Multi-Modal eHMIs: The Relative Impact of Light and Sound in AV-Pedestrian Interaction

DEBARGHA DEY, Jacobs-Technion Cornell Institute, Cornell Tech, USA and Eindhoven University of Technology, The Netherlands
TOROS SENAN, Sorama, the Netherlands
BART HENGEVELD, Eindhoven University of Technology, the Netherlands
MARK COLLEY, Institute of Media Informatics, Ulm University, Germany and Jacobs-Technion Cornell Institute, Cornell Tech, USA
AZRA HABIBOVIC, Scania CV, Sweden
WENDY JU, Jacobs-Technion Cornell Institute, Cornell Tech, USA

Fig. 1. We conducted a video-based study to test the effectiveness of 3 eHMIs: one visual (light-based: SPLB – Slow-Pulsing light Band), and two acoustic (sound-based: Bell eHMI that emits a dinging sound when the AV yields, and Drone eHMI that indicates a yielding intent by emitting a droning/ humming sound that changes pitch from high to low as the vehicle decelerates). Results showed that despite divergences in subjective opinions, the contrast between the eHMIs – individually or in combination with each other – had little effect in road-crossing decision objectively, indicating that once learned, eHMIs tend to work in general and design differences have relatively less impact. This leaves room for taking subjective user feedback into account in designing a pleasant user experience for eHMIs.

External Human-Machine Interfaces (eHMIs) have been evaluated to facilitate interactions between Automated Vehicles (AVs) and pedestrians. Most eHMIs are, however, visual/ light-based solutions, and multi-modal eHMIs have received little attention to date. We ran an experimental video study ($N = 29$) to systematically understand the effect on pedestrian's willingness to cross the road and user preferences of a light-based eHMI (light bar on the bumper) and two sound-based eHMIs (bell sound and droning sound), and combinations thereof. We found no objective change in pedestrians' willingness to cross the road based on the nature of eHMI, although people expressed different subjective preferences for the different ways an eHMI may communicate, and sometimes even strong dislike for multi-modal eHMIs. This shows that the modality of the evaluated eHMI concepts had relatively little impact on their effectiveness. Consequently, this lays an important groundwork for accessibility considerations of future eHMIs, and points towards the insight that provisions can be made for taking user preferences into account without compromising effectiveness.

CCS Concepts: • **Human-centered computing** → **Interaction design**; *Interaction techniques.*

Additional Key Words and Phrases: Automated vehicle, eHMI, vulnerable road user, VRU, pedestrian, vehicle-pedestrian interaction, multimodal interface

ACM Reference Format:

Debargha Dey, Toros 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 *CHI Conference on Human Factors in Computing Systems (CHI '24), May 11-May 16, 2024, Honolulu, Hawai'i, USA.* ACM, New York, NY, USA, 23 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

To facilitate seamless interactions between Automated Vehicles (AVs) and Vulnerable Road Users (VRUs) such as pedestrians or bicyclists [58] in all situations – including ambiguous ones – AVs may need solutions to bridge the communication gap arising from the absence of driver-centric communication (such as eye contact and gestures due to the absence of a human driver) [16, 68, 73, 81]. Prior work has shown that VRUs predominantly rely on vehicle kinematics to understand driving intent [34, 35, 75]. However, in situations when the intent of the AV is not clear enough from the kinematics alone, external Human-Machine Interfaces (eHMIs) were shown to be successful [2, 24, 27, 34, 44, 56, 59].

Most previous work on eHMIs focused on visual communication, employing abstract light patterns [65], anthropomorphic features [10], text [16], symbols [11], or projections [66]. However, relying solely on visual communication can have drawbacks for multiple reasons: VRUs may have permanent impairments, experience situational impairments (e.g., by being distracted, the view being occluded), or could just not be looking toward the AV [18]. Some work, therefore, also included personal devices such as the smartphone in communication [14]. However, relatively little research has been conducted with the auditory modality, or combined audio-visual multi-modal interfaces for communication between AVs and VRUs.

To address this research gap, we used a video-based study ($N = 29$) to systematically investigate the relative contribution of visual (light-based) and auditory (sound-based) communication in eHMIs in effectively communicating a vehicle's intention to yield as reflected in pedestrians' willingness to cross and subsequently explored subjective user preferences. We found that the presence of any eHMI, irrespective of modality, benefited pedestrians in understanding the yielding intention of a vehicle. However, in the controlled, experimental setting, there was no evidence that the modality of the eHMI played a role in objectively modulating pedestrians' willingness to cross the road. However, people had different subjective opinions and preferences for specific (combinations of) modalities. Interestingly, many people pointed out perceiving some forms of multi-modal eHMIs as overwhelming and unpleasant.

Taken together, this potentially indicates that eHMI modality does not need to be a determining factor in the development of a functional eHMI. There is no one optimal modality for eHMI effectiveness, making it possible, and perhaps critical, to cater to accessibility needs and user preferences. Although eHMI modality has an effect on user experience, there is no evidence of its impact on effectiveness in terms of communicating driving intent in neutral environments. This insight further adds to the discussion on the usefulness of multi-modal eHMIs from the perspective of accessibility – designing for accessible, multi-modal eHMIs may have the freedom to focus more on policy constraints and user preferences without a critical hindrance in terms of the effectiveness of conveying vehicle intention.

Contribution Statement. Previous work has mostly shown theoretical or anecdotal evidence regarding the benefits or drawbacks of multi-modality, and conclusive evidence has been missing. To address this, we contribute by exploring the effect of multi-modality through empirical tests of the *relative impact of an eHMI's modality* – visual, auditory, and a combination thereof – in communicating an AV's intent to yield. Using a video-based study (N = 29), we contribute by showing the objective response of pedestrians to multi-modal eHMIs in terms of their willingness to cross (i.e. effectiveness of visual, audio, and multi-modal eHMIs), as well as the subjective user preferences. Our insights highlight that eHMIs can be effective regardless of modality. This highlights the opportunity and feasibility of taking user preferences into account from an early stage in the eHMI development process, subsequently improving the interactions between AVs and pedestrians.

2 RELATED WORK

The mechanisms through which road users communicate today are diverse and include both explicit (e.g., hand gestures, turn indicator) and implicit (e.g., velocity) communication channels [78, 82, 85]. While the definition of these communication channels varies somewhat depending on the context and author (see e.g., [8, 38, 71]), it is important to note that these channels are often not mutually exclusive and are used in parallel.

Similar to the interaction between humans, one could expect that both implicit and explicit communication will be important for smooth interactions between AVs and VRUs in their vicinity. While repeated exposure might help VRUs to eventually learn to correctly interpret an AV's behavior based on its implicit communication, providing additional explicit signals of the AV's perceptual and cognitive capabilities is likely to help guide VRUs through the interaction [86]. This is also the reason why explicit communication from AVs using eHMIs with various modalities, either separately or in combination, has attracted considerable attention in the research community.

Visual eHMIs. In the state of the art of eHMI research, external visual communication is the primary modality of explicit communication in AVs with other road users. In the meta-analysis of the eHMI design space, Dey et al. [30] found that an overwhelming 97% of the coded concepts (68 out of 70) use a visual modality. Several studies show that the presence of a visual eHMI could aid the interactions between VRUs and AVs [24, 68]. However, there is currently no clear agreement which information is most beneficial and efficient, and how many visual signals might be suitable. Visual eHMIs have been envisioned in a variety of form factors, ranging from light bands [2, 34, 48, 51, 65], two-dimensional displays on grills and/ or windshields [6, 11, 32, 84], or projections on the street ahead [23, 36, 61]. Each of these form factors and placements has its own benefits and limitations, and there is currently no clear evidence which of them is most suitable. However, one can argue that the front of the vehicle matches current expectations of pedestrians who generally look towards the location of the driver's head or vehicle movement [37, 41]. Considering this, the visual eHMI used in this study is located on the vehicle grill.

Auditory eHMIs. Recent policy guidelines in Europe with regard to silent vehicles (including Electric Vehicles) mandate that all vehicles need to emit an auditory signal based on the European Union's regulation on auditory vehicle alert systems (AVAS) [5, 42]. With regard to auditory cues for AVs, research has been limited. Prior work has shown that auditory cues in the form of friendly messages to engage pedestrians, positive feedback to invite pedestrians to cross the street safely, urgent warnings or alarms are able to successfully attract pedestrians [22]. In a recent real-traffic study by [74] with a Wizard-of-Oz AV to assess the impact of audio interfaces on AV communication within the actual outdoor soundscape showed that sound designed to communicate the vehicle speed led pedestrians to have a clearer perception of the vehicle's intent and experience a better interaction quality. This is an example of a *continuous audio cue* that communicates the status/ intent of the vehicle at all times. In contrast, several studies used auditory icons as *discrete audio cues* to evaluate sound in AV communication. For instance, Böckle et al. [7] used a bell sound to indicate that the AV will start driving. Hudson et al. [60] proposed different sound cues including playing music or a verbal message to announce a vehicle's yielding intention. While music was not effective, a clear preference was reported for the verbal message "safe to cross"; a result corroborated by other studies that also validated the use of verbal messages by Mahadevan et al. [69], Mahadevan et al. [70], and Colley et al. [16]. However, spoken text has the disadvantage of language-dependence and distortion in a busy traffic environment. Deb et al. [25] also showed that people generally prefer a loud sound over a loudspeaker announcing safety. A recent study [76] tested different humming sounds, jingles, human-like utterances ("ahem"), horns, and bells for automated buses in real traffic, and found that auditory cues can be used to effectively engage with other road users, and that different sounds have different connotations and meanings. Work by Florentine et al. [47] shows that although music as audio cues was useful in a warning/ acknowledgement situation, people tend to prefer light-based eHMIs for AVs to communicate an intention to yield. This was contradicted by Merat et al. [73], who showed people's preference for auditory signals over visual ones in announcing situation awareness and detection. Deb et al. [27] also showed that, compared to sound (horn, music, and verbal warning saying "safe to cross"), visual eHMIs had a much larger effect on the willingness to initiate crossing. Besides experimental evaluations, recent patents demonstrating the way to generate acoustic feedback as a means for AVs to interact with pedestrians [50, 88] highlight the technological readiness of this communication mechanism.

