. Data are shown with 95% confidence intervals.

Hearing Group ($F = 5.042, df = 1, p = .040$; Modified ANOVA-type statistic). Post-hoc pairwise comparisons of *Background Noise* showed that participants gave a higher usefulness rating in the *Quiet* conditions ($M = 0.65, SD = 1.18$) than *Loud* conditions ($M = 0.26, SD = 1.21, p < .001$). Post-hoc analysis for the *Auditory Stimuli* showed that both *Bell* ($M = 0.58, SD = 1.19, p = .010$) and *Speech* ($M = 0.80, SD = 1.09, p < .001$) were better than *Baseline* ($M = −0.02, SD = 1.19$). Post-hoc analysis for the *Hearing Group* showed that *Hearing* participants ($M = 0.69, SD = 1.10$) gave significantly higher ratings than *DHH* participants ($M = −0.08, SD = 1.26, p < .001$).

We found interaction effects of *Hearing Group × Auditory Stimuli* ($F = 6.179, df = 1.907, p = .002$). Post-hoc results showed that (1) among *Hearing* participants, *Bell* ($M = 0.88, SD = 0.97, p = .005$) and *Speech* ($M = 1.16, SD = 0.88, p < .001$) were rated significantly higher than *Baseline* ($M = 0.04, SD = 1.14$). (2) Regarding *Bell*, usefulness ratings provided by *Hearing* participants ($M = 0.88, SD = 0.97$) were significantly higher than *DHH* participants ($M = −0.09, SD = 1.40, p = .019$). (3) Regarding *Speech*, usefulness ratings provided by *Hearing* participants ($M = 1.16, SD = 0.88$) were significantly higher than *DHH* participants ($M = 0.00, SD = 1.13, p < .001$). Figure 4 illustrates usefulness ratings for all conditions across *Hearing* and *DHH* participants.

Regarding **Satisfying**, the NPVA revealed no significant main effects. We observed an interaction effect of *Hearing Group × Auditory Stimuli* ($F = 4.671, df = 1.918, p = .010$). Post-hoc analysis showed that (1) among *Hearing* participants, *Baseline* ($M = 0.22, SD = 0.88$) was rated significantly lower than *Bell* ($M = 0.8, SD = 0.99, p = .011$) and *Speech* ($M = 0.83, SD = 0.89, p = .026$), (2) regarding *Speech*, *DHH* participants ($M = 0.05, SD = 1.08$) gave a significantly lower rating than Hearing participants ($M = 0.83, SD = 0.89, p = .003$). Data for each condition across *Hearing* and *DHH* participants can be found in Figure 5.

5.1.3 Perceived Safety. Figure 6 illustrates usefulness ratings for all conditions across *Hearing* and *DHH* participants. The NPVA showed a significant main effect of *Background Noise* ($F = 12.468, df = 1, p < .001$) and *Auditory Stimuli* ($F = 4.727, df = 1.94, p = .009$). Post-hoc analysis on main effect *Background Noise* confirmed that participants gave a higher rating in *Quiet* conditions ($M =$
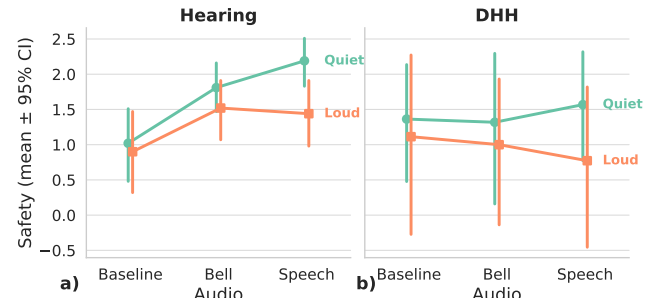
Figure 6: Mean Safety ratings of audio stimuli types under Quiet vs. Loud background noise across a) Hearing participants and b) DHH participants. Data are shown with 95% confidence intervals.

$1.59, SD = 1.30$) than *Loud* conditions ($M = 1.19, SD = 1.53, p < .001$). Post-hoc results of the main effect *Auditory Stimuli* showed that ratings for *Baseline* ($M = 1.05, SD = 1.55$) was significantly lower than *Bell* ($M = 1.15, SD = 1.34, p = .022$) and *Speech* ($M = 1.62, SD = 1.35, p = .002$).

We found a significant interaction effect of *Auditory Stimuli × Background Noise* ($F = 4.023, df = 1.94, p = .018$). Post-hoc pairwise comparisons analysis showed that (a) regarding *Bell*, ratings were higher in *Quiet* conditions ($M = 1.66, SD = 1.29$) than *Loud* conditions ($M = 1.36, SD = 1.38, p = .047$) (b) regarding *Speech*, ratings were also higher in *Quiet* conditions ($M = 2, SD = 1.09$) and *Loud* conditions ($M = 1.24, SD = 1.49, p < .001$). In addition, post-hoc results also yielded that under *Quiet* conditions, *Baseline* was rated significantly lower than *Speech* ($M = 2, SD = 1.09, p < .001$).

We also found a significant interaction effect of *Hearing Group × Auditory Stimuli* ($F = 10.791, df = 1.94, p < .001$). Post-hoc analysis showed that among *Hearing* group, *Baseline* ($M = 0.96, SD = 1.40$) was rated significantly lower than *Bell* ($M = 1.67, SD = 1.05, p = .005$) and *Speech* ($M = 1.82, SD = 1.10, p = .002$).

5.1.4 Mental Workload. The NPVA revealed a significant main effect of *Background Noise* ($F = 11.841, df = 1, p < .001$) on Mental workload. Post-hoc analysis confirmed that ratings were lower

Figure 7: Mean Mental Workload ratings of audio stimuli types under Quiet vs. Loud background noise across a) Hearing participants and b) DHH participants. Data are shown with 95% confidence intervals.
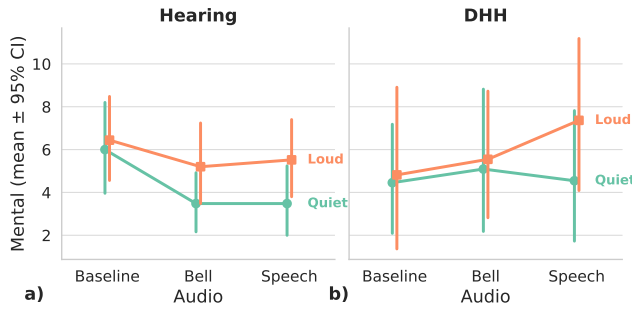
in *Quiet* conditions ($M = 4.44, SD = 4.75$) than *Loud* conditions ($M = 5.78, SD = 5.22, p < .001$). Figure 7 shows usefulness ratings for all conditions across *Hearing* and *DHH* participants.

The NPVA showed a significant interaction effect of *Auditory Stimuli × Background Noise* ($F = 4.169, df = 1.83, p = .018$). Post-hoc pairwise comparisons yielded that (a) for *Bell*, mental workload rating was lower in *Quiet* conditions ($M = 3.97, SD = 4.46$) than *Loud* ($M = 5.31, SD = 4.90, p = .0473$), (b) for *Speech*, mental workload rating was lower in *Quiet* conditions ($M = 3.81, SD = 4.60$) than *Loud* conditions ($M = 6.08, SD = 5.13, p < .001$). In addition, post-hoc results also showed that under *Quiet* background environment, mental workload was rated significantly lower for *Speech* ($M = 3.81, SD = 4.60$) than *Baseline* ($M = 5.53, SD = 5.12, p = .033$).

The NPVA also yielded a significant interaction effect of *Hearing Group × Auditory Stimuli* ($F = 8.844, df = 1.89, p < .001$) on mental workload. Post-hoc results showed that among *Hearing* participants, mental workload rating was higher for *Baseline* ($M = 6.22, SD = 5.30$) than *Bell* ($M = 4.34, SD = 4.24, p < .001$) and *Speech* ($M = 4.5, SD = 4.53, p = .007$).