Multi-modal eHMIs. Previous work has recognized the need to design for communication with accessibility in mind, and emphasized the necessity for multi-modal communication [4, 16, 17, 52, 67]. However, the review by Dey et al. [30] shows that only a fraction (approximately 30%) of the concepts in the literature are multi-modal. In their study, Dou et al. [39] designed 12 eHMI concepts where visual (LED-based smile/arrow), audio (human voice/warning sound) and vehicle body language (the approaching speed decreases gradually/remains unchanged) modalities were combined and evaluated in VRThey concluded that multi-modal eHMIs resulted in more satisfactory interaction and improved safety compared to the uni-modal eHMI, in addition to noting that the visual modality had greater impact than audio, especially when it comes to the warning sound. On contrary, results from the Wizard-of-Oz study by Ahn et al. [3] showed that auditory signals are advantageous over visual ones in cognitive response. They also concluded that a combination of audio-visual modality is most effective in understanding information. The eHMI with an audio-visual modality was also reported to be more appealing than the eHMI with a single modality in the VR study with 12 pedestrians by He et al. [54]; the pedestrians selected the combination of a symbol and anthropomorphic voice as preferable over other eHMI types. The importance of multi-modal eHMI was also highlighted by Mahadevan et al. [70] who designed four eHMI concepts combining at least two modalities: visual (on vehicle or street), auditory (on vehicle or pedestrians' cellphones) and physical (vibration on pedestrians' cellphone or moving hand on vehicle), and pointed out that each modality has specific trade-offs which designers should consider when making new interfaces. Insights on the effectiveness of multi-modal eHMIs are inconclusive. In this work, we attempt to address this through a systematic comparison of three different eHMIs to understand user preferences in AV-pedestrian interaction.

3 eHMI CONCEPTS

To investigate the user preferences of communication, we chose three different eHMIs: one visual, and two auditory.

Visual (light-based) eHMI: SPLB/ Slowly-Pulsing Light Band

A Slow-Pulsing Light Band (SPLB) was used as a representative of the visual eHMI. We picked the light band eHMI design since it is the most widely used/ proposed visual eHMI [30] due to its relative simplicity, ease of implementation, and abstract execution [1, 24, 33, 43, 51, 53, 55, 77]. We adapted the light band design by integrating insights from prior research [31, 43, 65], which show that a uniform pattern like a slow-pulsing animation in cyan color is a good solution for showing intention to yield. For our study, we designed the SPLB mounted on the bumper of the vehicle. When the AV cruises in the automated mode without an intention to yield, it glows in a solid, cyan color. When the AV intends to yield, it pulsates in a sinusoidal pattern (the entire light bar alternately dims and glows at a rate of 0.75 Hz). This pattern was chosen because it was found to be the most appropriate animation pattern to communicate yielding intention for an AV in prior research [31]. When the AV wants to start driving again, the light bar returns to a steady glowing state, indicating a state change to "driving in automated mode". In essence, the pulsating eHMI tells the pedestrian, "I intend to yield", while the steadily glowing eHMI communicates, "I intend to keep driving".

Auditory (sound-based) eHMI

As discussed in Section 2, two kinds of intention-communicating sound-based eHMI designs have been used in the space of auditory eHMIs so far without any conclusive evidence regarding their effectiveness. To this end, we chose two auditory eHMIs – one representative of each [74]: continuous and discrete.

Continuous sound-based eHMI: Drone. Derived from prior work [74], this eHMI produces a continuous, droning sound as an indication of the AV's driving intention, akin to AVAS (acoustic vehicle alerting systems) used in electric vehicles[1] [87]. The eHMI emits a three-voice square wave with the fundamental frequencies of 43, 65, and 87 Hz, with frequency variations set at a rate of 3% over fundamental frequency, for each fundamental frequency, for every 1 km/hr of speed change [74]. The droning sound depends on the AV's intent. The two states of the AV – "driving in automated mode" and "at rest" correspond to two levels of droning sounds – a higher pitched and a lower pitched drone, respectively. As the AV slows down to a complete stop, the pitch of the drone decreases continuously. The rate of change of the pitch is dependent on the deceleration of the AV – the faster the AV decelerates, the faster the pitch changes down. Essentially, this leads to the behavior that when the AV is driving/ cruising, the eHMI generates a hum of constant pitch that is independent of the speed of the AV. Once the AV starts to yield, the eHMI generates a hum of decreasing pitch corresponding to the vehicle's speed. Once the vehicle stops, the eHMI continues to emit a constant hum of a lower pitch, and this continues as long as the AV stays stopped (indicating AV "at rest"). When the AV intends to start driving again, the sound quickly returns to the high-pitched hum corresponding to the non-yielding intention (indicating the vehicle is "driving in automated mode").

Discrete sound-based eHMI: Bell. This eHMI uses auditory icons [40] to communicate the AV's intention. As opposed to music or spoken words, which have disadvantages as discussed in Section 2, we used an abstract auditory icon and chose the sound of a bell, as used in prior work [7]. The bell sound was generated from a single sample downloaded from an online audio repository[2] released under the Creative Commons license. The single bell sound sample is concatenated to form a repeating bell sound sequence. When the AV cruises without the intention to yield, there is no sound. When the AV intends to yield, the eHMI activates a bell dinging at a frequency of 0.75 Hz. This sound is played throughout the time that the AV slows down, stops, and stays at rest until it is ready to start driving again, at which point the bell sound of the eHMI deactivates.

No eHMI. As a baseline, we included a 'No eHMI' condition with no light or sound augmentation, and the AV operates without explicitly communicating its yielding or non-yielding intent.

[1] https://unece.org/press/new-un-regulation-keeps-silent-cars-becoming-dangerous-cars, last accessed: Apr 16, 2023
[2] https://freesound.org/, last accessed: Feb 27, 2023

Visual eHMI (Light)	Discrete auditory eHMI (Bell)	Continuous auditory eHMI (Drone)	
			No eHMI
		×	Drone
	×		Bell
	×	×	Bell + Drone
×			Light
×		×	Light + Drone
×	×		Light + Bell
×	×	×	Light + Bell + Drone

Table 1. We used a full factorial design using three eHMIs, resulting in 8 eHMI combinations (blocks)

4 RESEARCH QUESTION AND HYPOTHESES

Our goal was to investigate the effectiveness of acoustic cues in the communication of intent of AVs in AV-pedestrian interaction in contrast and combination with light-based cues. In a video-based study, we investigate the following research question:

Does the addition of acoustic cues to a light-band eHMI aid in the communication of yielding intent?

Our hypotheses were as follows:

- ***H1:*** The addition of *Bell* sound to SPLB will help pedestrians comprehend an AV's yielding intent sooner.
- ***H2:*** The addition of *Drone* sound to SPLB will help pedestrians comprehend an AV's yielding intent sooner.
- ***H3:*** The addition of the combination of *Bell* and *Drone* sounds to SPLB will help pedestrians comprehend an AV's yielding intent sooner.

We posit that multiple modalities will increase the salience of the eHMI and will enable a pedestrian to understand the vehicle's driving intent more effectively. We ground these hypotheses in Wickens' Multiple Resource Theory (MRT) model [90], which states that when a task requires different resources, (e.g., in this case, the task of crossing depends on visual and auditory perception of the vehicle's intent), they can be processed simultaneously and quicker.

5 METHOD

The eHMI concepts were evaluated in a video-based within-subject experiment, as this allowed for practicable lab conditions where any potential danger for participants can be avoided. The experiment was submitted to and approved by the ethical review board of the researchers' institution(s).

Task. In this video-based experiment, the participants watched 48 videos of an AV approaching them while they assumed the role of a pedestrian intending to cross the road. While watching the videos, the participant indicated their willingness to cross the road in real-time as the vehicle in the video approached them [89].

Participants. We conducted the study with university students and staff who were recruited through convenience sampling, via a variety of channels, including the university experiment participation database, social media, and word of mouth (N = 29, 8 male, 21 female; mean age = 29.83 years; SD = 6.91 years). Only individuals who had normal or corrected-to-normal vision were recruited. We implemented a within-subjects setup across the 8 evaluation conditions (Table 1).

5.1 Apparatus and Study Setup

For the experiment's video stimuli, we captured video clips of an approaching Toyota Prius from a pedestrian's perspective. The pedestrian location was at the curbside of a straight road that was free from any traffic or other road

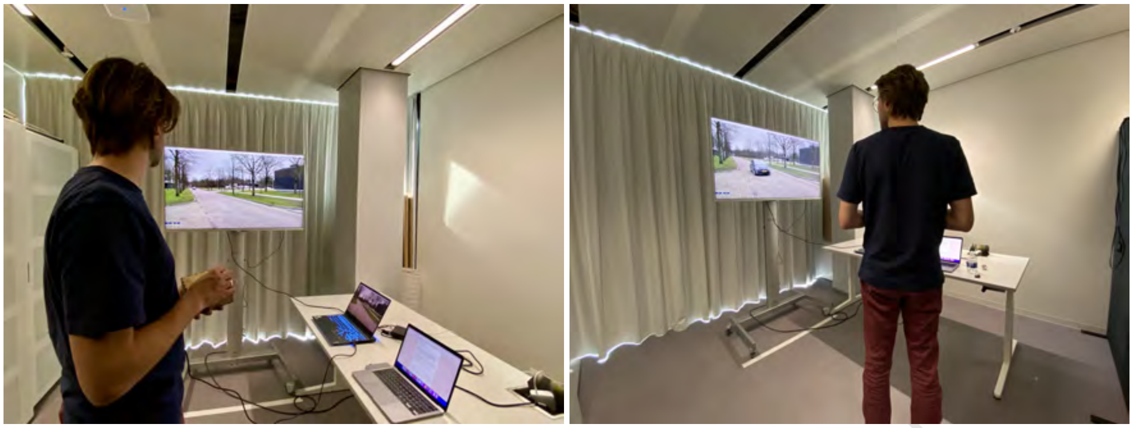


Fig. 2. The setup of the experiment: the participant stood sideways in front of a 55" screen where the video stimuli were presented.

Trial #	eHMIConcept	Behavior	Exposure #
1	No eHMI (Block 1)	Yielding (50km/h ↘ 0km/h)	1
2			2
3		Not Yielding (50km/h constant)	1
4			2
5		Not Yielding (50km/h ↘ 20km/h)	1
6			2
7	Light (Block 2)	Yielding (50km/h ↘ 0km/h)	1
8			2
9		Not Yielding (50km/h constant)	1
10			2
11		Yielding (50km/h ↘ 20km/h)	1
12			2
13	Bell (Block 3)	Yielding (50km/h ↘ 0km/h)	1
14			2
15		Not Yielding (50km/h constant)	1
16			2
17		Not Yielding (50km/h ↘ 20km/h)	1
18			2
19	Drone (Block 4)	Yielding (50km/h ↘ 0km/h)	1
20			2
21		Not Yielding (50km/h constant)	1
22			2
23		Not Yielding (50km/h ↘ 20km/h)	1
24			2
25	Light + Bell (Block 5)	Yielding (50km/h ↘ 0km/h)	1
26			2
27		Not Yielding (50km/h constant)	1
28			2
29		Not Yielding (50km/h ↘ 20km/h)	1
30			2
31	Light + Drone (Block 6)	Yielding (50km/h ↘ 0km/h)	1
32			2
33		Not Yielding (50km/h constant)	1
34			2
35		Yielding (50km/h ↘ 20km/h)	1
36			2
37	Bell + Drone (Block 7)	Yielding (50km/h ↘ 0km/h)	1
38			2
39		Not Yielding (50km/h constant)	1
40			2
41		Not Yielding (50km/h ↘ 20km/h)	1
42			2
43	Light + Bell + Drone (Block 8)	Yielding (50km/h ↘ 0km/h)	1
44			2
45		Not Yielding (50km/h constant)	1
46			2
47		Not Yielding (50km/h ↘ 20km/h)	1
48			2

Table 2. Study design: All participants experienced eight blocks of videos, each corresponding to an eHMI condition. The blocks were presented in a randomized order. We also randomized the order of stimuli within a block.

users. The interaction took place with no intersection or pedestrian crossing to ensure that the decision whether to cross the road is a direct result of the consideration of the AV's behavior and not from an expectation of right of way.

Vehicle behaviors: Our focus in this study was the interaction of pedestrians with the eHMI when the vehicle yields. However, to avoid a learning effect, we chose the AV to exhibit three different driving behaviors. In the first of the three behaviors – yielding behavior – the car slowed down to a complete stop in front of the pedestrian. The AV approached from a distance of 200 m at 50 km/h (standard city driving speed in Europe) and slowed down to a full stop at 5 m before the pedestrian. At 45 m away from the pedestrian, the car started braking gently but purposefully to indicate a deliberate yielding behavior, resulting in a total braking distance of 40 m and a literature-supported normal braking deceleration rate of 2.4 m/s^2 [28]. For the visual eHMI conditions, the SPLB eHMI starts indicating the yielding intention (pulsate) at a distance of 60 m. The second behavior was non-yielding. Here, the car approached and passed the pedestrian at a constant speed of 50 km/h. Additionally, we included a breaching behavior where the car slowed down but did not yield to the pedestrian (third of three behaviors). In this case, the AV slowed down from 50 km/h to 20 km/h and then kept driving without stopping. This is representative of ambiguous behavior. In this non-yielding behavior, the AV slowed down, which could confuse pedestrians into thinking that the AV is yielding to them, even though it does not intend to do so. (This could be an example of a behavior where an AV – aware of the presence of a pedestrian on the curbside – slows down as a measure of defensive driving, ready to stop if the pedestrian steps on the road, but without an active intention to yield).

For each behavior, the three eHMI conditions were applied as explained in Section 3. Apart from the eHMIs described, there was no further communication from the AV. We purposefully chose for the vehicle to not have any status lamp that communicated its automated driving mode to avoid confusion of a visual signal in the auditory-only eHMI conditions. Each stimulus was a video of the car from when it was ≈ 200 m away until either 3 seconds after having stopped for the pedestrian or until having passed the pedestrian without stopping. We recorded the pedestrians' *willingness-to-cross* to the yielding car from when the car was 12 seconds away from the pedestrian. For a yielding AV, we measured the pedestrians' willingness to cross relative to the 'Time-to-stop' (TTS) of the AV, which we defined as the moment when the AV comes to a complete stop in front of the pedestrian. For a non-yielding AV, we measured relative to the 'Time-to-arrival' (TTA) of the AV, which we defined as the moment when the front bumper of the AV reached the pedestrian's location.

Implementation: We used a Ghost Driver Wizard-of-Oz setup to hide the driver under a 'seat suit' and to create an illusion of an AV [80]. We captured the videos (4K resolution, 60 frames per second) during the daytime on an overcast day, which led to a uniformly lit environment devoid of starkly contrasting areas of direct sunlight and shadows. We augmented these videos with the proposed eHMI concepts. The visualizations of the light eHMI were added post-hoc using Adobe After Effects[3] and the audio of the sound-based eHMIs (Bell and Drone) was synthesized using MATLAB[4]. For both auditory eHMIs – continuous and discrete – the sound levels were increased logarithmically until the AV was closest to the pedestrian, accounting for the Doppler effect in the frequency shift [21].

We programmed the stimuli into a Processing[5] shell so that each video stimulus could be presented one after another, and the participant responses could be stored in a synchronized manner with the video. The video stimuli were presented to the participants on a 55-inch display in landscape orientation, as shown in Figure 2. To record the pedestrians' willingness to cross as a function of the AV's TTS or TTA, we used a slider device as input device as proposed by Walker et al. [89]. The participant could move the slider to indicate their willingness to cross the road. The two ends of the slider were mapped to 0 and 100 (corresponding to no willingness to cross and total willingness to cross), and the device recorded inputs at a rate of 10 Hz. We also instructed the participants that the continuum of the slider in between the ends can be used to express ambiguity regarding their decision.

5.2 Procedure

The experiment was conducted in a closed room at the researchers' institution. After each participant gave their informed consent at the start of the study, we asked them to stand in front of the display to watch the video stimuli. The participants stood sideways in front of the screen at a distance of approximately 1.5 m from the screen as shown in

[3] https://www.adobe.com/products/aftereffects.html, last accessed: Apr 16, 2023
[4] https://www.mathworks.com/products/matlab.html, last accessed: Apr 16, 2023
[5] https://processing.org/, last accessed: Apr 16, 2023

Figure 2. We asked them to imagine that they were standing at the curbside of a road that they would like to cross, and the road extended to their left in the screen (see Figure 2). The experiment was conducted in a controlled, silent laboratory room, and the audio signals (vehicle's engine sound and auditory eHMI) were transmitted through the speakers of the screen. The participants were informed that they would be encountering AVs in the study, and they were given a description of AVs and how they work adapted from Deb et al. [26]. Participants were instructed that they could always trust the eHMI message (i.e., there is no reason to fear a system failure). Before the measured trials began, the participant had the opportunity to experience three practice trials to familiarize themselves with the setup and the slider input device. The three stimuli used for the practice trial were the same as the videos with the 'No eHMI' condition, and the participants experienced each behavior once in a randomized order. After ensuring that the participants understood the task, they were allowed to proceed with the experiment.

Each participant experienced 48 trials (8 blocks $\times$ 3 behaviors $\times$ 2 exposures, see Table 2). The experiment conditions included the eight different eHMI concepts (see Table 1) and the three different behaviors of the car (yielding, not yielding with a constant speed, not yielding while slowing). We presented each set of stimuli pertaining to a certain eHMI concept block-wise to the participant (see Table 2). All eight blocks were presented in a randomized order to counterbalance any learning effects. For each condition of eHMI concept and yielding / non-yielding behavior, the participant experienced 2 exposures, which led to 6 video stimuli per block. Within each block, we also counterbalanced the order of presentation of the stimuli to avoid learning effects. Before a particular block of eHMI started, the experimenter showed the participant a video of the eHMI concept and explained it. We did this to ascertain that the participants understood the eHMI concepts and that the results of their responses were an accurate measure of the efficacy of the eHMI and not their intuitiveness. Once the participant confirmed that they understood the eHMI concept, they proceeded with the block. At the end of each block (corresponding to an eHMI condition), the experimenter asked them to fill out the standard, 26-item User Experience Questionnaire (UEQ) [64] for the eHMI concept. At the end of the experiment, the participants had to subjectively rank the three base (uncombined) eHMI conditions they encountered (*Light*, *Bell*, and *Drone*). Subsequently, the experiment concluded with a short semi-structured interview/discussion with the participant regarding how they perceived the crossing scenarios. They were asked to reflect upon how they decided to cross in front of the approaching AVs, as well as their impression of the different eHMI stimuli they experienced. We asked them to highlight if a particular concept stood out in a positive or negative way. The interview took approximately 10 minutes, and the entire experiment lasted approximately 60 minutes. Each participant was compensated with €15.00.

5.3 Measures

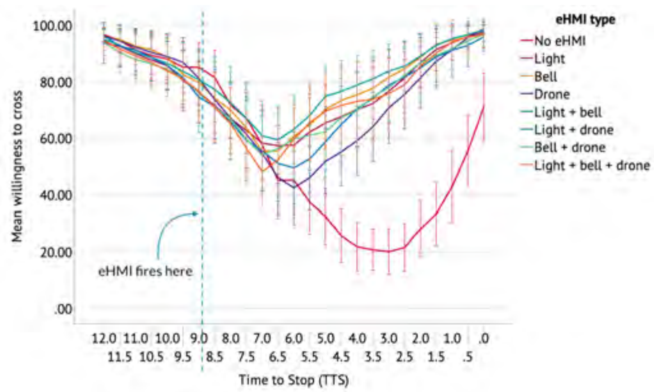
This study incorporated three different measures to evaluate the different eHMI concepts. Firstly, we used the *Willingness to Cross* data from the slider input device as an objective surrogate measure for the pedestrians' feeling of safety around the AV as described in prior work [89]. Secondly, we used the data of the 26-item *User Experience Questionnaire* that participants filled out for each eHMI. These data are transformed into the six User Experience factors attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. Finally, we used the participants' *Subjective ranking* data to determine any significant order of preference between the different kinds of eHMI concepts under investigation. For subjective ranking, we sought to simplify the task for the participants and requested them to rank the three base (primary) kinds of eHMIs (Light, Bell, and Drone) instead of asking them to rank all the eight combinations.

6 RESULTS

All statistical tests are reported at a 0.05 significance level for main effects with a Bonferroni post-hoc adjustment.

6.1 Objective Data

6.1.1 Willingness to cross. In our analysis, we extracted the *willingness-to-cross* values from an arrival time of 12.0 s in 0.5 s intervals and took the average of the values from both exposures. For each of these measurement points on the time scale, we conducted a repeated-measures ANOVA across the eight eHMI conditions. This essentially allowed us to compare the effectiveness of the eHMIs on Willingness to cross at 0.5 s intervals as the AV approached the pedestrian and its gap (measured in terms of Time-to-Arrival or Time-to-Stop) diminished.