5.2 Objective Crossing Behaviour

5.2.1 Eye Gaze Behaviour. We could not observe any significant effect of *Auditory Stimuli*, *Background Noise*, *Hearing Group* or their interactions among measurements of **Light Strip Duration**, **Display Duration**, **Whole Vehicle Duration**, **Active Visual eHMI** and **Active Visual eHMI Percentage**. Contrary to our expectations that *Auditory Stimuli* and *Background Noise* may modulate gaze behaviour, the results indicated that participants' gaze duration for these measurements seems relatively stable across conditions and between Hearing Groups (Hearing and DHH).

5.2.2 Movement behaviour. Both data sets below covered 432 trials (36 participants × 6 conditions × 2 repetition) of crossings. We did not observe a single crash during the study.

As for **Step Into the Road Time**, the NPVA revealed a significant interaction effect of *Hearing Group × Background Noise* ($F = 3.872, df = 1, p = .049$). However, post-hoc results did not yield any significance. Regarding **Early Step Into the Road Count**, we also could not observe any significant difference.

5.3 Necessity and Reasonability for Bell and Speech eHMIs

Regarding necessity, among Hearing participants, the *Bell* received an average score of 4.96 ($SD = 1.51$) and *Speech* received an average score of 5.76 ($SD = 1.36$). As for *DHH* participants, the *Bell* received an average score of 3.91 ($SD = 2.17$) and *Speech* received an average score of 4.64 ($SD = 1.63$). Regarding reasonability, among *Hearing* participants, the *Bell* received an average score of 5.16 ($SD = 1.40$) and *Speech* received an average score of 5.60 ($SD = 1.32$). As for *DHH* participants, the *Bell* received an average score of 4.18 ($SD = 1.94$) and *Speech* received an average score of 5.27 ($SD = 1.74$).

We were interested in understanding whether participants value the *Bell* and *Speech* differently and whether there was a significant difference between *Hearing* and *DHH* participants; therefore, we explored the impact of Auditory Stimuli (*Bell* and *Speech*) and Group (*Hearing* and *DHH*) on the necessity and reasonability ratings. The NPVA revealed a significant main effect of *Group* ($F = 5.023, df = 1, p = .025$) on necessity ratings. Post-hoc results confirmed that *Hearing* participants ($M = 5.36, SD = 1.48$) gave higher necessity ratings than *DHH* participants ($M = 4.27, SD = 1.91$). Regarding reasonability, the NPVA did not yield a significant difference.

5.4 Ranking

The ranking shows a preference for *Speech* ($M = 1.58, SD = 0.69$, 19 ranked it first, while 13 ranked it second). It was followed by *Bell* ($M = 1.72, SD = 0.66$, 14 ranked it first and 18 ranked it second), and *Baseline* was rated the worst ($M = 2.69, SD = 0.62$, 28 ranked it as the third). *Baseline* was largely disliked by *Hearing* participants (only 1 participant did not rank it as the third); opinions among *DHH* participants were varied. We noted that 3 *DHH* participants ranked *Baseline* as their first, 5 *DHH* participants ranked it as second, and only 4 out of 11 ranked it as the third option.

Therefore, we were interested in understanding whether participants rank the *Bell* and *Speech* differently and whether there was a significant difference between *Hearing* and *DHH* participants. We followed the same analysis as described in the Necessity and Reasonability subsection. The NPVA revealed a significant main effect of *Audio* ($F = 7.551, df = 1.959, p < .001$). Post-hoc pairwise comparisons suggested that *Bell* ($M = 1.72, SD = 0.66$) and *Speech* ($M = 1.58, SD = 0.69$) were ranked significantly better than *Baseline* ($M = 2.69, SD = 0.62$, both $p < .001$). We also observed a significant interaction effect of Group and Auditory Stimuli ($F = 5.079, df = 1.959, p = .007$), post-hoc showed that the results of the main effect only held true for *Hearing* participants (i.e., *Baseline* was ranked significantly worse than *Speech* [$p < .001$] and Bell [$p = .003$]).

5.5 Qualitative Results

The qualitative data were analysed using inductive thematic coding, which allowed themes to emerge from the participants' own descriptions. Two coders conducted the initial coding independently; Coder 1 designed the study and conducted the interviews, while Coder 2 was not involved in data collection. Afterward, both coders met to discuss, refine, and reconcile the themes to ensure consistency and reliability. We present anecdotal feedback and participants' opinions based on the specific questions asked during

the interviews. To facilitate descriptive comparison, we counted the frequency of mentions for each theme across participants, providing an overview of both common and contrasting views while maintaining the qualitative depth of interpretation.

5.5.1 Baseline. Slightly over half of the participants (N=20; Hearing: 14; DHH: 6) said they could hear clearly about the Baseline (AVAS and tire-pavement volume changes). Five participants (Hearing: 3; DHH: 2) mentioned they may have heard it, but they could not distinguish it from the background during the study. The remaining participants noted that they did not hear it. Several participants (N=25; Hearing: 17; DHH: 8) said there was an impact of the Background Noise, 8 explicitly mentioned that the louder background noise made it harder to hear and 1 mentioned that "the quiet environment helped me focus on the visual" [P7].

5.5.2 Bell. Most participants (N=27; Hearing: 21; DHH: 6) said they could hear the Bell under both quiet and noisy backgrounds. In addition, 7 participants said they could only hear (1) when the vehicle is fully stopped (DHH: 1) and (2) in a quiet environment (Hearing: 4; DHH: 2). Among the 34 people who could hear the Bell, 24 participants (Hearing: 17; DHH: 7) agreed that the Background Noise impacted their perception of it. Among these 24 participants, 18 participants (Hearing: 13; DHH: 5) said louder background noise made it harder to hear. Interestingly, participants (N=4) mentioned that they felt the Bell sound blended in the background, making it difficult to hear. In addition, we observed that 3 DHH participants' hearing aid filtered the Bell sound as part of the Background Noise, especially in a noisy environment.

Overall, 9 participants explicitly praised the Bell, perceived it to be "good", "nice", and "perfect". This could be due to its "audibility in quiet environment" [P22], "no language barrier" [P10], and choice of timing and frequency being "good pace", "well-timed pauses" [P4, P6, P27]. 7 (Hearing: 6; DHH: 1) participants felt positive about the tone of the sound, described it to be "pleasant", but 3 others (Hearing: 2; DHH: 1) perceived it to be "too calm", "too static", and "mismatch for AV context".

Four participants (all Hearing) argued that they found the use of Bell familiar in real-life, with a similar pattern has been used by "tram", "train level crossing", and is "similar to pedestrian crossing beeping". However, this was opposed by 15 others (Hearing: 10; DHH: 5), who mentioned that the "Bell has a weak association with traffic" and "its meaning is unclear". For instance, P1 said, "It sounds like a church bell, and I don't think it means it is about to slow, it doesn't link to it." P8 mentioned, "I wouldn't immediately associate the bell sound with a car". We followed up by asking what if Bell would be used for auditory eHMI, they suggested "it needs more sophisticated sound design to convey the message effectively" [P16] or "maybe use a beeping sound instead" [P19]. 5 participants (all Hearing) said the volume is too low and 6 participants (Hearing: 4; DHH: 2) said it was too low in the noisy environment. 3 participants (Hearing: 2; DHH: 1) explicitly said the Bell had "Blended into the background" [P7, P15, P36] and either "didn't stand out clearly" [P15] or "became indistinguishable" [36].

5.5.3 Speech. Most participants (N=30; Hearing: 25; DHH: 5) said they could hear the Speech eHMI under both noise levels. Unlike the Bell sound, all 25 Hearing participants were able to hear Speech.