(a) Pedestrians' willingness to cross as a function of the time-to-stopping of the yielding vehicle for different eHMIs. The 'No eHMI' condition registers the biggest drop in willingness to cross as the AV approaches the pedestrian. All other eHMI conditions show an increase in the willingness to cross after the eHMI fires.

Condition	F	Sig.	Effect size (η_p^2)
eHMI	25.70	<0.001	0.479
TTS	29.63	<0.001	0.514
eHMI * TTS	15.08	<0.001	0.350

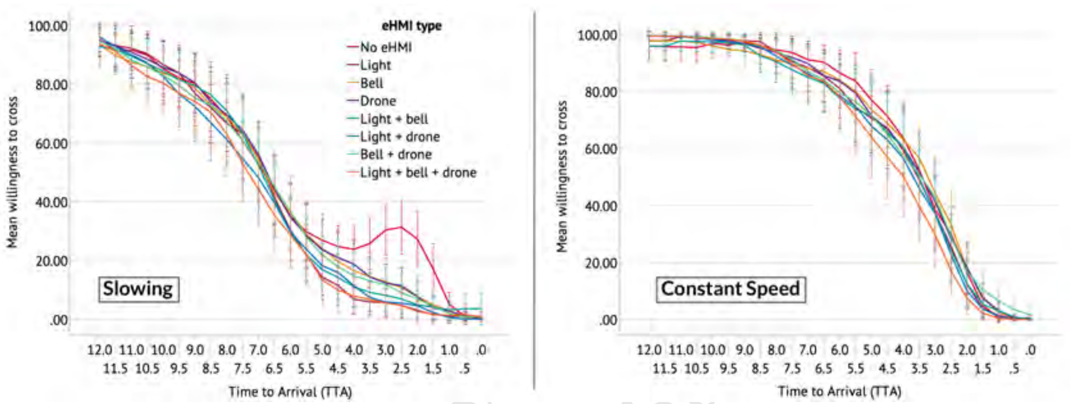
(b) Test statistics of the effects of *eHMI*, *Time-to-Stop*, and their interaction on pedestrians' willingness to cross.

Fig. 3. Performance of different eHMI concepts when the vehicle is yielding.

Yielding. Figure 3a shows the pedestrians' willingness to cross as a function of time (until the car comes to a complete stop) for the four eHMI conditions. We conducted a repeated-measures ANOVA across the eight eHMI conditions and the TTS for the vehicle (Figure 3b). Results show that the effect of *eHMI* was statistically significant and had a large effect size. As expected, *Time* had a highly significant effect on pedestrians' willingness to cross in all behaviors – it varied as the vehicle came closer (TTS decreased). Post-hoc tests found that the *No eHMI* condition was significantly different from all eHMI conditions ($p < 0.001$ for each comparison). However, there was no significant difference between the eHMI conditions (see Appendix A).

To investigate whether the different eHMI concepts had a significant effect at any specific TTS points in addition to its holistic effect across the entire experience, we conducted a repeated-measures ANOVA for each TTS point starting from 9.0 s (approximately the time when the eHMI – when present – activated to communicate yielding intention) in 0.5 s intervals. Appendix B shows the main effects of the eHMI in each measured TTS. The condition of sphericity was not met for any of these tests, so we report the test statistics with Greenhouse-Geisser correction [45, 46].

The results show a statistically significant effect of the eHMI as the AV comes closer, particularly from a TTS measurement of 6.0 s and less. Post-hoc tests show that the estimated marginal means for all eHMIs are statistically significantly different from the no eHMI condition - when eHMIs indicated that the vehicle was yielding, pedestrians' willingness to cross decreased less. However, the kind of eHMI did not have an effect on pedestrians' willingness to cross – for each of the analysis points, there was no significant difference in willingness to cross between the different eHMIs. In other words, there was no evidence that the *kind of eHMI* had a significant effect on the willingness to cross, although any eHMI performed significantly better than No eHMI. The pairwise comparisons are also reported in Appendix B.



(a) Pedestrians' willingness to cross as a function of the time-to-arrival for when the vehicle did not yield.

	Slowing			*Constant Speed*		
	F	Sig.	Effect size (η_p^2)	F	Sig.	Effect size (η_p^2)
eHMI	4.182	<0.001	0.130	1.348	0.230	0.046
TTA	199.114	<0.001	0.877	214.369	<0.001	0.884
eHMI * TTA	2.457	<0.001	0.081	1.404	0.246	0.048

(b) Test statistics of the effects of *eHMI*, *time to arrival*, and their interaction on pedestrians' willingness to cross for the two non-yielding conditions.

Fig. 4. Participants' responses of their willingness to cross for the non-yielding cases.

Not yielding. Our focus in this paper is on effective communication methods in multi-modal eHMIs for a *yielding* message, hence we present only a condensed analysis of the data for non-yielding behaviors. For each of the two non-yielding behaviors (1) maintaining a constant speed of 50 km/h and (2) slowing down from 50 km/h to a constant speed of 20 km/h the participants experienced the AV with all 8 combinations of eHMIs. In contrast to the yielding conditions, the eHMI remains the same when the vehicle does not yield (the light band on the bumper glows continuously; there is no bell; and the drone sound continues at a constant pitch). We conducted a repeated-measures ANOVA for each non-yielding behavior between each of the eHMI combinations every 0.5 s from 12.0 s of Time-to-Arrival (TTA) until 0 s (front bumper next to the pedestrian), and the test statistics are shown in Figure 4b.

As shown in Figure 4b, the eHMI had an effect on pedestrians' willingness to cross when the vehicle exhibited a slowing behavior. The plot of willingness to cross for a slowing vehicle (see Figure 4a) shows an interesting pattern: In the 'No eHMI' condition, the willingness to cross drops as the car approaches but rises again as it slows down –

pedestrians assumed that the slowing behavior meant that the vehicle was yielding to them. Only later, when they realized that the vehicle continued to drive, did they abruptly decide that they could no longer cross. In comparison, the pedestrians' willingness to cross stayed consistently lower from the TTA measurement of 5.0 s onward. In the presence of the eHMI, despite the slowing behavior of the car, there was less confusion about whether the car was yielding to them. Instead, the eHMI elucidated the car's intention to keep driving. Post-hoc tests of pairwise comparisons, however, reveal that the only significant difference was between the *No eHMI* condition and the *Light + bell + drone* condition. There was no observable difference between any other eHMI conditions. For the vehicle exhibiting a non-yielding behavior with constant speed, the eHMI did not have any significant effect on pedestrians' willingness to cross.

6.2 Subjective Data

6.2.1 ***User Experience Questionnaire.*** We used a repeated-measures ANOVA to test the effects of the different eHMI conditions for each of the six UEQ scales (a 7-point Likert scale from −3 to +3) to determine the overall user experience of each eHMI solution. We also included the *No eHMI* condition in the analysis as we also wanted to evaluate the overall experience of the approaching AV as a baseline. As the assumption of sphericity was violated for some of these tests and not for others, we uniformly report the multivariate tests as they do not assume sphericity and are more conservative [45, 46].

The tests of the main effects (5a) show that the effect of the different eHMI conditions is significant for each of the six UEQ scales. The effects are also shown in Figure 5b. Post-hoc tests show that the *No eHMI* condition performs significantly worse than any of the eHMIs in most of the six scales. The Drone eHMI did not perform significantly better than the No eHMI condition in terms of Attractiveness and Perspicuity. However, no statistically significant difference was found between the seven different kinds of eHMIs.

6.2.2 ***Subjective Ranking.*** The participants ranked the three base eHMIs (Light, Bell sound, and Drone sound) according to their preference. Descriptive statistics show that the mean order of preference of the three eHMIs was (1) Bell, (2) Light, and (3) Drone. The non-parametric Friedman's ANOVA found that there is no statistically significant order of preference for the three different kinds of eHMIs ($\chi^2(2) = 5.241$, $p = 0.073$).

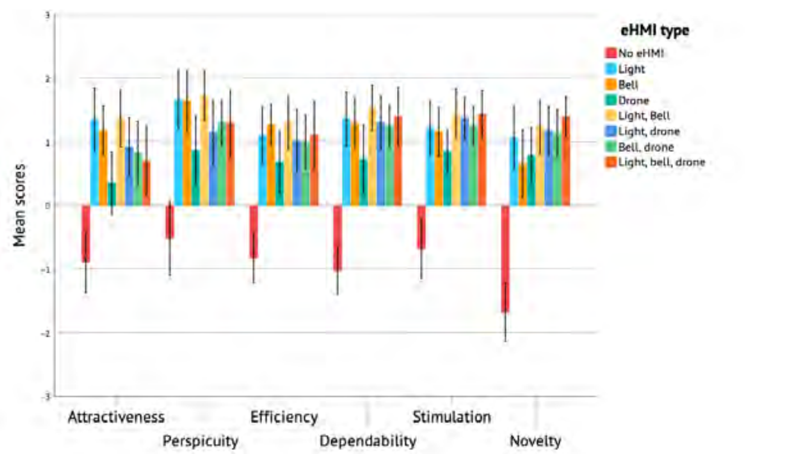
6.2.3 ***Qualitative Feedback.*** In addition to the quantitative data, we collected qualitative feedback through semi-structured interviews at the end of the experiment to gain insights from the subjective reasoning of the participants. Through a thematic analysis applied to the qualitative data through inductive coding, we outline the insights most relevant to the evaluated eHMIs, and in extension multi-modal eHMIs, along with selected participant quotes.

Reflections on visual eHMI. 6/29 participants (20.69%) explicitly mentioned liking the Light eHMI, commenting that it was "warm and welcoming" (P15) and "clear from a distance" (P16). However, there were also several participants (7/29) that remarked their difficulties with the Light eHMI, stating that it is "easy to miss if you blink" (P26), "hard to see" (P3), "can't see if it's blinking until it comes close" (P2, P7), and "difficult to distinguish if it was blinking or solid" (P8). Others also mentioned that this eHMI was "tiring" and required a "lot of effort to see if it was blinking" (P9), and "takes more attention because [they] really need to observe it" (P14). Others also noted that it was "counter-intuitive", because a blinking light had an association with warning (P9).

Reflections on acoustic eHMI. There were large individual differences between how participants perceived the different acoustic eHMIs. Participants expressed doubts if such solutions could work in a busy traffic environment with multiple vehicles (P27), and commented that it might be "difficult to distinguish where the sound comes from"

Attractiveness	F(7,22) = 8.45, p < .001, η_p^2 = .729
Perspicuity	F(7,22) = 6.61, p < .001, η_p^2 = .678
Efficiency	F(7,22) = 14.60, p < .001, η_p^2 = .823
Dependability	F(7,22) = 19.13, p < .001, η_p^2 = .859
Stimulation	F(7,22) = 12.23, p < .001, η_p^2 = .796
Novelty	F(7,22) = 20.81, p < .001, η_p^2 = .869

(a) Test results of main effects of eHMI condition across the six dimensions of UEQ.



(b) Mean score of each of the six UEQ dimensions clustered over the different eHMI conditions.

Fig. 5. Results of the UEQ analysis.

(P8). Several others also commented in a similar vein about the difficulty of localizing acoustic signals on specific vehicles if multiple vehicles are present – a problem that is mitigated by light signals: "How do I know it's the bell from that particular car?... the sound could be coming from anywhere, but you know the light is from that car" (P10). P13 mentions: "Now it's just one car, so you know of course that it is slowing down when you hear the bell, but if you are in a busy place with many cars and you don't know which one is making the sound... [if] the light is on the car, you at least know it is that car".

Several participants (12/29) explicitly mentioned liking the Bell eHMI. They characterized it as "charming" (P27), "friendly" (P15, P27), "nice" (P4), and "pleasant" (P4, P17), and commented that it was "distinctive and easy to identify (either there or not)" (P11, P28). P15 corroborated this by saying that it is "something that is simply not there if there is no intent to yield – a very clear instruction, like, if you don't hear anything, don't cross" (P15). However, 5/29 participants explicitly mentioned not liking the Bell eHMI. Some participants made associations with warnings and alarms (P10) and it "reminded [them] of the train tracks [level crossings]" (P14). A participant commented that they had to "remind [themself] about the bell that it means the car is stopping" (P10). Another participant (P15) remarked that in traffic "there are lots of bell-like sounds, and that might trigger you to think it I can cross". Interestingly, some participants also commented feeling "rushed to cross faster" (P20, P24), and being told to "move out of the way" (P25).

Concerning the Drone eHMI, 9/29 participants explicitly mentioned liking it. P12 stated that "the drone sound contributed, in my opinion, most to making the decision...it communicates the speed of the vehicle". Others remarked the signal as "intuitive" and requiring "least mental effort" (P9). One participant (P20) mentioned that from the pitch of the droning sound, they can extrapolate and "think how long it will take to slow down and stop". Others remarked about the novelty of the Drone ("fascinating sound" – P5). On the other side, 13/29 participants explicitly mentioned not liking the Drone eHMI. They commented on the need to take careful heed of the drone sound: "you have to listen very well" (P5, P18), "monitor change over time" (P26), pay attention to and think of it, and that is "extra processing" (P11). Several reported it to be "ominous" (P4, P26), "irritating" (P8), "annoying" (P14, P18), "depressing" (P27), "unpleasant" (P18), and "unfriendly" (P27). Interestingly, 6/29 participants associated the drone sound as a cognate to the vehicle's engine/ motor sound. Another interesting phenomenon was the tendency of some participants to conflate the pitch change of the drone with the Doppler effect due to the movement of the car – 8/29 participants explicitly mentioned confusion with regard to whether the perceived pitch change was due to an explicit intention to yield from the car, or simply a by-product of the car's approach.

Multi-modal eHMIs can offer assurance of vehicle intent. Several participants (11/29, 37.93%) explicitly reported that combinations of signals can complement each other and be reassuring, as it gives them a feeling of "having a confirmation" (P7, P10, P28). P21 mentioned, "When the light and drone came together, it was the best for me". Speaking of the Light + Bell combination, P11 commented that the combination was "actually a little bit assisting". P15 stated, "I think they nicely complement each other", while P16 remarked "it was reassuring... it was still nice to have the sound with the light because then your brain registers it all". Similarly, P18 gave a more elaborate justification in their comment: "the light is a good support for the bell. If you don't hear the bell, you can at least see the light. So when you're not looking well, you can hear the bell. And when you're not listening decently, then you can see the light". P15 also mentioned that the multi-modal signals changed their perception of the AV: "it's really trying to accommodate me in multiple ways, it's friendly" (P15). Only 3/29 participants explicitly claimed that they liked the combination of all three signals together because "it was very clear" (P19, P23) and it was "okay for me, because you couldn't miss it" (P25).

Multi-modal eHMIs may be overwhelming. Most participants did not like the combination of all three eHMIs. A majority of them (15/29) reported having focused on one signal and tuned out others. Which signal(s) they focused on and which one(s) they tuned out depended on individual preferences: "I sort of leaned on the light... I'm not sure I was putting attention on all because I just focused on the first one that I was completely certain about and then I sort of blocked out the rest" (P10). P11 mentioned, "When I see the light, I can immediately judge it and the others are... superficial... good to have, but not that necessary". P14 commented "it almost feels like for the combinations, there's always one that catches my attention first. And then I realize, oh, the others are changing as well". P14 additionally reported having "felt like I was unconsciously ignoring some of them" when multiple signals were presented together. P15 stated that they "had the clear idea that sound had a higher priority than the visual cue in [their] perspective... if all three were there, [they were] mostly focusing on 'do I hear bells, or not'?". P22 mentioned, "... if there is a combination of light and some sound, I will hear the sound, so I do not watch the light". Similarly, P23 reported to have"focus[ed] on one and blocked the others out".

Some participants did not easily tune out signals that were irrelevant to them. Several participants (14/29) explicitly reported multi-modal eHMIs where all three eHMIs were present together as unpleasant, and characterized them as "overwhelming" (P14), presenting "too much information" (P5, P10, P13, P14, 21), "chaotic" (P15), "confusing" (P26), "excessive" (P26), "not helpful" (P21), and as requiring a "lot more processing and attention" (P10), particularly when all three eHMIs activated together. P9 reported multi-modal eHMIs as "burdensome" and causing to feel "tired and stressed", emphasizing that "any kind of more than one signal was already too much". P5 mentioned that multi-modal eHMIs were "more fun, but also confusing" and that they "didn't know where to focus".

6.3 Evaluation of Hypotheses

Objective data showed no significant improvement of any form of multi-modal eHMIs over uni-modal ones. Therefore, we *reject H1 and H2*, that the addition of *Bell* sound and *Drone* sound respectively to SPLB did not enable pedestrians to comprehend the yielding intention of an AV sooner. However, participant interviews revealed that multimodality could offer a sense of reassurance and confirmation of a vehicle's yielding intent and aided crossing decisions, which points towards a potential for improved user experience. Similarly, we found that the addition of the combination of *Bell* and *Drone* sounds to SPLB (Light + Bell + Drone) did not have any objective effect in pedestrians' willingness to cross the road or the UEQ. However, a majority of participants reported it as being unpleasant. This leads us to *reject H3*, which hypothesized that the addition of the combination of *Bell* and *Drone* sounds to SPLB would enable pedestrians comprehend an AV's yielding intent sooner.

7 DISCUSSION

While most previous studies on eHMI for AVs focus on communication of yielding intent by means of a single modality, this study provides insights on both uni- and multi-modal eHMIs. Empirical results show that there is no clear winner with regard to the road-crossing decision-making performance. However, there are individual differences with regard to user preferences. We reflect on the nuanced implications of these findings with regard to the design of eHMI concepts.

Comparative effect of light and sound. Despite individual differences, there was no conclusive evidence of the visual or acoustic eHMIs outperforming the other. This finding does not support the insights from previous research such as Deb et al. [27], Florentine et al. [47], Merat et al. [73]. This can potentially be explained by prior findings by Pelikan and Jung [76], who posited that the timing of information is more important than the modality. In this study, as all eHMIs communicated the same information at the same time – although in different ways – this is a likely explanation for why there was no objective modulating effect on pedestrians' willingness to cross the road.

Subjectively however, most people preferred the combination of one visual and one acoustic signal, although the preference between the choice of acoustic signals varied. Some participants explicitly commented that the combination of the two acoustic signals did not help (P17, P18, P28). This is, therefore, a topic of future research. It is interesting to investigate whether multiple forms of communication using the same modality can have a detrimental effect on user perception. It is also interesting to note that different participants had completely different associations and mental models for the same signal. For instance, while many people perceived the bell as a calm, inviting, and friendly signal, others perceived it as being urgent and rushed and associated it with a warning. Similarly, while several people felt the drone was an intuitive and natural cognate for a vehicle's engine sound and speed, others found it unpleasant and burdensome.

In their study comparing auditory alarms in a clinical setting, Edworthy et al. [40] noted that auditory icon alarms outperformed tonal alarms. By this measure, the bell eHMI should have outperformed the drone eHMI, but this was objectively not the case. A potential reason for this could be the familiarity of bell-like sounds in the traffic setting (e.g., trams, level crossings, etc.), as pointed out by some participants, which points to the importance of context. Interestingly though, some participants mentioned that for them, the bell eHMI was easier to judge because it was "yes or no" (present or absent), so "when I hear the sound, I can go" (P11, P15). This was in contrast with the light or drone eHMI, where the difference between the communication of yielding and non-yielding intention was not as pronounced. This corroborates insights from prior research, which highlights the benefit of having a clear difference between the messages of communication in eHMIs [29].

Modality had little to no objective effect. Our results show that eHMIs can positively modulate pedestrians' willingness to cross by communicating an AV's intention to yield, as well as improving user experience. This is in line with a substantial corpus of previous research [24, 51, 57, 62]. However, we note that multi-modality did not significantly improve pedestrians' willingness to cross the road. This goes against our original hypotheses rooted in Wickens' Multiple Resource Theory [90], and further, also does not corroborate prior research in the field, which have tended to show a positive influence of multi-modality [3, 39, 70]. However, a potential theoretical explanation of why people found multi-modal eHMIs potentially overwhelming can be found in the Redundancy Principle of Mayer's theory of multimedia learning [72]. It suggests that redundant stimuli interferes with learning rather than facilitating it. When the same information is presented concurrently in multiple forms or is unnecessarily elaborated (in this case, through multi-modal stimuli), coordinating redundant information with essential information increases working memory load according to cognitive load theory, which may interfere with learning. However, we also note that despite variations in subjective preferences, multi-modality did not have an adverse effect on the objective comprehensibility of the eHMIs.