Additionally, 5 DHH participants were able to hear the Speech eHMI to some extent, either (1) an incomplete message (N=3), (2) some verbal sound but could not get what it said (N=1), or (3) clearly only under a quiet environment (N=1). Of the 35 participants who could hear or partly hear speech eHMI, 24 participants (Hearing: 17; DHH: 7) agreed that background noise impacted their ability to perceive speech eHMI. 23 out of 24 participants (Hearing: 16; DHH: 7) said louder noise made it harder.

Overall, 12 participants (Hearing: 10; DHH: 2) explicitly praised the Speech, perceived it to be "good", "best", "perfect". This could be because Speech helped with understanding vehicle intention (N=11; Hearing: 8; DHH: 3), for instance, "Hearing the speech gave clear identification that the vehicle was stopping" [P17]. Additionally, it could also be due to the voice helping build the trust (N=2; Hearing: 1; DHH: 1). 14 participants (Hearing: 11; DHH: 3) liked the message content design as they were "clear", "simple", "easily recognisable", "concise". Despite it being simple, 2 participants felt "odd" and "unnatural" to hear that the vehicle used the phrase "I am" [P13, P18]. They said they would prefer alternative wording in the third person like "the vehicle is". There were also some suggestions about the issue of delivery: 5 participants (all Hearing) said the designed gap (i.e., 0.5 second) was not enough and would be better to have a larger gap.

Although the choice of robotic women's voice was liked ("pleasant tone" and a "good delivery", "it is distinct from human voices") by 3 participants (Hearing: 2; DHH: 1), it also received criticism. P36 mentioned that "a high-pitched robotic female voice would be difficult for people with high-frequency hearing loss to hear." To improve this, the suggestion of "a deeper voice might be more accessible" by P34 and P36 could be explored. In total, 7 participants (Hearing: 5; DHH: 2) said the volume was not loud enough, with 3 of them (Hearing: 2; DHH: 1) explicitly said in the loud setting, they felt "hard to to hear" [P15, P17] or "only able to hear when it was close in noisy environment" [P35]. 5 out of 7 commented that they could not hear when the vehicle was at a distance, which they understood later, as it could be 10 meters or even farther away.

5.5.4 Real-life Sound Distraction. Siren signals—including police, ambulance, and fire sirens, as well as fire alarms—are most frequently mentioned by the participants (N=16; Hearing: 11; DHH: 5). These sounds were perceived as critical and urgent; they could indicate a potential danger hazard and a signal of need to stop. Vehicle awareness-inducing sounds (i.e., standard engine hum, tyre noise, slow approaches, and reversing announcements) were mentioned by 9 participants (Hearing: 7; DHH: 2). These were helpful for general awareness but not urgent. Reversing announcements, in particular, were described positively for their clarity without being intrusive. Aggressive driving behaviours—such as car horns and screeching brakes—were reported by 9 participants (Hearing: 6; DHH: 3). They perceived these sounds as "stressful" or "unpleasant", and often linked them to "reckless driving". Similarly, show-off driving behaviours, including loud acceleration, racing, and modified exhausts, were mentioned by 8 participants (Hearing: 5; DHH: 3) and were rated as highly distracting as they stood out in the soundscape but were less about imminent safety. 6 participants (Hearing: 4; DHH: 2) mentioned environmental noise like construction, weather, and shouting as these sounds could mask

more important auditory signals, forcing pedestrians to rely more heavily on visual cues.

5.5.5 Cross without Auditory Stimuli. Several participants (N=26; Hearing: 19; DHH: 7) said they would be able to cross without auditory eHMI, while 6 participants (Hearing: 4; DHH: 2) said visual eHMI alone is not enough, as they lacked trust in the use of visual eHMI alone. This statement was also mentioned by people who said they could cross with only visual eHMI, 11 out of 26 (Hearing: 10; DHH: 1) said auditory eHMI could help with "confidence", "confirmation", and create a "safety layer if visuals are missed due to distraction", therefore, they also explicitly expressed a preference for combined audio-visual cues.

6 Discussion

eHMIs have been researched for several years as a potential solution to improve communication between AVs and other road users. However, current research lacks focus in several critical areas: (1) evaluating these concepts with disabled pedestrians, particularly DHH people; (2) identifying accessibility barriers that auditory eHMIs may introduce; (3) understanding how DHH pedestrians interpret and interact with auditory cues, and ensuring that such cues do not inadvertently disadvantage them; and (4) examining how these concepts perform in realistic urban soundscapes, including how well different auditory eHMIs work under varying background noise conditions. Our study is the first VR simulation study that evaluates the effects of Background Noise on the perception of Auditory Stimuli among hearing and DHH people regarding their crossing experience (i.e., trust, acceptance, perceived safety, mental load) and behaviour (i.e., gaze behaviour, step-in road time, early step into the road count).

RQ1: How do the ratings for experience and behaviour differ between Hearing and DHH participants? To conclude **RQ1**, our results showed some measurements of experience (i.e., usefulness rating among all conditions, safety rating among Bell and Speech auditory eHMI) was rated significantly higher by Hearing participants than DHH participants. The significantly higher usefulness ratings and safety ratings among Hearing people could be because DHH people did not well perceive the auditory eHMIs— our interview confirmed that all Hearing participants were able to perceive one of the auditory eHMIs under both conditions fully, and 21 out of 25 were able to perceive both auditory eHMIs under both background noise conditions fully; however, only 7 out of 11 DHH participants were able to fully perceive one of the auditory eHMIs under both conditions, and only 2 DHH participants were able to perceive both auditory eHMIs under both background noise conditions fully (see Table 2).

Based on our statistical results on necessity ratings, another explanation could be that auditory eHMIs may be significantly less necessary for most DHH participants than for Hearing participants. However, our interview also revealed that four DHH participants would not be able to cross without auditory stimuli, as they believed that it gave them confidence and made them feel safer (e.g., P30: "Audio gives confirmation and builds confidence in what I'm already perceiving visually, especially with cues like the bell." P32: "I find that a safer thing than hearing the bell or the voice", and

P36: "I prefer having some sound to confirm that something is happening. It raises my situational awareness and makes me feel more confident.").

RQ2: What impact does the Background Noise have on pedestrians regarding experience and behaviour? To answer **RQ2**, we found that loud background noise impaired participants' crossing experience (trust, usefulness, safety, mental workload) but had no impact on the crossing behaviour. This finding corroborates the literature that loud background noise also has a negative impact on road crossing, in line with other daily activities such as office tasks [8]. Early studies [87] suggest that background noise caused participants to choose smaller crossing gaps, take more time to make crossing decisions, and be slower to respond to the crossing opportunity; these were not found in our work. A possible explanation could be that the background noise used in our study was more constant, there was no sudden and momentary sound that caused a sudden change in sound volume and frequency. Tapiro et al. [87] employed sounds that are sudden and momentary (e.g., shop alarm, cyclist passing, noisy siren), which were also found to be distracting to our participants based on our qualitative data. Future exploration on sudden and momentary sounds are needed.

Early studies on the effect of noise on gaze behaviour during conversation showed people may show increased attention to the mouth to compensate for ambiguous auditory input [42]. We hypothesised that participants might allocate more visual attention to active eHMI components under loud background noise conditions to compensate for hearing ambiguous audio. However, we found no significant differences between background noise conditions, indicating that gaze behaviour (duration and percentage) remains stable across different background conditions. This could be because road crossing is much more visually demanding compared to speech-focused conversation [76, 83, 84] where participants had already paid more attention to the vehicle and the active eHMIs. It could also be because there is less need for compensation via gaze shifts due to the designed auditory eHMI being concise, repeated, and predictable once heard.

RQ3: What impact do the Auditory Stimuli of the Audio-Visual eHMI have on pedestrians regarding experience and behaviour? Our results showed that additional auditory eHMIs like Bell and Speech would improve experience (trust, usefulness, safety). This finding supports the auditory eHMI literature conducted in video-based research [27], VR simulation research [18], and real-world Wizard-of-Oz research [9]. However, we did not observe a significant improvement in providing additional auditory eHMIs on the pedestrian's crossing behaviour, such as eye gaze behaviour and step into the road decision making. This does not support the real-world Wizard-of-Oz research [9], where participants made faster crossing decisions when the intention of the vehicle was played. To answer **RQ3**, providing additional auditory eHMIs, such as Bell and Speech, improves experience but does not impact behaviours.

Multi-modal eHMIs. Our participants and related works [18, 27] highlighted the benefits of multi-modal eHMI. In particular, our findings support the use of audio-visual eHMI. However, the auditory eHMI requires careful consideration to ensure it supports rather than overwhelms pedestrians. Transport noise already ranks among Europe's top three environmental health threats, with more

Table 2: Results of DHH Participants regarding whether or not they were able to hear the designed auditory eHMI among quiet or noisy background (ordered based on severity of the hearing loss).

ID	Left Ear Hearing Loss	Right Ear Hearing Loss	Identity	Preferred Communication	Bell Quiet	Bell Loud	Speech Quiet	Speech Loud
P31	Moderate	Severe	HoH	English	Yes	Yes	Yes	Yes
P32	Moderate	Severe	deaf	English	Yes	Yes	Yes	No
P26	Severe	Severe	HoH	English	Yes	Yes	Yes	No
P29	Severe	Severe	HoH	English	No	No	Yes	No
P34	Severe	Severe	deaf	English	Yes	No	Yes	Yes
P28	Profound	Severe	HoH	English	Yes	No	Yes	Yes
P36	Severe	Profound	deaf	English	Yes	Yes	Yes	No
P30	Profound	Profound	deaf	English	Yes	Yes	Yes	No
P25	Profound	Profound	Deaf	BSL	Yes	No	Yes	Yes
P27	Profound	Profound	Deaf	BSL	Yes	Yes	Yes	Yes
P33	Profound	Profound	deaf	BSL	No	No	No	No

than 20% of Europeans exposed to harmful levels [36]. Thus, auditory eHMI should be designed in line with the WHO's environmental noise guidelines [94], while also ensuring that signals are both perceptible under varying background noise environments [17]. It is worth noting that auditory eHMI may still fail to work for all people, e.g., P33 cannot hear all designed auditory eHMI even with the help of hearing aids who claimed during the interview that he could not access truck reverse sound as well in the real-life, which is typically loud (112 dB) and with a high frequency [30]. However, P33 agreed that auditory eHMI would still be useful for hearing people (e.g., "I think the speech is important for hearing people" but then said, because now he has become deaf (due to an accident), "it doesn't matter") and said he would be able to cross with purely visual eHMI. Future work could explore other modalities, such as haptics, to enable multi-modal eHMIs for people who were unable to access auditory eHMIs.

Auditory Stimuli and Hearing Technologies. Statistically, we could not conclude which auditory eHMI (Speech or Bell) was better, as we did not find any significant difference between them among all measurements we had (i.e., experience, behaviour, necessity, reasonability, and ranking). This could be because both auditory eHMIs communicated the intention at the same starting time (i.e., 1.5 seconds after the vehicle initiated the yielding), as prior work suggests the timing of information is more important [72]. We observed several concerns regarding each method during the interview. For Bell eHMI, (1) it is hard to understand the meaning/intention of the Bell at the beginning, (2) it has a weak association with traffic, and (3) there is a risk Bell would be masked with the background sound (potentially filtered by their hearing aids as described by participants but also supported by hearing aids review [57]).

As for Speech eHMI, a significant concern observed is that when DHH participants focused on crossing, they may only partially hear the message content (i.e., "Stopped" from "I'm Stopped"), which led to confusion. Early study [18] compared a low content message ("Cross") and a high content message ("I'm stopping, you can cross") and found that the high content message could reduce cognitive load for low vision or blind people. The high content message design

could cause more issues for DHH people, as there is a greater risk of missing critical details. Researchers and designers should consider working with DHH people and different hearing technologies to amplify auditory eHMI designs rather than filter out [66].

Variability within the DHH Group among Auditory eHMI. The DHH group exhibited greater variability than the hearing group. A key reason for this could be the different audibility of the auditory eHMIs in the presence of loud background noise. As shown in Table 2, DHH participants tend to find it easier to access auditory eHMIs in quiet environments, while the access to each eHMI under loud background environments varied from person to person (and not solely due to hearing loss level): some could hear the Speech eHMI but not the Bell eHMI (i.e., P25, P28, P34), while others perceived the opposite (i.e., P26, P30, P32, P36). The hearing group demonstrated a much more consistent trend. All hearing participants were able to hear the Speech eHMI in both quiet and noisy backgrounds (although louder noise required more effort). They were also able to hear Bell eHMI under quiet backgrounds, with only 4 out of 25 participants being unable to detect the Bell eHMI in the loud condition (compared to DHH participants, this is a much smaller percentage).

Another reason could be the hearing technology being used. Although everyone in our study uses hearing technologies, each device may have different functions and priorities (e.g., amplify speech, filter background noise, or a mixture of both [53, 75]). Future research with DHH people should make sure the details of the hearing technology brand, model, and key functions are collected for further analysis and record what functions are activated.

6.1 Practical Implications

Limited work explored the use of eHMI with disabled people [7, 18, 43], we **call for further research to involve disabled people** as we explored differences in crossing experiences between hearing and DHH people. Prior works with hearing eHMIs were usually tested in environments with limited to no soundscape, where participants had no issues assessing the auditory eHMI [18, 27]. We found that loud background noise had a negative impact on participants' perception of the crossing experience and sometimes led to

participants being unable to hear the auditory eHMI. Therefore, we suggest **further work in eHMIs should employ background noise** to ensure the auditory eHMIs would work under different environments (e.g., quiet rural area, quiet urban area, busy urban area).

We recommend that eHMI design and research should **enable audio-visual (multi-modal) eHMI** so that when pedestrians face situations where one modality was missed, they could still rely on the other. The reason for enabling audio-visual (multi-modal) eHMIs, such as the combination used in our study (Visual: Abstract Light + Text, Audio: Text or Bell), was that combined eHMIs significantly improve the crossing experience compared to just the visual eHMI and sound stimuli generated from the driving vehicle. Regarding which auditory eHMI would be the best, there is not enough evidence from our results to conclusively answer this question. Further research on auditory eHMIs is needed, and we suggest that such research focus on **creating hearing technologies-friendly auditory eHMI**. Based on our qualitative findings, some hearing technologies could mask and filter the Bell as part of the background noise. Although speech can be amplified, it could also cause issues for users who could not capture the entire phrase (i.e., a DHH participant only heard "Stop" from the phrase "I'm stopped"); which points to topics that need to be addressed in future work. This could be done by working with hearing technology companies to (1) ensure that eHMI messages are clearly delivered or (2) filter them if they are not clearly captured to avoid confusing the participants.

6.2 Limitation and Future Work

This research has some limitations, which can also serve as directions for future studies. As limited work has been conducted on auditory eHMI and especially with DHH people, we employed a simple environment with a single controlled scenario featuring a non-signalised crossing to remove most distractions and allow participants to concentrate on the features [18, 25]. This controlled setup prioritised internal validity by enabling control over key factors; however, while appropriate for this purpose, it may limit ecological validity [15]. Future work could explore the scenario where participants need to cross (1) with other pedestrians [15], (2) in mixed traffic with both manual and automated vehicles [17], or (3) at a controlled traffic or zebra crossing to explore how they might affect pedestrians' crossing experience and behaviours. Ultimately, exploring the eHMI in field studies to maximise the ecological validity [65].

The sample size for DHH people (N=11) is relatively small due to the difficulty in finding disabled participants and is uneven compared to the hearing group (N=25). The future work should include more participants and an even number of samples across the groups. In addition, low vision or blind people could heavily rely on auditory eHMI. Exploring other disabled groups in the future will enhance the overall comprehensiveness of our findings. Our study only involved a single experimental session and a single country. Future research could adopt a longitudinal design to test the findings in different countries and cultures, determining whether they are applicable in the long term across various contexts [16].