Another potential explanation for the apparent contradiction of our findings with previous studies could be cultural differences (past studies were done in China, Korea, and North America; while this study was conducted in Europe). Yet another explanation could be that past studies used other types of visual interfaces (e.g., arrows and icons). This highlights a design constraint: it is challenging to decouple form from function, as pointed out by Cefkin et al. [9], since by experiencing these interactions, people end up evaluating the specific interfaces. This makes it challenging to generalize the effect of multi-modality as a whole without tying it specifically to the tested interfaces. It is possible that with different metrics and different levels of granularity, nuanced differences between the eHMIs would emerge. However, for the eHMIs evaluated, the modality of communication did not play an objective role in affecting pedestrian's crossing behavior.

Need for multi-modality. Although no objective effect of multi-modality was observed, there are theoretical arguments for multi-modal communication in a traffic situation. One advantage of auditory communication is that it can function in the absence of visual attention. Therefore, auditory and multi-modal eHMI might excel in situations where VRUs are unable to notice an approaching AV immediately or are distracted while approaching the curb [14, 63]. Furthermore, auditory or multi-modal eHMIs can be critically important in addressing accessibility concerns of interactions between AVs and individuals with visual impairment [16]. Critically, the multi-modal eHMIs we tested did not perform worse than uni-modal eHMIs – it does not add to distraction or confusion, even though there are variations in individual preferences. This provides the foundational insight that multi-modal eHMIs are not detrimental to the effectiveness of eHMIs.

However, an auditory signal also comes with its challenges, which must be considered when designing multi-modal eHMIs. First, it is unknown how the directionality of sound affects perception in the wild – do participants recognize vehicles more quickly if they communicate via sound, compared to only visual communication? Second, the question arises – how auditory or multi-modal eHMIs will work in the usually busy and dynamic traffic environment when multiple AVs communicate simultaneously (see scalability challenges [12, 19, 36]). In such situations, pedestrians need to distinguish where a specific auditory signal originates and decipher conflicting messages if contradictory signals are heard from different AVs in traffic.

Implications for user experience and eHMI research. Our findings show that when the purpose of an eHMI is understood, its objective effectiveness is not hindered – despite differences in subjective preferences – highlighting the opportunity to consider the aspects of aesthetics, accessibility, and varying mental models in the design process. It points to the insight that once learned, eHMIs tend to work in general (which corroborates prior research) and that design differences have less of an objective effect. This leaves room for taking user preferences into account in the design of a pleasant and acceptable eHMI for a higher user experience. Since multi-modality did not have adverse effects on those with good vision or hearing, and does not hinder eHMI efficacy, there is a potential to use multi-modality to make information accessible to a more extensive population – leaving room for a viable design opportunity for accessible eHMIs. While multi-modality may be less important for individuals with good vision and hearing, delivering information through various modalities enhances accessibility for a broader population. Consequently, future research should focus more on the aspects of longer-term effects and cultural differences in the perceptions of eHMI rather than proposing further novel eHMIs. This paves the path for furthering the field of AV-pedestrian interaction research towards addressing issues such as accessibility and scalability concerns as pointed out in prior research [13, 17, 20, 30].

Limitations and Future Work. In our study, we chose a controlled setting to ensure that the different participants experienced the stimuli similarly to more clearly distinguish the relative impact of each stimulus. To limit confounding factors, we conducted the experiment in a simplified traffic scenario involving only one vehicle and one pedestrian on a straight and empty road devoid of any other traffic. Our findings provide insights with regard to the specific eHMIs we tested in such a neutral baseline scenario. However, the isolated laboratory setting also eliminates many factors of normal traffic environments, which would affect how people respond to any of the stimuli [15], so follow-up studies would need to be conducted to understand how the stimuli would be experienced in a real environment. Future work, therefore, must look into other situations or design implementations of eHMIs in more dynamic scenarios involving multiple vehicles and pedestrians – the complexity of such an environment may either improve the modulating effect of multi-modality, or prove to be further ineffective due to the added complexity within an already chaotic environment.

Another potential limitation is inherent to the video-based approach used in the study. While videos allow for a simple and quick proof-of-concept validation to occur under safe circumstances, it is possible participants exhibited more risk-taking behavior than they would in an environment with more potential for physical harm. However, previous research suggests that time-to-arrival estimates hold between video and real-life situations [79], and similar setups have been used with success in prior studies to study AV-pedestrian interactions. Additionally, existing literature shows that for vehicle-pedestrian interaction scenarios, the effects of a vehicle on pedestrians as experienced through videos are comparable with real life. Shen et al. [83] developed a video-based assessment tool to gauge (young) pedestrians' street crossing safety and concluded that video-based tests were valid and reliable. Fuest et al. [49] conducted a comparison study to evaluate the influence of an AV's driving behavior on pedestrians between real life (Wizard of Oz), Virtual Reality, and Video, and concluded that a video Wizard of Oz-based video setup (which this study used) can reproduce the critical crossing rate of a pedestrian from a real-world scenario with a difference of $\Delta < 1\%$. [49]. Consequently, we believe that the results remain ecologically valid.

8 CONCLUSION

This study presents a video-based experiment that investigated the user preferences with regard to multi-modal eHMIs – one visual and two auditory – when an AV wants to communicate its intention to yield. Our results show that eHMI modality had little objective effect in modulating pedestrians' willingness to cross a road. However, there were large individual differences in terms of subjective preferences of specific eHMI implementations. When it comes to multi-modal eHMIs, many people indicated strong preferences that differed from one another, but with a common thread that they liked some form of a combination of audio and visual eHMI (*Light+Bell* or *Light+Drone*, even though this preference did not have an objective impact on their willingness to cross). However, there was a thin line in this preference, and most people also found the case of *Light+Bell+Drone* unpleasant. This shows that design of multi-modal eHMIs, while potentially beneficial, is also extremely nuanced. The take-away message is that when it comes to eHMI multi-modality, *more is not necessarily better*. Our insights call attention to the need for carefully taking into consideration the user preferences in the design of multi-modal eHMIs from an early stage of the eHMI development process for a holistically optimal user experience for AV-pedestrian interaction.

ACKNOWLEDGMENTS

This work was supported by a fellowship of the German Academic Exchange Service (DAAD).

REFERENCES

[1] Claudia Ackermann, Matthias Beggiato, Sarah Schubert, and Josef F Krems. 2019. An experimental study to investigate design and assessment criteria: What is important for communication between pedestrians and automated vehicles? *Applied Ergonomics* 75 (2019), 272–282. https://doi.org/10.1016/j.apergo.2018.11.002

[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 *CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu, 1–14. https://doi.org/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] 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 *Designing Interactive Systems Conference 2021* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 339–351. https://doi.org/10.1145/3461778.3462068

[5] Sina Azizi Soldouz, Md Sami Hasnine, Mahadeo Sukhai, and Khandker Nurul Habib. 2020. Looking through the Perceptions of Blinds: Potential Impacts of Connected Autonomous Vehicles on Pedestrians with Visual Impairment. *Transportation Research Record* 2674, 5 (2020), 183–195. https://doi.org/10.1177/0361198120914299

[6] Pavlo Bazilinskyy, Dimitra Dodou, and Joost de Winter. 2019. Survey on eHMI concepts: The effect of text, color, and perspective. *Transportation Research Part F: Traffic Psychology and Behaviour* 67, April (2019), 175–194. https://doi.org/10.1016/j.trf.2019.10.013

[7] 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 - AutomotiveUI '17*. ACM, Oldenburg, Germany, 136–140. https://doi.org/10.1145/3131726.3131765

[8] C. Breazeal, C.D. Kidd, A.L. Thomaz, G. Hoffman, and M. Berlin. 2005. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, New York, NY, USA, 708–713. https://doi.org/10.1109/IROS.2005.1545011

[9] Melissa Cefkin, Jingyi Zhang, Erik Stayton, and Erik Vinkhuyzen. 2019. Multi-methods Research to Examine External HMI for Highly Automated Vehicles. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 11596 LNCS. Springer Verlag, Cham, 46–64. https://doi.org/10.1007/978-3-030-22666-4_4

[10] Chia-Ming Chang, Koki Toda, Daisuke Sakamoto, and Takeo Igarashi. 2017. Eyes on a Car: An Interface Design for Communication between an Autonomous Car and a Pedestrian. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Oldenburg, Germany) *(AutomotiveUI '17)*. Association for Computing Machinery, New York, NY, USA, 65–73. https://doi.org/10.1145/3122986.3122989

[11] Michael Clamann, Miles Aubert, and Mary L. Cummings. 2017. Evaluation of Vehicle-to-Pedestrian Communication Displays for Autonomous Vehicles. , 13 pages.

[12] 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. https://doi.org/10.1145/3596248

[13] Mark Colley, Julian Czymmeck, Mustafa Kücükkocak, Pascal Jansen, and Enrico Rukzio. 2024. PedSUMO: Simulacra of Automated Vehicle-Pedestrian Interaction Using SUMO To Study Large-Scale Effects. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24)*. ACM, New York, NY, USA, 6. https://doi.org/10.1145/3610977.3637478 March 11–14, 2024, Boulder, CO, USA.

[14] Mark Colley, Surong Li, and Enrico Rukzio. 2021. Increasing Pedestrian Safety Using External Communication of Autonomous Vehicles for Signalling Hazards. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 20, 10 pages. https://doi.org/10.1145/3447526.3472024

[15] 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. https://doi.org/10.1145/3409120.3410646

[16] Mark Colley, Marcel Walch, Jan Gugenheimer, Ali Askari, and Rukzio 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, Hawaii USA) *(CHI '20)*. ACM, Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3313831.3376472 Accepted.

[17] 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. https://doi.org/10.1145/3349263.3351521

[18] 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. https://doi.org/10.1145/3349263.3351523

[19] Mark Colley, Marcel Walch, and Enrico Rukzio. 2020. Unveiling the Lack of Scalability in Research on External Communication of Autonomous Vehicles. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3334480.3382865

[20] Mark Colley, Marcel Walch, and Enrico Rukzio. 2020. Unveiling the Lack of Scalability in Research on External Communication of Autonomous Vehicles. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, Hawaii USA) *(CHI '20)*. ACM, Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3334480.3382865

[21] G. Cook and T. Lesoing. 1991. A simple derivation of the Doppler effect for sound. *American Journal of Physics* 59, 3 (mar 1991), 218–220. https://doi.org/10.1119/1.16565

[22] Giorgio Costa. 2017. *Designing framework for human-autonomous vehicle interaction.* Master's Thesis. Keio University.

[23] Daimler AG. 2015. The Mercedes-Benz F 015 Luxury in Motion. https://www.mercedes-benz.com/en/mercedes-benz/innovation/research-vehicle-f-015-luxury-in-motion/https://ars.electronica.art/postcity/en/f015/https://www.youtube.com/watch?v=_cVN1yMJgWshttps://www.youtube.com/watch?v=IWB4xj7EILg

[24] 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* (2019). https://doi.org/10.1177/0018720819836343

[25] Shuchisnigdha Deb, Sushil Poudel, Suyogya Bhandari, and Brandon Warner. 2016. Identification of external design preferences in autonomous vehicles. In *Proceedings of the Industrial and Systems Engineering Research Conference, ISERC 2016*. Institute of Industrial Engineers, Norcross, GA 30092, USA, 470–476.

[26] Shuchisnigdha Deb, Lesley Strawderman, Daniel W. Carruth, Janice DuBien, Brian Smith, and Teena M. Garrison. 2017. Development and validation of a questionnaire to assess pedestrian receptivity toward fully autonomous vehicles. *Transportation Research Part C: Emerging Technologies* 84 (2017), 178–195. https://doi.org/10.1016/j.trc.2017.08.029

[27] 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 (nov 2018), 135–149. https://doi.org/10.1016/j.trf.2018.08.016

[28] Stavroula Panagiota Deligianni, Mohammed Quddus, Andrew Morris, Aaron Anvuur, and Steven Reed. 2017. Analyzing and Modeling Drivers' Deceleration Behavior from Normal Driving. *Transportation Research Record: Journal of the Transportation Research Board* 2663, 1 (jan 2017), 134–141. https://doi.org/10.3141/2663-17

[29] Debargha Dey, Azra Habibovic, Melanie Berger, Devanshi Bansal, Raymond H. Cuijpers, and Marieke Martens. 2022. Investigating the Need for Explicit Communication of Non-Yielding Intent through a Slow-Pulsing Light Band (SPLB) eHMI in AV-Pedestrian Interaction. In *Main Proceedings - 14th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2022*. ACM, New York, NY, USA, 307–318. https://doi.org/10.1145/3543174.3546086

[30] Debargha Dey, Azra Habibovic, Andreas Löcken, Philipp Wintersberger, Bastian Pfleging, Andreas Riener, Marieke Martens, and Jacques Terken. 2020. Taming the eHMI Jungle: A Classification Taxonomy to Guide, Compare, and Assess the Design Principles of Automated Vehicles' External Human-Machine Interfaces. *Transportation Research Interdisciplinary Perspectives* 7 (Aug. 2020), 100174. https://doi.org/10.1016/j.trip.2020.100174

[31] Debargha Dey, Azra Habibovic, Bastian Pfleging, Marieke Martens, and Jacques Terken. 2020. Color and Animation Preferences for a Light Band eHMI in Interactions Between Automated Vehicles and Pedestrians. In *CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376325

[32] 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. https://doi.org/10.1145/3409120.3410642

[33] Debargha Dey, Marieke Martens, Chao Wang, Felix Ros, and Jacques Terken. 2018. Interface Concepts for Intent Communication from Autonomous Vehicles to Vulnerable Road Users. In *Adjunct Proceedings of the 10th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '18)*. ACM, New York, NY, USA, 82–86. https://doi.org/10.1145/3239092.3265946

[34] Debargha Dey, Andrii Matviienko, Melanie Berger, Bastian Pfleging, Marieke Martens, and Jacques Terken. 2020. Communicating the Intention of an Automated Vehicle to Pedestrians: the Contributions of eHMI and Vehicle Behavior. *Information Technology* Special Issue: Automotive User Interfaces in the Age of Automation (2020), 123–141. https://doi.org/10.1515/ITIT-2020-0025

[35] Debargha Dey and Jacques Terken. 2017. Pedestrian Interaction with Vehicles: Roles of Explicit and Implicit Communication. In *AutomotiveUI '17 ACM 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Oldenburg). ACM, New York, NY, USA, 109–113. https://doi.org/10.1145/3122986.3123009

[36] Debargha Dey, Arjen Van Vastenhoven, Raymond H. Cuijpers, Marieke Martens, and Bastian Pfleging. 2021. Towards scalable eHMIs: Designing for AV-VRU communication beyond one pedestrian. In *Proceedings - 13th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2021*. ACM, Leeds, United Kingdom, 274–286.

[37] 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 - 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2019*. ACM, New York, NY, USA, 369–378. https://doi.org/10.1145/3342197.3344523

[38] Joshua E. Domeyer, John D. Lee, and Heishiro Toyoda. 2020. Vehicle Automation–Other Road User Communication and Coordination: Theory and Mechanisms. *IEEE Access* 8 (2020), 19860–19872. https://doi.org/10.1109/ACCESS.2020.2969233

[39] Jinzhen Dou, Shuanguang 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), 687.

[40] Judy Reed Edworthy, Cassie J. Parker, and Emily V. Martin. 2022. Discriminating between simultaneous audible alarms is easier with auditory icons. *Applied Ergonomics* 99, February 2021 (2022), 103609. https://doi.org/10.1016/j.apergo.2021.103609

[41] 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), 13. https://doi.org/10.3390/info11010013

[42] European Commission. 2014. On the sound level of motor vehicles and of replacement silencing systems, and amending Directive 2007/46/EC and repealing Directive 70/157/EEC. https://eur-lex.europa.eu/eli/reg/2014/540/oj

[43] Stefanie M. Faas and Martin Baumann. 2019. Yielding Light Signal Evaluation for Self-driving Vehicle and Pedestrian Interaction. In *In: Ahram T., Karwowski W., Pickl S., Taiar R. (eds) Human Systems Engineering and Design II. IHSED 2019. Advances in Intelligent Systems and Computing*, Vol. 1026. Springer International Publishing, Cham, 189–194. https://doi.org/10.1007/978-3-030-27928-8_29

[44] 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 (jan 2020), 171–186. https://doi.org/10.1016/j.trf.2019.12.009

[45] Andy Field. 2018. GLM 4: Repeated-Measures Designs. In *Discovering Statistics Using IBM SPSS 5th Edition*. SAGE Publications, Thousand Oaks, California, USA, 649–702.

[46] Andy Field. 2018. GLM 5: Mixed Designs. In *Discovering Statistics Using IBM SPSS 5th Edition*. SAGE Publications, Thousand Oaks, California, USA, 703–734.

[47] 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 - HAI '16*. Association for Computing Machinery, New York, NY, USA, 387–392. https://doi.org/10.1145/2974804.2974833

[48] Ford Motor Corporation. 2017. Ford, Virginia Tech Go Undercover to Develop Signals That Enable Autonomous Vehicles to Communicate with People. https://media.ford.com/content/fordmedia/fna/us/en/news/2017/09/13/ford-virginia-tech-autonomous-vehicle-human-testing.html

[49] Tanja Fuest, Elisabeth Schmidt, and Klaus Bengler. 2020. Comparison of methods to evaluate the influence of an automated vehicle's driving behavior on pedestrians: Wizard of Oz, virtual reality, and video. *Information (Switzerland)* 11, 6 (jun 2020), 22 pages. https://doi.org/10.3390/INFO11060291

[50] Donald K Grimm, Raymond J Kiefer, Linda S Angell, Richard K Deering, and Charles A Green. 2009. Vehicle to Entity Communication. https://patentimages.storage.googleapis.com/43/32/62/9d5f1f478b0f4a/US20110090093A1.pdf

[51] Azra Habibovic, Victor Malmsten Lundgren, Jonas Andersson, Maria Klingegård, Tobias Lagström, Anna Sirkka, Johan Fagerlönn, Claes Edgren, Rikard Fredriksson, Stas Krupenia, Dennis Saluäär, and Pontus Larsson. 2018. Communicating Intent of Automated Vehicles to Pedestrians. *Frontiers in Psychology* 9, August (2018), 15 pages. https://doi.org/10.3389/fpsyg.2018.01336

[52] 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 (sep 2022), 19 pages. https://doi.org/10.1145/3546717

[53] Michael Hamm, Wolfgang Huhn, and Johannes Reschke. 2018. Ideas for Next Lighting Generations in Digitalization and Autonomous Driving - Technocal Paper 2018-01-1038. *SAE Technical Papers* 2018-April (2018), 1–8. https://doi.org/10.4271/2018-01-1038

[54] 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: Proceedings of the AHFE 2021 Virtual Conferences on Usability and User Experience, Human Factors and Wearable Technologies, Human Factors in Virtual Environments and Game Design, and Human Factors and Assistive Technology*. Springer, Cham, 707–714. https://doi.org/10.1007/978-3-030-80091-8_84

[55] Ann-Christin Hensch, Isabel Neumann, Matthias Beggiato, Josephine Halama, and Josef F. Krems. 2019. How Should Automated Vehicles Communicate? – Effects of a Light-Based Communication Approach in a Wizard-of-Oz Study. In *AHFE 2019. Advances in Intelligent Systems and Computing, vol 964*, Vol. 964. Springer International Publishing, Cham, 79–91. https://doi.org/10.1007/978-3-030-20503-4_8

[56] Kai Holländer. 2019. A Pedestrian Perspective on Autonomous Vehicles. In *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion* (Marina del Ray, California) *(IUI '19)*. ACM, New York, NY, USA, 149–150. https://doi.org/10.1145/3308557.3308725

[57] Kai Holländer, Ashley Colley, Christian Mai, Jonna Häkkilä, Florian Alt, and Bastian Pfleging. 2019. Investigating the Influence of External Car Displays on Pedestrians' Crossing Behavior in Virtual Reality. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (Taipei, Taiwan) *(MobileHCI 2019)*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3338286.3340138

[58] Kai Holländer, Mark Colley, Enrico Rukzio, and Andreas Butz. 2021. A Taxonomy of Vulnerable Road Users for HCI Based On A Systematic Literature Review. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, Yokohama, Japan, 1–13. https://doi.org/10.1145/1122445.1122456

[59] Kai Holländer, Philipp Wintersberger, and Andreas Butz. 2019. Overtrust in External Cues of Automated Vehicles: An Experimental Investigation. In *11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '19)*. ACM, New York, NY, USA, 211–221. https://doi.org/10.1145/3342197.3344528

[60] Christopher R. Hudson, Shuchisnigdha Deb, Daniel W. Carruth, John McGinley, and Darren Frey. 2019. Pedestrian Perception of Autonomous Vehicles with External Interacting Features. *Advances in Intelligent Systems and Computing* 781 (2019), 33–39. https://doi.org/10.1007/978-3-319-94334-3_5

[61] Jaguar Land Rover. 2019. Jaguar land rover lights up the road ahead for self-driving vehicles of the future. https://www.jaguarlandrover.com/news/2019/01/jaguar-land-rover-lights-road-ahead-self-driving-vehicles-future

[62] Lars Kooijman, Riender Happee, and Joost C.F. de Winter. 2019. How do eHMIs affect pedestrians' crossing behavior? A study using a head-mounted display combined with a motion suit. *Information (Switzerland)* 10, 12 (dec 2019), 386. https://doi.org/10.3390/info10120386

[63] Mirjam Lanzer, Ina Koniakowsky, Mark Colley, and Martin Baumann. 2023. Interaction Effects of Pedestrian Behavior, Smartphone Distraction and External Communication of Automated Vehicles on Crossing and Gaze Behavior. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 768, 18 pages. https://doi.org/10.1145/3544548.3581303

[64] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *HCI and Usability for Education and Work*, Andreas Holzinger (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 63–76.