As for the background noise perspective, we only explored background noise with good representations of standard urban environments (confirmed to be realistic by our participants in both the iterative testing and the formal study). We did not include typical distracting sounds, such as sirens, aggressive driving sounds, and sounds induced by show-off driving behaviours, as mentioned in Section 5.5.4. This is because these sounds also involve visual distractions that add complexity to the scenario design, while we want to focus on the background noise as a more controlled starting point. Future work could explore how these sounds would impact DHH people's crossing experience with auditory eHMIs. Regarding auditory eHMI, we only utilised existing ones from prior work [27] for our studies; we did not employ a comprehensive sound creation process to explore all the characteristics of auditory eHMI. Additionally, the interval between each auditory eHMI would also impact pedestrians' experience. As we discussed in **Auditory Stimuli and Hearing Technologies**, there is also a need for further in-depth investigation on auditory eHMI design.

7 Conclusion

Through a VR simulation, this research investigated the effect of background noise (quiet and loud) with auditory stimuli (baseline, bell, speech) for AV-pedestrian communications. We also explored the crossing experience and crossing behaviours between Hearing participants (N=25) and DHH participants (N=11) with the intention of understanding how we can better design eHMI for DHH people. Our results draw three conclusions based on the crossing scenario we evaluated: (1) Auditory stimuli should be carefully designed with consideration for DHH people and ensure that those auditory stimuli can be captured by their hearing technologies, so that their experience and behaviour are not impaired. (2) Loud background noise level would significantly impact pedestrians' crossing experience, but not the crossing behaviour we measured. (3) Providing additional auditory stimuli (i.e., Bell or Speech) could improve crossing experience, but has no impact on crossing behaviour. We also proposed four practical implications that pave the way for inclusive eHMI design and research.

Acknowledgments

References

[1] ISO/TR 23049:2018. 2018. *Road Vehicles: Ergonomic Aspects of External Visual Communication from Automated Vehicles to Other Road Users.* Standard. International Organization for Standardization.

[2] Sander Ackermans, Debargha Dey, Peter Ruijten, Raymond H. Cuijpers, and Bastian Pfleging. 2020. The Effects of Explicit Intention Communication, Conspicuous Sensors, and Pedestrian Attitude in Interactions with Automated Vehicles. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376197

[3] Seonggeun Ahn, Dokshin Lim, and Byungwoo Kim. 2021. Comparative Study on Differences in User Reaction by Visual and Auditory Signals for Multimodal eHMI Design. In *HCI International 2021 - Posters*, Constantine Stephanidis, Margherita Antona, and Stavroula Ntoa (Eds.). Springer International Publishing, Cham, 217–223.

[4] Javier Albusac, D. Vallejo, J.J. Castro-Schez, and C. Gzlez-Morcillo. 2018. An expert fuzzy system for improving safety on pedestrian crossings by means of visual feedback. *Control Engineering Practice* 75 (2018), 38–54. doi:10.1016/j.conengprac.2018.03.008

[5] Savvas Argyropoulos, Konstantinos Moustakas, Alexey A. Karpov, Oya Aran, Dimitrios Tzovaras, Thanos Tsakiris, Giovanna Varni, and Byungjun Kwon. 2008. Multimodal user interface for the communication of the disabled. *Journal on Multimodal User Interfaces* 2, 2 (01 Sep 2008), 105–116. doi:10.1007/s12193-008-0012-2

[6] Ashratuz Zavin Asha, Owen Brierley, Sowmya Somanath, Patrick Finn, and Ehud Sharlin. 2022. Towards Designing Audio Interactions with Autonomous Vehicles: A Hearing-Enhanced Pedestrian Story. In *AutomationXP22: Engaging with Automation, Workshop at the ACM CHI Conference on Human Factors in Computing Systems (CHI '22)* (2022-04-30). New Orleans, LA, USA.

[7] Ashratuz Zavin Asha, Christopher Smith, Georgina Freeman, Sean Crump, Sowmya Somanath, Lora Oehlberg, and Ehud Sharlin. 2021. Co-Designing Interactions between Pedestrians in Wheelchairs and Autonomous Vehicles. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 339–351. doi:10.1145/3461778.3462068

[8] Simon Banbury and Dianne C. Berry. 1998. Disruption of office-related tasks by speech and office noise. *British Journal of Psychology* 89, 3 (1998), 499–517. doi:10.1111/j.2044-8295.1998.tb02699.x

[9] Janina Bindschädel, Pauline Weimann, and Andrea Kiesel. 2023. Using eHMI, acoustic signal, and pitch motion to communicate the intention of automated vehicles to pedestrians: A Wizard of Oz study. *Transportation Research Part F: Traffic Psychology and Behaviour* 97 (2023), 59–72. doi:10.1016/j.trf.2023.06.017

[10] Avram Block, Swapna Joshi, Wilbert Tabone, Aryaman Pandya, Seonghee Lee, Vaidehi Patil, Nicholas Britten, and Paul Schmitt. 2023. The Road Ahead: Advancing Interactions between Autonomous Vehicles, Pedestrians, and Other Road Users. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE Press, Busan, South Korea, 16–23. doi:10.1109/RO-MAN57019.2023.10309535

[11] Marc-Philipp Böckle, Anna Pernestål Brenden, Maria Klingegård, Azra Habibovic, and Martijn Bout. 2017. SAV2P: Exploring the Impact of an Interface for Shared Automated Vehicles on Pedestrians' Experience. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct* (Oldenburg, Germany) *(AutomotiveUI '17)*. Association for Computing Machinery, New York, NY, USA, 136–140. doi:10.1145/3131726.3131765

[12] Edgar Brunner, Sebastian Domhof, and Frank Langer. 2002. *Nonparametric Analysis of Longitudinal Data in Factorial Experiments.* Wiley, New York. 288 pages.

[13] California Public Utilities Commission. 2013. *Pacific Gas and Electric Company's Cressey-Gallo 115 kV Power Line Project: Final Mitigated Negative Declaration and Supporting Initial Study.* Technical Report Application A.11-11-020. State of California Public Utilities Commission. https://ia.cpuc.ca.gov/environment/info/aspen/cresseygallo/toc-fmnd.htm Filed June 7, 2013.

[14] Angela Carlin, Marie H. Murphy, and Alison M. Gallagher. 2016. Do Interventions to Increase Walking Work? A Systematic Review of Interventions in Children and Adolescents. *Sports Medicine* 46, 4 (2016), 515–530. doi:10.1007/s40279-015-0432-6

[15] Mark Colley, Julian Britten, and Enrico Rukzio. 2023. Scalability in External Communication of Automated Vehicles: Evaluation and Recommendations. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 2, Article 51 (jun 2023), 26 pages. doi:10.1145/3596248

[16] Mark Colley, Daniel Kornmüller, Debargha Dey, Wendy Ju, and Enrico Rukzio. 2024. Longitudinal Effects of External Communication of Automated Vehicles in the USA and Germany: A Comparative Study in Virtual Reality and Via a Browser. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 4, Article 176 (Nov. 2024), 33 pages. doi:10.1145/3699778

[17] Mark Colley and Enrico Rukzio. 2020. A Design Space for External Communication of Autonomous Vehicles. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Virtual Event, DC, USA) *(AutomotiveUI '20)*. Association for Computing Machinery, New York, NY, USA, 212–222. doi:10.1145/3409120.3410646

[18] Mark Colley, Marcel Walch, Jan Gugenheimer, Ali Askari, and Enrico Rukzio. 2020. Towards Inclusive External Communication of Autonomous Vehicles for Pedestrians with Vision Impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376472

[19] Mark Colley, Marcel Walch, Jan Gugenheimer, and Enrico Rukzio. 2019. Including people with impairments from the start: external communication of autonomous vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 307–314. doi:10.1145/3349263.3351521

[20] Mark Colley, Marcel Walch, and Enrico Rukzio. 2019. For a better (simulated) world: considerations for VR in external communication research. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 442–449. doi:10.1145/3349263.3351523

[21] Mario Covarrubias, Elia Gatti, Monica Bordegoni, Umberto Cugini, and Alessandro Mansutti. 2014. Improving manual skills in persons with disabilities (PWD) through a multimodal assistance system. *Disability and Rehabilitation: Assistive Technology* 9, 4 (2014), 335–343. doi:10.3109/17483107.2013.799238 PMID: 23692410.

[22] Koen de Clercq, Andre Dietrich, Juan Pablo Núñez Velasco, Joost de Winter, and Riender Happee. 2019. External Human-Machine Interfaces on Automated Vehicles: Effects on Pedestrian Crossing Decisions. *Human Factors* 61, 8 (2019), 1353–1370. doi:10.1177/0018720819836343 PMID: 30912985.

[23] Kátia M. de Paiva Vianna, Marília R. Alves Cardoso, and Regina M. Rodrigues. 2015. Noise pollution and annoyance: an urban soundscapes study. *Noise & Health* 17, 76 (May–Jun 2015), 125–133. doi:10.4103/1463-1741.155833

[24] Shuchisnigdha Deb, Daniel W. Carruth, and Christopher R. Hudson. 2020. How Communicating Features can Help Pedestrian Safety in the Presence of Self-Driving Vehicles: Virtual Reality Experiment. *IEEE Transactions on Human-Machine Systems* 50, 2 (2020), 176–186. doi:10.1109/THMS.2019.2960517

[25] Shuchisnigdha Deb, Lesley J. Strawderman, and Daniel W. Carruth. 2018. Investigating pedestrian suggestions for external features on fully autonomous vehicles: A virtual reality experiment. *Transportation Research Part F: Traffic Psychology and Behaviour* 59 (2018), 135–149. doi:10.1016/j.trf.2018.08.016

[26] Debargha Dey, Kai Holländer, Melanie Berger, Berry Eggen, Marieke Martens, Bastian Pfleging, and Jacques Terken. 2020. Distance-Dependent eHMIs for the Interaction Between Automated Vehicles and Pedestrians. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Virtual Event, DC, USA) *(AutomotiveUI '20)*. Association for Computing Machinery, New York, NY, USA, 192–204. doi:10.1145/3409120.3410642

[27] Debargha Dey, Toros Ufuk Senan, Bart Hengeveld, Mark Colley, Azra Habibovic, and Wendy Ju. 2024. Multi-Modal eHMIs: The Relative Impact of Light and Sound in AV-Pedestrian Interaction. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 91, 16 pages. doi:10.1145/3613904.3642031

[28] Debargha Dey, Francesco Walker, Marieke Martens, and Jacques Terken. 2019. Gaze Patterns in Pedestrian Interaction with Vehicles: Towards Effective Design of External Human-Machine Interfaces for Automated Vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 369–378. doi:10.1145/3342197.3344523

[29] Jinzhen Dou, Shanguang Chen, Zhi Tang, Chang Xu, and Chengqi Xue. 2021. Evaluation of Multimodal External Human–Machine Interface for Driverless Vehicles in Virtual Reality. *Symmetry* 13, 4 (2021), 17. doi:10.3390/sym13040687

[30] Economic Commission for Europe, Inland Transport Committee, World Forum for Harmonization of Vehicle Regulations. 2022. *Proposal for a new UN Regulation on Reverse Warning: Submitted by the Working Party on Noise and Tyres (GRBP) — 187th session of the World Forum for Harmonization of Vehicle Regulations, Geneva, 21-24 June 2022.* ECE/TRANS/WP.29/2022/88E. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2022-07/ECE_TRANS_WP.29_2022_88E.pdf Item 4.12.1 of provisional agenda; adopted at GRBP 75 (ECE/TRANS/WP.29/GRBP/73, para. 28)..

[31] Economic Commission for Europe, Inland Transport Committee, World Forum for Harmonization of Vehicle Regulations. 2024. *Reports of the World Forum for Harmonization of Vehicle Regulations on its 194th session: Administrative Committee of the 1958 Agreement (eighty-eighth session), Executive Committee of the 1998 Agreement (seventy-first session), Administrative Committee of the 1997 Agreement (fifteenth session).* Technical Report ECE/TRANS/WP.29/1181. United Nations Economic Commission for Europe, Geneva. 49 pages. Distr.: General, 11 December 2024.

[32] Y. B. Eisma, S. van Bergen, S. M. ter Brake, M. T. T. Hensen, W. J. Tempelaar, and J. C. F. de Winter. 2020. External Human–Machine Interfaces: The Effect of Display Location on Crossing Intentions and Eye Movements. *Information* 11, 1 (2020), 18. doi:10.3390/info11010013

[33] Yke Bauke Eisma, Lucas van Gent, and Joost de Winter. 2023. Should an external human-machine interface flash or just show text? A study with a gaze-contingent setup. *Transportation Research Part F: Traffic Psychology and Behaviour* 97 (2023), 140–154. doi:10.1016/j.trf.2023.07.005

[34] Robert Wall Emerson and Dona Sauerburger. 2008. Detecting Approaching Vehicles at Streets with No Traffic Control. *Journal of Visual Impairment & Blindness* 102, 12 (2008), 747–760. doi:10.1177/0145482X0810201203 PMID: 20126414.

[35] Md Atik Enam, Ananta Bastola, and Julian Brinkley. 2024. Are the External Human-Machine Interfaces (eHMI) Accessible for People with Disabilities? A Systematic Review. In *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS)*. IEEE Press, Toronto, Canada, 1–6. doi:10.1109/ICHMS59971.2024.10555703

[36] European Environment Agency. 2025. *Environmental Noise in Europe 2025.* EEA Report 05/2025. European Environment Agency, Luxembourg: Publications Office of the European Union. doi:10.2800/1181642 Report published 24 June 2025; based on Environmental Noise Directive data from the 2022 reporting round.

[37] Stefanie M. Faas, Andrea C. Kao, and Martin Baumann. 2020. A Longitudinal Video Study on Communicating Status and Intent for Self-Driving Vehicle Pedestrian Interaction. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376484

[38] Stefanie M. Faas, Lesley-Ann Mathis, and Martin Baumann. 2020. External HMI for self-driving vehicles: Which information shall be displayed? *Transportation Research Part F: Traffic Psychology and Behaviour* 68 (2020), 171–186. doi:10.1016/j.trf.2019.12.009

[39] Evelyn Florentine, Mark Adam Ang, Scott Drew Pendleton, Hans Andersen, and Marcelo H. Ang. 2016. Pedestrian Notification Methods in Autonomous Vehicles for Multi-Class Mobility-on-Demand Service. In *Proceedings of the Fourth International Conference on Human Agent Interaction* (Biopolis, Singapore) *(HAI '16)*. Association for Computing Machinery, New York, NY, USA, 387–392. doi:10.1145/2974804.2974833

[40] Thomas Golob and Will Recker. 2001. Relationships Among Urban Freeway Accidents, Traffic Flow, Weather and Lighting Conditions. *Institute of Transportation Studies, UC Berkeley, Institute of Transportation Studies, Research Reports, Working Papers, Proceedings* 129 (01 2001), 24. doi:10.1061/(ASCE)0733-947X(2003)129:4(342)

[41] Kamer Gür, Eda Kılınç, Enes Günaydın, and Hilal Gülsün. 2021. The epidemiology of injury among adolescents with hearing loss, health beliefs regarding injury and associated factors. *Disability and Health Journal* 14, 2 (2021), 100994. doi:10.1016/j.dhjo.2020.100994

[42] Lauren V. Hadley, W. Owen Brimijoin, and William M. Whitmer. 2019. Speech, movement, and gaze behaviours during dyadic conversation in noise. *Scientific Reports* 9, 1 (2019), 10451. doi:10.1038/s41598-019-46416-0

[43] Mathias Haimerl, Mark Colley, and Andreas Riener. 2022. Evaluation of Common External Communication Concepts of Automated Vehicles for People With Intellectual Disabilities. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 182 (Sept. 2022), 19 pages. doi:10.1145/3546717

[44] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, Amsterdam, Netherlands, 139–183. doi:10.1016/S0166-4115(08)62386-9

[45] Zhifan He, Zhengyu Tan, Ruifo Zhang, Yanyan Li, and Bin Liu. 2021. How Pedestrian-AV Interaction Is Affected by the eHMI: A Virtual Reality Experiment. In *Advances in Usability, User Experience, Wearable and Assistive Technology*, Tareq Z. Ahram and Christianne S. Falcão (Eds.). Springer International Publishing, Cham, 707–714.