[65] Yee Mun Lee, Ruth Madigan, Chinebuli Uzondu, Jorge Garcia, Richard Romano, Gustav Markkula, and Natasha Merat. 2022. Learning to interpret novel eHMI: The effect of vehicle kinematics and eHMI familiarity on pedestrian' crossing behavior. *Journal of Safety Research* 80, December (2022), 270–280. https://doi.org/10.1016/j.jsr.2021.12.010

[66] Andreas Löcken, Carmen Golling, and Andreas Riener. 2019. How Should Automated Vehicles Interact with Pedestrians? A Comparative Analysis of Interaction Concepts in Virtual Reality. 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, 262–274. https://doi.org/10.1145/3342197.3344544

[67] Andreas Löcken, Andrii Matviienko, Mark Colley, Debargha Dey, Azra Habibovic, Yee Mun Lee, and Andreas Riener. 2022. Accessible Automated Automotive Workshop Series (A3WS): International Perspective on Inclusive External Human-Machine Interfaces. In *Adjunct Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, New York, NY, USA, 192–195.

[68] 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? Springer International Publishing, Cham, 485–497. https://doi.org/10.1007/978-3-319-41682-3_41

[69] Karthik Mahadevan, Elaheh Sanoubari, Sowmya Somanath, James E Young, and Ehud Sharlin. 2019. AV-pedestrian interaction design using a pedestrian mixed traffic simulator. In *DIS 2019 - Proceedings of the 2019 ACM Designing Interactive Systems Conference*. ACM, New York, NY, USA, 475–486. https://doi.org/10.1145/3322276.3322328

[70] 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 - CHI '18*. University of Calgary, ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174003

[71] G. Markkula, R. Madigan, D. Nathanael, E. Portouli, Y. M. Lee, A. Dietrich, J. Billington, A. Schieben, and N. Merat. 2020. Defining interactions: a conceptual framework for understanding interactive behaviour in human and automated road traffic. *Theoretical Issues in Ergonomics Science* 21, 6 (2020), 728–752. https://doi.org/10.1080/1463922X.2020.1736686 arXiv:https://doi.org/10.1080/1463922X.2020.1736686

[72] Richard E. Mayer. 2009. *Multimedia learning, second edition*. Cambridge University Press, Cambridge, England. 1–304 pages. https://doi.org/10.1017/CBO9780511811678

[73] Natasha Merat, Tyron Louw, Ruth Madigan, Marc Wilbrink, and Anna Schieben. 2018. What externally presented information do VRUs require when interacting with fully Automated Road Transport Systems in shared space? *Accident Analysis and Prevention* March (2018), 0–1. https://doi.org/10.1016/j.aap.2018.03.018

[74] Dylan Moore, Rebecca Currano, and David Sirkin. 2020. Sound Decisions: How Synthetic Motor Sounds Improve Autonomous Vehicle-Pedestrian Interactions. In *Proceedings - 12th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2020*. ACM, New York, NY, USA, 94–103. https://doi.org/10.1145/3409120.3410667

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

[76] Hannah R. M. Pelikan and Malte 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 (HRI '23)*. Association for Computing Machinery, New York, NY, USA, 172–182. https://doi.org/10.1145/3568162.3576979

[77] Tibor Petzoldt, Katja Schleinitz, and Rainer Banse. 2018. Potential safety effects of a frontal brake light for motor vehicles. In *IET Intelligent Transport Systems*, Vol. 12. Wiley Online Library, Hoboken, New Jersey, USA, 449–453. https://doi.org/10.1049/iet-its.2017.0321

[78] Amir Rasouli, Iuliia Kotseruba, and John K Tsotsos. 2017. Understanding pedestrian behavior in complex traffic scenes. *IEEE Transactions on Intelligent Vehicles* 3, 1 (2017), 61–70.

[79] Miguel Ángel Recarte, Ángela Conchillo, and Luis Miguel Nunes. 2005. Estimation of arrival time in vehicle and video. *Psicothema* 17, 1 (2005), 112–117. http://www.redalyc.org/articulo.oa?id=72717118

[80] Dirk Rothenbücher, Jamy Li, David Sirkin, Brian Mok, and Wendy Ju. 2016. Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles. In *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016*. IEEE, New York, NY, USA, 795–802. https://doi.org/10.1109/ROMAN.2016.7745210

[81] Anna Schieben, Marc Wilbrink, Carmen Kettwich Ruth Madigan, Tyron Louw, and Natasha Merat. 2018. Designing the interaction of automated vehicles with other traffic participants: A design framework based on human needs and expectations. *Cognition, Technology & Work* 0, August (2018), 0. https://doi.org/10.1007/s10111-018-0521-z

[82] 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. https://doi.org/10.1016/j.trf.2009.02.003

[83] Zhuo Shen, Jinfei Ma, and Ning Wang. 2023. Development and validation of a video-based assessment tool for children's street-crossing safety. *Journal of Transport & Health* 33 (2023), 101716. https://doi.org/10.1016/j.jth.2023.101716

[84] Ye Eun Song, Christian Lehsing, Tanja Fuest, and Klaus Bengler. 2018. External HMIs and their effect on the interaction between pedestrians and automated vehicles. In *Advances in Intelligent Systems and Computing*, Vol. 722. Springer International Publishing, Cham, 13–18. https://doi.org/10.1007/978-3-319-73888-8_3

[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. https://doi.org/10.1016/j.aap.2017.02.018

[86] Sam Thellman and Tom Ziemke. 2021. The Perceptual Belief Problem: Why Explainability Is a Tough Challenge in Social Robotics. *J. Hum.-Robot Interact.* 10, 3, Article 29 (jul 2021), 15 pages. https://doi.org/10.1145/3461781

[87] UNECE: United Nations Economic Commission for Europe: Inland Transport Committee. 2015. *Proposal for a new Regulation concerning the approval of quiet road transport vehicles (QRTV): ECE/TRANS/WP.29/2016/26*. Technical Report. United Nations.

[88] Emar Vegt and Lenja Sorokin. 2016. Method and Control Unit for Communication between an Autonomous Vehicle and a Road User. https://patentimages.storage.googleapis.com/24/b8/12/0a5eb583bf5a1c/US20160362045A1.pdf

[89] Francesco Walker, Debargha Dey, Marieke Martens, Bastian Pfleging, Berry Eggen, and Jacques Terken. 2019. Feeling-of-Safety Slider : A Platform for Measuring Pedestrian Comfort in Field Interactions with Vehicles. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312880

[90] Christopher D Wickens. 2002. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science* 3, 2 (2002), 159–177. https://doi.org/10.1080/14639220210123806

A PAIRWISE COMPARISON OF eHMI EFFECT ON WILLINGNESS TO CROSS

	No eHMI	Light	Bell	Drone	Light+Bell	Light+Drone	Bell+Drone	Light+Bell+Drone
No eHMI	–	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Light	–	–	1.000	1.000	1.000	1.000	1.000	1.000
Bell	–	–	–	0.340	1.000	1.000	1.000	1.000
Drone	–	–	–	–	0.117	1.000	1.000	1.000
Light+Bell	–	–	–	–	–	0.559	0.130	0.140
Light+Drone	–	–	–	–	–	–	1.000	1.000
Bell+Drone	–	–	–	–	–	–	–	1.000
Light+Bell+Drone	–	–	–	–	–	–	–	–

Table 3. Pairwise post-hoc comparisons between eHMI conditions with regard to their overall effect on pedestrians' willingness to cross with a Bonferroni correction applied at a significance level of 0.00178 are reported. All observed significant differences were between the No-eHMI condition and other eHMI conditions. No significant differences between visual and auditory eHMIs were found.

B EFFECT OF eHMI AT DIFFERENT TTS POINTS FOR A YIELDING VEHICLE

TTS	F	Sig.	η_p^2	Pairs of significant differences
9.0	1.926	0.108	0.064	
8.5	1.865	0.107	0.062	
8.0	1.075	0.376	0.037	
7.5	1.472	0.209	0.050	
7.0	1.246	0.292	0.043	
6.5	2.277	0.050	0.075	
6.0	5.329	**<.001**	0.160	
5.5	10.958	**<.001**	0.281	(1, 2), (1, 3), (1, 5), (1, 8)
5.0	17.055	**<.001**	0.379	(1, 2), (1, 3), (1, 5), (1, 6), (1, 7), (1, 8)
4.5	22.317	**<.001**	0.444	(1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8)
4.0	29.909	**<.001**	0.516	(1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8)
3.5	37.886	**<.001**	0.575	(1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8)
3.0	44.150	**<.001**	0.612	(1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8)
2.5	42.379	**<.001**	0.602	(1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8)
2.0	38.424	**<.001**	0.578	(1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8)
1.5	38.484	**<.001**	0.579	(1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8)
1.0	31.078	**<.001**	0.526	(1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8)
0.5	22.376	**<.001**	0.444	(1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8)
0.0	13.142	**<.001**	0.319	

Legend
1 – No eHMI
2 – Light
3 – Bell
4 – Drone
5 – Light + bell
6 – Light + drone
7 – Bell + drone
8 – Light + bell + drone

Table 4. Main effects of different eHMIs across different Time-to-stop (TTS) measuring points for a yielding vehicle. The TTS points where the eHMI had a significant effect are highlighted in bold. Any corresponding significant differences from pairwise post-hoc comparisons with a Bonferroni correction applied at a significance level of 0.00178 are reported. All observed significant differences were between the No-eHMI condition and other eHMI conditions. We did not find significant differences at any measurement point between conditions where some form of eHMI was present.