[46] Marion Hersh, James Ohene-Djan, and Saduf Naqvi. 2010. Investigating Road Safety Issues and Deaf People in the United Kingdom: An Empirical Study and Recommendations for Good Practice. *Journal of prevention & intervention in the community* 38 (10 2010), 290–305. doi:10.1080/10852352.2010.509021

[47] Marion. A. Hersh. 2016. Improving Deafblind Travelers' Experiences: An International Survey. *Journal of Travel Research* 55, 3 (2016), 380–394. doi:10.1177/0047287514546225

[48] Isti Hidayati, Wendy Tan, and Claudia Yamu. 2021. Conceptualizing Mobility Inequality: Mobility and Accessibility for the Marginalized. *Journal of Planning Literature* 36, 4 (2021), 492–507. doi:10.1177/08854122211012898

[49] Christopher R. Hudson, Shuchisnigdha Deb, Daniel W. Carruth, John McGinley, and Darren Frey. 2019. Pedestrian Perception of Autonomous Vehicles with External Interacting Features. In *Advances in Human Factors and Systems Interaction*, Isabel L. Nunes (Ed.). Springer International Publishing, Cham, 33–39.

[50] Lykke Muller Iversen, Gerd Marbjerg, and Hans Bendtsen. 2013. Noise from electric vehicles - 'State of the art' literature survey. *INTER-NOISE and NOISE-CON Congress and Conference Proceedings* 247, 8 (15 Sep 2013), 267–271. https://www.ingentaconnect.com/content/ince/incecp/2013/00000247/00000008/art00032

[51] Kari Kallinen. 2002. Reading news from a pocket computer in a distracting environment: effects of the tempo of background music. *Computers in Human Behavior* 18, 5 (2002), 537–551. doi:10.1016/S0747-5632(02)00005-5

[52] Sukran Karaosmanoglu, Katja Rogers, Dennis Wolf, Enrico Rukzio, Frank Steinicke, and Lennart E. Nacke. 2021. Feels like Team Spirit: Biometric and Strategic Interdependence in Asymmetric Multiplayer VR Games. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 443, 15 pages. doi:10.1145/3411764.3445492

[53] James M. Kates. 2002. *Signal Processing for Hearing Aids*. Springer US, Boston, MA, 235–277. doi:10.1007/0-306-47042-X_6

[54] Andrew J. Kolarik, Brian C. J. Moore, Pavel Zahorik, Silvia Cirstea, and Shahina Pardhan. 2016. Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, & Psychophysics* 78 (2016), 373–395. doi:10.3758/s13414-015-1015-1 Published online 2015 Nov 20.

[55] Moritz Körber. 2019. Theoretical Considerations and Development of a Questionnaire to Measure Trust in Automation. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*, Sebastiano Bagnara, Riccardo

[56] Tartaglia, Sara Albolino, Thomas Alexander, and Yushi Fujita (Eds.). Springer International Publishing, Cham, 13–30.

[56] Charles Korte and Rosalyn Grant. 1980. Traffic Noise, Environmental Awareness, and Pedestrian Behavior. *Environment and Behavior* 12, 3 (1980), 408–420. doi:10.1177/0013916580123006

[57] Stefan Launer, Justin A. Zakis, and Brian C. J. Moore. 2016. *Hearing Aid Signal Processing*. Springer International Publishing, Cham, 93–130. doi:10.1007/978-3-319-33036-5_4

[58] Abigail R. I. Lee, Justin M. Ales, and Julie M. Harris. 2019. Speed change discrimination for motion in depth using constant world and retinal speeds. *PLoS ONE* 14, 4 (2019), 21. doi:10.1371/journal.pone.0214766

[59] Victor Malmsten Lundgren, Azra Habibovic, Jonas Andersson, Tobias Lagström, Maria Nilsson, Anna Sirkka, Johan Fagerlönn, Rikard Fredriksson, Claes Edgren, Stas Krupenia, and Dennis Saluäär. 2017. Will There Be New Communication Needs When Introducing Automated Vehicles to the Urban Context?. In *Advances in Human Aspects of Transportation*, Neville A. Stanton, Steven Landry, Giuseppe Di Bucchianico, and Andrea Vallicelli (Eds.). Springer International Publishing, Cham, 485–497.

[60] Jörgen Lundälv. 2004. Self-reported experiences of incidents and injury events in traffic among hearing impaired people as pedestrians and cyclists. A follow-up study of mobility and use of hearing equipment. *International Journal of Rehabilitation Research* 27 (03 2004), 79–80. doi:10.1097/01.mrr.0000119280.12233.9d

[61] Karthik Mahadevan, Elaheh Sanoubari, Sowmya Somanath, James E. Young, and Ehud Sharlin. 2019. AV-Pedestrian Interaction Design Using a Pedestrian Mixed Traffic Simulator. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. Association for Computing Machinery, New York, NY, USA, 475–486. doi:10.1145/3322276.3322328

[62] Karthik Mahadevan, Sowmya Somanath, and Ehud Sharlin. 2018. Communicating Awareness and Intent in Autonomous Vehicle-Pedestrian Interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3173574.3174003

[63] Abby McCormack and Heather Fortnum. 2013. Why do people fitted with hearing aids not wear them? *International Journal of Audiology* 52, 5 (2013), 360–368. doi:10.3109/14992027.2013.769066 PMID: 23473209.

[64] H. McElrea and L. Standing. 1992. Fast music causes fast drinking. *Perceptual and Motor Skills* 75, 2 (October 1992), 362. doi:10.2466/pms.1992.75.2.362

[65] Dylan Moore, Rebecca Currano, G. Ella Strack, and David Sirkin. 2019. The Case for Implicit External Human-Machine Interfaces for Autonomous Vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 295–307. doi:10.1145/3342197.3345320

[66] Dylan James Moore. 2019. *The design of implicit pedestrian–autonomous vehicle interactions*. Ph. D. Dissertation. Stanford University, Stanford, California. Electronic resource; monographic; 1 online resource.

[67] Kimihiro Noguchi, Yulia R. Gel, Edgar Brunner, and Frank Konietschke. 2012. nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments. *Journal of Statistical Software* 50, 12 (2012), 1–23. doi:10.18637/jss.v050.i12

[68] Adrian C. North, David J. Hargreaves, and Sarah J. Heath. 1998. Musical Tempo and Time Perception in a Gymnasium. *Psychology of Music* 26, 1 (1998), 78–88. doi:10.1177/0305735698261007

[69] Marie-Agnès Pallas, Julien Cesbron, Sergio Luzzi, Lucia Busa, Gianfrancesco Colucci, and Raffaella Bellomini. 2020. *LIFE E-VIA : Review on electric vehicles and their noise emission*. Technical Report. IFSTTAR - French Institute of Science and Technology for Transport, Development and Networks. 98 pages.

[70] Jenna Rachel Panter and Andy Jones. 2010. Attitudes and the Environment as Determinants of Active Travel in Adults: What Do and Don't We Know? *Journal of Physical Activity and Health* 7, 4 (2010), 551 – 561. doi:10.1123/jpah.7.4.551

[71] Dario Pecchini and Felice Giuliani. 2015. Street-Crossing Behavior of People with Disabilities. *Journal of Transportation Engineering* 141, 10 (06 2015), 04015022. doi:10.1061/(ASCE)TE.1943-5436.0000782

[72] Hannah R. M. Pelikan and Andreas F. Jung. 2023. Designing Robot Sound-In-Interaction: The Case of Autonomous Public Transport Shuttle Buses. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (Stockholm, Sweden) *(HRI '23)*. Association for Computing Machinery, New York, NY, USA, 172–182. doi:10.1145/3568162.3576979

[73] Nathaniel R. Peterson, David B. Pisoni, and Richard T. Miyamoto. 2010. Cochlear implants and spoken language processing abilities: Review and assessment of the literature. *Restorative Neurology and Neuroscience* 28, 2 (2010), 237–250. doi:10.3233/RNN-2010-0535 PMID: 20404411.

[74] Tibor Petzoldt. 2014. On the relationship between pedestrian gap acceptance and time to arrival estimates. *Accident Analysis & Prevention* 72 (2014), 127–133. doi:10.1016/j.aap.2014.06.019

[75] Reinier Plomp. 1994. Noise, Amplification, and Compression: Considerations of Three Main Issues in Hearing Aid Design. *Ear and Hearing* 15, 1

(1994), 2–12. https://journals.lww.com/ear-hearing/fulltext/1994/02000/noise,_amplification,_and_compression_.2.aspx

[76] Brian J. Pugliese, Benjamin K. Barton, Shane J. Davis, and Gerardo Lopez. 2020. Assessing pedestrian safety across modalities via a simulated vehicle time-to-arrival task. *Accident Analysis & Prevention* 134 (2020), 105344. doi:10.1016/j.aap.2019.105344

[77] Amir Rasouli, Iuliia Kotseruba, and John K. Tsotsos. 2018. Understanding Pedestrian Behavior in Complex Traffic Scenes. *IEEE Transactions on Intelligent Vehicles* 3, 1 (2018), 61–70. doi:10.1109/TIV.2017.2788193

[78] Shima Reddy. 2016. 4 Major Types of Noise Hazard. https://www.environmentalpollution.in/noise-pollution/4-major-types-of-noise-hazard/1093. Accessed: 2025-03-27.

[79] Jan Ole Rixen, Teresa Hirzle, Mark Colley, Yannick Etzel, Enrico Rukzio, and Jan Gugenheimer. 2021. Exploring Augmented Visual Alterations in Interpersonal Communication. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 730, 11 pages. doi:10.1145/3411764.3445597

[80] Zhanna Sarsenbayeva, Niels van Berkel, Eduardo Velloso, Vassilis Kostakos, and Jorge Goncalves. 2018. Effect of Distinct Ambient Noise Types on Mobile Interaction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2, Article 82 (July 2018), 23 pages. doi:10.1145/3214285

[81] S. Schmidt and B. Färber. 2009. Pedestrians at the kerb – Recognising the action intentions of humans. *Transportation Research Part F: Traffic Psychology and Behaviour* 12, 4 (2009), 300–310. doi:10.1016/j.trf.2009.02.003

[82] Mariola Sliwinska-Kowalska and Adrian Davis. 2012. Noise-induced hearing loss. *Noise and Health* 14, 61 (Nov./Dec. 2012), 274–280. doi:10.4103/1463-1741.104893

[83] Francisco Soares, Emanuel Silva, Frederico Pereira, Carlos Silva, Emanuel Sousa, and Elisabete Freitas. 2020. The Influence of Noise Emitted by Vehicles on Pedestrian Crossing Decision-Making: A Study in a Virtual Environment. *Applied Sciences* 10, 8 (2020), 15. doi:10.3390/app10082913

[84] Francisco Soares, Emanuel Silva, Frederico Pereira, Carlos Silva, Emanuel Sousa, and Elisabete Freitas. 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 (2021), 202–220. doi:10.1016/j.trf.2021.08.014

[85] Matus Sucha, Daniel Dostal, and Ralf Risser. 2017. Pedestrian-driver communication and decision strategies at marked crossings. *Accident Analysis & Prevention* 102 (2017), 41–50. doi:10.1016/j.aap.2017.02.018

[86] Rouxian Sun, Xiangling Zhuang, Changxu Wu, Guozhen Zhao, and Kan Zhang. 2015. The estimation of vehicle speed and stopping distance by pedestrians crossing streets in a naturalistic traffic environment. *Transportation Research Part F: Traffic Psychology and Behaviour* 30 (2015), 97–106. doi:10.1016/j.trf.2015.02.002

[87] Hagai Tapiro, Tal Oron-Gilad, and Yisrael Parmet. 2018. The effect of environmental distractions on child pedestrian's crossing behavior. *Safety Science* 106 (2018), 219–229. doi:10.1016/j.ssci.2018.03.024

[88] The Waymo Team. 2022. *New Waymo One features inspired by DOT's Inclusive Design Challenge*. Google Waymo. https://waymo.com/blog/2022/08/new-waymo-one-features-inspired-by-dots Concept car / prototype.

[89] Tram Thi Minh Tran, Callum Parker, and Martin Tomitsch. 2021. A Review of Virtual Reality Studies on Autonomous Vehicle–Pedestrian Interaction. *IEEE Transactions on Human-Machine Systems* 51, 6 (2021), 641–652. doi:10.1109/THMS.2021.3107517

[90] Jinke D. Van Der Laan, Adriaan Heino, and Dick De Waard. 1997. A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies* 5, 1 (1997), 1–10. doi:10.1016/S0968-090X(96)00025-3

[91] Holly S. Vitense, Julie A. Jacko, and V. Kathlene Emery. 2002. Multimodal feedback: establishing a performance baseline for improved access by individuals with visual impairments. In *Proceedings of the Fifth International ACM Conference on Assistive Technologies* (Edinburgh, Scotland) *(Assets '02)*. Association for Computing Machinery, New York, NY, USA, 49–56. doi:10.1145/638249.638260

[92] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research Contributions in Human-Computer Interaction. *Interactions* 23, 3 (apr 2016), 38–44. doi:10.1145/2907069

[93] World Health Organization. 2021. *World Report on Hearing*. World Health Organization, Geneva. 252 pages. https://www.who.int/publications/i/item/9789240020481

[94] World Health Organization Regional Office for Europe. 2018. *Environmental Noise Guidelines for the European Region*. Technical Report. World Health Organization, Copenhagen, Denmark. https://www.who.int/publications/i/item/9789289053563 WHO Regional Office for Europe guideline.

[95] Wenge Xu, Foroogh Hajiseyedjavadi, Kurtis Weir, Chukwuemeka Eze, and Mark Colley. 2026. Towards Inclusive External Human-Machine Interface: Exploring the Effects of Visual and Auditory eHMI for Deaf and Hard-of-Hearing People. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain* (Barcelona, Spain) *(CHI '26)*. Association for Computing Machinery, Barcelona, Spain, 21 pages. doi:10.1145/3772318.3790738

[96] Jingyi Zhang, Erik Vinkhuyzen, and Melissa Cefkin. 2018. Evaluation of an Autonomous Vehicle External Communication System Concept: A Survey Study. In *Advances in Human Aspects of Transportation*. Springer International Publishing, Cham, 650–661. doi:10.1007/978-3-319-60441-1_63

[97] Yunfang Zheng, Jacob Swanson, Janet Koehnke, and Jianwei Guan. 2022. Sound Localization of Listeners With Normal Hearing, Impaired Hearing, Hearing Aids, Bone-Anchored Hearing Instruments, and Cochlear Implants: A Review. *American Journal of Audiology* 31, 3 (2022), 819–834. doi:10.1044/2022_AJSLP-22-00006